SURGICAL MANAGEMENT OF GASTROESOPHAGEAL REFLUX DISEASE IN CHILDREN

RISK STRATIFICATION AND PREDICTION OF OUTCOMES

CANDIDATE	MISS EVA W. MACHARIA-COATES
DEGREE	DOCTOR OF PHILOSOPHY
INSTITUTE	GREAT ORMOND STREET INSTITUTE OF CHILD HEALTH
	UNIVERSITY COLLEGE LONDON
SUBMITTED	FEBRUARY 2019

DECLARATION

‘I, Eva Wambui Macharia confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.'

ABSTRACT

Introduction: Since the 1980s fundoplication, an operation developed for adults with hiatus hernia and reflux symptoms, has been performed in children with gastroesophageal reflux disease (GORD). When compared to adult outcomes, paediatric fundoplication has resulted in higher failure and revision rates.

In the **first** chapter we explore differences in paradigm, patient population and outcomes. Firstly, symptoms are poorly defined and are measured by instruments of varying quality. Secondly, neurological impairment (NI), prematurity and congenital anomalies (oesophageal atresia, congenital diaphragmatic hernia) are prevalent in children.

Purpose: To develop methods for stratifying paediatric fundoplication risk and predicting outcomes based on symptom profile, demographic factors, congenital and medical history.

Methods: Study objectives are addressed in three *opera*: a symptom questionnaire development (TARDIS:REFLUX), a randomised controlled trial (RCT) and a retrospective database study (RDS).

TARDIS: REFLUX: In the **second** chapter, digital research methods are used to design and validate a symptom questionnaire for paediatric GORD. The questionnaire is a market-viable smartphone app hosted on a commercial platform and trialed in a clinical pilot study.

RCT: In the third chapter, the REMOS trial is reported. The trial addresses the subset of children with NI and feeding difficulties. Participants are randomized to gastrostomy with or without fundoplication. Notably, pre- and post-operative reflux is quantified using pH-impedance.

RDS: In the **fourth** chapter, data mining and machine learning strategies are applied to a retrospective paediatric GORD database. Predictive modelling techniques applied include logistic regression, decision trees, random forests and market basket analysis.

Results and conclusion: This work makes two key contributions. Firstly, an effective methodology for development of digital research tools is presented here. Secondly, a synthesis is made of literature, the randomised controlled trial and retrospective database modelling. The resulting product is an evidence-based algorithm for the surgical management of children with GORD.

IMPACT STATEMENT

This project contributes to the knowledge-base of gastro-oesophageal reflux disease (GORD) through methodology, publications and product.

Firstly, using digital research methods, we have devised a novel app-based questionnaire for tracking symptoms in relation to GORD.

A second contribution is the data mining approach taken to knowledge discovery in GORD research. Embracing big data approaches, we have used descriptive and predictive modelling to extract information from retrospective data which would remain otherwise under-utilised.

Another methodological contribution is the analysis of the conduct and outcomes of the REMOS trial. From this trial, a pragmatic approach to clinical research, specifically in surgery, is described.

This thesis has resulted in three key publications: two articles in major paediatric surgery journals and publication of a software application- the TARDIS:REFLUX app. In the journal articles, we have defined the role of the upper gastrointestinal contrast study in the investigation of GORD. We have also described the role of fundoplication in children with ventilator dependency. The TARDIS:REFLUX app was published on a commercial platform and disseminated around the world. To demonstrate its efficacy as a research tool, it has been applied in a clinical research pilot study at Great Ormond Street Hospital. It also has value as a public engagement tool, enabling the public to participate in the shaping of research and clinical instruments.

Several presentations at national and international conferences have arisen from this work. Presented topics include modelling methodology, data architecture of app development and role of gastrostomy and jejunostomy tubes in children with GORD.

This project has two tangible products with clinical significance. Firstly, the TARDIS:REFLUX app which has value as a specific research tool and as a proof-of-concept for digital research tools. Secondly, the Paediatric GORD Surgical Risk Stratification Algorithm which arises from the synthesis of knowledge discovered in producing this opus. This algorithm provides an evidence-based pathway for surgeons to assess patients referred for fundoplication.

ACKNOWLEDGEMENTS

Firstly, I express sincere gratitude to my supervisor Professor TJ Cole. You continually challenged me to grapple with my data and engage with the code. Your personal warmth and honesty provided an invaluable steer.

I would also like to thank my supervisor Dr. S. Eaton for his faith and effort. You have provided the thread of continuity that kept this thesis project alive through a period of change at the Institute.

I also acknowledge Professors Pierro and De Coppi, who were responsible for REMOS, the inception project.

I also thank Richard Groves for his time and patience. Working on the TARDIS:REFLUX app with you was a stimulating period of discovery.

This project is dedicated in gratitude and respect to my parents, Professor JMZ Kamau and Associate Professor JW Kamau. Through teaching and example, you have engendered a deep love of knowledge and a culture of quiet industry in all your children.

Lastly, I thank my husband Tony Coates. You are my steady state, my equilibrium and my rock. This PhD is very much the product of our relationship, as are our beautiful boys, Mark and Jonathan.

TABLE OF CONTENTS

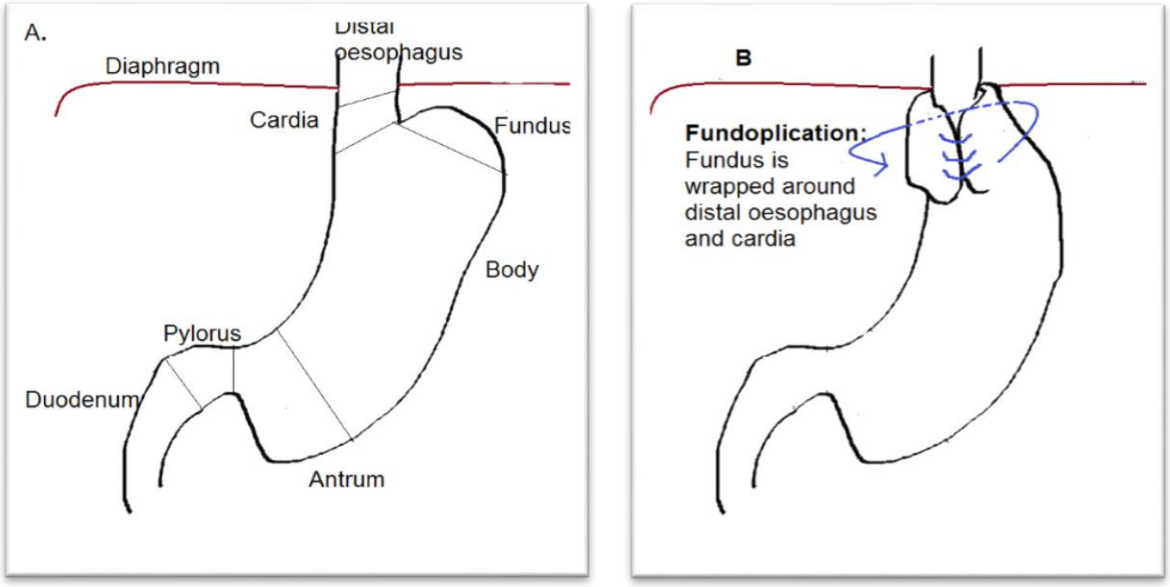
SECTION I: INTRODUCTION

CHAPTER 1: GASTROESOPHAGEAL REFLUX

Gastro-oesophageal reflux (GOR) is the retrograde passage of gastric contents from the stomach into the oesophagus. Food and fluid boluses are transmitted in an antegrade direction along the oesophagus due to peristaltic waves of oesophageal muscles and, more passively, gravity. The distal oesophagus passes between the crural muscles of the diaphragmatic hiatus. At the gastro-oesophageal junction (GOJ), the oesophagus meets the stomach at an acute angle, the angle of His. GOR occurs when the normal antegrade flow is reversed resulting in the passage of gastric contents into the oesophagus.

In children, first line of treatment for GOR is medical therapy. This involves a staged approach from feeding interventions e.g. thickening milk, to gastric acid suppression medication. The mainstay of surgical management of gastroesophageal reflux is the fundoplication. Fundoplication is an operation in which the fundus of the stomach is folded over the cardia, then stitched (plicated) into place.

Figure 1: Anatomy of the distal oesophagus and stomach. B. Fundoplication



Fundoplication is the third most common abdominal operation performed in children(1). Fundoplication carries a remarkably high re-operation rate, ranging from 10-30% in published literature(2)·(3). To understand how this operation has emerged as a treatment for paediatric reflux, we must first understand the underlying disease and treatment paradigm.

GASTRO-OESOPHAGEAL REFLUX DISEASE: A SHIFTING PARADIGM

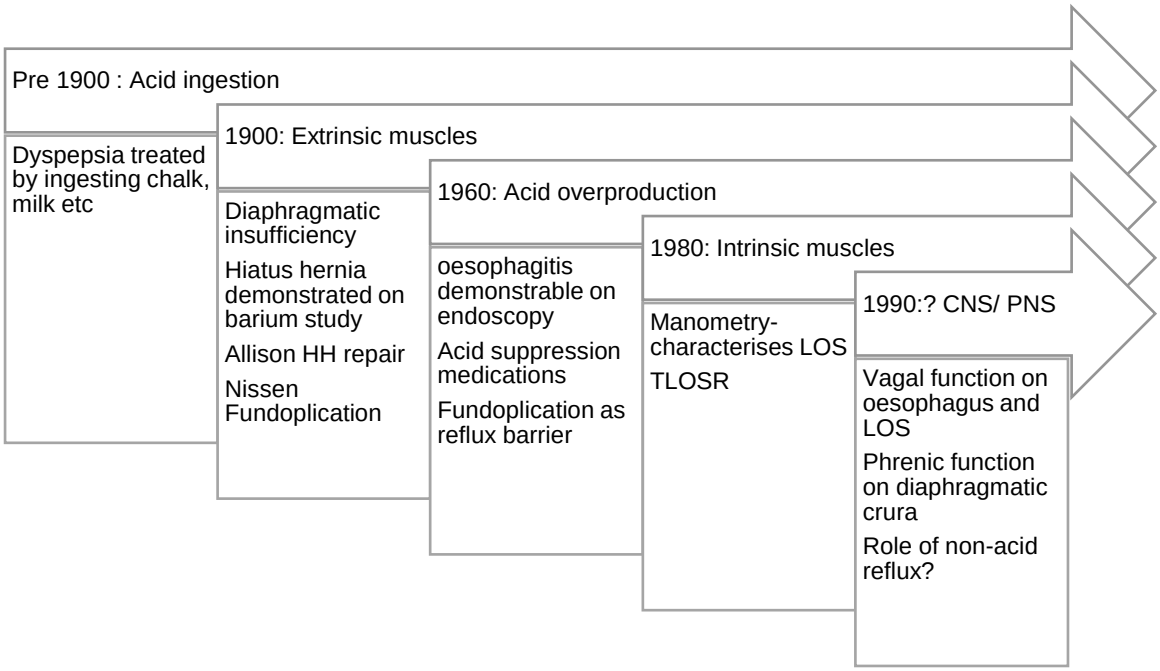
Historically, it was unclear which organ GOR symptoms arose from. In some accounts, reflux symptoms were described as an "imbecility of the stomach" (4). In others, as reflux pain was felt in the chest, symptoms were related to the heart. Consequently, it became known as cardiodynia or cardalgia(4) and, more latterly, heartburn. Dyspepsia, a common term synonymous with heartburn, was linked the ingestion of certain acid-containing foods and propensity to regurgitation. By the 19th Century, a link between the oesophagus and these visceral symptoms was made. Case reports of ulcers in the distal third of the oesophagus date back to further to 1833(5).

The relationship between gastric acid and dyspepsia was understood if not proven. Indeed, in the 17th century, chalk was described as a therapy to quell the acidic nature of dyspepsia(4). Bland diets, milk and white oxide of bismuth were advocated as therapies for dyspepsia. Post mortem examinations of animals had demonstrated the acidic content of gastric juices(4). In 1910, Schwartz published his theory linking excess gastric acid to gastric ulceration, launching the famous dictum: "no acid, no ulcer"(4,6).

In the 20th century, reflux emerged as the mechanism linking gastric acid to oesophageal ulceration. Friedenwald and Feldman(7) described the ulceration of the distal third of the oesophagus as similar to that seen in the stomach and duodenum. Crucially, they discriminated between 'peptic' i.e. due to gastric acid and other causes of ulceration e.g. carcinoma, foreign body. In 1906, Tilleston, a Harvard pathologist, collated case histories and post mortem specimens of patients with oesophageal ulcers(8). He described 12 different causes of oesophageal ulceration (including syphilitic and tuberculous ulcers). Of oesophageal ulceration, Tilleston writes: "in order that the peptic ulcer should be formed, it is evidently necessary that the cardia should be insufficient, allowing regurgitation of the gastric juice into the oesophagus". In the early 20th Century, insufficiency of the gastric cardia was understood to be the underlying cause of GOR.

The paradigm shifted with the development of a method for demonstrating reflux(9). In 1902, Bradford Cannon (1871-1945) , using recently developed x-ray imaging, captured images of food infused with heavy metal contrast media (bismuth, barium). He visualized the passage of boluses from the stomach into the oesophagus of an anaesthetized cat. He also observed that the cardia would contract and relax allowing passage of refluxate. Cannon's observations were important in three ways. Firstly, it introduced the idea that GOR was a physiological phenomenon with appropriate contraction and relaxation at the GOJ. Secondly, he demonstrated neuromuscular mediation indicating that the lower oesophageal sphincter (LOS) was more than an anatomical sphincter. Lastly, it established an imaging method to demonstrate instances of reflux. Indeed, for the next 40 years, barium contrast studies became the standard method for the diagnosis of reflux.

Figure 2: Shifting paradigm of gastroesophageal reflux disease



Use of barium studies led to increased and definitive diagnoses of hiatus hernia and a further shift in the GORD paradigm. Hiatus hernia is the abnormal cranial movement of the stomach into the thoracic cavity through an abnormally capacious or insufficient diaphragmatic hiatus. Symptoms of hiatus hernia (HH) coincide with those of reflux e.g. heartburn, regurgitation, early satiety and dysphagia. Hiatus hernia became synonymous with reflux(10), with coincident treatments. It followed, therefore, that initial surgical attempts to correct GOR focused on HH repair and reconstruction of the angle of His. Allison, who later became renowned for a method of hiatal hernia repair(11) postulated a pinchcock mechanism to explain GOR / HH. Allison, wrote of the diaphragm: 'first, it compresses the walls of the oesophagus from side to side, and second, it pulls down and increases the angulation of the oesophagus'(10). Surgical treatment for reflux was, therefore, hiatal hernia repair. Repair techniques reduced the hernia, lengthened the intra-abdominal oesophagus, included gastropexy to anchor the stomach and enhanced the oesophagogastric angle of His(10).

It was in this knowledge environment that German surgeon Rudolf Nissen (1896-1981) first performed fundoplication (1936). The anti-reflux effect of fundoplication was discovered perhaps serendipitously, but certainly accidentally. A 28-year-old male presented with a hiatus hernia and distal oesophageal ulcer that was eroding the pericardium. Nissen resected the ulcerated segment, anastomosing the distal oesophagus to the gastric cardia. Concerned about subsequent anastomotic leak, he folded the fundus over the anastomosis as a kind of patch. At follow-up, he noted that the patient's reflux symptoms had resolved.

The hiatus hernia paradigm of GORD was dominant in the first half of the 20th century. However, a 20-year prospective study laid the groundwork for a change of paradigm. Palmer(12) reported that patients with hiatus hernia had neither reflux symptoms nor oesophagitis. Furthermore, many patients with oesophagitis did not have hiatus hernia(12). Wenkelstein, writing in 1934, noted that patients could

have heartburn in the absence of oesophageal ulceration(13). He presented barium contrast and rigid oesophagoscopy of five cases of patients with chronic symptoms of heartburn. He was able to demonstrate that although radiological findings demonstrated reflux, endoscopic findings demonstrated 'diffuse inflammation without a definite ulcer'. These observations led to a further shift in the disease paradigm. In 1946, Allison introduced the term 'reflux oesophagitis', replacing 'peptic ulcer of the oesophagus.'

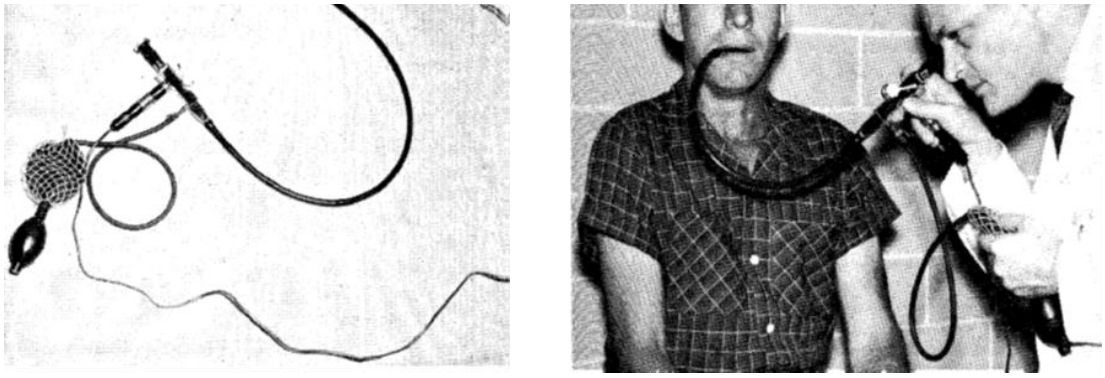
The reflux oesophagitis paradigm was contingent upon demonstration of oesophageal exposure to *gastric* acid. Bernstein and Baker(14)'s famous acid perfusion test instilled hydrochloric acid into a study subject's oesophagus and triggered heartburn, thus making the link between heartburn and oesophageal acid exposure. Tuttle and Grossman were first to report use of pH measurements in 1958(10) to demonstrate oesophageal acid exposure. They demonstrated transient episodes of 'oesophageal acidification'. Johnson and DeMeester (1974) are widely credited with the development of 24-hour oesophageal pH monitoring into a diagnostic tool for GORD(15) . The technique evolved from oesophageal placement of diodes in the oesophagus to the present i.e. antimony probes. Use of prolonged monitoring overcame a limitation of barium contrast studies i.e. transient capture of events which may or may not be physiological.

Another key shift in the paradigm was triggered by endoscopic examination of the oesophagus. Endoscopes were first invented and applied in human body cavities in the early 19th century(13). By the 1930's they were rigid instruments with an electric light attached. The operator peered down one end of a telescope, hoping to visualise the body region in question at the other end. It was described as looking 'through a glass darkly on the distant boundary"(10). Inspection of the stomach for ulcers was developed as diagnostic in the 1940's and 1950's. However, the rigid instruments posted a great risk of perforation to the narrower calibre oesophagus.

The invention and application of the flexible endoscope by Hirschowitz in 1957 revolutionised upper gastrointestinal diagnostics(16). The flexible endoscope made oesophageal examination relatively easier and safer. Inspection of the oesophagus became a standard part of the procedure of GI endoscopy. Pinch biopsy sampling of the oesophagus could be done by passing instruments within a canal in the flexible endoscope. This technical advance led to the development of corresponding macroscopic and microscopic diagnostic standards for oesophagitis.

The reflux oesophagitis paradigm captured the idea that gastric acid cased mucosal injury to the oesophagus, resulting in symptoms. Endoscopy and biopsy enabled both macroscopic and microscopic visualization of the sign (oesophagitis) secondary to the symptom (reflux). Allison and Johnstone (1953) presented a series of 7 patients with reflux oesophagitis and distal oesophageal ulceration lined with columnar rather than stratified epithelium(17).

Figure 3: Hirschowitz's flexible endoscope, an obliging patient is examined. Reproduced with permission form Morgenthal et al. Surgical endoscopy, 2006.



In 1960, Hayward's landmark paper in Thorax reported a large series of 200 patients with distal oesophageal columnar epithelium. Hayward presented the view that columnar epithelium represented transformation of the mucosa from stratified to columnar due to injury. Injury in turn was secondary to gastroesophageal reflux(10). Barrett (editor of the Thorax at the time) initially famously opposed this view. He believed that the observation was of ulcerated stomach displaced into the chest. Barrett would later accept that this columnar epithelium was observed in the oesophagus, rather than the stomach. Despite his initial opposition, this observation came to be known as Barrett's oesophagus. Importantly, this change was demonstrable both macroscopically and microscopically. Barrett's oesophagus was later confirmed as a predisposing factor for oesophageal adenocarcinoma. Fundoplication transformed from a symptom control procedure to potential neoadjuvant therapy for oesophageal adenocarcinoma in patients with Barrett's oesophagus.

A link between oesophageal acid exposure and symptoms of heartburn was posited by Carney et al. Developing on seminal work by Ismail-Beigi et al., these authors demonstrated that oesophageal acid exposure led to dilatation of intercellular space in the rabbit oesophageal mucosa(10). This increase in mucosal permeability was thought to expose sensory nerve endings within the submucosa to noxious acid, leading to pain i.e. heartburn. Endoscopy enabled clinicians to take oesophageal biopsies and demonstrate microscopic changes in patients with no macroscopic change. This further extended the disease paradigm to encompass patients with symptoms, macroscopically normal oesophagi but early changes on microscopy.

A pivot in the management of GORD arrived midway through the 20th century(10). In 1956, Code posited the role of the histamine receptor as the final common pathway in gastric acid secretion. Targeted treatment followed with the discovery of a histamine receptor antagonist in 1964. Histamine-2 (H2) receptors on gastric parietal cells interact with circulating histamine leading to downstream upregulation of a proton/potassium pump. This pump actively exchanges intraluminal potassium for intracellular protons in gastric luminal fluid, leading to acidification of gastric secretions. Cimetidine, a histamine (H2) receptor agonist, was introduced in 1973 and was the first in a line of acid suppression medications for GORD. More efficacious drugs were to follow. The discovery of proton pump inhibitors (PPIs) in the late 1970s enabled downstream targeting of the proton pump that acidified gastric secretion. This resulted in more effective neutralisation and even alkalinisation of gastric pH(10). To date, PPIs have been proven to be more effective than H2 antagonists for GORD symptom control(18). Evidence for

efficacy in promoting healing of gastric ulcers and maintenance therapy to prevent ulcer recurrence is somewhat equivocal(18,19). Acid suppression of gastric contents is a mainstay of GORD therapy.

In 1955, Nissen revisited the fundoplication as an operation for reflux oesophagitis. He performed the operation- with good results- in a series of patients with GOR symptoms but no hiatus hernia(4). Oesophageal manometry, developed in the late 1950's, further denuded the hiatal hernia paradigm of GORD. Manometry demonstrated unequivocally that the LOS, a zone of smooth muscle exerting sphincter pressure at the gastroesophageal junction, did indeed exist(10). Until this juncture, Cannon's barium findings demonstrating a LOS in cats had not been reproduced in humans. These manometry findings made it clear that this was a limitation of modality rather than proof of absence of the LOS. In a landmark review (1954), Ingelfinger wrote "the bulk of evidence favours the existence of a LOS"(20). Until the unequivocal demonstration of the lower oesophageal sphincter (LOS) in the 1950's, it was thought to be a purely anatomical mechanism(4). Code and Fyke(21) reported physiological pressure measurements at the gastroesophageal junction during swallowing and positioning manoeuvres. They suggested that dysfunction of the intrinsic LOS was the cause of reflux disease.

The LOS dysfunction paradigm faced some challenges. Dent et al(22) observed that only a minority of patients with GOR episodes on pH recordings had LOS incompetence on manometry. Therefore, LOS incompetence was not the full story(23). In healthy volunteers, GOR episodes were unrelated to low LOS pressure but correlated with transient relaxation of the LOS (TLOSR)(23). The neural mediation of reflux utilised the same reflex pathway as the swallow reflex. However, unlike swallowing, GOR can occur when relaxation of the LOS is not directly preceded by the swallowing reflex(24). A new paradigm emerged in which *inappropriate* TLOSR was recognized as the major contributory mechanism causing GORD.

The GOR paradigm thus shifted from a primary mechanical insufficiency of the diaphragmatic crura to a complex neuromuscular phenomenon. This was foreshadowed by Tilleston (1906) who wrote "normally the cardia is closed, owing to the tonic contraction of circular fibres. Relaxation takes place through inhibitory impulses from the vagus fibres, and occurs with the act of swallowing, and with pyrosis, belching and vomiting". It was not till the 1980s, that role of vagal nerve modulation in LOS tone and TLOSR was better understood.

In the 1990s, the development of multi-channel intraluminal pH-impedance(pH-MII) further challenged the established paradigm(25,26). The pH study is underpinned by the idea that gastric reflux is of fluids with pH <4. This concept was challenged by Attwood et al(27) who suggested that oesophagitis arose from injury to both acid and alkali pH-MII. This established the concept of non-acid gastric reflux. Use of pH-MII enabled the separation of measurement of bolus refluxate versus ph. Studies demonstrated that GOR was a phenomenon of both acid and non-acid refluxate. This raised key questions. Firstly, which was physiological – acid reflux, non-acid reflux both? Secondly, which kind of reflux led to symptoms? Complete answers to these two questions are yet to be found.

It was against this historical background that typical reflux syndrome was defined. GORD was defined in 1997 by the Second Canadian Consensus Conference on the Management of Patients with Gastroesophageal Reflux Disease(28) of as

“Demonstrable reflux, symptoms and/or mucosal damage”.

However, there continued to be patients with reflux symptoms and no micro or macroscopic changes on endoscopy. In adult patients, evidence of erosive oesophagitis is found in 2 -13% of patients with GOR symptoms(29,30). In 2006, a working group of 44 experts from 18 countries published a series of statements defining GORD in adults(22). A Delphi process was used to develop consensus on these statements. In this document, GORD was defined as:

“...a condition that develops when the reflux of stomach contents causes troublesome symptoms and/or complications.”

Absent in this updated definition is the requirement for mucosal damage. This exclusion reflects the fact that reflux symptoms can manifest without confirmatory evidence of oesophagitis. Indeed, the International classification of disease (ICD-10) describes GORD as “a disorder marked by frequent or severe heartburn”. It is further classified:

K21: Gastroesophageal reflux disease with oesophagitis

K21.9 Gastroesophageal reflux disease without esophagitis

This wider definition also allows for inclusion of extra-oesophageal symptoms associated with reflux. Mucosal breaks are addressed in the Montreal consensus definition of reflux oesophagitis. Reflux esophagitis as the presence of endoscopically visible breaks in the oesophageal mucosa at or immediately above the GEJ

CURRENT UNDERSTANDING OF GORD

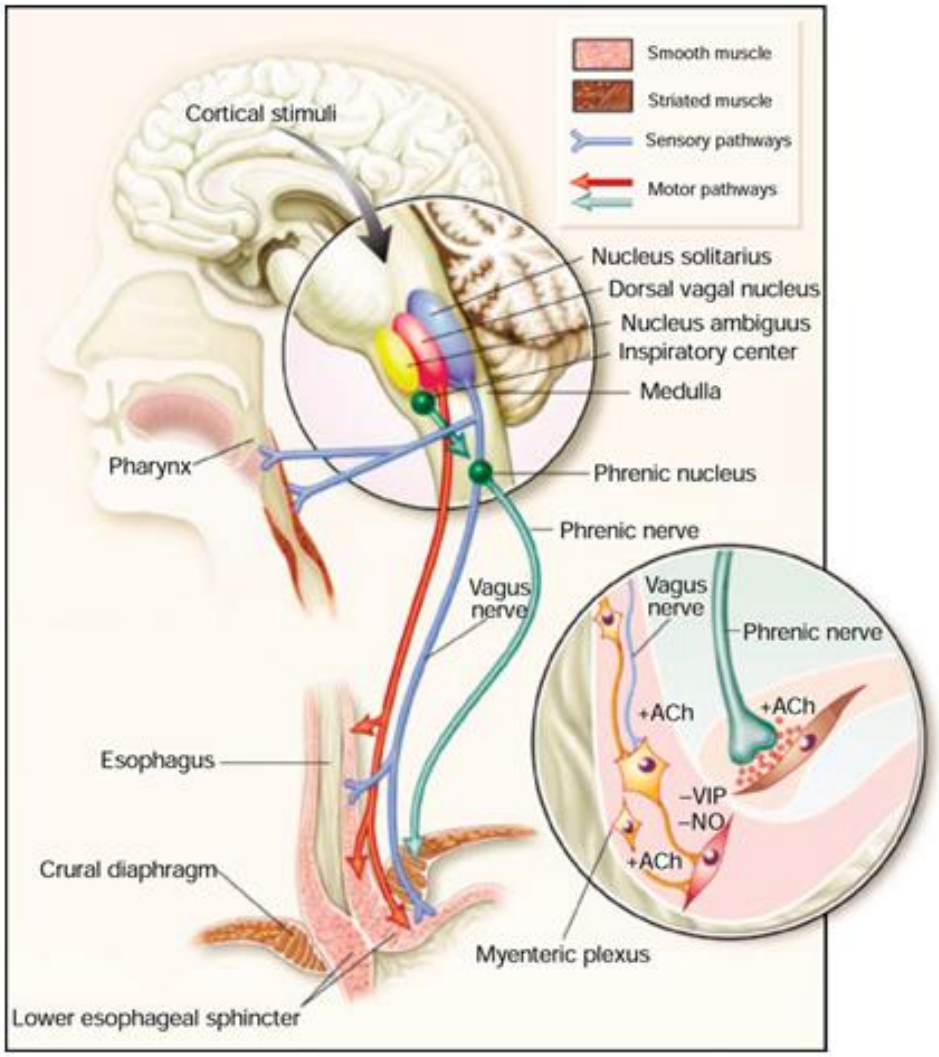
Refluxate may be of liquid (refluxate, regurgitate or vomit) or air (belching or burping). To pass into the stomach, a bolus must have sufficient pressure to overcome tone at the GOJ. Tone at this junction is maintained in several ways(31).

1. Acute angle of His
2. Extrinsic pressure arising from the diaphragmatic crura
3. Pressure gradient between intra-thoracic and intra-abdominal portions of the oesophagus
4. A high-pressure zone of smooth muscle i.e. the lower oesophageal sphincter (LOS)

Several mechanisms have been suggested to explain the phenomenon of GORD:

1. *Loss of extrinsic diaphragmatic crura pressure*. This has both a mechanical and a physiological element. Structural defects e.g. hiatus hernia, congenital diaphragmatic hernia lead to loss of the mechanical extrinsic pressure of the diaphragmatic crura. However, the role of diaphragmatic neuromodulation was increasingly understood by the end of the 20th century. TLOSR episodes coincide with selective inhibition of the diaphragmatic crura(24) via the phrenic nerve. The vagus nerve is the primary mediator of both swallowing and reflux. The afferent fibres deploy sensation from the oropharynx and oesophagus to the brainstem nucleus solitarius. The efferent fibres arise from the dorsal vagal nucleus. With acetylcholine as the neurotransmitter, the efferent vagal fibres mediate peristalsis of the oesophagus and contraction of the LOS. Relaxation of the oesophageal and LOS muscles is mediated by efferent vagal fibres releasing nitric oxide (NO) and vasoactive intestinal peptide (VIP)(32).

Figure 4: Neuromodulation at the gastroesophageal junction. Reproduced with permission from GI motility online (May 2006) doi:10.1038/gimo14



2. *Excessive intra-abdominal pressure* e.g. chronic coughing, obesity

Inappropriate relaxation of the lower oesophageal sphincter (LOS): The LOS relaxes transiently as a bolus is propagated from the proximal oesophagus to the distal oesophagus. This transient relaxation of the lower oesophageal sphincter (TRLOS) is a physiological terminal part of the swallowing reflex(33). The LOS also relaxes to allow belching. The synaptic mediation of LOS relaxation is not fully understood. However, atropine has been demonstrated to decrease the frequency of TRLOS(34). Novel pharmacokinetic therapies e.g. GABA-B agonists and mGluR5 antagonists are under development to modulate GOR by modulating the tone at the LOS (28). Inappropriate relaxation occurs extraneous to a swallow or belch event. Oesophageal manometry, has demonstrated that inappropriate TRLOS is more frequent in patients with GOR (27).

In summary, the GORD paradigm has shifted from insufficiency of extrinsic diaphragmatic muscles, to hiatus hernia, to insufficiency of the LOS, to inappropriate TLOSR. Despite these shifts of paradigm, fundoplication has remained the mainstay of anti-reflux surgery since the late 1950s. It has also been adopted as the operation of choice for paediatric GORD despite the obvious differences in disease paradigm, which will be discussed next.

CHAPTER 2: PAEDIATRIC GORD

Compared to GORD in adults, paediatric GORD has some peculiarities that further challenge the disease paradigm. Typical reflux syndrome, as found in adults, is characterized by heartburn with or without regurgitation. However, “typical reflux syndrome” cannot be diagnosed in infants and younger children who lack the cognitive ability to reliably report symptoms”(35).

In infants and children, particularly those with neurological impairment, extra-oesophageal symptoms may encompass non-specific presentations e.g. crying and arching. In fact, the ICD-10 K21 code for GORD specifically excludes GORD in the newborn. A separate entity, newborn GORD (P78.83), is defined in the ICD as: “neonatal gastroesophageal reflux applicable to newborns/ age 0 years.” It is worthwhile noting that this circuitous definition does not define what the actual diagnostic standard is.

The adult paradigm has been based on controlling oesophageal acid exposure. However, pH-impedance has revealed the high prevalence of non-acid reflux in children (26,36). This revelation further challenges the adult-based paradigm as applies to children. Another peculiarity is mucosal sequelae. Cumulative mucosal damage, uncommon in adults, is even less common in the paediatric population, where oesophagitis is found in 1.2-4% of symptomatic patients(37,38).

Recognising that an adult paradigm has little utility for paediatric patients with GORD, and that the Montreal GORD consensus takes no account characteristics of GORD as it affects children, a paediatric group was convened. Using a Delphi process inspired by the Montreal consensus, a group of 8 paediatric gastroenterologists developed statements defining and describing GORD in children(22,35).

“GORD is present is present when reflux of gastric contents is the cause of troublesome symptoms and / or complications(22)’(35).”

This definition then creates a diagnostic burden. Firstly, the diagnosis depends on demonstrating an association between reflux episodes and symptoms. Secondly, causality is introduced. This definition relies on demonstrating both the strength of association and the direction of causality between reflux episodes and symptoms.

EPIDEMIOLOGY

The incidence of GORD varies with age (39). Daily regurgitation is present in 40-60% of infants and considered normal. By the age of one year, less than 5% of infants regurgitate daily. In a questionnaire survey of parents, Hegar et al(40) found that 73% of Indonesian infants had daily regurgitation at the age of 1 month, whereas the age of a year, only 4% of infants had daily regurgitation. Van Howe and Storms conducted a prospective cohort study of 128 infants born consecutively(41)(9). Symptoms were assessed with the validated and revised infant gastro-oesophageal reflux questionnaire (I-GERQ-R)(42) . They found that symptoms e.g. regurgitation, crying and fussiness, decreased in frequency over 6 months.

Only 2% of 2-year-olds have regurgitation, and only 2% of 10- to 17-year-olds require treatment for heartburn with anti-acid medications(39). By the age of 17 years, 5% of patients report heartburn, and 1.4% report having regurgitation(39,43). In adults, daily heartburn is reported in 5–10% of respondents.

A key to understanding paediatric GORD is the concept of physiological versus pathological reflux(22). Reflux symptoms in a newborn e.g. posseting, are considered physiological. When the same is observed in an older child, it is considered pathological. Therefore, defining the boundary between physiological and pathological GOR is a key challenge for paediatric practitioners.

GORD may be more prevalent in Western countries. A recent systematic review of the epidemiology of GORD identified the prevalence of heartburn as 2.5% in China and 19.8% in the USA(44). A cross-sectional study of the French population identified a GOR prevalence of 10.3%(39). Systematic comparisons of geographical prevalence data in paediatric patients are not available. Passive smoking is an environmental factor that has been associated with childhood GORD(45).

Childhood obesity is also implicated as a risk factor(46,47). GORD also has a higher reported prevalence in children with neurological impairment and congenital anomalies e.g. oesophageal atresia, tracheal anomalies and congenital diaphragmatic hernia.

COMORBIDITIES ASSOCIATED WITH GORD

Studies have identified cohorts of patients who appear to have a higher incidence of GORD. In adults, key comorbidities are obesity and smoking. In children, these conditions include neurological impairment (NI), prematurity and congenital anomalies of the diaphragm, trachea and oesophagus.

Neurological impairment

In paediatric practice, NI is an umbrella term for a heterogeneous group of co-morbidities underpinned by congenital or acquired anomalies of neurological or cognitive function. Cerebral palsy comprises the largest proportion of this group(48). Cerebral palsy describes a group of disorders of muscle tone, strength and coordination secondary to peri-partum brain injury. Other co-morbidities characterized as NI include pervasive global developmental delay, Down's syndrome, congenital myopathy, myotonic dystrophies and inborn errors of metabolism(48,49). The prevalence of neurological impairment (NI) is around 0.2% of live births. The rate of NI in neonates appear to be increasing(50) and this may be due to increased survival of premature infants. NI is found in 25% of infants with extreme prematurity (born at or before 28 weeks' gestation)(51).

GORD is a complex problem in children with NI(52). It has a reported prevalence of 14-75%(49,53,54). Several explanations have been advanced for the higher prevalence of GORD in children with NI. Children with NI may have dysfunctional sucking and swallowing, and remain on milk feeds beyond the age of weaning(48,55). Children with NI may have a higher propensity towards inappropriate relaxation of the lower oesophageal sphincter, due to immaturity of the vagal enteric nervous system(48,55). Seizures cause spasticity, which in turn may increase abdominal pressure and cause episodes of aspiration. Seizures may cause a loss of tone, which may cause inappropriate relaxation of the lower oesophageal sphincter (LOS), thus permitting reflux episodes(52).

Children with NI and feeding difficulties may have a nasogastric or gastrostomy tube for feeding. It is unclear whether tube feeds are neutral or exacerbating to GORD(56–59). Fundoplication in NI children is often performed in the context of a pre-existing gastrostomy tube(60). When gastrostomy tube placement is indicated, a key question is whether a concomitant fundoplication should be performed(52,61).

It is important to note that the fundoplication procedure, developed for NN adults, was not developed in a context where there was a high incidence of tube feeding. This dilemma is peculiar to the paediatric fundoplication and hence the paediatric surgeon.

Fundoplication is more common in children with NI due to higher rates of failure of acid suppression medication(62). More children with NI receive operative intervention for GORD(2,63). Adverse outcomes(64,65), including mortality(66), occur more frequently in NI children(67). In a retrospective review of 7467 children undergoing fundoplication, Fonkalsrud et al(63) found that 44% were described as having NI. Recurrence of symptoms occurred in16% of NI children compared to 5% in the neurologically normal children. In particular, retching and bloating post fundoplication are reported to be higher in NI children. This has led some to suggest variations in procedure e.g. the Rossetti modification of the Nissen fundoplication(68), and partial anterior fundoplication(69). Failure of fundoplication is higher in NI children(67,70). Pearl et al(2) found that the re-operation rate for fundoplication was 19% compared to 5% in neurologically normal (NN) children. A systematic review of laparoscopic fundoplication reported rates of re-operation in NI at 15% versus 7% in NI versus NN children(71).

Children with NI represent a large proportion of patients with GOR. They have a higher risk of operative management and a higher burden of morbidity. To understand the relationship between NI and GORD, a finer characterization of NI is required.

Oesophageal atresia

In patients with a history of oesophageal atresia with or without tracheoesophageal atresia (OA±TOF) the reported prevalence of GORD ranges from 40-80%(72). Rates of fundoplication are higher in this subgroup with 40% of patients with GOR needing an fundoplication(73) post OATOF repair. Long-term follow-up studies have demonstrated increased risk of gastric (Barrett's oesophagus(74,75)) and intestinal metaplasia(76) which in turn increase risk of oesophageal adenocarcinoma. Patients with OATOF form a significant proportion of patients receiving fundoplication. Koivusalo and Pakarinen(77) reported that 22% of 217 patients undergoing primary fundoplication had OATOF. Notably, they also reported a 15% failure of fundoplication that was predicted by history of OATOF(77).

Anomalous oesophageal development, leading to oesophageal atresia, may also responsible for functional dysmotility of the oesophagus(78). Recently, impedance manometry has been used to characterise GORD in this subgroup(79). Motility studies in patients with OA±TOF have also demonstrated lower LOS pressures, longer transition zones (between pharyngeal and vagal innervation of the oesophagus) and dysmotility across the transition zone(78). GORD may also be secondary to congenital foreshortening of the oesophagus(80). Mobilisation and anastomosis under tension during surgery may also draw the distal oesophagus cranially, further into the chest(80). This is particularly true of patients with long-gap OATOF. Gastric dysmotility has also been posited as an underlying mechanism(81).

Following OATOF repair the risk of respiratory morbidity is high. Chetcuti et al(82) found that 44% of patients post OATOF repair were subsequently hospitalized for respiratory illness. Both pulmonary hypoplasia and GORD have been suggested as underlying mechanisms for this phenomenon.

Peetsold et al(83) followed up 31 adolescents with a history of neonatal repair of OATOF and subsequent GORD. Eleven patients had a history of fundoplication and 20 patients did not. Their lung function tests were compared to 13 adolescent controls with a history of GORD only. Compared to non-OATOF GORD, patients with a history of OATOF repair had significantly lower forced expiratory volume (FEV1) and total lung capacity. This suggests an element of restriction in lung capacity not present in adolescents with non-OATOF GORD. The authors suggest that possible causes include pulmonary hypoplasia, pleural scarring from neonatal thoracotomy and scoliosis associated with thoracotomy(83).

OATOF is also associated with a higher incidence of acute life-threatening events (ALTE)(80). The pathophysiology is not clear but both tracheomalacia and GORD may be indicated. Tracheomalacia is treated surgically with aortopexy using a thoracic approach. Some surgeons propose aortopexy prior to fundoplication(84), whilst others advocate fundoplication alone(85). Indeed, some surgeons advocate both aortopexy and fundoplication in this subgroup(86).

Congenital diaphragmatic hernia

Neonates with congenital diaphragmatic hernia (CDH) are born with a diaphragmatic aperture (usually left-sided) secondary to anomalous development of the lung and the diaphragm. This developmental anomaly of the diaphragm is secondary to pulmonary hypoplasia. The reported incidence of GORD in infants with CDH ranges from 30-60%(87–89). Medical therapy for GORD in CDH is required for 24-60% of infants(90). Adult survivors have high rates of oesophagitis and Barratt's oesophagus(87). For these reasons, fundoplication rates are higher in this subgroup, ranging from 10 -15% of CDH cohorts(87) .

Prophylactic fundoplication has been advocated for this subgroup. Reasons cited include the high rates of GORD and obviated risk of future laparotomy(91). The propensity towards GORD may be the lack of LOS reinforcement by a normally-formed diaphragm.

Prematurity

NI overlaps with prematurity: infants surviving pre-term birth comprise 25% of children with NI(51). GORD is more prevalent and persistent in infants with a history of pre-term birth. More preterm infants require operative management. Prematurity is a risk factor for failure of operative management requiring revision of fundoplication(92).

Suggested mechanisms that predispose premature infants to GORD include delayed gastric emptying(93). Omari et al(94) demonstrated normal gastric emptying times in preterm infants with GORD symptoms. This is indeed the rationale for the use of prokinetics in infants with GORD. Mechanical ventilation(95), large volume milk feeds(78) and apnoea(96,97) are other posited risk factors.

Chronic lung disease

Prematurity also overlaps with chronic lung disease (CLD)(98). CLD (also known as bronchopulmonary dysplasia) describes the requirement for respiratory support beyond 28 days of life. Incidence is higher in premature and low birthweight infants. GORD is considered a risk factor in chronic respiratory illnesses. pH and pH-MII studies are pathological in patients with chronic cough and asthma. A symptom correlation has been identified in children with cough and reflux episodes on pH-MII(99). Both

acid(100) and non-acid reflux(101) are implicated. Following fundoplication, chronic cough is a risk factor for recurrence of GORD(102). Conversely, there are reports of improvement of respiratory symptoms following fundoplication in preterm neonates with CLD(103).

The direction is not clear on objective measures. Blondeau et al(104) used impedance manometry used to objectively record cough and reflux episodes in 26 children with chronic unexplained cough. They identified that both acid and non-acid reflux preceded coughing episodes. However, the reverse was not demonstrated i.e. cough did not precede reflux.

A recent Cochrane review by Chang et al(105) of chronic cough and GOR identified 5 prospective cohort studies and 1 randomised control trial (RCT) assessing treatment of cough with improvement of GOR symptoms as the outcomes. The RCT did not identify any benefit to treating couch symptoms with a PPI. Conversely, there were increased adverse events in infants treated with a proton pump inhibitor, compared to controls.

Asthma

In older children, the relationship between GORD and asthma has been remains unclear. In a recent systematic review, Thakkar et al(106) identified 5 studies comparing prevalence of GORD in patients with asthma against controls. Pooled data suggests that the prevalence of GORD was 22% in patients with asthma and 4.8% in the control groups. The strength of association between symptoms of asthma and episodes of GOR is also unclear. In a systematic review of the association between respiratory symptoms and GOR, Tolia and Vandenplas(107) found that the cumulative risk of GOR in asthmatic children was nearly six times that of normal controls. However, Condino et al(108) using pH/impedance, found no relationship between asthma symptoms and reflux episodes.

Cystic fibrosis

Cystic fibrosis is a genetic condition characterised by viscid secretions of exocrine organs. This leads to mucous build up in affected organs particularly the lung, pancreas and intestines. Pulmonary CF leads to chronic infection which, where lung transplantation is not possible, leads to early mortality. A key priority in the management of cystic fibrosis (CF) management is pulmonary hygiene. Hence great focus is placed on preventing the deleterious effects of reflux and aspiration on the lung. The incidence of GORD in children with CF is reported to be 35- 81%(37). In published fundoplication series, the proportion of children with CF is as high as 8%(37). The direction of causality is not clear, i.e. whether pulmonary dysfunction leads to GORD or whether GORD exacerbates pulmonary dysfunction and no randomised controlled data are available. CF has characteristics that would dispose to GORD e.g. chronic cough, expanded lung volumes. GORD appears to be risk factor for severity of CF i.e. pulmonary function deterioration, respiratory exacerbations. GORD may contribute to infection by lowering mucosal barrier integrity and altering the pulmonary microbiome. Recent studies have identified a positive correlation between non-acid reflux episodes and pulmonary *Pseudomonas aeruginosa* colonisation(109). Studies suggest that >50% of CF patients are treated with PPIs(110). Even after lung transplantation, GORD appears to persist(111). Therefore, testing and treating GORD prior to lung transplantation is a priority. Fundoplication is reported to reduce use of anti-reflux medication, improve nutritional indices(112) and pulmonary function tests in children with CF(113).

Complication rates are similar to those for children without CF(114). However, some studies have found no improvement in nutrition or pulmonary status, with a high incidence (48%) of recurrence of GERD(114). RCT trial data is required to further define patient selection, benefits and risks of fundoplication in children with CF.

Swallowing dysfunction

Swallowing dysfunction in infants and children is often overlaps with NI, chronic lung disease and prematurity(98) (115). The direction of effect is not clear. Does GORD exacerbate swallowing dysfunction? Or does swallowing dysfunction exacerbate GORD? Small series demonstrate improvements in swallowing function metrics after medical(116) and surgical treatment(116) of GORD. Both NI and GORD have been identified as risk factors for development of bronchiectasis in patients with underlying swallowing dysfunction(115).

Tracheal anomalies

Trachea anomalies may increase the risk of GORD. The Children's hospital of Philadelphia identified GORD in 23% of 56 children with airway anomalies(117). GORD is considered a predisposing factor in patients with acquired tracheal stenosis(118,119).

Dental erosions

GORD is a risk factor for dental erosions. Children with developmental delay appear to have an increased risk(120). This risk may be related to acid exposure. RCT support a protective effect of PPIs on dental erosions(121). A pH-MII study in a small group found a positive correlation between reflux index and degree of dental erosion(122).

Musculoskeletal anomalies

There is an overlap of patients with NI, neuromuscular disease and musculoskeletal anomalies. Key conditions in this encapsulated in this term include:

- neuromuscular orthopaedics e.g. myotonic dystrophy, myasthenia gravis, Duchenne's muscular dystrophy. spinal deformity e.g. scoliosis
- developmental dysplasia of the hip (DDH)
- osteogenesis imperfecta

There is a crossover in coverage of patients with skeletal anomalies. For example, a patient with muscular dystrophy may be known to both the orthopaedic, neurosurgery and neuromuscular specialists at an institution. However, with an isolated orthopaedic problem (e.g. DDH) is not at increased risk of GORD. However, acknowledging this overlap, analysis of a cohort of GORD patients should be inclusive of children with musculoskeletal anomalies at the risk of including children with isolated orthopaedic conditions.

CHAPTER 3: DIAGNOSIS OF GORD

SYMPTOMS

What are the key symptoms and signs of gastro-oesophageal reflux disease? The answer to this question depends on who is asked. In reviewing the literature, it becomes clear that cohort sampling has resulted in an unclear picture of the symptomatology of GOR. For example, a survey of infants feeding at home reported regurgitation and crying as the most prevalent symptoms (30). In series reporting GORD in children with neurological impairment, key symptoms are feeding difficulties and failure to thrive (18, 24). A review of patients with respiratory symptoms and GOR will link chronic cough as a key symptom. In reviewing surgical literature, key symptoms are apparent life-threatening events (ALTE) and recurrent chest infection (29) (Figure 2).

Equally, there are reported indices of GOR, where the patient or care-giver reports symptoms. These symptoms are then attributed by the patient, care-giver or clinician to GOR. Systematic methods of collecting symptoms and signs in patients with GOR have been developed. Well-validated questionnaires e.g. Infant GERD score, GERD-9, 17 have been developed to enable scoring by severity and, hence, risk stratification. However, questionnaires in younger children are reliant on a history from the care-giver. No questionnaire has been validated for use in patients with GOR less than 9 years old. Furthermore, a questionnaire accounting for important co-morbidities in a quantitative way e.g. neurological impairment has not been developed. Therefore, at present, the use of these tools is limited.

As stated succinctly by Hassall(123):

"infants have a limited of behaviours in response to pathological processes, among the most common being irritability, arching, unexplained crying, apparent discomfort."

Therefore, assessing GORD using questionnaire instruments e.g. Infant-GORQ is handicapped by this pitfall.

Vomiting, regurgitation and GORD

GER is classically defined as effortless regurgitation of gastric contents into the oesophagus(124). However, many authors use vomiting and GORD interchangeably. The vomiting child is often considered to have GORD.

Richards et al(125) present a succinct differentiation between reflux and emesis. When observed, gastroesophageal efflux appears to be effortless vomiting or regurgitation. However, when the emetic reflex is activated, other sympathetic signs can be observed. These are pallor, sweating, retching and forceful vomiting(125).

Figure 5: A summary of common indications for fundoplication. Reproduced with permission from: Pacilli et al (2005). The surgical treatment of gastroesophageal reflux in neonates and infants. Seminars in Paediatric Surgery (2005)14:34-41(29)

Table 1 Review of the literature: symptoms and indications for fundoplication in selected series of neonates and infants with gastro-esophageal reflux

Study	Number of patients (age group)	Symptoms and indications (%)					
		Apnea and/or bradycardia and/or ALTEs	Aspiration and/or pneumonia	BPD and/or RDS	Failure to thrive	Severe emesis	Stricture or esophagitis
Randolph, 1983[50]	72 (≤1 yr)	11	33	-	49	-	7
St Cyr et al, 1986[23]	45 (≤6 mo)	17	44	16	20	2	-
Hrabovsky and Mullett, 1986[75]	17 (prem infants)	29	35	82	-	6	-
Giuffre et al, 1987[74]	9 (prem infants)	-	100	100	100	-	-
St Cyr et al, 1989[81]	51 (≤2 yr)	18	55	27	-	-	-
Justo and Gray, 1991[77]	11 (prem infants)	27	82	54	-	-	-
Kazerooni et al, 1994[20]	160 (<2 yr)	30	53	-	68	58	8
Krishnamoorthy et al, 1994[73]	39 (LBW infants)	64	31	31		23*	-
Thompson et al, 1996[55]	25 (<1 yr)	-	8	-	92	-	-
Rowe et al, 1995[78]	21 (prem infants)	58	19	-	14	-	-
Zamir et al, 1997[58]	11 (<2 yr)	9	82	-	36	27	-
Kubiak et al, 1999[82]	66 (<4 mo)	24	29	-	52	39	11
Fonkalsrud et al, 1999[85]	110 (<3 mo)	52	13	44	37	56	-
Somme et al, 2002[57]	53 (<1 yr)	21	-	-	96	-	-
Barnes et al, 2003[76]	10 (8 prem infants, 2 term infants)	100	100	60	100	100	-

Abbreviations: ALTEs, apparent life threatening events; BPD, bronchopolmunary dysplasia; LBW, low birth weight; prem, premature; RDS, respiratory distress syndrome.
*23% of the patients had failure to thrive or severe emesis.

These variations reflect differences in sampling by different investigators. A prospective, population-based cohort study with a long follow-up period would be the soundest way to determine the true prevalence of symptoms of GOR.

Vomiting is the forceful expulsion of stomach contents into the mouth. By definition, vomiting results in retrograde bolus movement across the gastroesophageal junction. In contrast to GOR, however, vomiting may be triggered centrally, where a noxious stimulus acts on the chemoreceptor trigger zone ion the area postrema of the brain. Determining whether a patient is vomiting due to such a noxious stimulus e.g. concurrent gastro-intestinal infection, or due to inappropriate relaxation of the GOJ, is challenging. To further characterise the vomiting associated with GOR, guidance is sought from the expert consensus statement on GOR. This document states that:

“Bilious vomiting should not be recognised as gastro-oesophageal reflux disease. “

Retching is the forceful increase of abdominal pressure against a closed glottis. It is the first step of a normal vomit. However, retching can be pathological when repetitive and not followed by vomiting. Retching is particularly prevalent in children with neurological impairment. It may present as part of cyclical vomiting syndrome, where a pattern of upper gastro-intestinal dysmotility is triggered by anomalous activity in the vomiting centre. Retching has been reported to be a symptom of GOR. It has also been reported as cause of GOR, as retching episodes raise intra-abdominal pressure (32). As there is a paucity of data on the aetiology of this phenomenon, no conclusions can be drawn on its association with GOR.

Regurgitation is the effortless return of stomach contents into the oesophagus and oropharynx(126). It is physiological in newborns and infants, with 40-60% of infants having daily episodes of regurgitation. The natural history of regurgitation is resolution with age (8, 11, and 31). These data suggest that there is a threshold to be found between physiological regurgitation and pathological regurgitation.

Apparent life-threatening events

Apparent life-threatening events (ALTE) are defines as episodes of respiratory and or cardiac instability resulting in near death. These events are rare (0.0026% of live births) but potentially catastrophic, with 38% of all ALTEs resulting in death (33). The natural history of ALTEs is resolution. However, ALTEs result in death in 0.001% of live births (33). Therefore, the mortality rate amongst infants with ALTE is ~40%- this relative risk is great indeed. Predicting and eliminating risk factors for ALTE is therefore a priority.

The association between ALTEs and GOR remains unclear. Several authors list ALTEs as an indication for fundoplication. However, in a systematic review of published reports of extra-oesophageal symptoms of GOR, Tolia and Vandenplas (31) found little evidence for a causal relationship between ALTE and episodes of GOR. Using impedance, Di Fiori (2010) et al addressed the temporal and causal relationship between ALTEs and episodes of GOR. This study demonstrated a temporal relationship between ALTE and GOR for only one third of all ALTE events. Whilst GOR preceded ALTE in some cases (3%), more ALTEs preceded GOR (9%).

Lopez-Alonso et al.(96) failed to identify a temporal association between episodes of GOR and episodes of ALTE (34-36). Peters et al. (2002) found no association between apnoea and episodes of GOR in 21 premature infants. Mousa et al.(127) reviewed apnoea in infants and found no relationship between apnoea and GOR (37). Similarly, Wenzl et al. (97) investigated a cohort of 22 infants and found only 30% of reflux episodes of reflux were linked to 5-10 second apnoea episodes.

Although ALTEs are a dangerous phenomenon with potentially catastrophic consequences, the assumption of a link between GOR and ALTE is not safe. The use of ALTEs as a marker of severity of GOR and an indication for fundoplication is, consequently, also unsafe.

In contrast, Monasterio et al(128) presented a series of 18 patients with Pierre Robin sequence. This condition is characterized by micrognathia and glossoptosis with associated swallowing difficulties and upper respiratory obstruction. Oesophageal pH measurements of GORD appeared to resolve along with metrics of apnoea severity after treatment of micrognathia with distraction osteogenesis. Although anecdotal, the resolution of GORD with treatment of the primary airway problem supports the thesis that, in some cases, apnoea may trigger reflux episodes.

Wenzl et al (97) assessed 22 infants with respiratory symptoms and recurrent regurgitation. Mean gestational age was 38±3 weeks. The age at investigation ranged from 30 to 127 days. As the aim of the study was apnoea detection, all infants were admitted for overnight study in a sleep laboratory. The mean time in apnoea associated with reflux was significantly greater than mean time in apnoea without reflux. GER episodes lasting more than 30 seconds were more likely to be associated with apnoea episodes compared to shorter ones. The authors suggested a temporal association between apnoea episodes and reflux episodes.

This study is marred by definitions. In the study by Wenzl et al (97), apnoea was defined as breathing arrest of 5s. In contrast, apnoea is defined by the American Academy of Pediatrics (129)as follows:

"*an unexplained episode of cessation of breathing for 20 seconds or longer, or a shorter respiratory pause associated with bradycardia, cyanosis, pallor, and/or marked hypotonia*."

This is the accepted definition in clinical practice. In this study, 96% of apnoea episodes lasted less than 10 seconds and the longest apnoea episode lasted 12 seconds. The study does not detail whether there were associated symptoms of bradycardia, cyanosis or hypotonia. Therefore, based on this study, it cannot be concluded that there is a relationship between apnoea (as defined clinically) and reflux episodes. A larger study with more standardised definitions would provide useful evidence.

Magista et al(130) reviewed impedance studies of 6 premature infants (median gestation 31 weeks (27-36 weeks). They found that apnoea episodes were more frequently associated with reflux than without reflux. An important caveat must be stated. Apnoea of prematurity is a well-recognized phenomenon. They are often central and related to immaturity of central nervous system respiratory control. Reflux events are physiological in infants and probably more so in premature infants. The coincidence of these to conditions of prematurity does not indicate causality.

In contrast to the findings above, Mousa et al (127) investigated apnoea in infants and found no relationship between apnoea and reflux episodes. Indeed, only 15% of apnoea episodes in this study were temporally associated with reflux. They also did not identify a proportional difference in acid versus non-acid reflux association with apnoea.

Aspiration and chest infection

Recurrent chest infection secondary to suspected aspiration is an often-quoted indication for fundoplication. In a review of their practice over 8 years, Pimpalwar and Najmaldin (39) reported recurrent chest infection to be an indication for surgery in 81% of patients. Aspiration is thought to be the mechanism linking reflux and recurrent chest infections.

Aspiration is the inhalation of foreign material into the airways. Aspiration occurs when the protective mechanisms of the vocal cords that keep tracheal and oesophageal contents separate are overcome. Aspirate can arise from oropharyngeal secretions. This is colloquially known as 'over the top' aspiration. Aspirate can also be of gastric content, a.k.a. 'down-under' aspiration. GORD is the mechanism thought to be responsible for gastric aspiration(126).

Aspiration can lead to pneumonitis (chemical inflammation), particularly if the aspirate is of low pH or contains particulate matter e.g. food. Aspiration can also lead to pneumonia i.e. an infective process. The clinical distinction between the two can be difficult and the processes can coincide. However, pneumonitis tends to present more acutely than pneumonia. Long term, both processes can lead to pulmonary fibrosis(126). Aspiration can be silent and unwitnessed. Therefore, where suspected, objective tests for aspiration can be of use e.g. video fluoroscopy. However, given the invasive nature of these tests, management of often based on clinical suspicion.

The causes of aspiration are impaired gag reflex and GI motility disorders e.g. oesophageal dysmotility, GORD, gastroparesis, bowel obstruction. Aspiration is a particular concern for children with NI where there is impairment of gag reflexes. Aspiration can also occur when consciousness is depressed e.g. apnoea, seizures, general anaesthesia.

INVESTIGATIONS

Several techniques have emerged in GORD diagnostics.

1. pH study
2. pH/impedance study (pH-MII)
3. Upper gastro-intestinal contrast study (UGIC)
4. Upper gastrointestinal endoscopy
5. Oesophageal manometry

In addition, related to GORD, methods to diagnose aspiration and measure gastric emptying have been developed and are briefly discussed.

24-hour ambulatory pH study

During a pH study, an antimony probe is placed at the distal oesophageal junction (DOJ). A small current is passed through the antimony probe. The conductive properties of antimony vary with the pH of the solution in which it is placed. As the pH at the DOJ changes, so does the current. The resulting trace of these changes is the essence of a pH study.

As discussed earlier, Tuttle and Grossman first describe oesophageal pH measurement in 1958. Johnson and Demeester(15) are credited for developing prolonged pH testing protocol. The pH study was developed for use in adults with GORD at a time when the disease paradigm was of oesophageal acid exposure. In paediatric patients the use of the pH study was described by Sondheimer et al in 1980(131). On the pH scale, the threshold between acid and alkali, the neutral point, is pH 7.

The Johnson and Demeester protocol defined a reflux episode as any drop as <pH 4. This diagnostic threshold was based on the observation that the pepsin-producing crypt cells in the stomach are activated below a pH of 4. The rationale was therefore, that gastric pH, at least in the post-prandial period, was less than 4. Solutions with a pH >4 in the GOJ were not considered gastric in origin and, therefore, not reflux. The reflux index (the percentage of time that the pH falls below 4) is a standardised metric of GORD severity. In adult patients for example, a reflux index >6% when supine is considered pathological.

The use of pH-metry in adults did not translate smoothly into paediatric practice. An illustration of this is the quest for normative data. Normative data for paediatric patients has been developed. Due to the invasive nature of the test, early studies providing were of asymptomatic older children and often limited by the number of subjects investigated e.g. Boix-Ochoia et al (1978) (132) (n=20). The largest contributory study was by Vandenplas *et al*(133) who studied 509 healthy infants. These infants were screened as part of a risk assessment for sudden infant death syndrome. The cut-off for pathological reflux was defined as reflux index above the 95th centile for this population. This corresponded to a reflux index of 10%. In contrast to adults, reflux index appeared to decreased with age; the normal reflux index was 13% at birth, 8% at 1 year of age(134).

Table 1: Normative data for pH studies in paediatric patients.

Patients	Reflux index (%)	Reflux episodes/ 24hrs.	Author	Sample size
Infants	10	72	Vandenplas et al (1991)(134) (n=509)	509
6m – 6 yrs.	6	-	Boix-Ochoia et al (1978) (132) (n=20)	20
0-6 m	1.7	10	Cucchiara et al(135) (52) (n=63 controls)	63
6-24 m	1.2	12		
>24 months	1.3	10		

It is clear from these data that reflux index varies within and between studies. The North American Society for Pediatric Gastroenterology, Hepatology and Nutrition (NASPGHAN) practice guideline (136) on pH-metry describes the following normal values.

Table 2: Normal values of paediatric pH study as described by NASPGHAN.

Reflux index	Recommendation
<3%	Normal
>7%	Pathological

By the committee's own admission *'normal values should be regarded as suspect'*. There are several reasons for this. Firstly, early normative data were established using glass diodes(136) which correlate poorly to antimony probes used more latterly and cited most in literature. Secondly, 'normal' test populations varied from asymptomatic infants (137) who did not vomit or regurgitate (134). The committee(136) offers a guideline, rather than a diagnostic threshold.

Another area where pH-metry fails to translate into paediatric practice is risk stratification. Johnson and Demeester(15) performed pH tests in 50 asymptomatic adult volunteers and had created a scoring system for GORD severity. Six pH-study parameters were selected to contribute to the score. They are

1. Number of reflux episodes
2. Total time the pH is < 4 or > 7
3. Upright time in reflux
4. Supine time in reflux
5. Number of reflux episodes > 5 minutes
6. Duration of longest reflux episode

Not all six parameters were always abnormal. It was clear that some were more important than others. Johnson and Demeester used standard deviation of the mean of these six parameters to weight individual factors. A cumulative score was achieved by adding the individual parameter values. The 95th centile for the cumulative score marked the border between normal and abnormal.

Although the score is well validated, it has some internal limitations. Standardisation of the 6 parameters is based on the assumption of normal distribution. However, Jalal et al(138) assessed the Kurtosis and skewness of the six parameters in 45 asymptomatic adult patients (mean age 66 ±5.6 years). They found that the 6 components used in the JD score were not normally distributed.

The JD score is therefore limited by the underlying assumption of normality of key parameters. A scoring system that weights individual parameters in a non-parametric way would be an improvement. However, the sample size of 50 is too small to base such a scoring system upon. Despite these limitations, the score was modified for use in paediatric patients by Boix-Ochoa(132) in 1980. Modification was necessary because of observed differences in the 6 score parameters i.e.:

1. Children had a lower percentage time in reflux in the upright position
2. Children had a higher percentage time in reflux in the supine position
3. Children spent more time in the prone position
4. Children had fewer reflux episodes in the monitoring period
5. Children had longer and more frequent reflux episodes over 5 minutes.

Normal values for the Boix-Ochoa score were based on the investigation of 20 asymptomatic children and 103 symptomatic children. As with the Johnson and Demeester score, parameter values were standardised using standard deviation and assuming normal distribution of the parameter. A composite score was obtained by adding the individual's scores for key parameters. As with the Johnson and Demeester score, the threshold for composite score normality was set at the 95th Centile for the asymptomatic group (n=20).

The Boix-Score is therefore limited by the same factors as the Johnson Demeester Score. This score has limited application in risk stratification of paediatric GORD. Currently, by convention(136), reflux indices of <5%, 5-10% and >10% respectively have been termed mild, moderate and severe GORD. However, there are no symptomatic or histological(135) correlates for this convention.

The height of the reflux column was also considered to be a metric of GORD severity i.e. the higher the reflux column, the more severe the GORD. This led to the measurement of pH in the proximal and distal oesophagus(139). It was therefore a natural transition to measurement of pH along the whole oesophagus, as is done with pH-MII.

The pH continuum?

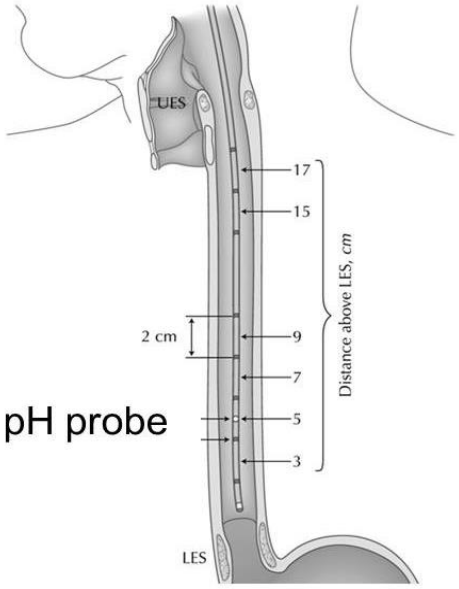
Perhaps the key flaw of pH testing is the fact of the pH continuum. The pH test is based on assuming that pH<4 refluxate is gastric in origin and pH>4 is not. This was based on the observation that there is minimal pepsin activity at pH 4 and none at 5(140). Further, symptoms in adults were observed at pH 4 or below, but not above(14). For the pH study, it is assumed that the normal pH of the oesophagus is >pH 4. Infact, the normal pH of the oesophagus is closer to pH7. Secondly, it is assumed that refluxate of pH 4-7 is not pathological.

Combined pH and multi-channel intraluminal Impedance (pH-MII) monitoring of the oesophagus challenged these assumptions by enabling independent measurement of reflux episodes and pH of refluxate.

Multi-channel pH impedance measurement

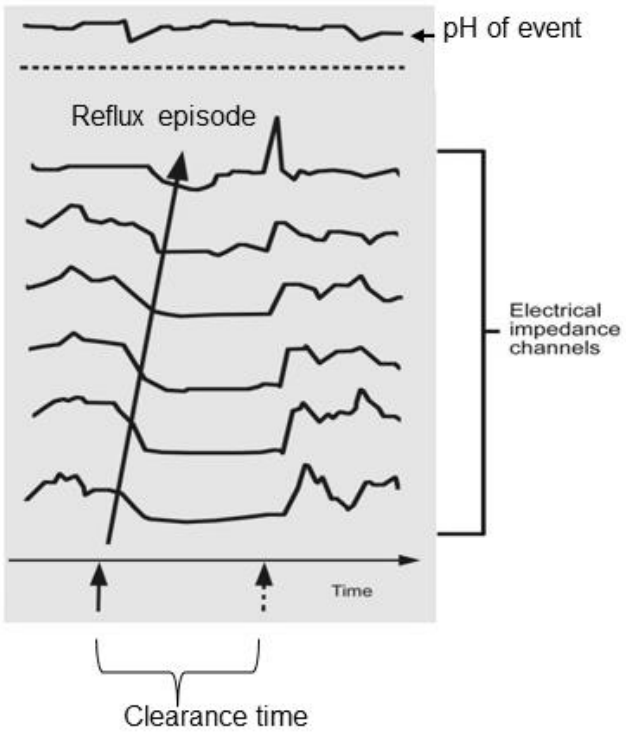
A catheter in introduced into the oesophagus. This catheter has several (usually 7-10) impedance rings which lie 2 cm apart along the oesophagus. The catheter also has an antimony pH sensor, which lies near the distal tip and is ideally positioned just above the gastro-oesophageal junction.

Figure 6: Schematic of combined pH and impedance oesophageal catheter



The catheter is connected to a power source and a piezoelectric current is passed through the impedance rings. Resistance across these rings can be measured, based on the impedance to current by whatever lies within the rings. Air has poor conductivity. Therefore, when current passes through it, resistance is high. Liquid has good conductivity. Therefore, when current passes through it, the resistance is low. These changes in current are visually represented on a trace of current against time. The visual representation on the trace will be a decrease in resistance.

Figure 7: An array of impedance rings and a pH sensor within the oesophagus.



Measurable impedance parameters are reflux, bolus clearance time, bolus velocity and height of reflux bolus. Matched with pH information yields the further metrics i.e. acid reflux, non-acid reflux, acid clearance time.

As with any new technology, standardised diagnostic criteria must be established. In November 2002, a group of experts met in Oporto , Portugal, to developed diagnostic guidelines for pH-MII(141). The group of 11 experts included specialists working in paediatrics. The resulting guidelines were known as the Porto consensus.

Below is an illustrative sample of the consensus rules that emerged.

1. Bolus reflux: a retrograde drop in resistance ≥50% of the baseline in at least two distal impedance channels.

2. Acid reflux: reflux episode associated with a drop in pH to <4, or a drop in impedance during a period when oesophageal pH was already <4.0.

3. Non-acid reflux: a retrograde drop in impedance to >50% of the baseline impedance value in at least two most-distal channels. PH remains above 4.

The consensus was summarised by this statement:

"Reflux is best detected by impedance and its acidity characterised by pH-metry(141)*".*

The 2009 consensus on paediatric GORD by the North American and European societies for gastroenterology (NASPHGAN and ESPGHAN) says of pH-MII(35):

"This test detects acid, weakly acid and non-acid reflux episodes. It is superior to pH monitoring alone for evaluation of the temporal relation between symptoms and GER."

However, the consensus statement also focuses on the utility of severity scales with prognostic value. The statement continues:

"Whether combined oesophageal pH and impedance monitoring will provide useful measurements that vary directly with disease severity, prognosis, and response to therapy in paediatric patients has yet to be determined."

Non-acid reflux

pH-MII has further transformed the paediatric GORD paradigm. A key finding was that non-acid reflux is present and at least as frequent as acid reflux in children. pH-metry alone was underdiagnosing the scale of this condition. Del Buono et al focused-on children with neurological impairment and found that non-acid reflux (NAR) comprised a full 50% of all reflux episodes (58). Lopez Alonso et al (96) focused on new-borns and found that NAR was twice as common as acid reflux (AR). Furthermore, this study reported that the pattern of NAR was more proximal, suggesting a higher risk of aspiration (35). Pilic et al (142) reviewed 700 children presenting with symptoms suggestive of GER, including 329 children with pulmonary symptoms, 325 with gastrointestinal symptoms, and 46 with neurologic symptoms. 45% of patients with GOR would not have been recognised on pH metry alone (57).

The acid-to-non-acid reflux ratio is not static: it changes with age and feeding pattern. In a group of 34 infants, Condino et al (108) found that 47% of the reflux episodes were acid, and 53% were non-acid. Furthermore, with increasing age and decreasing frequency of feeds, the proportion of acid reflux increased as post-prandial buffering effect of milk was lost.

Symptom / reflux association

Another key development was the possibility of estimating association between symptoms and acid/ non-acid reflux. Several metrics have been developed to associate symptoms with impedance reflux episodes. These are symptom index (SI), symptom sensitivity index (SSI) and symptom association probability (SAP).

The SI quantifies the proportion of symptom episodes concurrent with reflux. For a given symptom e.g. cough, SI is defined as:

$$Symptom\ index\ (SI) = \frac{Total\ reflux\ associated\ symptoms\ (RAS)}{Total\ symptoms\ in\ 24\ hours} \times 100\%$$

SI values of 50% are considered abnormal(143).

The SSI quantifies a patient's sensitivity to reflux episodes. SSI is defined as:

$$Symptom\ sensitivity\ index\ (SSI) = \frac{Total\ reflux\ associated\ symptoms\ (RAS)}{Total\ reflux\ events\ in\ 24\ hours} x100\%$$

SSI values of >10% are considered abnormal(144).

There are limitations to these parameters. SI over-emphasizes association when total symptoms are few. A patient with 2 RAS out of 3 symptoms will have a higher SI than a patient with 2 RAS out of 4 symptoms. Therefore, this parameter is highly sensitive but poorly specific(145). Regarding the SSI, a

patient with 4 RAS and 10 reflux events in 24 hours will have a higher SSI than a patient with 4 RAS and 25 reflux episodes. The SSI is therefore poorly sensitive as the numbers of reflux events increase.

Acknowledging these limitations, Weusten et al(146) developed the symptom association probability.

The symptoms and reflux events both temporarily associated and otherwise are documented in a contingency table. The probability (P) of observing this particular set of numbers and marginal totals follows a hypergeometric distribution. This is a probability distribution describing the probability of 'successful' outcomes from a defined number of draws. The draws are finite and made without replacement. Fisher's exact test is used to estimate the probability that the reflux event and symptom event are associated.

Figure 8: Contingency table for estimating symptom association probability

	Symptom (S+)	No Symptom (S-)	Total
Reflux (R+)	S+R+	S-R+	12
	2	10	
No Reflux (R-)	S+R-	S-R-	10
	4	6	
Total	6	16	22

In the example above, the probability of S+R+ can be estimated using Fisher's exact test: the two-tailed value of P is 0.35. The SAP is then calculated as:

$$SAP = (1 - P) \times 100\%$$

$$SAP = (1 - 0.35)x100\% = 65\%$$

SAP is said to be positive when >95%(146).

Some authors have attempted to apply these parameters in the study of paediatric subjects with GORD. Rosen et al (147) reviewed MII results of children aged 3 months to 18 years with unexplained respiratory symptoms resistant to acid suppression therapy. Although they found SI and SSI were considered normal(147), logistic regression was conducted to assess factors predictive for RAS. Younger age was associated with increased risk of RAS. MII parameters positively associated with RAS were non-acid and liquid reflux.

Antegrade bolus studies

In patients with GOR, retrograde bolus flow across the GOJ is the datum of interest. However, in a subset of patients with oesophageal motility disorders, information about antegrade bolus flow is also relevant. GOR affects 40 -60% of patients with previous oesophageal atresia. Di Pace et al(148) evaluated 30 children with suspected GOR. Fifteen of these children had been treated at birth for oesophageal atresia with no associated malformation. The other 15 had no congenital malformations, but had suspected reflux. In this limited study, they found dramatic differences in impedance parameters between the two groups. Patients with dysphagia (n=15) on average had significantly longer bolus and segment transit times. Bolus clearance was longer in this group. Although more studies are needed, it

is clear MII will help delineate the symptomatology of GOR against that of oesophageal dysmotility in complex patients.

Limitations of pH-MII

Risk stratification of severity using pH-MII

Historically, height of reflux column has been considered a severity measure. This metric is particularly simple to extract with pH-MII due to the intraluminal impedance rings arrayed in the oesophagus. The height of the reflux column can be estimated using the probe at the GOJ as a reference. Condino et al (108) found an association between symptoms fussiness/pain, arching and burping and the height of the reflux column in infants (2 weeks to 1 year) undergoing MII. Borelli et al (149) found that NAR episodes and the height of the column were positively correlated with a finding of LLMs in twenty-one children studied.

pH-MII reports present data on multiple parameters. There is a tension between pH parameters and MII parameters and it remains unclear which parameters best correlate to symptoms and severity of GORD. For example, Greifer et al(150) reviewed records of 63 children with extra-oesophageal signs of GORD. Of these, 6 patients met criteria for a pathological study based on Demeester score. There were 10 patients who met the criteria for pathological study based on impedance criteria, specifically reflux episodes. There were 7 patients with a SI. However, the mean DeMeester score for this subgroup was normal. Given these multiple criteria, which parameters carry greatest import? The attempt to identify pH-MII metrics that correlate with symptoms and severity highlights a key limitation of pH-MII i.e. absence of normative data.

Normative data for pH-MII

To develop a body of normative data, a standardised protocol of pH-MII is required. Initially, the recently updated(151) Porto consensus was adopted wholesale for paediatric practice(94,145,152). However, it was clear that, given multiple companies providing tools for pH-MII, various protocols had emerged. In an attempt to standardise, the European paediatric impedance working group (Euro-PIG) was convened(145). In 2012, a consensus statement was released (142) and some key definitions emerged. pH findings were categorised as acid (pH<4), weakly acid (pH 4-7) and non-acid (pH >7). A gas belch was defined as a sharp increase in resistance above 3000 Ohm. Controversially, a symptom/ reflux association window of ±2 minutes was set. By their own admission, this window was based on consensus rather than evidence(145).

As with pH-metry, no good data normative data exist for paediatric pH-MII testing. Equally, there are no good data correlating test parameters with severity of symptoms and/or mucosal damage. This dearth is due to the invasive nature of these investigations(94). The pH-MII is an invasive and oft-unpleasant test that is difficult to justify in non-consenting children, even if their parents were to agree(153). During the period of this PhD study, normal values were simply not available. Indeed, all the data available were studies in adults and preterm neonates.

Adult pH-MII data arose from a study by Shay et al(26) for normative data. This multicentre study included 60 healthy volunteers (median age 39 (22-62) years). The recommendation from this study was accepting the values greater than the 95th percentile as pathological. It is therefore from this study

that the critical threshold of 73 reflux episodes/24 hours emerged. It is important to appreciate that this adult threshold was adopted in interpretation of MII studies at our institution.

The results from the study(26) above can be summarized as follows:

Table 3: pH-MII findings in healthy adult volunteers adopted as normative data

Impedance parameter	Median (IQR)	95th percentile
% acid exposure	1.2 (0.3-2.5)	6.3
Total reflux episode in 24 hours	30 (18-45)	73
Acid reflux episode in 24 hours (pH≤4)	18 (7-31)	55
Weakly acid reflux episode in 24 hours (4≤pH≤7)	9 (6-15)	26
Non-acid reflux episode in 24 hours (pH ≥7)	0 (0, 0)	1

Other authors have used data from adult patients to interpret paediatric pH-MII results. For example, in 2010 Pilic et al (142) and the German Paediatric Impedance Group (G-PIG) guided by data from adult studies set a threshold of >70 episodes in 24 hours in patients aged ≥1 year. In patients aged less than a year, in recognition that pH study data suggested a higher frequency of physiological reflux, the threshold for reflux episodes was set at >100 episodes. As yet, these thresholds have not been applied and validated in larger groups of patients.

Some paediatric data has been published. Lopez-Alonso et al(96) modified nasogastric feeding tubes to allow oesophageal impedance monitoring of 21 asymptomatic preterm neonates. They identified a reflux episode frequency of 71 episodes over 24 hours. They also found that non-acid reflux episodes were twice as common as reflux episodes. The methodology in this study is noteworthy. However, this study population is atypical in many ways. Firstly, the median gestation at birth was 32 weeks (interquartile range (IQR) 30-34)). The median weight at testing was 1.7kg (IQR 1.5-1.9 kg). The median age at testing was 12 days (IQR 9-17.5 days). Reflux in this age group is physiological. Lastly, although these patients are described as ‘healthy preterm neonates’, this author hesitates to accept low birthweight preterm babies as an appropriate control population to provide ‘normative’ data.

In 2014, Mousa et al(36) published data that the authors describe as ‘normative’. However, the methodology used to achieve this data calls this definition into question. This study is a retrospective database review of pH-MII test results of investigated infants. Children included in the study were symptomatic, justifying the investigation. The acid reflux and symptom association parameters of the study were reviewed. Where the acid reflux threshold did not meet 6% for children and 3% for infants, studies were included. Other inclusion criteria were absence of temporal association between GER and symptoms. There is a high variability in non-acid reflux episodes in children due to the buffering effect of milk feeds. Therefore, including children who had low reflux indices on test date might simply be bias for children on milk feeds on the test date. These data are a summary of non-acid reflux episodes in

symptomatic children who, at least on test date, did not demonstrate high oesophageal acid exposure or symptom association.

Caveats accepted, these 'normative' paediatric data can be summarized in the tables below(36):

Table 4: Normative data for Infants. Sample median age was 4.8 months (3 weeks -11.9 months)

	Median	**95th Centile**
Percentage acid reflux episodes	0.6 (0.3-0.9)	1.4
Acid reflux episodes/ 24 hours	20 (11-26)	48
Percentage non-acid reflux episodes	0.7 (0.5-1.2)	2.5
Non-acid reflux episodes/ 24 hours	32 (16-45)	67

Table 5: Normative data for children. Sample median age was 7.2 (1.3-17) years)(36)

	Median	95th Centile
Percentage acid reflux episodes	0.4 (0.2-0.8)	1.3
Acid reflux episodes/ 24 hours	14 (11-15)	55
Percentage non-acid reflux episodes	0.1 (0-0.3)	1
Non-acid reflux episodes/ 24 hours	6 (3-11)	34

Another criticism of these data, and the field as a whole, is the arbitrary categorization of data into infant and child. Continuous / scalar data appears to be coerced into categories. Although we observe a trend towards reflux improving with age, no dramatic change in physiology takes place on a child's first birthday. Feeding patterns change in the first year from milk fed to weaning. Position of the child changes from mostly supine to often upright. However, these changes occur at variable times in different children and are simply not captured by an artificial infant/child duality.

Rather than use measures of central tendency e.g. mean, median, to set 'normal thresholds' a percentile approach can be taken. Simply stating percentiles for each parameter, or setting a 95% percentile as normal. The clinician can therefore interpret where their patient lies compared to other children. Validation studies of this approach would be necessary.

Another fundamental issue facing pH-MII is study reproducibility. The reproducibility of MII studies remains an open question. Vandenplas et al(133) found good reproducibility between two consecutive 24-hour studies in the same patient. However, Dalby et al (154) found the reproducibility of impedance is poor, with a variability of reflux index sometimes by a factor of 3.

In summary, pH-MII is a transformative technology in the study of paediatric GORD. It has become clear that NAR reflux was a significant factor that was hitherto unquantified. As this technology is relatively recent, normative data, risk stratification parameters and corresponding validation studies are

urgently needed. In my opinion, this work is necessary before pH-MII can be declared the 'gold standard'(153) in the diagnosis of paediatric GORD.

Oesophageal manometry

As mentioned earlier, the development of manometry was crucial to the understanding of the mechanism of gastroesophageal reflux. Cannon first demonstrate the LOS in cats in 1902, but radiological evidence for the same mechanism in humans was lacking for many a year. The debate on the existence of an anatomical or functional sphincter persisted. Code(21) was able to demonstrate that there was indeed a functional LOS in the 1950s.

Oesophageal manometry is an invasive study that involved introduction of a solid or liquid state into the oesophagus. The patient is requested to perform various swallowing and positioning procedures and the pressure profile of the oesophagus is mapped. This test is uncomfortable for a cooperating adult. It is of limited value in a child unable to cooperate either due to young age of NI. In children with structural anomalies e.g. oesophageal stricture, achalasia, manometry under sedation may provide useful information. Oesophageal manometry is not routinely used in the diagnosis of GORD in children.

Upper GI contrast (UCIC)

Historically, the UGIC was initially used to diagnose hiatus hernia in adults at a time when HH and GORD were considered synonymous(11). However, the advent of the pH study and its establishment as the gold-standard of reflux diagnosis diminished the role of the UGIC(10).

The UGIC is a radiological investigation. Contrast is introduced orally as first preference. However, in children disordered swallowing and at risk of aspiration, contrast may be introduced via a nasogastric tube or gastrostomy tube. The volume administered is a quantity sufficient to fill the stomach. Serial X-ray is use to visualize the flow of contrast.

Historically, dilute barium was the contrast medium. However, when aspirated, barium is highly corrosive to the lungs and results in a chemical pneumonitis. To mitigate this risk water-soluble, iodine-based contrast medium e.g. Omnipaque® was introduced. During the study, provocation manoeuvres can be utilised. These included e.g. distension of the stomach with water, lateral and prone positioning. The examination and images were reported by a senior radiologist. The presence or absence of reflux, and the height of the reflux column (to the distal, middle, upper oesophagus, or aspirated) are described.

UGIC and GORD

Compared to the pH study, the specificity and sensitivity of the UGI study for detecting reflux is poor. Askglaad et al(155) prospectively assessed reflux in 21 infants less than 1 year, using UGIC (barium) and pH probe. They demonstrated that contrast had a specificity of 50% and a sensitivity of 29%, as compared to 24-h pH monitoring. In our own series(156), we compared pH-MII against UGIC in 66 children. UGIC was poorly sensitive at 43.2%, with a negative predictive value of 24%. Although it is well recognised that the UGI contrast is a poor investigation for detecting reflux(157)(65), our series demonstrated that radiologists continued to commented on, and even be provoke reflux episodes during the UGIC(156).

Numerous studies(95,156,158–160) categorise GOR severity based on the oesophageal height of the reflux column. Distal, mid and upper oesophageal height corresponds to mild moderate and severe GOR. Aspiration is also considered a sign of reflux severity. This approach has its limitations. In vitro demonstration of high reflux during a half-hour UGI contrast study may not correspond to in vivo GORD. The utility of this metric is further compromised by provocation manoeuvres.

In a retrospective review of patients under 1 year undergoing fundoplication, Sharif et al classified reflux as severe if oesophagitis was present *and* if there was reflux to the upper oesophagus/ aspiration during the upper GI contrast study(159). Inherent in this classification are assumed correspondences between a morphological and radiological marker of GORD.

Salient here is the question of normative paediatric data. This lack is brought into sharp relief when considering the diagnosis of GORD in neonates. Alvares et al(158) provide data on 41 preterm neonates with symptoms of GORD and a pH-metry reflux index >10%. UGIC was positive in 23 patients (56%). Shah et al(160) reported on a series of 122 children weighing less than 5kg who had fundoplication. In this series, 50 (41%) children in this series had a history of prematurity. The mean age at surgery was 3 months. UGIC was the diagnostic modality used was the UGIC in 44% of patients. GOR is often a physiological finding in infants. Is an UGIC finding of reflux at this age also physiological? Normative data on UGIC findings by age of patients would provide a useful comparator.

UGIC and structural anomalies

In addition to GORD diagnostics, the UGIC is also used to detect structural anomalies i.e. malrotation and hiatus hernia(161). The gold standard for diagnosing malrotation is an intra-operative inspection of the duodenojejunal junction. Such an assessment is possible during fundoplication. Therefore, pre-operative UGIC's only utility is informing the operative plan. A recent retrospective review of 843 patients who underwent fundoplication investigated whether pre-operative UGIC informed operative management(157). There were 656 patients who had an UGIC. Malrotation was identified in 16 patients (2.4%) and incorrectly identified in 6 patients. Malrotation was also identified in 4 patients who had undergone a previous Ladd's procedure for malrotation. In these patients, prior knowledge was available rendering the test unnecessary. Therefore, the UGIC informed operative management for 4.5% of patients. Subsequently, the UGIC information was disproved or unnecessary in half those patients. Compared to pH study, the sensitivity of the UGIC study for reflux was 30.8% in this study.

In our series, we reviewed reports of 116 children with GORD undergoing UGIC. We identified low rates of malrotation (0.9%) and hiatus hernia (1%). Alvares et al(158) identified hiatus hernia in 1 of 41 preterm neonates studied with UGIC.

In summary, the use of the UGIC in paediatric GORD is likely a hangover from adult practice where it was used to demonstrate hiatus hernia. Compared to pH-metry and pH-MII it is poorly sensitive for reflux. Although correlations between UGIC findings and histological/ morphological markers of GORD are made, these correlations are not justified by data. Lastly, the use of the UGIC to demonstrate structural anomalies is low yield and likely unjustified.

Upper GI endoscopy (UGIE)

Accompanying the development of flexible oesophageal endoscopy for oesophagitis was the development of diagnostic criteria for oesophagitis. Of note are the Savary Miller classification, Los Angeles and Hertzel Dent classifications. These classifications are based on macroscopic visualisations of erosions and breaks in the oesophageal mucosa. During endoscopy, pinch biopsies are taken of the distal oesophagus. Histological classifications for severity of mucosal changes have been developed e.g. the Knuff and Leap classification.

In children particularly, endoscopy requires a general anaesthetic. Combined with the risk of perforation, this demands that the investigation be high yield, sensitive and specific. However, endoscopy is of limited utility in paediatric practice. Firstly, a predictive relationship between symptoms and endoscopic and histological findings has not been identified(162)(145). There is no criterion for the minimum grade of histological injury required to diagnose GORD. Therefore, the symptoms triggering endoscopy are at the discretion of the endoscopist and, hence, subject to variation.

Secondly, endoscopy is a low yield test. In adult practice, approximately half of patients with GORD symptoms are found to have erosive oesophagitis on endoscopy(163). This figure is even lower in paediatric studies. A multicentre survey by Gilger et al(37) identified macroscopic 'erosive oesophagitis '(Los Angeles classification) ' rates of 12% in 7188 children's endoscopy reports reviewed retrospectively. Biopsy findings were not evaluated. However, not all patients studied had endoscopy primarily for GORD symptoms. Therefore, this figure could be an understatement.

Thirdly, macroscopic findings of erosive oesophagitis are vulnerable to intra and inter observer variability. Furthermore, correlation between endoscopic and histological pathology is poor(163). Vieira et al(164) found that 72% of 167 infants who had normal findings on endoscopy had histological evidence of oesophagitis. Conversely, of the three patients with Grade III changes on histology, 5 had normal appearances on endoscopy.

Lastly, Barrett's oesophagus, a condition resulting from chronic and cumulative injury, is rare in children. Furthermore, historical studies identifying Barrett's mucosa did not sufficiently exclude hiatus hernia as the cause for identifying gastric-type mucosa in the oesophagus(136).

In summary, paediatric endoscopy is recommended where it is important to exclude other causes of symptoms e.g. eosinophilic oesophagitis, or for follow-up of previously confirmed histological reflux oesophagitis(163).

Markers of aspiration

Aspiration is a key concern for children with NI and GORD. There are several diagnostics for aspiration, but none are in routine use due to invasive nature and low specificity. Biomarkers of aspiration include lipid- and fat-laden macrophages, tracheal pH and pepsin. Macrophages are the inflammatory cells that absorb foreign material including aspirate. Borelli et al(149) demonstrated that macrophages had a higher concentration in bronchial aspirates of patients with more proximal NAR. Contrarily, Rosen et al(165) did not identify a correlation between LLM and height of reflux.

Krishnan et al(166) also found that LLM macrophages were poorly sensitive and poorly specific when bronchioalveolar lavage in children with and without GORD was compared. Other tests that have been used to confirm aspiration are milk scintigraphy, tracheal ph. and tracheal pepsin.

Markers of gastric emptying

Delayed gastric emptying has been posited as a mechanism for GORD risk. Measures of gastric emptying include milk scan scintigraphy and scintigraphy, manometry and electromyography. None of these methods is routinely used for GORD diagnosis.

Summary of investigations

In summary, the use pH-metry is limited in its ability to test acid reflux and foundational errors in severity scoring. The UGIC was a carry-over from adult practice where they are useful for excluding hiatus hernia. In children however, this is a low yield investigation for structural anomalies with poor sensitivity and specificity for GORD. The pH-MII has conceptually transformed our understanding of paediatric GORD. However, lack of validation of protocols, normative data and severity metrics are challenges that require urgent inquiry.

What of the children who have had fundoplication based on the information gleaned from these studies?

Surgeons may justify prior decisions to operate based on a synthesis of *both* symptoms and investigations. Furthermore, a surgeon may argue that, for some at least, operative invention was offered only after non-operative interventions failed to control reflux.

Fundoplication is indicated when management with other therapies have failed. Various pharmacological and non-pharmacological interventions have been applied with mixed results.

NON-PHARMACOLOGICAL THERAPY

Positioning is the most basic intervention. For term and preterm infants, sleeping in the prone or left lateral position was associated with fewer reflux episodes on pH monitoring(167,168). However, the association between prone sleeping and sudden infant death has limited the recommendation to just the left lateral position(167,169). In infants, elevation of the cot end of the bed is not beneficial(170). For older children, fewer GOR episodes were observed when the child was upright or sleeping with 30-degree cot elevation at the head end(171).

In a study by Orenstein and MacGowan(172), parents were given a package of non-pharmacological interventions. These included avoiding tobacco smoke exposure, switching formula fed babies to semi-elemental formula, encouraging breast-feeding mothers to avoid cow and soy milk in diet, thickening feeds with rice cereal and optimizing position of feeding. Symptoms were tracked using the infant gastroesophageal reflux questionnaire-revised (I-GERQ). Authors(172) report that 78% of patients experienced some improvement in scores from baseline to 2 weeks.

Thickening of feeds using corn starch, carob gum, galactomannan etc. has been shown to reduce observed symptoms of reflux(170). However, a systematic review of randomised controlled trials (RCTs) by Horvarth et al(173) demonstrated no reduction in objective pH measures of GOR. Alarmingly, reports of necrotising enterocolitis in premature infants after use of xanthan gum thickeners led practitioners to eschew thickeners in this population(174,175).

Feeding strategies e.g. continuous versus bolus feeds(94), low rate/long duration vs high rate/short(170) duration have also been investigated. There is evidence to support the use of small volume but frequent of continuous feeds in infants(170).

PHARMACOLOGICAL THERAPIES

In paediatric practice, barrier agents are a first line therapy. Sodium and magnesium alginate products e.g. Gaviscon® Infant decease reflux episodes by increasing feed viscosity(124). Due to its sodium content (0.92 mmoL Na^+/dose), Gaviscon is used with caution in children who require close sodium monitoring e.g. prematurity, renal disease, congestive cardiac failure, ileal stomas. Lastly, there are anecdotal reports of intestinal obstruction in adult patients receiving Gaviscon. However, the paediatric Gaviscon was reformulated in 1999 to exclude aluminium(124). This was due to concerns about aluminium accumulation particularly in children with chronic renal failure. Theoretically, this also removes the risk of aluminium salt agglutination and secondary intestinal obstruction. However, co-administration of alginates with thickening agents should be avoided(176).

Acid suppression medications used in paediatric practice are H2 antagonists (Ranitidine) and PPIs (Omeprazole, Lansoprazole, Pantoprazole etc.). The efficacy of both groups of drugs to alter gastric pH has been proven in a sizeable series of RCTs(124). In neonates, ranitidine is used with caution given the observation of an six fold increase rate of necrotising enterocolitis in very infants(177). PPIs work by raising gastric pH and decreasing volume of acid secretions(74). PPIs can raise gastric pH >6 and

have become the mainstay of treatment in infants and children. However, lansoprazole has been associated with an increase in lower respiratory tract infection. The rise in gastric pH is thought to encourage bacterial overgrowth, particularly C. Difficile. In infants, RCT evidence is limited by the physiological versus pathological classification. For example, it is not clear whether symptoms of fussiness / crying would improve with increasing age despite treatment(178,179).

Prokinetic motility agents have also been used as treatment of GORD. The underlying rationale is increasing gastric emptying reduces the gastric content available to reflux, thereby reducing episodes. Cisapride acts upon a serotonin receptor which actives cholinergic neurons stimulating gastric smooth muscle to contract, increasing peristalsis. Some studies demonstrated reductions in reflux index with Cisapride(180). However, systematic review identified no symptom control benefit over placebo, no treatment, thickening agents or alginates(180). Furthermore, it is associated with cardiac arrhythmia (long QT syndrome, sudden cardiac death) leading to its restricted use.

Domperidone is active upon a dopamine 2 receptor blocking the antagonistic action of dopaminergic neurons on cholinergic smooth muscle. Domperidone has also been associated with cardiac arrhythmia and its use restricted in children with co-existing cardiac disease(124). Metoclopramide is another prokinetic agent working on both serotonin and dopaminergic receptors. Paediatric application is limited by both the FDA[1] and EMA[2] due to serious adverse effects reported in 11-34% of treated children(124). Adverse side effects include oculogyric crises, trismus and tardive dyskinesia. Erythromycin is a macrolide antibiotic that appears to have a prokinetic effect.

Erythromycin acts directly on motilin receptors on gastrointestinal smooth muscle. Its agonist effect promotes peristalsis and thus gastric emptying. Motilin, a part of the motilin/ghrelin family of peptides, is a hormone that is released by gastric cells post-prandially. This hormone acts on GI smooth muscle motilin receptors to activate the migrating motor complex. The resulting gastric emptying is thought to reduce episodes of GOR. Although the use of erythromycin in GORD is reported, there are no randomised controlled trials (RCT) examining its efficacy(124).

SURGICAL OPTIONS

Fundoplication has remained the mainstay surgical approach to treating reflux since the last half of the 20th Century. This is remarkable considering the changes in disease paradigm that have taken place in that time.

Other procedures proposed as alternatives to fundoplication include gastrojejunostomy tube, endoscopic fundoplication and endoscopic fundoplication.

Fundoplication

Fundoplication begins with a laparotomy or a laparoscopic approach into the abdomen. The fundus of the stomach is mobilised by dividing the short gastric vessels and mobilising the distal oesophagus through the diaphragmatic crura. The fundus is wrapped around the distal oesophagus and cardia, then

[1] Food and Drug Administration (USA)
[2] European Medicines Agency

sutured in place creating creates a 1-2 cm cuff around the gastric cardia. The posterior crural hiatus is closed.

Varying degrees of wrapping have been described for various indications. Proponents include Nissen, Nissen-Rosetti, Thal, Toupet-Rosenthal and Hill. Prospective studies of partial and complete wraps in children have not identified significant differences in key outcomes of rGERD and RF(181,182). However, the Nissen fundoplication appears to have a higher risk of early post-operative dysphagia(182).

Table 6: A variety of fundoplication methods

Type	Proponent	Degree of wrap	Direction of wrap
Complete	Nissen	360	Posterior
Partial	Toupet	270	Posterior
	Thal	180	Anterior
	Belsey (thoracic approach)	270	Anterior and posterior
	Dor	180	Right anterolateral
	Watson	120	Left anterolateral
	Hill	180-210	Posterior gastropexy

The Nissen fundoplication is the most common operation performed for fundoplication in children. In 1998 Fonkalsrud et al published a multicentre retrospective review of fundoplication performed for symptomatic GORD across 7 children's hospitals in the USA. They reported on 7467 patients and found that open Nissen fundoplication was performed in 64% of patients(63). In 2011, LaRiviere et al(183) surveyed 121 surgeons working for 42 Child Health Corporation of America hospitals. They identified 120 surgeons who performed the Nissen procedure and 78% of these performed it laparoscopically.

Data from human and animal studies(184,185) suggest that fundoplication 4 mechanically effects of fundoplication (186):

1. Correcting hiatal herniation
2. Lengthening the intra-abdominal portion of the oesophagus
3. Tightening the crura
4. Increasing external pressure at the level of the LOS.

Beyond a mechanical valve effect, the fundoplication is thought to decrease frequency of TRLOS(187). Suggested mechanisms are fundal denervation during surgery and reduced fundal volume. Analogies have been made with the action of 5HT antagonist Sumatriptan(187). Sumatriptan reduces fundal tone, extending gastric meal-induced relaxations. This action corresponds to observed reductions in TRLOS. The histological effects of fundoplication have also been studied in the ferret model(185).

Understanding of the indications, complications and outcomes of fundoplication has benefited from large published series. Fonkalsrud et al reported indications are summarised by system affected in table

below. In 1998, Fonkalsrud et al reported on 7467 children who had mostly open fundoplication (97%) across 7 children's centres in the USA. In 2004, Gilger et al(3) published a retrospective review of fundoplication in 198 children in Texas between 1996-1998. In 2005, Diaz and colleagues published a retrospective review of 762 patients comparing OF versus NF(92). Similarly, Baerg et al(188) assessed risk factors for RF in a retrospective cohort study of series of 832 children. In 2005, Rothenberg(189) published the largest single surgeon series of paediatric LF (n=1050). A decade later, he updated the series, reporting on 2008 LFs performed or supervised by a single surgeon(190).

Our understanding of the effect of fundoplication in children is limited by ethical concerns. Due to the invasive nature of investigations, there are few studies with complete samples of pre-and post-fundoplication pH or pH-MII findings. In a systematic review of fundoplication efficacy, Mauritz et al(191) include 17 studies of 1280 children who had had investigations before and after fundoplication. In 8 of thes studies, reflux index decreased after fundoplication. However, manometry findings did not demonstrate any increase in lower oesophageal sphincter pressure after fundoplication in the three studies reporting this investigation.

Indications for fundoplication

In literature, indications for fundoplication vary greatly but can be broadly categorised as:

1. Symptom control:
2. Prophylaxis:

Symptom control

Most cases are for symptom control. However, few published reports document indications for surgery(71).

Fundoplication for respiratory symptoms is problematic. Firstly, few authors systematically describe and diagnose respiratory indications for fundoplication. Secondly, there are mixed results. Cheung et al(192) found no difference in pre- and post-operative rates of pneumonia in 20 children undergoing followed up for a median of 3.5 years after fundoplication.

Table 7: Indications for fundoplication for symptom control

Indication	System	Indication	References
Reflux symptom control	Airway	Aspiration pneumonia(	(193)
	Respiratory	Recurrent chest infections/ recurrent aspiration(3)	
		Apnoea	(77,193)
		Nocturnal cough	(3)
		Respiratory symptoms	(77,194)
	Gastrointestinal	Regurgitation	(77)
		Vomiting	(3,193)
		Haematemesis	(193)
		Gastric outlet obstruction	(193)
		Paraoesophageal hernia	(193)
		Feeding difficulty (pain with feeding, feed refusal)	(3,193)
		Oesophageal stricture	(77,193)
		Failure to thrive	(77),(193)
		Gastrointestinal symptoms	(77,194)
	Other	Neurological impairment	(194)

Prophylactic fundoplication

Prophylactic fundoplication has been reported for patients with caustic strictures(193), OATOF(193), CDH(193,194) and abdominal wall defects(193). Where a patient has CDH or OATOF, prophylactic fundoplication occurs at the time of correction of the primary defect i.e. in the neonatal period.

There is great debate on whether all patients with OATOF / CDH should have prophylaxis. For CDH, prophylactic fundoplication is considered by some to anticipate fundoplication later. Rates of fundoplication in older children vary (11%(90,194), 12%(195), 21%(89), 23%(196)). Some predictive factors have been suggested e.g. liver in chest(89)(194), need for a patch(194,196), diaphragmatic

defect >75%(90). Some authors advocate prophylactic fundoplication to anticipate growth failure particularly in children who require a patch repair of CDH(91).

Diamond et al(194) identified features of CDH (liver in chest, need for patch closure) that were predictive of subsequent need for fundoplication in a cohort of 86 children(194). However, failure of fundoplication is also higher, suggesting that fundoplication is a poor substitute for a structurally normal diaphragm in the prevention of GORD(65,90).

Furthermore, GORD may also improve with time in these patients. One study with long-term follow-up (median age at follow-up was 12.1years; range 6-17) suggested that incidence of GORD within a CDH cohort also decrease with increasing age (39% to 12%)(87). Another identified the prevalence of typical reflux symptoms in children (median age 6) and adults (median age 21) with CDH history to be 41% and 15% respectively(197). Pathological pH findings were identified in 58% of children and 33% of adults(197).

Maier et al(198) performed a patient-blinded RCT of CDH repair with or without fundoplication found a trend towards worse GORD symptoms at 6 months in infants who did not receive prophylactic fundoplication. However, this trend did not meet significance and no difference between groups was detectible at 12- or 24-months follow-up. Notably, only left sided CDH and Thal fundoplication was performed. Therefore, this RCT may be non-contributory by excluding right-sided CDH (liver in chest (89),(194) and by performing 180 degree (Thal), rather than 360 degree (Nissen) fundoplication.

In children with OATOF, prophylactic fundoplication has few proponents because the OATOF repair is performed through a thoracic approach. However, given the high prevalence of Barrett's oesophagus in this subgroup, longitudinal investigation of the protective effect of fundoplication is required.

Variations in practice

In adapting the procedure to paediatric practice, some surgeons have made modifications to the classical procedure described by Nissen(199). Paediatric surgeons will apply the Rosetti modifications i.e. a short (2-3cm), floppier wrap. Other modifications described by Rosetti are minimal/ no crural dissection, liver retraction without mobilisation, leaving short gastric vessels intact(199).

Pyloroplasty

Another area of varied practice is the role of pyloroplasty. Pyloroplasty is advocated by some in children with evidence of delayed gastric emptying. However, pyloroplasty has recently fallen out of surgical vogue. Wockenforth et al(200) reported on a single surgeon practice in the north of England. There were 255 patients who underwent pyloroplasty between 1998- 2008. Authors report that pyloroplasty was abandoned after 9 procedures due to severity of post-operative dumping syndrome.

Laparoscopic versus open fundoplication

Perhaps the greatest schism in practice occurred in the transition from open (OF) to laparoscopic fundoplication (LF). The first LF was described by Roberts and Cuschieri in 1991(201). In children, the first LF was reported in a 10 year old obese boy from Memphis, USA in 1993(202) .

With any new surgical innovation, initial adoption will be cautions. Therefore, early case series are likely based on selection of ideal candidates. Research data are also limited by vague patient selection criteria

for LF vs. OF(71). A systematic review of studies focusing on outcomes of LF noted that a majority of studies included were prone to biases of “selection, detection, reporting an attrition” (71). Furthermore, comparative studies seldom controlled for underlying comorbidities or the confounding effect of the laparoscopic learning curve(71). Comparative data arise from prospective and retrospective(188) cohort studies and some RCTs(203,204) . There are large numbers available from case series of LF published by Rothenberg(189,190). However, case series tend to be biased in favour of the described intervention.

The largest comparative series is a retrospective database review from the Paediatric Health Information System, a 42-hospital collaboration in the US(205). Authors identified 3978 LF and 3105 OF for comparison. LF was associated with shorter hospital stay and lower rates of hospital acquired infection. Surgical complications were fewer with LF. OF was typically more expensive than LF due to length of stay. Mattioli et al(206) also found shorter length of stay after LF.

Other comparative studies have found faster recover with LF patients feeing earlier than OF patients after surgery(207).

LF performed less well by some measures. Franzen et al(208), in an RCT, reported higher rates of post-operative dysphagia and poorer patient satisfaction with symptom control after laparoscopic surgery. In a large retrospective cohort study (832 patients), Baerg et al(188) found that open fundoplication was not associated with an increased risk of redo surgery. Diaz et al also reported higher RF rates following LF(92). Controlling for laparoscopic learning curve appeared important. Rothenberg had a conversion rate of 7.5% and a complication rate of 7.45 in his first series(189). In his second series approximately 10 years and 1000 patients later, the conversion rate was 0.1% and the complication rate was 3.1%(190). Initial studies found a longer operating time. Operating time had fallen from over 2 hours to 35 minutes(190).

In an RCT at our institution, McHoney et al(209) compared 20 children who had OF with 19 children undergoing LF. There was no difference in morphine requirement in the first 48 hours. Laparoscopic operative time was significantly longer (160 versus 80 minutes on average). No significant difference was observed in time to feed and time to discharge. The primary outcome for this study was both clinical and metabolic. There were similar rates of post op rGERD and RF. This study is of limited generalizability due to design. Clinical outcomes were secondary. The primary outcome measure on which the sample size was based was energy expenditure. Therefore, the study likely underpowered to detect differences in clinical outcomes e.g. time to feeds, analgesia requirement.

In another RCT, Knatten et al(210) randomised children to either laparoscopic or open Nissen fundoplication. A key finding was a higher rate pf post fundoplication GORD in the laparoscopic group at median follow-up of 4 years(210). There are several methodological problems. Firstly, this study was performed between 2003 and 2009. In keeping with contemporaneous practice, pH-metry was the diagnostic standard for GORD. However, where pH-metry was not performed, upper GI contrast was used as a diagnostic and inclusion criterion for GORD. As these diagnostics are not equivalent, they cannot be used interchangeably and define control intervention groups. Furthermore, the approach to patients with gastrostomy is problematic. There are 11 children (5 NI and 6 NN) who had a gastrostomy

concomitantly. These children's data is amalgamated with that of the non-gastrostomy patients. Thirty-six children with NI were mostly tube fed. Crucially, 6 had a pre-existing gastrostomy or nasogastric tube pre-operatively. Gastrostomy related complications are simply excluded from analysis of post-operative complications.

Furthermore, post-discharge complications were inexplicably excluded when assessing complications. These exclusions are made to coerce the laparoscopic and open fundoplication groups into a comparable frame. However, the said same exclusions render comparison of NI and non-NI groups illegitimate.

In a subsequent paper based on the same cohort, authors have compared pre-operative status and outcomes in NI and NN children. The same group(60) report that that GORD recurrence, early complications and long-term patient satisfaction were similar for NI and non-NI patients. In this comparison, "early outcome in NI and in NN patients were compared regardless of surgical technique". NI was used as a minimizing criterion in the prior study which randomised with control. Therefore, comparison of NI versus NN patients in this RCT should be treated as subgroup analysis.

Currently, the laparoscopic versus open fundoplication debate has moved from being a patient-centred issue to a surgeon-specific issue. Holden et al found that, comparing surveys performed 10 years apart, surgeons admitting to wholly laparoscopic practice had increased from 4% to 21%. For my generation of surgeons, fundoplication is taught as a laparoscopic procedure with conversion to OF as an occasional complication. The LF vs. OF duality informs on a surgeons training trajectory rather than the patients risk factors. In assessing retrospective data, it is not possible to control for this factor. Therefore, in this thesis we shall remain agnostic to LF or OF and pool indications and outcomes.

Fundoplication with / without gastrostomy a paediatric debate

Unlike adults with GORD, children with GORD may require tube feeding for uncoordinated swallow, failure to thrive etc. Before PEG placement became common place in children, the surgical gastrostomy was the procedure of choice. The question then arose: should concomitant fundoplication be performed at the time of gastrostomy.

In patients with clear signs and symptoms of GORD, one might think the answer is obvious. However, there may be overlap in symptoms and attributing causality is often impossible. For example, a child with failure to thrive may require a gastrostomy rather than a fundoplication, or both. Furthermore, it remains unclear if GT improves, worsens or is neutral to GORD status. Data from literature are contradictory. Launay et al(211) reported that reflux index improved or normalised after a surgical gastrostomy. More recently, Toporowskja-Kowalska et al(212) measured pre- and post-gastrostomy reflux episodes in 15 patients with neurological impairment. The mean time that pH-MII was performed after gastrostomy formation was 7 months. They found no significant difference in pre- and post-gastrostomy MII parameters.

In the surgical community there is no common agreement on this issue. Fox et al(205) surveyed 6 paediatric surgeons working at a tertiary referral children's hospital. They asked surgeons to review indications for fundoplication in 166 patients referred for gastrostomy. Concomitant fundoplication was

performed in 52% of patients. A review by Vernon-Roberts and Sullivan(52) identified few limited comparative studies and no RCTs in this area.

This variation in practice belies a central dilemma: in NI children with gastrostomy, is fundoplication or medication is better at controlling GORD symptoms. REMOS (Reflux medical or surgical) was an RCT study established at GOSH in 2008. The full title was "Gastrostomy with medical treatment versus gastrostomy with fundoplication in children with neurological impairment". This study is discussed extensively in Section 4.

Outcomes of fundoplication

Successful fundoplication should result in the resolution of reflux. Rates of reflux resolution vary. Key outcome measures are recurrence of GORD symptoms (rGERD) and revision/redo fundoplication. Gilger et al(3) reported a rGERD rate of 63% at follow-up 2 months after fundoplication in 198 cases reviewed.

Fonkalsrud et al(63) report a 94% cure rate. Koivusalo and Parkinen report symptom resolution in 90%(77). Baerg et al(188) identified a RF (RF) rate of 12% in a retrospective review of fundoplication in 832 children <18 years at two centres in the USA . The mean age at first fundoplication was significantly lower in children who went on to have redo surgery compared to these who had only one fundoplication. Other modifiable risk factors for failure included retching OR: 3.59 (95% CI: 1.56-8.25) and hiatal dissection OR: 8.45 (95% CI: 2.45-29.11). Unmodifiable factors included male gender (p=0.008). In this cohort, 18 children (2%) had three fundoplication (second redo) and 4 (0.4%) had four fundoplication. Ngercham et al (2007)(102) identified 116 children who had RF and matched them with 209 children who had primary fundoplication. Matching was based on surgeon, laparoscopic versus open approach and type of fundoplication. Risk factors identified were age (<6 years), hiatus hernia, pre-operative retching and oesophageal dilatation. Notably, NI was not a risk factor for RF.

Other authors also admit to redo of RF. Gilger et al(3) published a case series of 199 fundoplication procedures in children performed between 1996-1999. The first revision rate in this series was 18%. The second revision rate was 2%.

Table 8: Summary of incidence of RF

Cohort	Type of study	Primary fundoplication	1st revision (%)	2nd revision (%)	3rd revision (%)
Retrospective GOR	Retrospective	1008	66(6.5)	7(0.7)	2 (0.2)
Rothenberg (2005)(189)	Retrospective, laparoscopic	1050	32(3)	3(0.3)	0
Shariff et al (2010)(159)	Retrospective, infants 0.5-11 months, LF	79	14 (18)		
Diaz et al(92) (2005)	Retrospective, LF and OF	762	55		
Gilger et al(3) (2004)	Retrospective	198	35(18)	4(2)	
Fonkalsrud (1998)(63)	Retrospective	7467	611 (8)		
Baerg et al.(2013)(188)	Retrospective	823	100(12)	18(2.2)	4(0.5)
St Peter et al. (2007)(213)	Retrospective, all laparoscopic	273	21(7.8)	4(1.5)	-
St Peter et al. (2011)(214)	Prospective RCT, all laparoscopic	177	19(10)	2(1.1)	1(0.6)
Shah et al. (2010)(160)	Retrospective, laparoscopic, <5kg	122	2(1.6)		-

Many authors have attempted to identify factors predicting RF risk. Kimber, Kiely and Spitz reviewed operative notes of 66 patients who had RF over a 15-year period at GOSH. The primary mechanism (45%) was herniation of an intact wrap into the posterior mediastinum (15%), suggesting crural hiatus repair failure. The third most frequent mechanism was disruption of the wrap (215). A combination of wrap disruption and wrap herniation was found in 30% of patients. A tight wrap was implicated in 7% of patients.

Complications of fundoplication

To obtain informed consent, a surgeon must explain both the benefits and the risks of surgery. Both LF and OF can result in intra- and post-operative complications. The key complication of concern is post-fundoplication syndrome(186). It is also described as gas-bloat syndrome. This is a constellation of symptoms that include gagging, retching, food refusal, nausea and abdominal distension. Post-fundoplication syndrome has been attributed to many factors including decreased motility, decreased gastric accommodation and gastric hypersecretion(186). Complications can be understood in terms of proximity from the surgery.

Table 9: Complication rates following open fundoplication

Timing	Complication	Incidence
Intra-operative	Hepatic vein laceration	6%(193)
Early post-operative	Jejunal perforation	3%(193)
	Pneumothorax	6%(193), 0.4%(194)
	Pneumonia	0.4%(194)
	Wound infection	3%(194)
	Gastric outlet obstruction	3%(193)
	Wrap too tight, dysphagia	1%(194), 4%(3)
	Wrap disruption	7%(63), 38%(2) 7(194)
	Respiratory complications	4.4%(3,63)
	Gagging/retching	33%(3)
	Gas bloat	3.6%(63), 6%(3)
	Dumping syndrome	3%(3,194)
	Sacral decubitus	3%(193)
	Wound breakdown	1%(194)
	Incisional hernia	1%(194)
	Urosepsis	0.4%(194)
Late post-operative	Aspiration	9%(2)
	RF	28%(2), 18%(3)
	Adhesive intestinal obstruction	2.6%(63), 3%(193) 2%(194)
	Gastrostomy-related complications	2%(194)
	Recurrent reflux symptoms requiring medication post ARS	63%(3)

Laparoscopic complication results in some idiosyncratic complications e.g. pneumothorax. Other complications are comparable to the open technique. Rates of complication appear lower. However, as

discussed prior, historic data are skewed as surgeons will often select the best candidates for the laparoscopic approach.

Table 10: Complications following laparoscopic fundoplication

Complication	Incidence
Wrap failure	5%(189), 7%(190)
Oesophageal perforation	<1%(189)
Gastric perforation	<1%(189)
Conversion to open	<1%(189)
Liver Haemorrhage	<1%(159,189)
Haemorrhage from other vessels	<1%(217)
Pneumothorax	<1%(159,189)
Gastroparesis	<1%(189)
Dysphagia	<1%(189)
Gastrostomy tube complication	<1%(189)
Wound infection	<1%(189)
RF	<1%(189)
Incarcerated paraoesophageal hernia	<1%(189)

Mortality following fundoplication, often defined as within 30 post-operative days, has also been reported. Rates range between <1%(189) and 5%(50).

Table 11: Reported causes of mortality following fundoplication

Cause	Incidence
All-cause	0.8%(63), 20%(159)
Multi-organ failure	3%(216)
Sepsis	3%(216)
Acute renal failure	3%(216)
Aspiration	3%(216)

Gastrojejunal tube

As early as 2002, Wales et al(217) compared a retrospective cohort of 111 patients treated with either fundoplication or GJT. The 48 patients who had GJT had an increased risk of bowel obstruction and

intussusception post procedure. They were also more likely to continue taking 'anti-reflux' medication. However, crucially, 14.5% of GJT patients had symptom improvement and had their GJT removed. This suggests that a trial of GJT in a patient with a pre-existing GT is a reasonable alternative to fundoplication.

Srivastava et al(218) reviewed data from a large administrative database to identify 43 children with neurological impairment who had of first GJT and 323 cases of first fundoplication. They used propensity scoring to retrospectively pseudo-randomise children taking into consideration baseline characteristics e.g. age, previous aspiration pneumonia, chronic lung disease, cardiac comorbidities. They found no significant difference in post-operative rates of aspiration pneumonia or mortality.

Total oesophagogastric dissociation and other anti-reflux procedures

In 1997, Bianchi proposed the use of total oesophageal gastric dissociation (TOGD) in children with neurological impairment as an alternative to fundoplication(219). NI can result in pharyngeal muscle incoordination necessitating GT feeding. TOGD involves disconnection of the oesophagus entirely from the stomach. The oesophagus is reconnected to the jejunum. The stomach and duodenum become the Y-limb of the Roux-en-Y loop and are anastomosed to the jejunum. A gastrostomy is formed to enable feeding through intermittent or permanent catheterisation. Pyloroplasty is performed if there is evidence of delayed gastric emptying or suspicion of vagal nerve injury.

Figure 9: Schematic illustrating total oesophagogastric dissociation for reflux

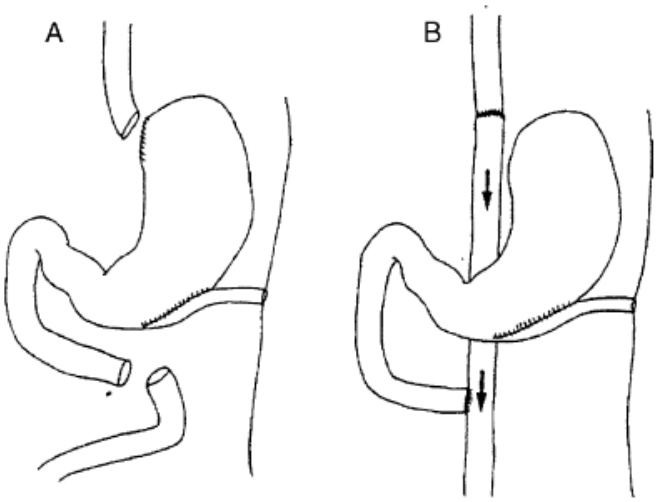


Fig. 1 The esophagus is transected above the gastroesophageal junction and the stomach oversewn (A). An isoperistaltic antire-fluxing Roux-en-Y loop of jejunum is brought without tension, through the transverse mesocolon, passing behind the stomach to anastomose with the lower esophagus. Bowel continuity is established by end-to-side jejunojejunostomy at 40 cm from the esophagojejunal anastomosis (B).

A TOGD procedure obviates the risk of RF which is high in this subgroup. The key benefit is elimination of gastro-oesophageal reflux. The proximal stomach is a blind ending loop. The distal oesophagus is anastomosed to the jejunum which has no reservoir capacity.

The procedure has been used as a rescue procedure after failed fundoplication. However, Morabito et al(219), the Bianchi group, proposed its use as a primary anti-reflux procedure. However, TOGD is certainly a greater surgical undertaking compared with fundoplication. It requires extensive mobilisation and multiple anastomoses with the potential to leak. Reported complications appear severe i.e. oesophagojejunal dehiscence, 18% early re-operation rate, internal bowel hernia and 4 / 26 deaths that are 'unrelated to surgery'(219,220). It also remains unclear how patient selection for primary TOGD would be performed. De Lagausie et al(221) applied this approach in the special circumstance of 13 children with colonic transposition following OATOF or tracheal cleft repair. Half the selected patients had persistence of reflux after previous fundoplication with recurrence of symptoms. Importantly, 12 / 13 patients were neurologically normal. There is evidence that uptake of this procedure amongst surgeons is poor. A systematic review Peters et al(220) identified 181 cases in 14 years of published reports, with 117 primary and 64 as rescue.

CHAPTER 5: RISK STRATIFICATION

THE SCOPE OF THE PROBLEM

We can summarise the central question of this opus as 'the surgeon's dilemma'.

Imagine a surgeon standing in a waiting room, surveying a roomful of children with GORD. The children have a variety of symptoms and comorbidities. They have had an idiosyncrasy of tests performed. Some come with an UGIC; some have had a pH study. Some have had an endoscopy. Surveying the room, how can the surgeon choose which child will have greatest benefit for lowest risk from fundoplication?

The job of a surgeon is to choose the right operation for the right patient at the right time. In this opus I apply this dictum to fundoplication.

- Is it the right operation for paediatric GORD?
- Who are the right children to have fundoplication?
- When is the right age to perform fundoplication?

Below, we discuss special characteristics of paediatric GORD i.e. aetiology, symptomatology, natural history. These characteristics must be understood in assessing a child's suitability for fundoplication.

Risk stratification

Risk stratification is increasingly a feature of clinical practice. A powerful early driver was the need for health management organisations to cost and compare patients by premorbid conditions and outcomes. Also, governmental oversight mandates required the comparison of institutional outcomes. For example, findings of excessive post-cardiac surgery mortality at Bristol Children's Hospital led to a public inquiry [1] and the institution of mandatory national outcome reporting. All congenital cardiac surgery centres have participated in reporting of data to the UK Congenital Heart Audit since the year 2000. Mortality is analysed by the National Institute for Cardiovascular Outcomes Research (NICOR). Comparison of institutional performance is stratified by instituting a review of paediatric cardiac surgery. In this review, the UK government compared outcomes of 12 institutions undertaking cardiac surgery in children. A key question was whether the mortality outcomes at Bristol were due to operations performed on children with poorer pre-morbid condition and more complex cardiac anomalies. To enable a fair risk-adjusted comparison between institutions, case mix measures were developed. These measures have evolved over the years to include well validated factors e.g. univariate heart [2], lower weight and younger age [3].

In addition to comparing institutional performance, risk stratification is also useful for benchmarking performance over time. Individual patient risk stratification can also be extremely valuable. Understanding and quantifying how pre-morbid conditions may influence outcomes will help patients and families make informed treatment choices. Care can also be customised to the patient's individual risk profile.

In considering the paediatric gastro-oesophageal reflux, several issues emerge. Firstly, diagnosis has been based historically on an adult patient paradigm. However unlike adults, children have a lower burden of acid reflux therefore, diagnosis based on acid reflux has probably been misleading. Secondly,

mucosal damage is cumulative and probably acid reflux related. Using mucosal damage as a diagnostic marker of severity is therefore inappropriate in a paediatric population.

Risk stratification by premorbid condition

Coarse stratifications into risk groups for GOR have emerged. Taking the largest at-risk group i.e. children with neurological impairment, it is clear that this is a coarse stratification which encompasses a heterogeneous group of patients. It is not known whether the aetiology and pathophysiology of GOR is different in patients with spastic cerebral palsy, compared with a flaccid muscle wasting disease.

Risk stratification by symptoms

There are numerous reported associations between GOR and extra-oesophageal symptoms e.g. cough. Although the association is often based on clinical suspicion, few studies have rigorously tested the association. New technologies e.g. pH impedance now allows greater interrogation of symptoms and reflux episodes. On the whole, data suggest that traditional associations made between apnoea, ALTE and GOR episodes are not borne out on analysis of temporal relationships. Application of these new technologies will help determine which symptoms are rightly associated with GOR, the strength of this association, and the direction of causality.

Risk stratification by pre-operative investigations

At present, classification and scoring systems for GOR use either UGI parameters or pH parameters. On UGI study, the height of reflux is often associated to reflux severity. However, the UGI is a poorly sensitive study for reflux. Furthermore, with the application of provocation manoeuvres, the study becomes operator dependent and, therefore, the height of reflux as a marker of severity is brought into question.

Scoring based on pH study is limited by the fact that, increasingly, data suggest that non-acid reflux features largely in paediatric patients. Equally, a historical analysis of the origins of these scores suggests that they were generated on small sample sizes with extrapolations made from adult to paediatric practice. Traditional, pH study-based scoring systems for reflux severity need urgent updating.

The plan of work

The foundational study driving this opus is the REMOS trial. This is a randomised clinical trial investigating whether children with NI should have gastrostomy with or without fundoplication. In addition to this, we will investigate the symptoms of GORD in children. We plan to review current symptom measurement instruments (questionnaires) and work to improve these tools.

We shall also investigate comorbid risk factors affecting fundoplication. This will encompass the risk of fundoplication as well as the risk of failure of fundoplication. This last investigation will involve data mining of a retrospective database of children with GORD.

Lastly, a synthesis of these opera will be presented. An approach to surgical risk stratification will be proposed based on the knowledge discovery from the studies detailed here.

SECTION II: DEVELOPMENT OF A SMARTPHONE APP TO RECORD SYMPTOMS OF GASTROESOPHAGEAL REFLUX

CHAPTER 1: INTRODUCTION

The definition and diagnosis of GORD i.e. “a condition that develops when the reflux of stomach contents causes troublesome symptoms and/or complications (1)” requires examination of the patient’s experience of symptoms.

Typically, reflux syndrome is described as ‘symptoms of retrosternal burning pain which is worse after meals’ (4). However, this paradigm of acid reflux-causing symptoms, whilst well established in the adult population, cannot be extrapolated to the paediatric population (3). Preverbal infants and neurologically impaired children, in which reflux is prevalent, will not be able to describe the precise location, or timing of pain, so the diagnosis of reflux may rely on other symptoms or signs.

In children, numerous other symptoms have been anecdotally associated with GOR (1, 4). Common symptoms reported include regurgitation, vomiting, cough, recurrent chest infections, pain, nausea, feeding difficulties and failure to thrive (5). Rarer symptoms include dental erosions, bronchial and tracheal inflammation (6). Delineating normal behaviour and physiology of infancy and symptoms of GORD can be difficult. For example, normal infants can cry for up to two hours daily (7) and at least one daily episode of regurgitation is observed in 50% of infants less than 3 months and 67% of infants at 4 months (8).

Inconsistent symptom-based definitions (2) of GORD have driven the search for objective physiological measures of GOR events. Most recently, the pH impedance study has been used to demonstrate temporal association of individual symptoms and reflux episodes. The strength of symptom-reflux association can be statistically tested. Statistical associations have been reported between GOR episodes and apnoea, apparent life-threatening events and exacerbations of asthma (9). Even with symptoms associated with reflux, the direction of causality can be unclear. For example, reflux episodes have been demonstrated to trigger coughing episodes. However, coughing episodes have been demonstrated to trigger reflux (11). In children with neurological impairment e.g. cerebral palsy, clinicians may attribute dystonic movements to expression of reflux pain (e.g. Sandifer syndrome). Equally, clinicians may posit that reflux results from dystonic contractions of abdominal muscle or hypotonia of the gastro-oesophageal sphincter during a seizure episode (12). Notably, pH impedance testing is an invasive test requiring nasogastric intubation. Therefore, in uncomplicated GORD, national guideline-providing bodies recommend diagnosis of GORD in infants and children based on clinical symptoms (10).

To have a symptom-based diagnosis of GORD, it is important to define:

- symptoms contributing towards diagnosis
- frequency and severity thresholds of these symptoms

In this chapter, symptom data capture methods used, their advantages and disadvantages are described. I propose and describe the development of a novel method of symptom data capture using a smartphone application.

CHAPTER 2: MEASURING SYMPTOMS OF GASTROESOPHAGEAL REFLUX

To understand the verity and reliability of symptom data, it is important to understand how these data are captured. Two approaches predominate. These are:

- Patient history from the caregiver
- Symptom questionnaires

CAREGIVER HISTORY

GOR is prevalent in preverbal infants. Therefore, to understand symptoms of reflux in infants, a clinician must ask the caregiver (usually a parent) a series of questions. When GOR persists in older children it is prevalent in children with neurological impairment who are often unable to articulate their symptoms. Therefore, the clinician is dependent on an accurate and comprehensive history from the caregiver and the correct interpretation of this history.

The questions asked are tailored to the clinical condition, taking co-morbidities and mode of feeding into account. Each clinician tailors the questions based on their knowledge of GOR and its symptoms and their clinical expertise. Each clinician has their own formulation of questions, which can be modified based on a holistic understanding of the child's illness. Each caregiver will respond idiosyncratically during each encounter. Therefore, caregiver histories are inherently variable and anecdotal. Not all caregivers will be accurate historians. Furthermore, the attribution of causation is particularly problematic, giving rise to considerations of validity of information. Caregivers may selectively report symptoms that they consider to be related to episodes of GOR. However, this association is difficult to demonstrate objectively and leads to reporting biases. GORD may be diagnosed if either a clinician feels strongly that it may be present. Questions may be presented in a leading way. Caregiver bias may also lead to a history may be inaccurately recalled or presented.

Appreciation of the inherent limitations of caregiver histories has led to the use of symptom questionnaires / symptom diaries. Symptom questionnaires are designed to address accuracy, reliability and validity of data collected.

SYMPTOM QUESTIONNAIRE

A symptom questionnaire is a measuring instrument designed to objectively measure symptoms of a particular condition. There are several questionnaires available for assessing symptoms of GOR in children. Of note are the I-GERQ (Infant Gastroesophageal Questionnaire) and the GSQ (GERD symptom questionnaire).

The I-GERQ questionnaire was developed and validated by Orenstein et al (1993, 1996) (13, 14). It was validated for children aged 1-14 months. The gold standards used were abnormal pH probe studies and abnormal oesophageal biopsies. It is a 138-item inventory of various symptoms of GORD. The length of this questionnaire limited its use to *ad hoc* or one-off assessments. There was no pre-defined recall period. This questionnaire was not designed to measure patient progress or response to interventions.

Addressing these issues, this inventory was revised by Kleinman et al (15) (16, 17) in 2006, resulting in the IGERQ-R. The IGERQ-R was developed by identifying the most discriminating questions in the

138-item IGERQ inventory from the findings of a validation study (18). From these questions, a 12-item questionnaire was developed. The IGERQ-R is the most commonly used and well-validated questionnaire in this genre. It has been used in several studies to assess symptom frequency and severity. Although this iteration of the IGERQ-R is well validated, its use remained limited by the absence of a defined recall period. Further revision of this questionnaire (IGERQ-R) addresses this, validating use to weekly intervals (15, 19).

The median time for completion of the I-GERQ is 20 minutes. This limits its use as a prospective serial assessment tool (14). There is a shorter version (IGER-SF). This, however, lacks documented validation (20) at this time. A common limitation of all I-GERQ iterations is the limited symptom severity assessments. Lack of severity calibration makes response to treatment difficult to assess.

The frequency and severity of GERD symptoms vary with age. Both I-GERQ and IGER-SF are targeted to infants less than 1 year of age. Neither questionnaire version contains questions specific to the 1 through 4-year-old cohort. The GSQ questionnaire developed by Deal et al (21) addresses this limitation by having a version for infants (0-11 months: GSQ-I) and a version for young children (1-4 years: GSQ-YC). The gold standard for validation was a diagnosis of GORD made by a paediatric gastroenterologist. Although commonly used, this validation standard I.e. expert opinion, may be an imperfect.

The GSQ assesses symptom frequency and severity in the seven days prior to the completion of the questionnaire. Respondents are also asked to rate severity from 1 (least severe) to 10 (most severe). In analysing the data, the symptom frequency and severity are combined to give a symptom score. A comparison of symptom scores between individuals is confounded by the subjective symptom severity element. The main drawback of this questionnaire is the retrospective assessment of symptom frequency, rendering the data prone to recall bias. A second limitation is the combination of symptom severity and frequency. Lastly, comparing symptom scores between infants and children is also problematic as infant severity scores are caregiver ratings, while scores for older children are self-rated.

The review of existing symptom questionnaires and their limitations are summarised in the Table 1 below.

Table 12: Questionnaires used for the assessment of GOR in children

Questionnaire	Year	Gold standard for validation	Target Respondent	Age of Child Assessed	Domains addressed	Number of items	Recall period	References	Limitations
I-GERQ	1993, 1996	Abnormal pH probe studies and / or abnormal oesophageal biopsies	Caregiver	1 – 14 months	Demographics Symptoms Associated events	138	No defined recall periods	(13-15, 19)	Subjective assessment of symptoms and severity. No defined recall periods. Not suited to therapeutic response evaluation.
I-GERQ-R	2006	Physician global assessment of symptoms	Caregiver	< 18 months	Frequency Severity	12	7 days		Validated against subjective gold standard. Subjective assessments of symptoms and severity.
GSQ- I	2005	Paediatric gastroenterologist assessment	Caregiver	0-11 months	Frequency Severity	9	7 days	(22)	Validated against subjective gold standard. Composite score combining symptoms and severity. Subjective assessments of symptoms and severity.
GSQ-YC			Child	1-4 years		6			

The existence of symptom capture tools does not necessarily lead to their usage. Symptoms are a blind spot, particularly in the surgical literature. In a systematic review of paediatric laparoscopic fundoplication literature, Martin et al (1) found that only 36% of studies described the symptoms that led to surgery. One explanation for this is the hierarchy of referrals. Patients are often referred for surgery by paediatricians (e.g. neurologists and gastroenterologists) who have exhausted the full range of non-surgical options for symptom control. Surgeons may fail to focus on the specific symptom as they trust that the patient has been appropriately triaged to their service, and may also use objective measures such as pH impedance studies.

The questionnaires currently used have some limitations specific to the use of paper as a medium of delivery. For paper questionnaires, long completion times and the inconvenience of carrying paper questionnaires limits assessment to snapshots taken at intervals. This can render data capture retrospective. The serial completion of paper questionnaires can be tedious, inconvenient and inefficient.

Addressing these limitations, it is clear that prospective, continuous and ambulatory monitoring solves the problems of recall. Changing the medium to smartphone based questionnaires addresses the problem of inconvenience as devices are ordinarily carried in caregiver's purses and pockets.

SMARTPHONE QUESTIONNAIRES IN HEALTHCARE

A government survey (YouGov Smartphone Mobile Internet Experience Study 2011), found that 35% of adults own a smartphone (23). Survey data reveals that the top three functions used by adult smartphone owners are email, internet banking and social networking respectively (24). Therefore desirability and acceptability of connecting to the internet on a mobile smartphone is established. Consumers trends demonstrating a willingness to share personally and commercially sensitive information over mobile data networks.

There are numerous applications available aimed at personal health for smartphone users. In reviewing the field, I categorised apps according to tools offered. These are summarised as follows:

- Symptom tracking: Allows input of symptoms and when they occurred
- Reporting tools: Allows creation of a personal health record specific to that disease or condition
- Advice/coaching: Health information is provided within the app. Links to further helpful information online are often available.
- Remote monitoring: Transmits data to participating clinicians
- Interventions: Responds to data input with an interventional response e.g. an input of a high blood glucose leads to an automated response from the software suggesting increased insulin dosage.
- Remote monitoring: Clinicians / researchers can remotely monitor data input in a live way. They may respond with advice or recommendation e.g. an input of a high blood glucose leads to a response from the clinician suggesting increased insulin dosage.

Table 13: Summary of smartphone app genres

Type of instrument	Examples	Description	Symptom Tracker	Reporting tools	Advice/ Coaching	Intervention	Remote symptom monitoring	Central data repository
Simple symptom tracker	Glaucoma	Track weight, blood glucose and insulin levels.	Yes	Yes	No	No	No	No
Advice/ Coaching	AsthmaMD	Patient inputs performance on peak flow metres and intake of medications. Reports include thresholds suggesting therapeutic encounters.	Yes	Yes	Yes	No	Yes	No
	NHS Quit Smoking	Details money saved since quitting, health benefits and top tips to encourage adherence to smoking cessation.	Yes	Yes	Yes	No	No	No
	NHS choices drinks tracker	Tracks alcohol unit intake and provides personalised feedback on drinking.	Yes	Yes	Yes	No	No	No
Remote monitoring	eCAALYX	Remote monitoring systems for older people with multiple chronic diseases.	Yes	Yes	Yes	Yes	Yes	No
	EMA	Using mobile phones to measure adolescent diabetes adherence to a diabetes self-care programme.	Yes	Yes	Yes	No	No	Yes
Intervention	Mobile diabetes	Mobile coaching and patient self-management intervention. A randomised controlled study comparing the mobile application coaching against standard treatment demonstrated a significant difference (1.2%, $p< 0.001$) in mean Hba1C over 12 months.	Yes	Yes	No	Yes	Yes	Yes
Epidemiological	TARDIS	Data collection tool for symptoms of GOR. Central data repository with potential for population-based data assessment of GOR.	Yes	Yes	Yes	Potentially	Potentially	Yes

THE TARDIS:REFLUX APP

In this chapter I describe the development of a smartphone app for accurate and prospective collection of GORD symptom data. This smartphone symptom tracker has been named the TARDIS app as it **T**racks **A**ctivity in **R**elation to **Dis**ease. To our knowledge, this is the first symptom questionnaire for gastro-oesophageal reflux in children delivered on a smartphone. It is a simple questionnaire recording symptom timing and associated events only.

We believe it provides a convenient way to record multiple symptoms and generate an accurate symptom profile for each patient. To generate a symptom profile for the population, the smartphone app facilitates immediate, electronic and secure transmission of data to a central repository. Clinicians populate the central repository with data on key variables e.g. comorbidities, investigations and outcome. This process builds a rich dataset from which a robust epidemiological model of GOR may emerge.

The design is simple and includes four standard elements:

- Symptom tracker: This is a simple symptom frequency questionnaire
- Reporting tools: reports of symptoms recorded are available for consumption by the user
- Advice: Help pages are available both within the app and on the project website www.ucl.ac.uk/tardis
- Central repository: data are collected remotely and transmitted to a central database. This innovation transforms a simple symptom tracker into an epidemiological tool, with the potential to collect symptom data on a population level.

The development of the app is a central part of the PhD opus. It was a collaborative project requiring involvement of software developers. In my role as project lead, I was responsible for inception, user interface design, coordination and deployment onto the app store. In subsequent sections, I describe the elements of development that I was responsible for.

Aim

The objective of this pilot study is two-fold.

Firstly, to demonstrate the application of smartphone technology in the study of GOR in children.

Secondly, to demonstrate the efficacy of a smartphone application as a novel epidemiological and research tool. Such tools may be extended to the study of other conditions e.g. constipation, asthma and epilepsy. Epidemiological apps like TARDIS will supplement existing and future research projects by providing useful and cost-effective research tools.

Rationale

In the study of gastroesophageal reflux disease, there are several factors that render apps an attractive medium for data collection:

- **Ambulatory data recording:** As smartphones are ordinarily carried by caregivers, caregivers can record symptoms prospectively as they occur, eliminating recall bias. This can be done in conjunction with other ambulatory tests e.g. pH impedance.

- **Remote data collection:** GORD is a disease that evolves in the community. Wireless transfer of information over the internet would ensure that events that occur at home may be recorded and in close to real-time. Events can be assessed by the physician in hospital without the necessity of the patient visiting the hospital. A future application of smartphone symptom questionnaires as part of a telemedicine portfolio of tools is envisaged.
- **Confounding severity and frequency of symptoms:** a smartphone application can conveniently incorporate a visual analogue scale to rate severity for each symptom when it is recorded. This would allow convenient and separate analysis of symptom severity and frequency data as separate dimensions of GOR symptoms.

The generic use of apps as clinical and research adjuncts present several potential advantages and applications:

- **Convenience**: Most adults in the UK will carry a mobile phone devise of some sort. Mobile devices are portable and are already commonly used for tracking daily activities such as fitness. For chronic conditions, carrying paper questionnaires to complete over long capture periods is impractical.
- **Reduction of error and bias:** Caregiver histories entail several biases e.g. recency, attribution, confirmation bias, etc. Errors e.g. chronology of events, inaccurate estimation of frequency, may also alter the impression taken from the caregiver history. Real-time recording of symptoms and events, with data validation protocols, can mitigate these issues. Other features to improve data fidelity include pre-validated data fields, error alerts for non-valid entries, notifications services to remind users to record symptoms, progress updates to reinforce data collection, etc. Digital storage and transmission can be more convenient, secure and immediate than sending a completed paper questionnaire through the post, with secondary transfer of data to an electronic container.
- **Cost**: Beyond the initial cost of developing a suitable app, costs are minimised by the use of patient-held devices. The distribution model for the app is a one-to-many relationship with no cost limits on how many applications may be downloaded and used. Considering the prohibitive cost of procuring and developing hardware and software for the NHS, a data collection tool that entails minimal outlay for the NHS is certainly attractive.
- **Language**: a symptom frequency tracker utilising visual analogues would be simpler to translate and validate in multiple languages than a 138-item inventory such as the IGERQ.

Summary

The development of apps as clinical and research adjuncts is in its nascence. In reviewing the academic literature in 2011, we found several examples of apps contributing to this area. However, there was no clear methodology described. As such, there was no protocol to follow or equipment list to review to identify requirements.

The methods described are therefore, by necessity, a post-hoc summary of the steps taken. One of the aims of this project was to establish a methodology for app development in a healthcare research context. The degree of attainment of objective can be judged by the coherence and reproducibility of the methods presented below.

The waterfall model of software development was used as a loose framework for this project. The model enumerates five steps:

1. Specification
2. Design
3. Implementation/ Build
4. Verification/ Testing
5. Maintenance

Each of these steps will be described in the sections to follow.

CHAPTER 3: SPECIFICATION

The process of software development begins with specifying the requirements of the user. In research, the analogue to this process is defining variables and outcomes to be observed. The researcher defines data sources, data to be collected and elaborates on the process.

DEFINING USERS

From the outset, it was clear that not all data can be reliably reported by all users. For example, a clinician is well-placed to report the date and time of a procedure e.g. oesophagogastric dissociation. However, a parent is better placed to report a condition experienced at home e.g. nosebleed. Based on this analysis, we defined roles for reporting of data.

Four reporting roles emerged. These are coded 1 to 4 and are detailed in **Table 14** below. The fourth role= "All" denotes data that can be reliably reported by parents, patients, clinicians and the researchers.

Table 14: Four reporting roles are defined.

Roles	
ID	ROLE
1	Parent
2	Clinician
3	Researcher
4	All

DEFINING DATA FIELDS

Symptom data

A qualitative systematic review of the PubMed literature database was conducted to identify list of possible symptoms to include. The Boolean search terms of the PubMed database is detailed below:

Search

- **"gastroesophageal reflux"[MeSH Major Topic])**
- **AND "all child"[Filter] AND "all infant"[Filter] "adolescent"[Filter] AND "newborn"[Filter]) AND ("2006"[Date - Entrez] AND: "2012"[Date - Entrez])) AND "English"[Language]**

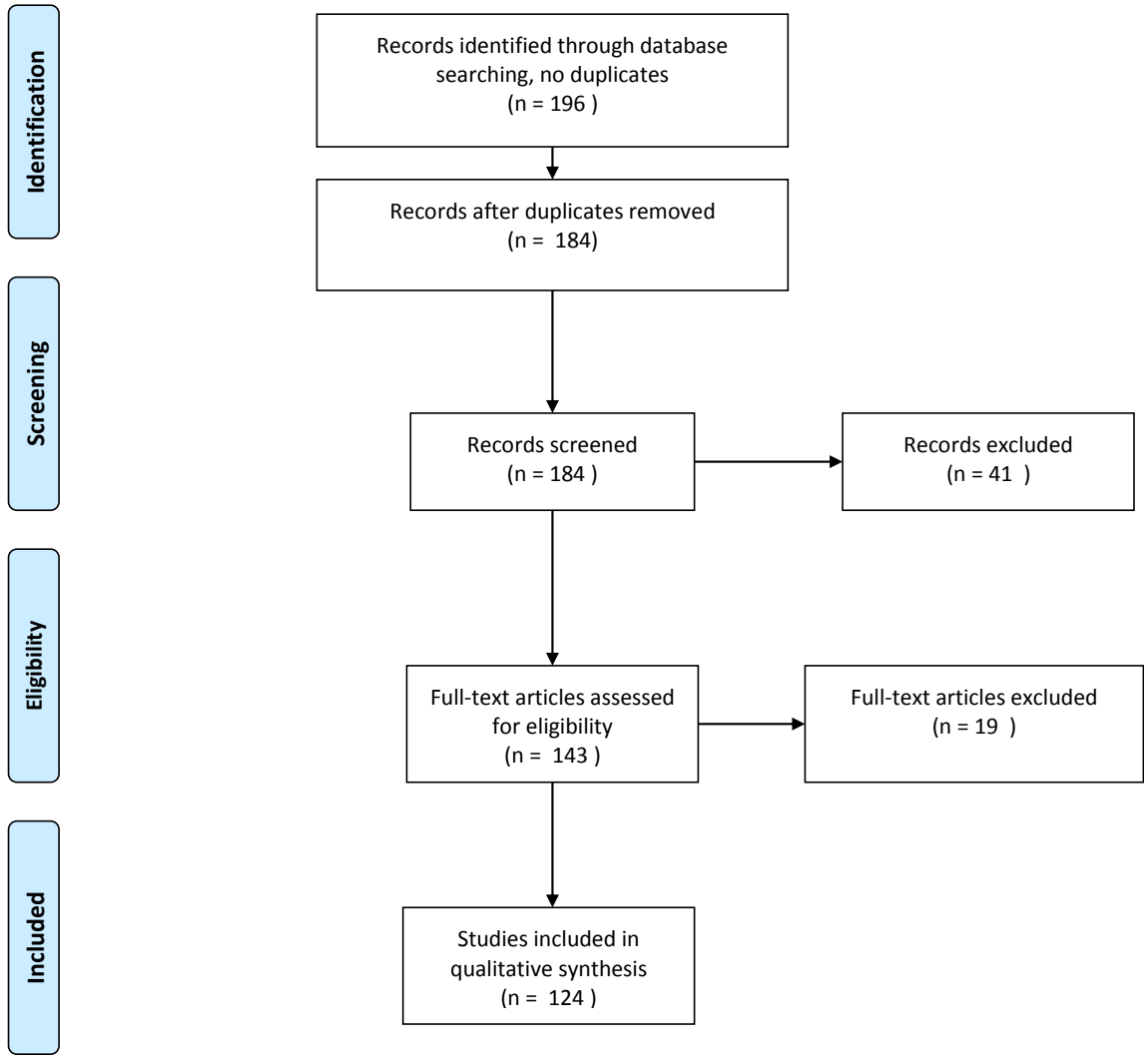
We included articles focused primarily on the gastroesophageal reflux in children. Therefore, the key search criterion was MeSH major topic of gastroesophageal reflux. The search was limited by age to select articles in infants and children. Furthermore, the search was limited by publication date to select recent (past 8 years) of articles.

Records that met the following criteria were excluded:

- Duplicated records or duplicated reporting on cohorts (n=12)
- Screened records in which the primary topic was not gastroesophageal reflux (n=41)
- Full articles in which symptoms of gastroesophageal reflux were not described. Examples include
  - Feeding protocol after total oesophagogastric dissociation procedure.
  - Secondary lesions in infants with laryngomalacia

The chart below describes inclusion and exclusion steps in keeping with PRISMA guidelines for systematic review.

Figure 10: PRISMA flow diagram describing screening, eligibility and inclusion of journal articles for meta-analysis



Included articles were then reviewed for symptoms. Symptoms mentioned were matched to World Health Organisation International Classification of Diseases 2012 (WHO ICD-10) codes. Mention of each symptom was counted. Finally, the results were tabulated and symptoms were ranked in order of frequency, with the most frequent symptom achieving the highest rank.

ICD-10 descriptions may not be immediately recognisable to individuals without a background in medicine as a subject. For example, patients are more likely to report a nosebleed than an episode of epistaxis. Therefore, for each ICD-10 symptom, a patient-readable version for each was specified.

Table 15: Symptoms coded and ranked in order of frequency

ICD 10 code	ICD10 Symptom	Patient-readable description	Rank	Role
P92.1	Regurgitation and rumination in newborn	Regurgitation	26	4
P92	Vomiting in a newborn	Vomiting	25	4
P92.9	Feeding problem of newborn, unspecified	Food refusal	24	4
R68.12	Irritable infant	Crying	23	4
R68.11	Excessive crying of infant	Irritability	22	
530.81, 723.5	oesophageal reflux and torticollis (Arching associated with reflux, aka Sandifer syndrome)	Arching	21	4
R06.6	Hiccough	Hiccoughs	20	4
J39.2	Other diseases of the pharynx, hyperactive gag reflex	Gagging	19	4
R13.1	Dysphagia	Feeding discomfort	18	4
R05	Cough	Coughing	17	4
R14.2	Eructation	Belching	16	4
T18	Choking (food)	Choking	15	4
R06.2	Wheezing	Wheezing	14	4
R06.7	Sneezing	Sneezing	13	4
R06.1	Stridor	Noisy breathing	12	4
R06.0	Dyspnoea	Short of breath	11	4
R06.91	Apnoea	Apnoea or breath holding spell	10	4
T17.81	Gastric contents in other parts of respiratory tract	Suspected aspiration	9	4
R13.0	Aphagia	Unable to swallow	8	4
R11	Nausea	Nausea	7	4
R12	Heartburn	Heartburn	6	4
R10.4	Other and unspecified abdominal pain	Stomach ache	5	4
R14.0	Abdominal distension (gaseous)	Bloated stomach	4	4
R14.1	Gas pain	Bloating with colicky pain	3	4
R19.3	Abdominal rigidity	Hard stomach	2	4
R11.14	Bilious vomiting	Green vomit	1	4

Event data

Event data are defined as procedures, medications, investigations or diagnoses that occur in patients with gastroesophageal reflux. Events differ from symptoms as they occur less frequently. Event data also comprise interventions and important diagnoses which may reflect the behaviour of clinicians or prevailing practice at an institution. Symptom data reflect the natural evolution of the condition in an individual patient. Event data may sometimes be a response to symptom data.

Based on clinical experience, we summarised the key events that occur in patents with GOR. The summary yielded a total of 24 events. Although not all events were suitable for reporting by parents, the list was considered too long to be displayed in a single selection menu. Therefore, we organised events into event categories. These are Diagnosis, Investigation, Medication and Procedure.

We defined the events that would be of interest to be recorded in a patient-readable way. We categorised data according to the role of the person reporting the event. A dichotomy emerged in the reporting roles. Events requiring clinical acumen and investigations to accurately describe were reserved for the clinician (2). For example, a gastrostomy site infection is diagnosed by demonstrating purulent discharge and with laboratory confirmation of microorganisms and inflammatory cells. All other events were suitable for reporting by all roles (4). Therefore, it was unnecessary to further specific roles for parents (1) or the researcher (3).

Table 16: Events data, data categories, patient-readable descriptions and reporting roles

ICD10 Code	Event Category	ICD10 Event	Patient-readable description	Role
K59	Diagnosis	Constipation	Constipation	4
R19.7	Diagnosis	Diarrhoea, unspecified	Diarrhoea	4
K94.22	Diagnosis	Infection of gastrostomy	Gastrostomy infection	4
R00.	Diagnosis	Epistaxis	Nose Bleed	4
J18	Diagnosis	Pneumonia unspecified organism	Pneumonia	4
R09.3	Diagnosis	Abnormal sputum	Productive cough	4
G93.82	Diagnosis	Death, cause not specified		2
799.82	Diagnosis	Apparent life-threatening event	ALTE	2
J69.0	Diagnosis	Pneumonitis due to inhalation of food and vomit	Aspiration pneumonia	2
K94.22	Diagnosis	Gastrostomy site infection	Gastrostomy site infection	2
J18	Diagnosis	Pneumonia unspecified organism	Pneumonia	2
R09.2	Diagnosis	Respiratory arrest	Respiratory Arrest	2
R63.8	Investigation	Other symptoms and signs concerning food and fluid intake	Weight	4
BW00	Investigation	Imaging of Anatomical Region > Abdomen	Abdominal X-ray	2
BW03	Investigation	Imaging of Anatomical Region > Chest	Chest X-ray	2
BB24	Investigation	Imaging, Anatomical Regions, Computerized Tomography > Chest and Abdomen	CT Chest	2
BD15	Investigation	Upper GI Contrast Study	Upper GI study	2
97.51	Procedure	Non-operative; Removal of gastrostomy tube	Gastrostomy Removal	4
0DH67DZ	Procedure	Insertion of Intraluminal Device into Stomach, Via Natural or Artificial Opening	Nasogastric tube	4
0DHA7U	Procedure	Insertion of Intraluminal Device into Jejunum, Via Natural or Artificial Opening	Nasojejunal tube	4
44.66	Procedure	Other Procedure for Creation of Oesophagogastric Sphincteric Competence; Fundoplication, Gastric Cardioplasty, Nissen Fundoplication, Restoration of Cardio-oesophageal Angle	Fundoplication	2
0D977ZZ	Procedure	Drainage of Stomach, Pylorus, Via Natural or Artificial Opening	Gastrostomy (Laparoscopic/Open)	2
0D840ZZ	Procedure	Division of Oesophagogastric Junction, Open Approach	Oesophageal dissociation procedure	2
0D9740Z	Procedure	Drainage of Stomach, Pylorus with Drainage Device, Percutaneous Endoscopic Approach	PEG insertion	2

Medication events

Children with gastroesophageal reflux disease are commonly prescribed the following medications: Proton pump inhibitors (e.g. omeprazole, lansoprazole); H2-receptor antagonists (Ranitidine) is the most commonly prescribed H2-receptor antagonist. Domperidone is the most commonly prescribed motility agent. According the data specification, the required information on medications was which medication was being taken, and any dose changes. In paediatric practice, medication is prescribed based on weight. We determined that asking app to enter the precise dosage of medication would be tedious and prone to error. The precise dosage of medication, where required, was information best obtained from the prescriber. Therefore, for the Parent Role, we decided to require information on the qualitative change e.g. increased dose of omeprazole (Table 17).

Table 17: Patient-readable medication events available for reporting in the app

Medication information
Started Omeprazole/Lansoprazole
Increased Dose Omeprazole/Lansoprazole
Reduced Dose Omeprazole/Lansoprazole
Stopped Omeprazole/Lansoprazole
Started Ranitidine
Increased Dose Ranitidine
Reduced Dose Ranitidine
Stopped Ranitidine
Started Domperidone
Increased Dose Domperidone
Reduced Dose Domperidone
Stopped Domperidone

Diagnosis events

We reviewed ICD-10 to identify comorbidities associated with GOR and GORD. We identified 13 diagnoses of interest (**Table 18**). Of these, only 6 were suitable for reporting by parents. These were included in the app.

Table 18: Patient-readable diagnosis events available for reporting in the app

ICD10 Code	Event Category	ICD10 Event	Patient-readable description	Reporting Role
K59	Diagnosis	Constipation	Constipation	4
R19.7	Diagnosis	Diarrhoea, unspecified	Diarrhoea	4
K94.22	Diagnosis	Infection of gastrostomy	Gastrostomy infection	4

R00.	Diagnosis	Epistaxis	Nose Bleed	4
J18	Diagnosis	Pneumonia unspecified organism	Pneumonia	4
R09.3	Diagnosis	Abnormal sputum	Productive cough	4

Investigations

We reviewed ICD-10 to identify investigations important for patients with GORD. We identified 5 investigations of interest (**Table 19**). Only one investigation was deemed suitable for parental reporting i.e. Weight. Parents will encounter more frequent opportunities to measure the child e.g. at home, with the health visitor, at the GP practice. Measurement of weight is commonplace and we expect parents to reliable record their child's weight. We acknowledge there will be some measurement error as multiple devices may be used to obtain the weight.

Table 19: Patient-readable investigation available for parental reporting in the app

ICD10 Code	Event Category	ICD10 Event	Patient-readable description	Reporting Role
R63.8	Investigation	Other symptoms and signs concerning food and fluid intake	Weight	4

Procedural events

We reviewed ICD-10 to identify procedures important for patients with GORD. Seven procedures were identified (Table 20).

- Nasogastric tube insertion/change: this is frequently done by parents at home, once training has been given.
- Nasojejunal tube: Although this is done in hospital, parents would be present for the procedure to give consent. It is usually done in the radiology department. Although the clinicians responsible for the child's care might request the test, they would most likely not be present for the test. In this scenario, the parent is the most reliable witness to this event.
- Gastrostomy tube removal: This procedure is done when the child no longer requires the tube. Depending on the kind of tube used, this is done in hospital under anaesthetic or in any general clinic by a nurse, GP or hospital doctor. In the latter scenario, the clinician responsible for care may not be present. The parent is the most reliable witness to this event.

These 3 procedures were made suitable for reporting by parents, clinicians and the researcher. As parents use feeding tubes daily, parents would be in the best position to detect tube complications and instigate management by clinicians.

Table 20: Patient-readable procedures available for parental reporting in the app

ICD10 Code	Event Category	ICD10 Event	Patient-readable description	Reporting Role
97.51	Procedure	Non-operative; Removal of gastrostomy tube	Gastrostomy Removal	4
0DH67DZ	Procedure	Insertion of Intraluminal Device into Stomach, Via Natural or Artificial Opening	Nasogastric tube	4
0DHA7U	Procedure	Insertion of Intraluminal Device into Jejunum, Via Natural or Artificial Opening	Nasojejunal tube	4

SPECIFYING THE DATA MODEL

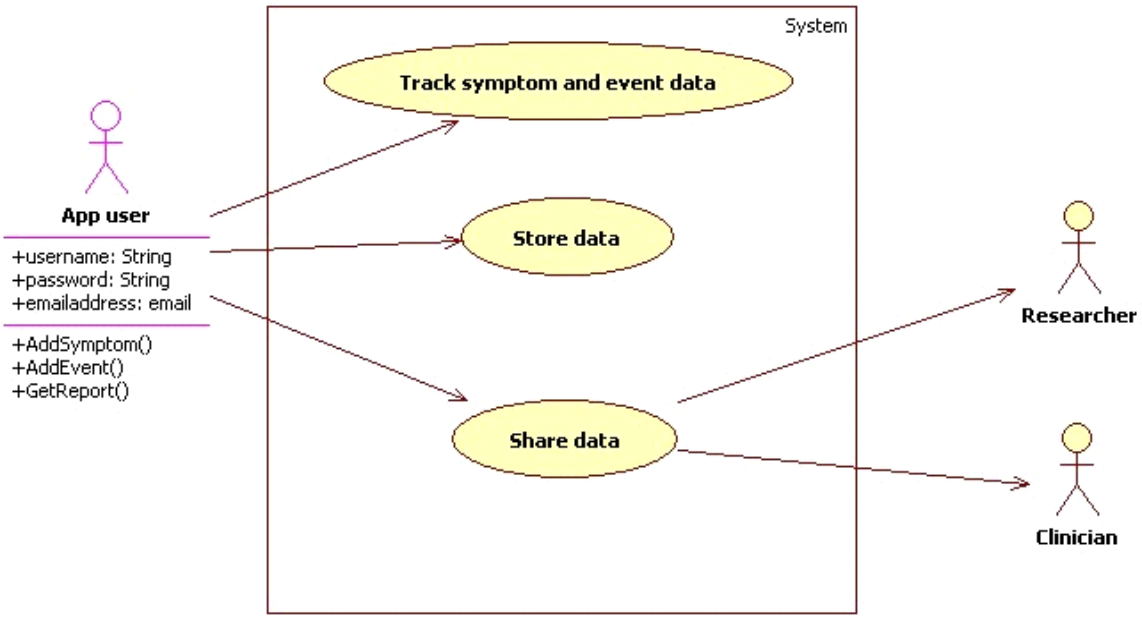
A data model can be defined as a description of information and objects in a system, their characteristics and relationships. Specifying a data model is an integral part of project specification. Standardised approaches have been developed to enable specification to be read by any developer and the vision translated into a product through code. Unified Modelling Language (UML) diagrams are commonly used to provide stylised, consistent formats and annotations. UML diagrams were used to specify the data model.

Specifying the use case

A use case defines the interaction between the “actors” using a system and the system itself. In this project, we identified three actors- the researcher, the clinician and the app user. App users are parents of children with gastroesophageal reflux disease.

The system was designed to allow the app user to track symptom and event data, store data and share data with researchers and clinicians. These three steps or activities are illustrated in the use case diagram below (Figure 1). The System comprises the software and hardware used to carry out the three steps and will be defined further in later sections (25).

Figure 11: Defining the use case for the app user



Specifying data objects

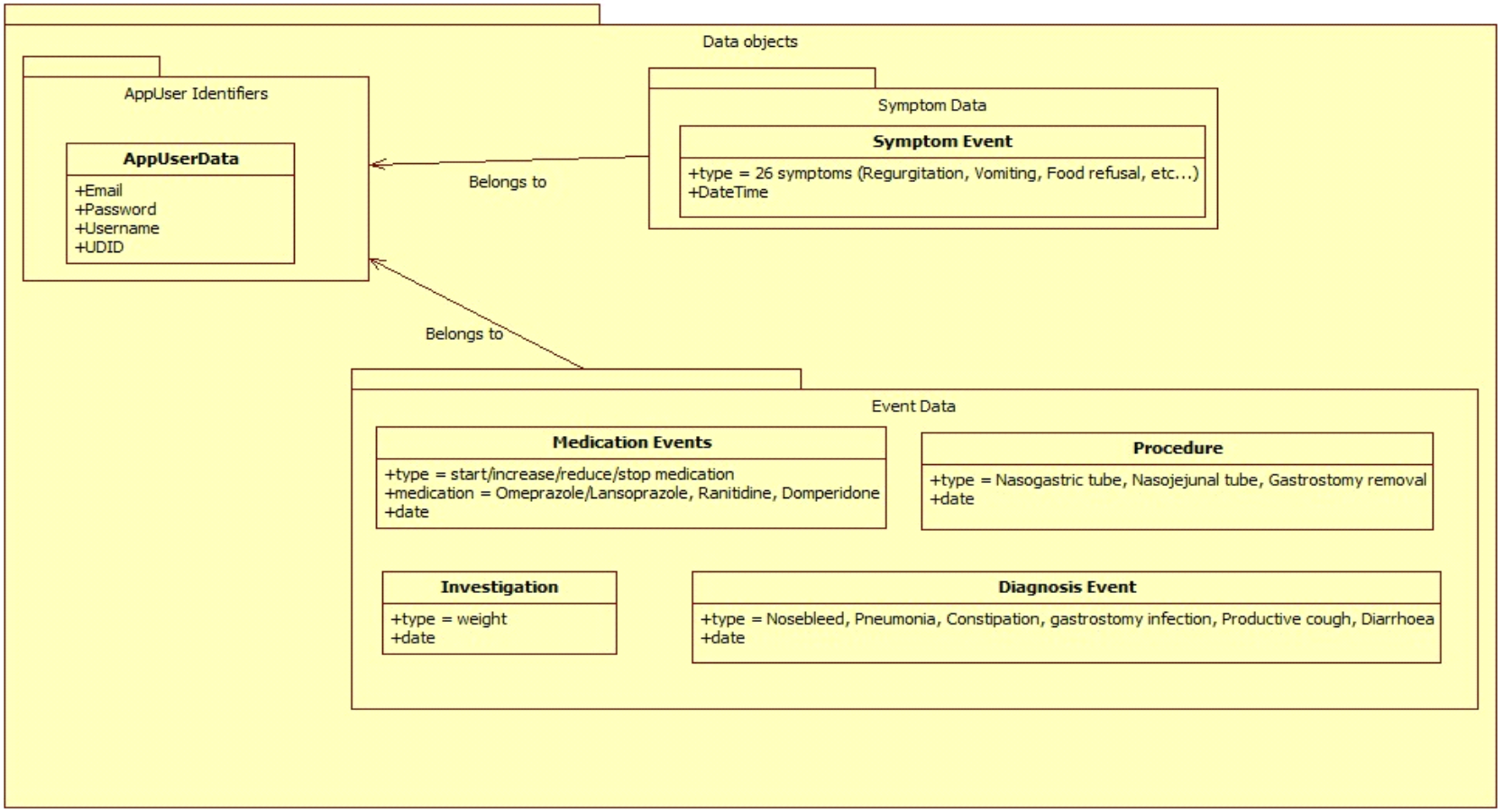
The data to be transmitted from the app user to the clinician and researcher are the data defined as suitable for roles 1 and 4. This data are described in Data requirements section above. The process of data specification summarises the data requirements in an object framework. This process is necessary for defining the objects upon which the coded instructions will act. For object-based programming, this is a useful approach.

We defined a package of 3 data objects. The first object is App User Data. These are the identifiers that define a unique app user. The app user's iPhone contains a unique identifier – the UDID.

Specifying data flows and security

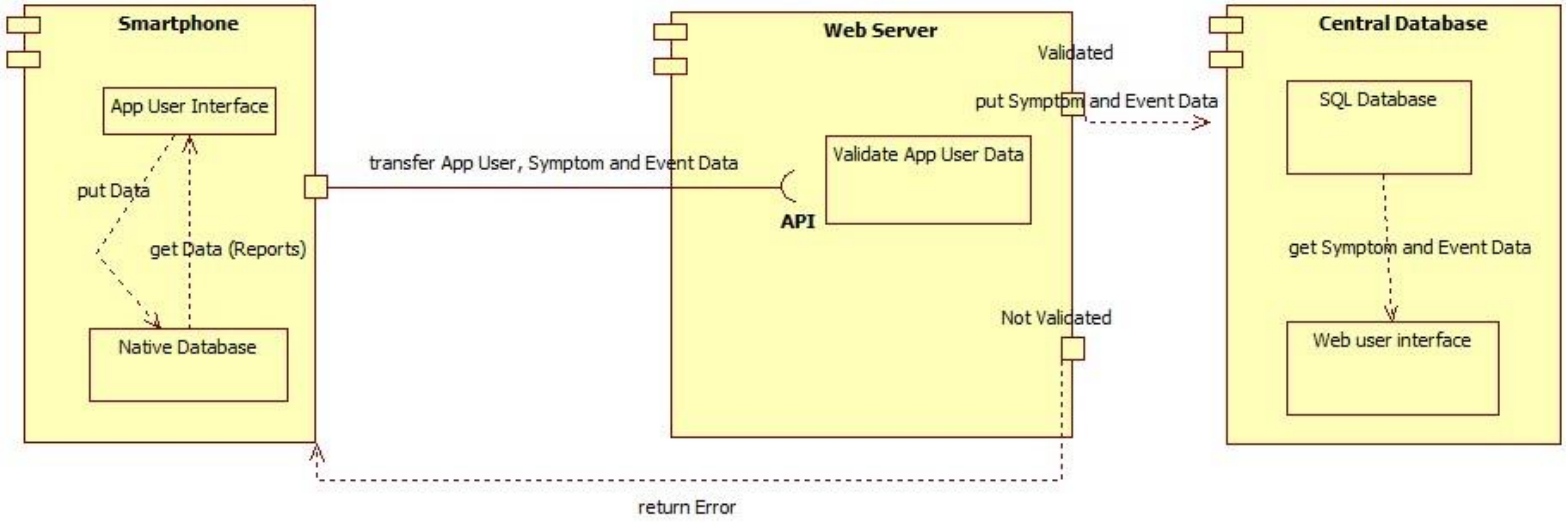
Having identified the data objects, the next step was to specify the transfer of data between the different users and components of the system. The components of the system are the software and hardware elements that participate in the execution of a system. The diagram below depicts the components of the system and data flows between them.

Figure 12: The diagram below illustrates the data object definitions.



The TARDIS system has three main components: a smartphone app, an application programming interface (API) hosted on a web server and a central database.

Figure 13: Components of the TARDIS system and data flows between them.



The main constraint to consider when establishing the data architecture for this system was the data protection requirements this project must meet. The principles of the Data Protection Act 1998 were observed. The following principles were particularly relevant to this project:

"*Principle 3*: Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed.

Principle 7: Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data."

We defined the personal data to be collected and, in keeping with Principle 3, minimised this where possible. Principle 7 necessitates a robust and coherent method for transfer and processing of the data.

The app would be used to collect the following person-identifiable data (PID):

Participant name

App user email address: an email address was a necessary registration detail to enable communication with app users.

Unique Device Identifier (UDID): This is a unique 40-digit alphanumeric string that is assigned to a unique iOS device. It is used by Apple to identify user's devices. To deploy the app to testers, it was noted that UDID would be collected by the dissemination platform e.g. Apple Store, Testflight. This information is a potential identifier. Apps can use the UDID to track app user behaviour. Many mobile telephone companies use UDIDs to generate profiles on their users' behaviour and preference. Although a UDID cannot be used to trace a specific user without access to the UDID provider's records of private ownership, UDID's may be used to generate market intelligence on a user's behaviour.

We would have liked to record demographic information important to the understanding of this condition e.g. data of birth of patient and gestation. However, for the purposes of a pilot study with yet unproven feasibility and security, it was considered that collecting person-identifiable data at this level was not justifiable.

WIREFRAMING THE USER INTERFACE

"Wireframing" is a term that arises from website design. It is the generation of a visual guide or blueprint that encapsulates the layout of the website. Wireframing is used to describe the key design components and interactions of a website i.e. Information design, navigation design, and interface design (26).

The wireframe may be further detailed and annotated to capture the desired functions, content and behaviours of the page elements. The images have no underlying code or functionality. The purpose of the wireframe is to act as a scaffold onto which the developer can add the code to deliver function. We applied wireframing to this project as it offers great utility. We found wireframing to be a very useful tool in enabling the conversation between the researcher (EM) and the app developer (RG) and Java developer (ABC). In its most basic form, wireframing can be done by sketching images on a sheet of paper. We used custom software to generate the wireframes (Balsamiq®, Balsamiq Studios LLC Sacramento CA, USA). Balsamiq allows users to create low-fidelity images. These images are purposely designed to look like sketches. The rationale is to signal that the wireframes are guidelines for the work in progress, rather than rigid requirements to be delivered.

Information design for the TARDIS:REFLUX app

Based on the requirements specified, we used wireframes to describe the information to be carried in the user interface. For example, in the wireframe below (**Figure 14**) we indicate that the user interface should contain the information for consent.

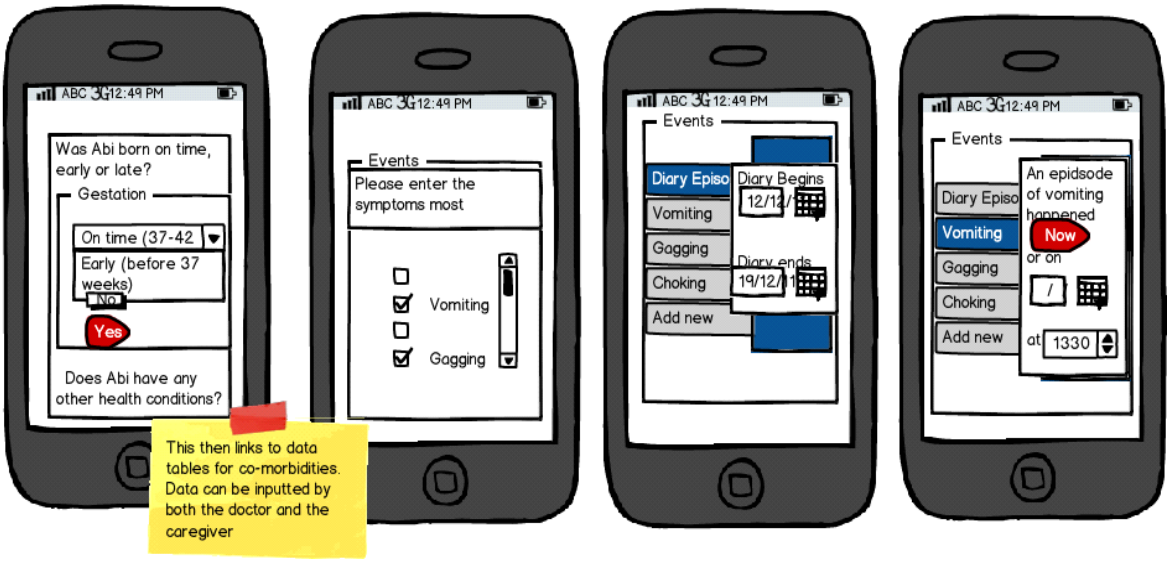
Figure 14: Wireframing the information to be contained in the consent and registration process.

Navigation design for the TARDIS:REFLUX app

The researcher can visualise the user's progress through the views of the app. The user's progress through views is also known as workflow. The basis of the workflows was an understanding of clinical workflows. When taking a clinical history, a clinician works logically through the scenario, starting with open questions and gathering detailed information from focused, closed questions. The questionnaire devised for the TARDIS:REFLUX app contained some focused, closed questions. These were presented in a workflow logic resembling the clinical interview.

In the sequence of images below, the researcher envisions the workflow for navigation through the registration process. The wireframe diagrams are annotated to further detail transmission of data from this user view to the central database.

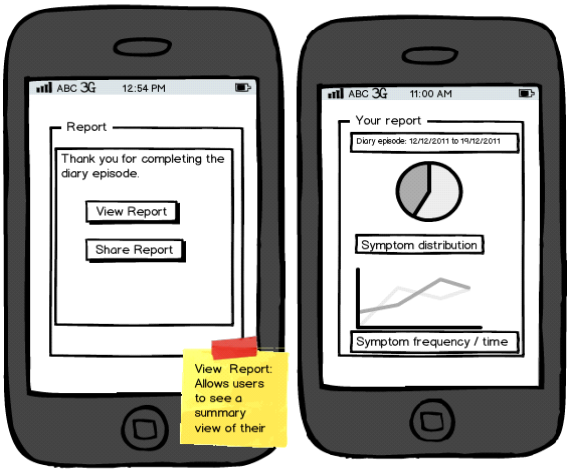
Figure 15: Wireframe describing the envisioned navigation design



Interface design for the TARDIS:REFLUX app

The researcher also specified elements of user interface design through wireframes. In this example, the researcher specifies the appearance of in-app reports. The goal was user-generated graphical summaries of the inputted data, stored in a native database. The illustration is a pie-chart summarising the burden of symptoms and a line chart showing symptom frequency over time.

Figure 16: Wireframe illustrating user interface design.



Summary

Specifying and wireframing the project was analogous to defining variables and outcomes. In contrast to standard experimental approaches, the researcher specifies the project whilst keeping in mind that

another actor- the developer- will implement the vision. Therefore, project specification is of necessity a process of translation: experimental measures and design are translated into the language of software development.

Based on the specifications above, the developer was able to begin work on coding the user interface.

CHAPTER 4: DESIGN AND IMPLEMENTATION

In previous sections, we have described the specification of the TARDIS:REFLUX app. In this section, we describe the design of the user interface. We also describe how the app software was mounted on the specified hardware i.e. the iPhone smartphone.

This stage of software development was an iterative conversation between the app developer (RG) and researcher (EWM). The researcher provided a data model and wireframe suggestions for the user interface. The developer, working on the iOS XCode® platform, would write code in Objective C to generate user interfaces.

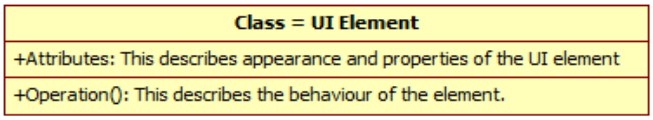
Using the XCode testing platform, the app developer would then share the latest 'build' of the app with the researcher. The researcher would download and review this build and respond with feedback. Issues were tracked using an online issue tracker (Trac). Further consultation and modifications would follow. Subject to satisfactory progress, the next step of views would be attempted. This step of the process took 4 months to complete.

The user interface (UI) elements, their characteristics and interactions are described below. As the researcher was specifying this research tool to app industry professionals, it was necessary to apply the digital tools that are the industry standard. The researcher learned and utilised Unified Modelling Language (UML) diagrams. UML is a stylised way of describing the characteristics and behaviours of elements in a UI.

The UI elements are also known as objects. Based on common characteristics, these objects can be grouped in classes e.g. Navigation bar, Picker, etc. For each class of objects, attributes are described e.g. size, background colour etc. The operation of each class is also described. Operations describe the behaviour of a class e.g. tapping on a navigation bar allows the user to navigate to another view. UI elements in this app will be used to "get" views, and "set" the value of inputs to be used in subsequent steps of the workflow.

This information about objects, classes, attributes and characteristics is presented in a standardised way within a UML diagram. The class of a UI element is described in the top row. The second row contains the attributes of the class. The third row contains the operational characteristics of the class.

Figure 17: Unified Modelling Language (UML) diagrams, specifically components of class diagrams are used to describe user interface (UI) elements.



The StarUML application(222) was used to implement UML diagrams. To populate the user views, we selected elements from a large menu of UI elements available in the iOS framework. These include:

- Navigation View Controllers
- Navigation Bar
- UI Buttons

- Alerts
- Pickers
- Date and Time Picker
- Text Picker
- Number Picker
- Action Buttons

USER INTERFACE DEVELOPMENT

The user interface (UI) of an app is comprised of a series of screens or views. To complete tasks e.g. data entry, the user navigates through views and interacts with elements within the views. We developed the user interface by dividing the work into a number of views. The views were developed in the following sequence:

- Start page user interface
- “Add symptoms” user interface
- “Add views user” user interface
- “View report” user interface
- Help pages
- User login views

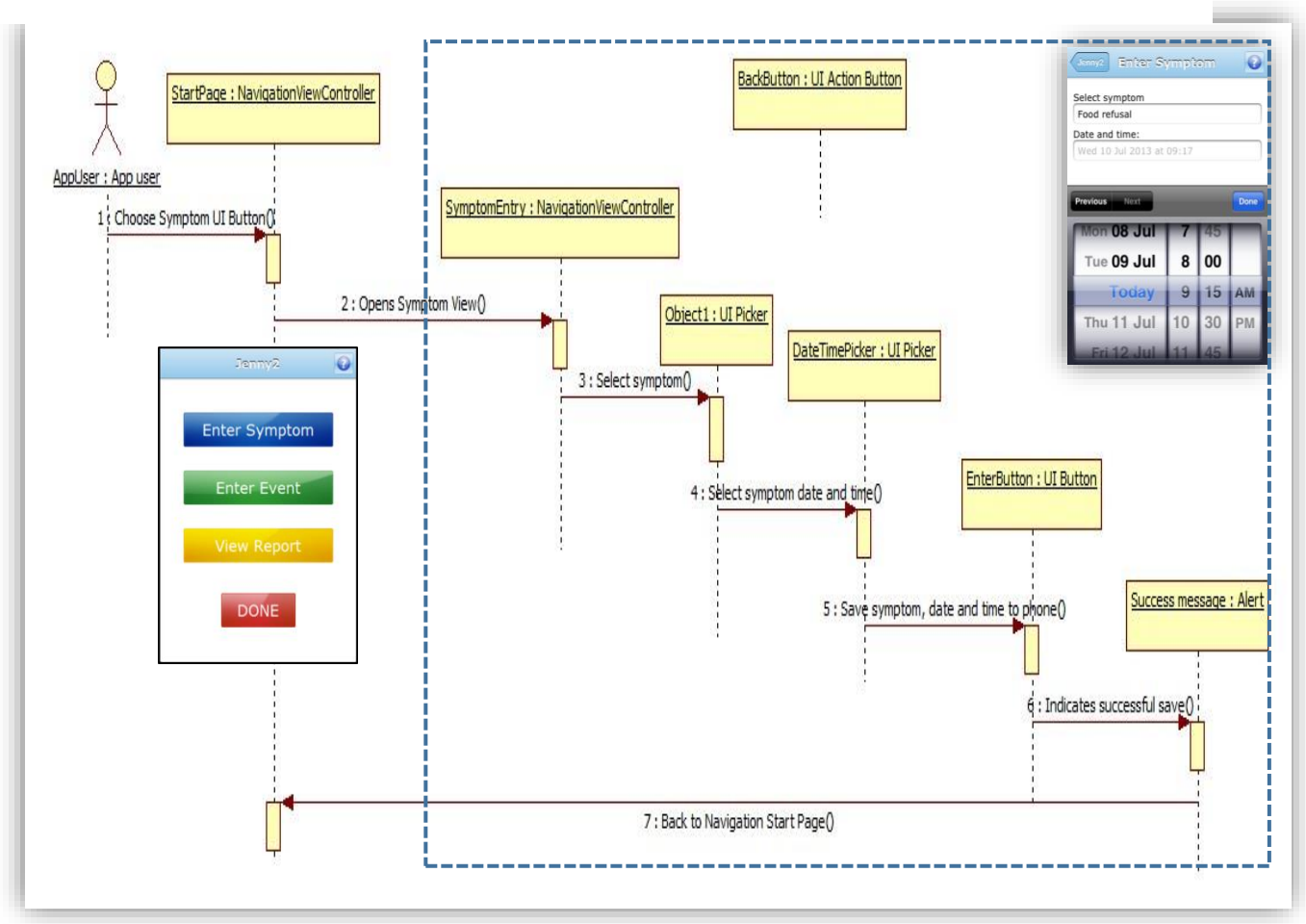
The login views were completed last because these were dependent on defining user management and security. User management in turn was dependent on app-server interactions and the existence of an application programming interface.

Start Page UI

This page is the launch page from which users can access all the functions of the app. Navigation starts here. We chose to have a simple page with three buttons reflecting the three key functions of the app i.e. adding symptoms, adding events and generating reports. Sequence diagrams were used to specify the navigation process envisioned from page to page. Aesthetic element of the user user interface were also carefully specified.

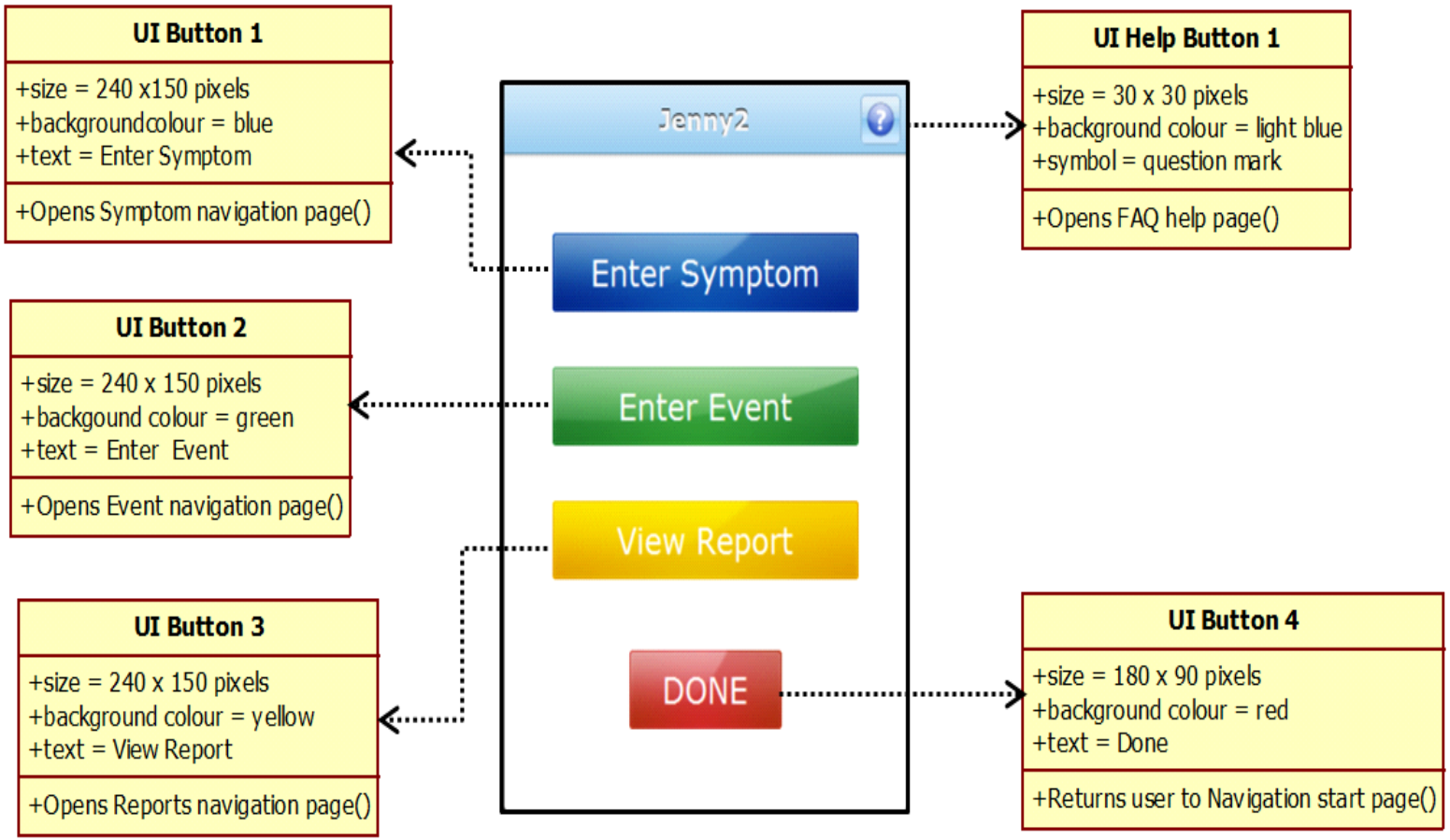
In the UML diagram below, the placement, soze and colours of the buttons on the start page are specified. Furtehr, the behaviour of the button is specified e.g. the “Enter symptom” button is speficied to [+Open Symptom navigation page].

Figure 18: UML sequence diagram demonstrating the procedure for navigating from the start page to selecting date and time.



Aesthetic elements of the user interface were also carefully specified. In the UML diagram below, the placement, size and colours of the buttons on the start page are specified. Further, the behaviour of the button is specified e.g. the “Enter symptom” button is specified to [+Open Symptom navigation page].

Figure 19: User interface (UI) elements in the navigation start page.

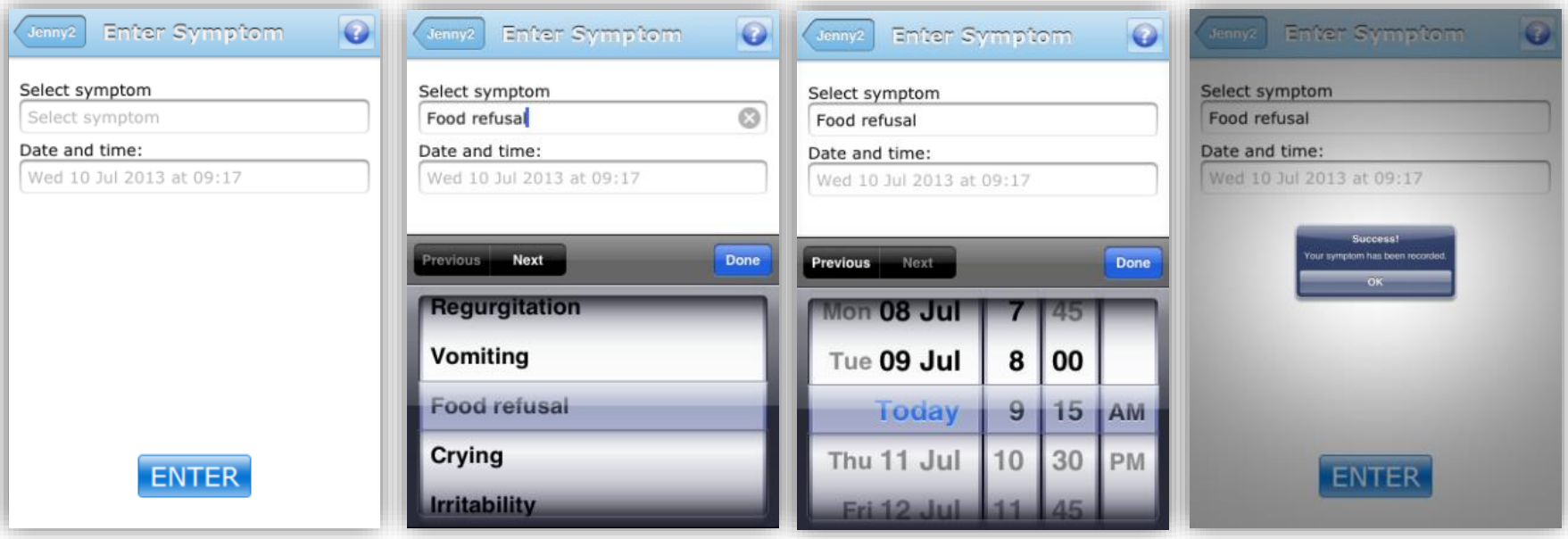


We selected different colours to delineate each key function. For simplicity, we chose primary colours where possible. This also gave the app a “playground” feel utilised extensively in products aimed at children. The user view has been annotated with the behaviours and attributes of each button. These behaviours are specified in sequence diagrams e.g. Figure 18.

Symptom UI

We specified user navigation from the Start Page to the Symptom View. The images in Figure 20 are taken from the app and demonstrate the resulting workflow for adding symptoms. It describes the seven steps required to select a symptom, attribute a date and time, and save this data to the native database.

Figure 20: Resulting UI for symptom entry



Events UI

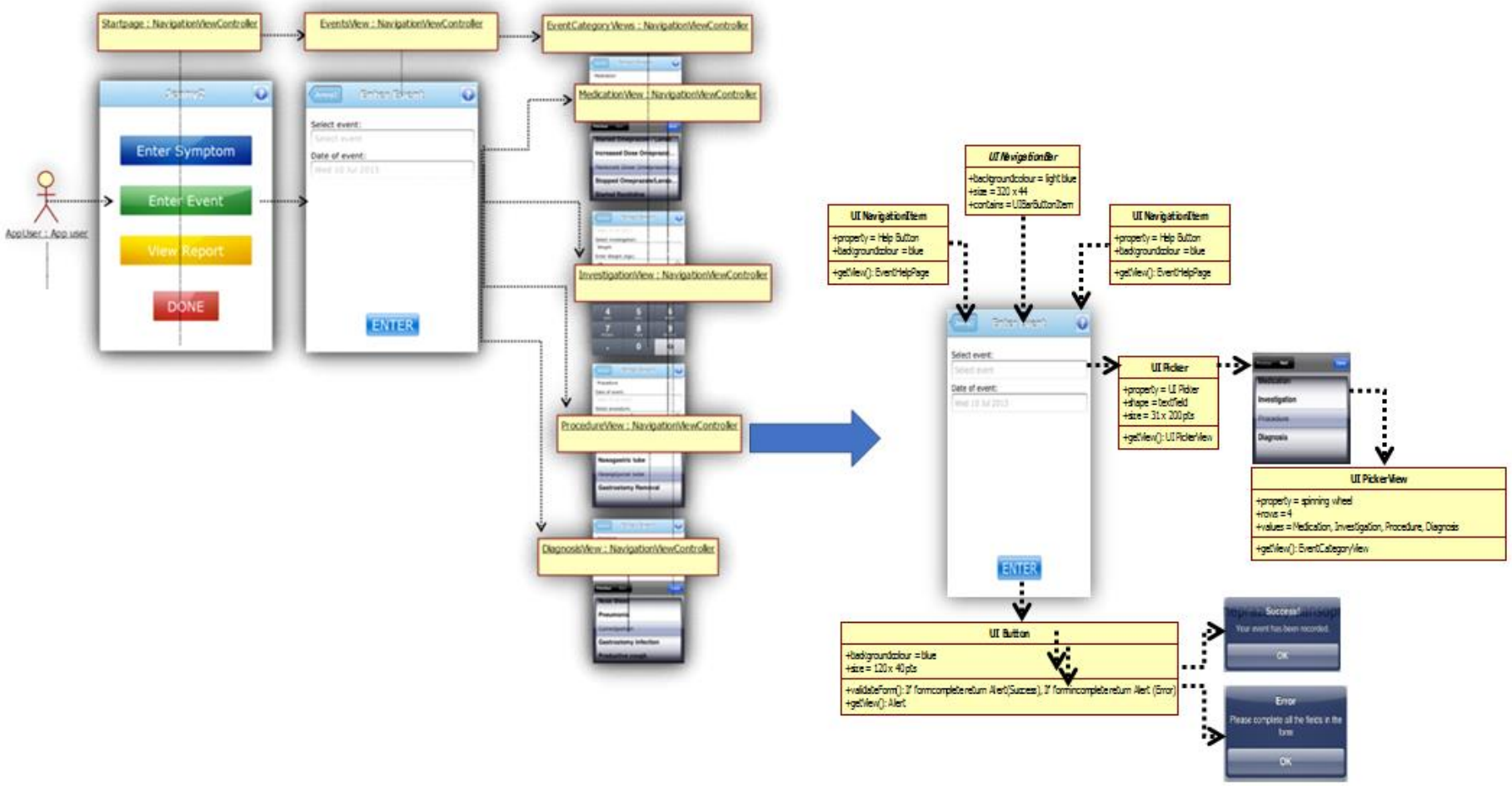
Events were categorised according to the classification described in the data specification.

There were 4 event categories i.e.

- Medication
- Investigation
- Procedure
- Diagnosis

Each of the four categories each had sub-menus. This resulted in a multi-level workflow illustrated below (Figure 21). The UML specification of the workflow for ‘Procedures’ is illustrated below. Other workflows are sufficiently similar not to require reproduction.

Figure 21: Navigation through events multi-level UI. In this workflow, a 'Procedure' event is selected

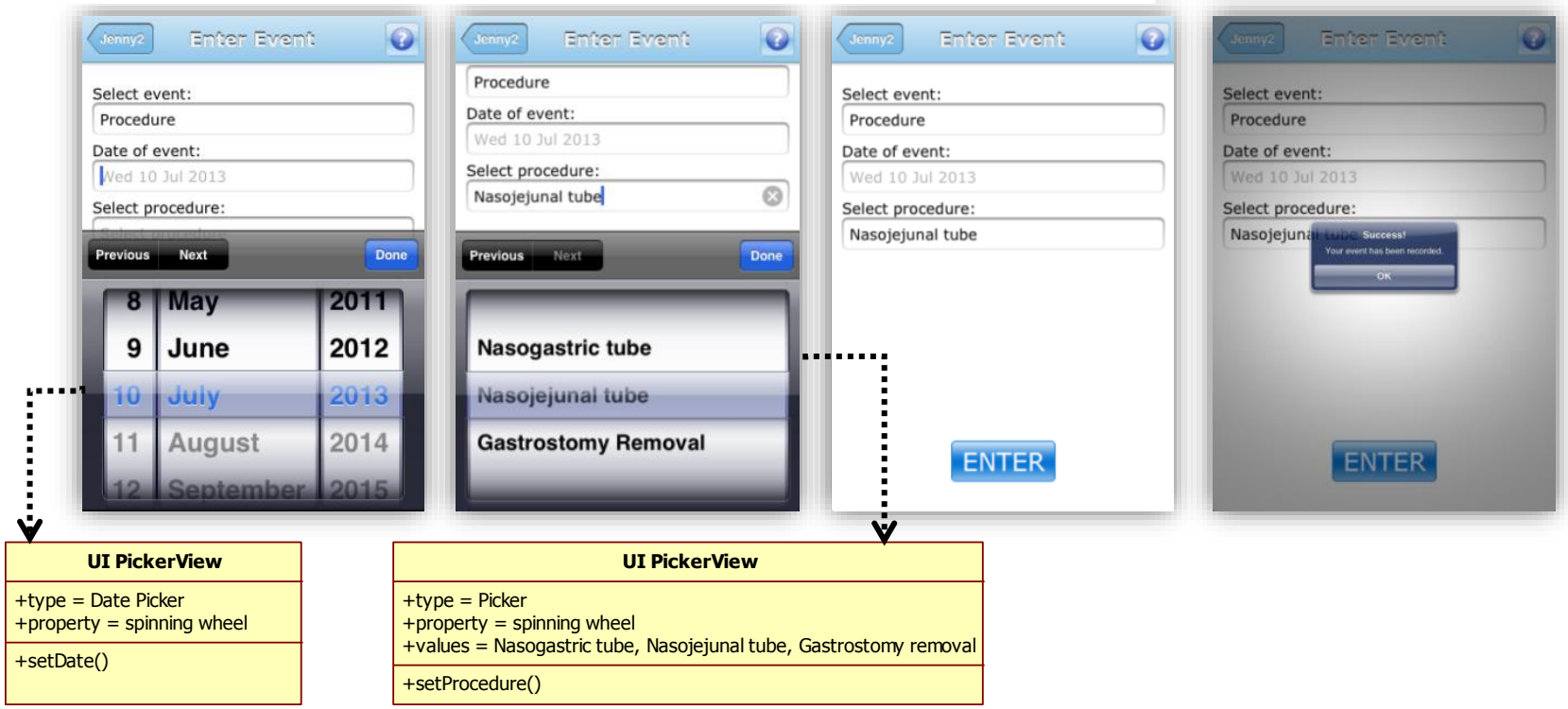


For ‘Procedures’ , those suitable for caregiver recoding (category 4) were:

- Nasogastric tube insertion/change:
- Nasojejunal tube:
- Gastrostomy tube removal

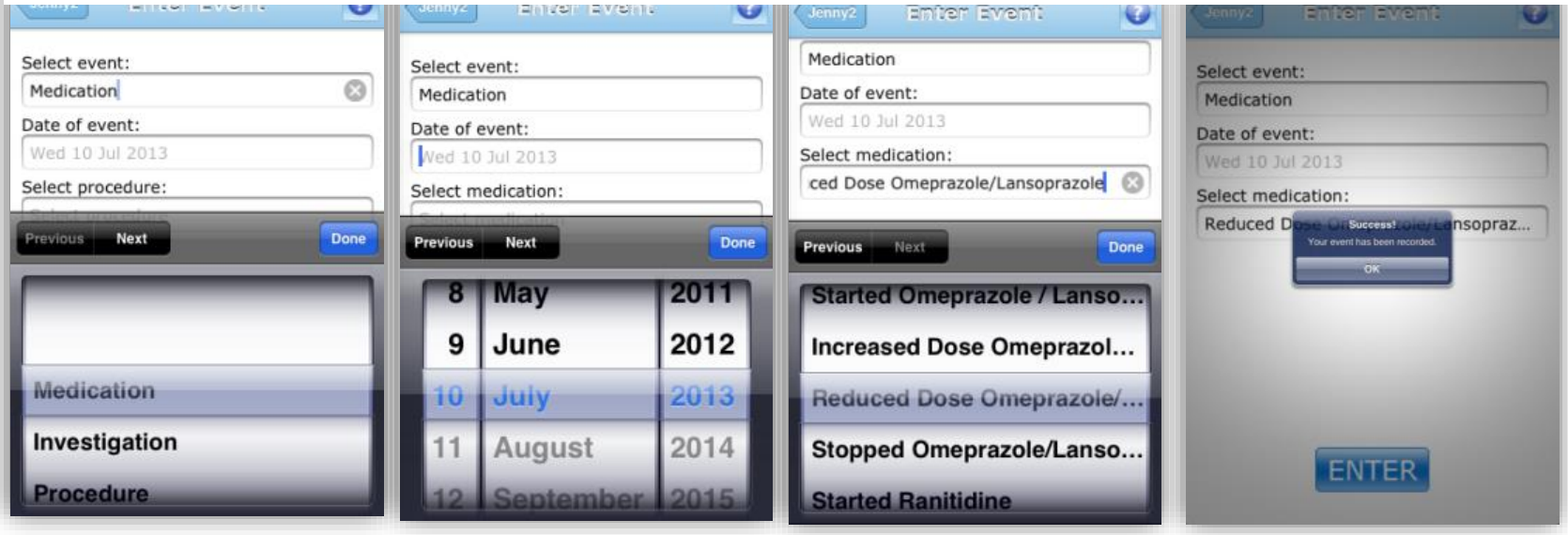
The resulting UI is illustrated in Figure 25 below.

Figure 22: Screenshots demonstrating UI for recording procedures



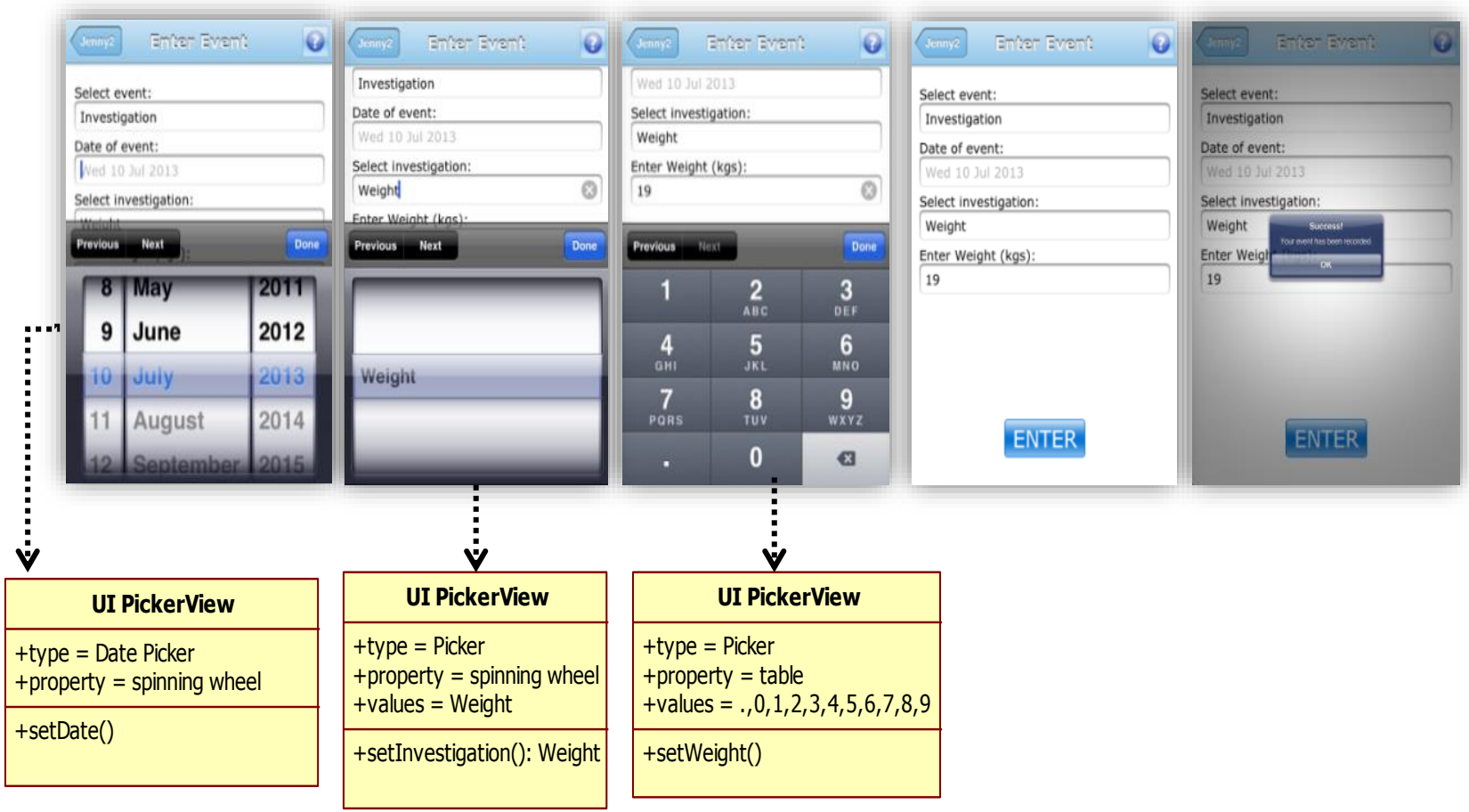
For medications, the drug was specified and a qualitative assessment of the dose of dose changes was made (Figure 23).

Figure 23: Screenshots of the UI for recording medication events



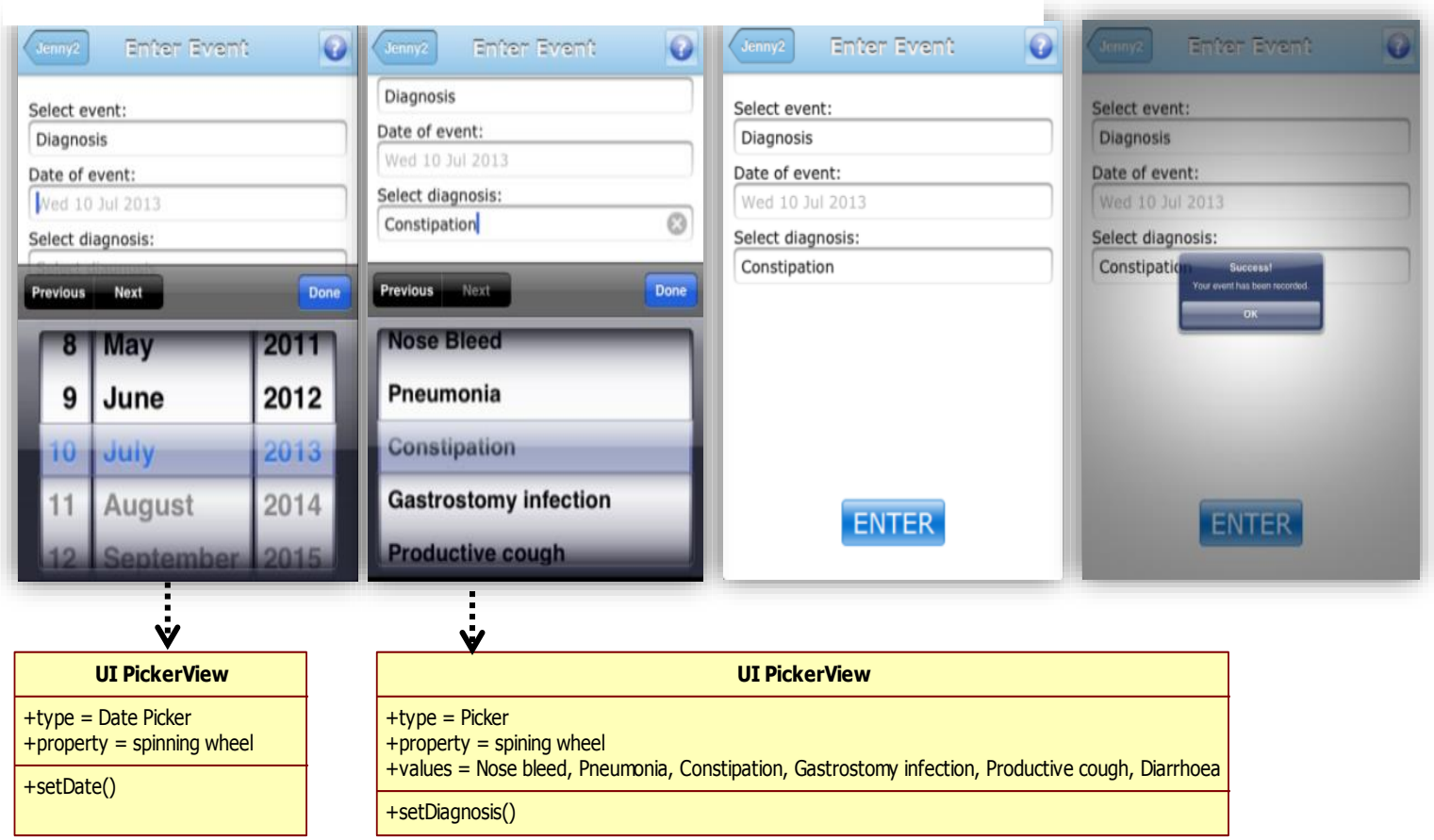
For the ‘Investigation’ event category , app users enter weight as an integer. A screenshot for weight entry UI is shown in Figure 24.

Figure 24: Screenshot of the Investigations UI for weight entry



‘Diagnosis’ events were list of conditions related to GOR, specified in the previous section (Events data). The UI workflow is illustrated in Figure 25 below.

Figure 25: User interface views demonstrating the workflow for recording a diagnosis.



Reports UI

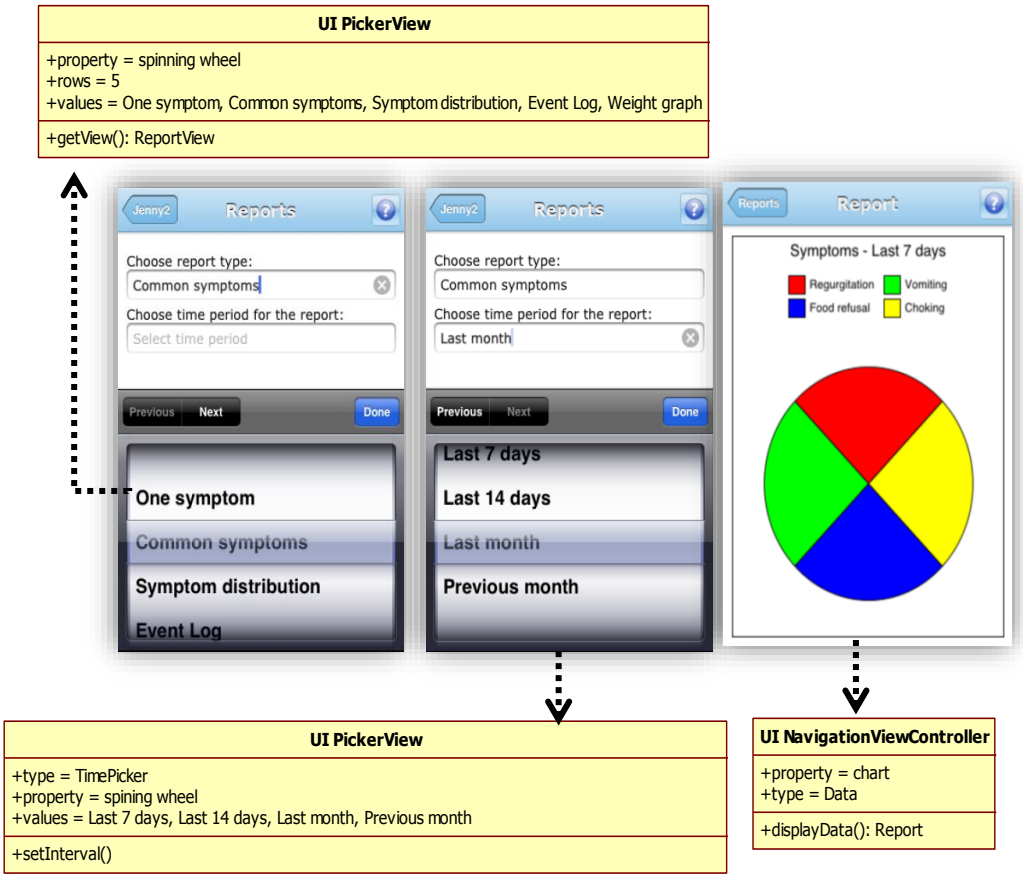
From the StartPage navigation view controller, the third navigation option for users to select is “View Reports”.

We designed reporting views based on the data available and the feasibility of generating views. The reports were based on an algorithm that summarises symptoms or events for a defined time period, then generates a view. Reports summarised a single symptom, common symptoms, symptom distribution, events and weight. The time intervals were pre-defined as:

- Last 7 days
- Last 14 days
- Last month
- Previous month

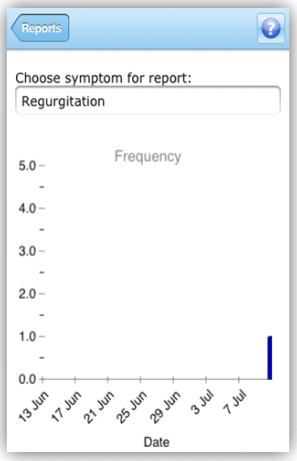
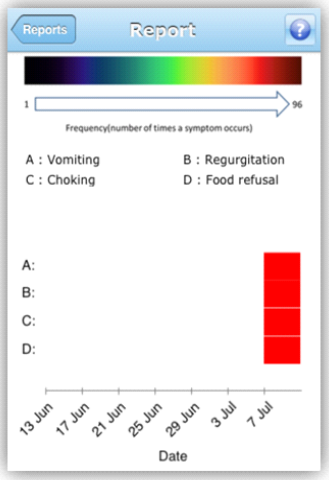
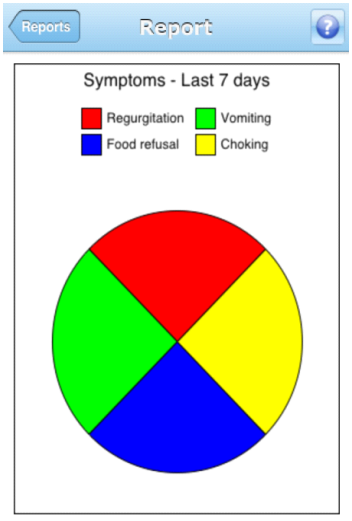
The screenshots in Figure 26 illustrate the workflow that enables app users to generate various reports. UI elements are annotated.

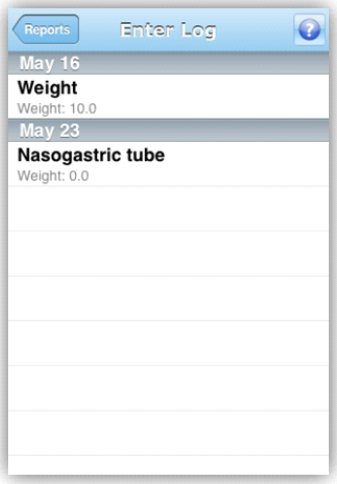
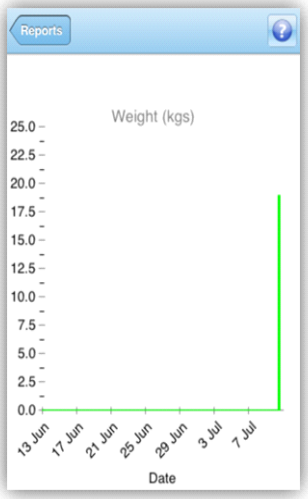
Figure 26: UI views demonstrating workflow used to generate and view reports.



There were five different reports available to users.

Table 21: Reporting views

Report	Screen shot
Single symptom report : This is a frequency diagram which charts the number of times a single symptom is experienced per day within a certain time interval.	
Common symptoms heat map: This heat map /chloropleth map is based on a matrix with symptom frequency on the Y axis and date on the X axis. Symptom frequency is then represented as a colour. We used a full spectral colour progression to represent symptoms ranging from 1-96.	
Symptom distribution pie chart: This is enabled visualisation of the various symptoms experienced in a period of time.	

Report	Screenshot
Event log: This is a table that lists events occurring within the defined time interval.	
Weight chart: This is a line graph with weight on the Y-axis and date on the X-axis. This chart enables users to view the trend in weight over a defined time interval.	

USER MANAGEMENT

Access to and use of the app were restricted. In the data specifications, we required users to identify using three points of identification. These identifiers are an email address, username and password.

The 3-points of identification were validated at two levels:

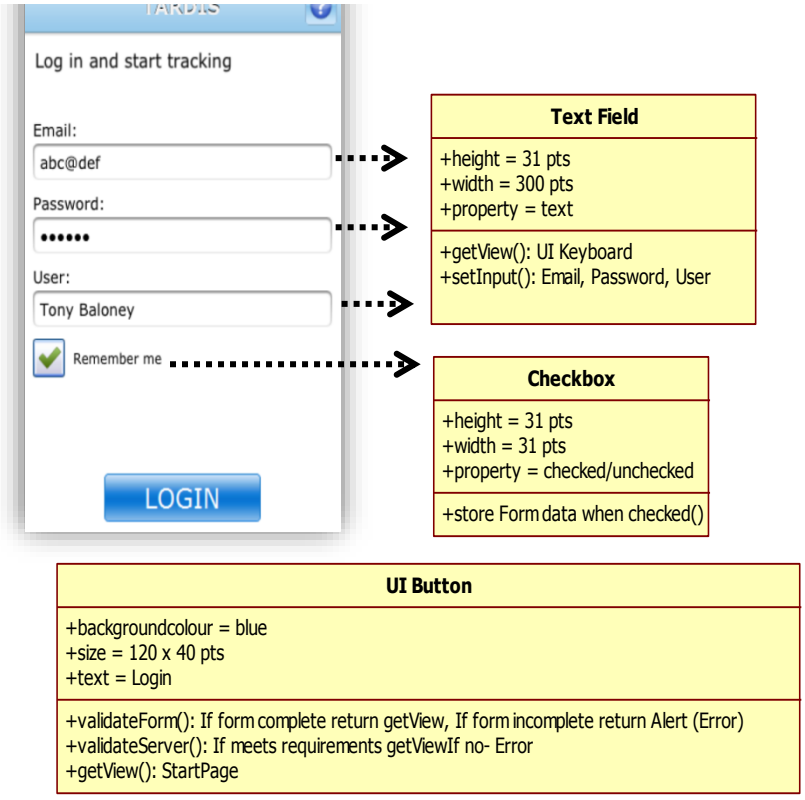
1. App validation: user is validated on the app if the individual dialogues are correctly populated.
   - Email: must be in the format *@*.
   - Password: an 8-digit string that can be chosen by the user.
   - username: no validations. Can be a patient's name or nickname. This is potentially person identifiable data.
2. Application programming interface (API) authentication. that transmits data between the app and the online database. If app validation criteria are met, the unique data object is transmitted to the API for authentication. The API is further described in the next section on Data Storage and Transfer.

The 3 individual identifiers do not- individually needs to be unique to each app user. However, the wrapped package of 3-point identifiers serves as a unique data object that identifies and defines individual app users and corresponds to an individual account. Therefore, parents with more than one child with GORD could set up an account for each using the same app and smartphone.

User Login views

The specifications of the user login resulted in the UI below.

Figure 27: Login page view annotated to demonstrate UI elements, appearance and behaviour



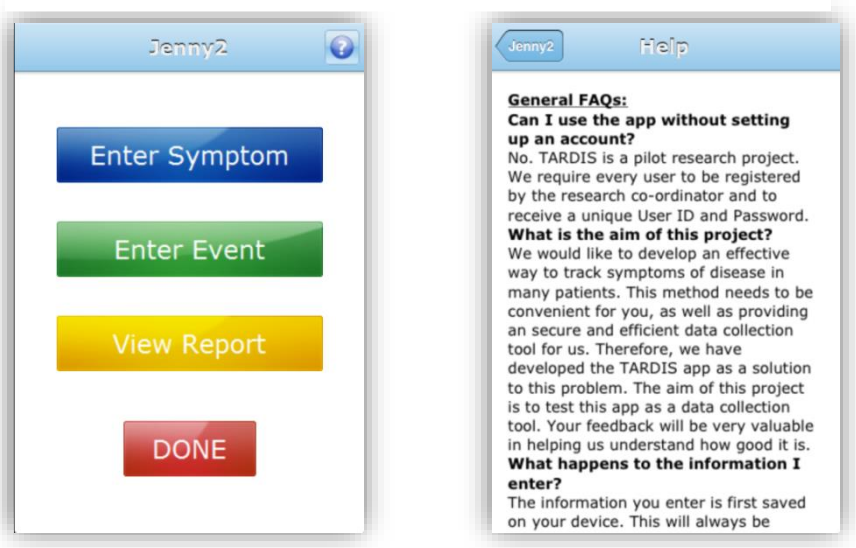
A “Remember me” checkbox was added to the login page. Checking this box generated a ‘cookie’ for this form. A cookie is a small data file that is generated and stored on a web browser. Data contained in this instance are the login text that the user has previously entered. Maintaining user login information streamlines the login process and user interaction.

Help pages

We anticipated that users may need guidance when navigating through app views. We designed help pages to correspond with each view. Care was taken to ensure that patient-readable terms were used. Help pages are accessed on clicking the"?” icon in the top right-hand corner of the screen.

The login help page contains notes describing the information required for each field. Importantly, this page also contains contact information for users to reach the researcher and a link to the project website. It also contains user management information i.e. help for users who have forgotten their username and password. Help pages for the Start Page navigation view (Figure 28), Symptom and Event pages as well as Report views were designed and populated.

Figure 28: Help page associated with Start Page navigation view



DATA STORAGE AND TRANSFER

The data inputted by app users was stored on the native SQLite database on the iPhone. iOS includes a SQLite relational database engine. The app developer creates and configures a database object when initialising the application. Data submitted through the user interface are stored in this native database.

As defined in the data requirements, a method was required to transmit data from individual app user’s devices to a central database. An application programming interface (API) is a protocol that specifies how software components interact with each other. The protocol may contain a pre-defined set of routines to complete a specified task. For the TARDIS API, the components communicating were the app and a central database. Transmission of data between the app native database and the project database server was achieved using an application program interface (API). A second iterative conversation between the research and the API developer (AC) was required to achieve this goal.

DOCUMENTATION

In software development, the purpose of documentation is two-fold. Firstly, there is developer-facing documentation which annotates the code within the project. This is a series of explanatory notes which developers can revert to when reviewing or updating the software. The documentation for the project is available in a series of annotated UML diagrams summarised in the appendix. The full source code for the app and the API are published and publicly available on a web-based open-source project hosting platform Github (27).

Secondly, there is consumer-facing documentation. These are documents created to help the consumer use the product. For the TARDIS:REFLUX app, we deployed consumer-facing notes on two platforms. Firstly, the help pages within the app give descriptions of functions and guidance for password recovery. Secondly, we placed more detailed help-pages on the project website (www.ucl.ac.uk/tardis), with screenshots of each page of the app and explanations of data flows.

Summary

The TARDIS:REFLUX app user interface was developed through an iterative process. The developers translated the researcher's specifications into a navigation views in a series of 'builds. Builds were implemented on the iOS devices held by the development and the researcher. The iterative design and implantation of the TARDIS:REFLUX app was the 'alpha testing' phase. Alpha–testing can be defined as a trial of software conducted by users in the developing organisation early on in the development cycle. At the end of the process, a viable app was produced and made ready for testing with end users. Testing a software product late in the development cycle, with testers external to the developing organisation, is known as beta testing. In the next section, we describe beta testing the TARDIS:REFLUX app.

CHAPTER 5: BETA-TESTING TARDIS:REFLUX

Once the 'proof-of-concept' prototype was available, the next step was to test the app with its intended audience. In the software release cycle, beta testing is analogous to a field test, in which the software product is tested by a selection of potential consumers. The aim of the beta test is to identify performance issues and assess usability.

In 1981 Al-Awar et al(223) observed that "Technical advances in the past 20 years have brought sophisticated computing power within reach of a large population of large population of occasional, casual or discretionary users". In the 21st century, we acknowledge that this increase in computing sophistication and power requires commensurate improvements in systems usability testing.

Assessing usability

The study of human interaction with software interfaces is well established. The International Standards Organisation for human-system interfaces (ISO 9241) is a multi-part guideline that encapsulates the key aspects of this interaction. The revisions made to this standard over time provide footnotes to the narrative of evolving interfaces and expanding consumer expectations. This standard initially began as a guide to ergonomic application of computer screens in the office. The standard was most recently revised in 2008, and has evolved into a compendium of guidelines which covers systems ranging from software to hardware. Ergonomic considerations include visual, tactile and haptic human-system interactions. The sub-standard that is most relevant to the software interface – and hence app development is ISO 9241-110.

Software testing frameworks have evolved in step with ISO standards. The ISO 9241-11 is a standard developed to describe ergonomic standards for human-computer interaction. The standard (revised in 2006) suggests the following measurable dimensions when assessing a system:

- suitability for the task (usability)
- suitability for learning (learnability)
- suitability for individualisation
- conformity with user expectations
- self-descriptiveness
- controllability
- error tolerance

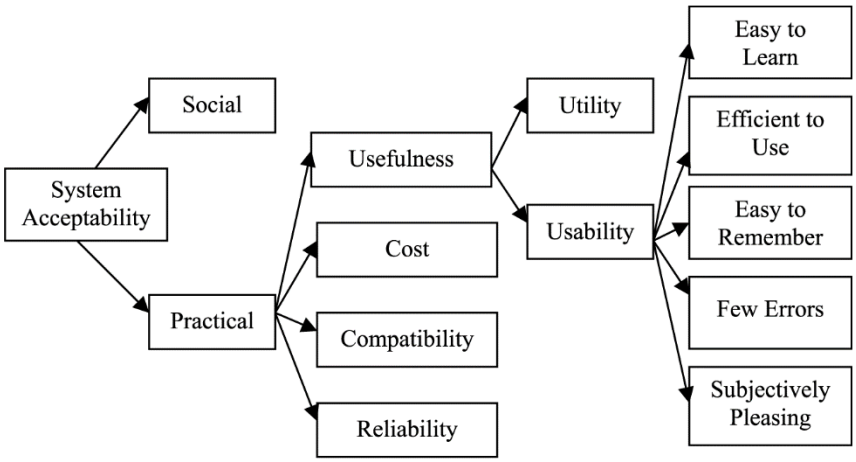
The software engineering body of knowledge (SWEBOK)(224) details the knowledge area of software testing and provides the framework for testing. The SWEBOK standard is written by the ISO collaboration with the leading professional body in engineering, the Institute of Electronic and Electrical Engineers Computer Society.

System testing can be approached on four levels: unit testing, integration testing, system testing and acceptance testing. Unit, integration and systems testing are achieved in a development environment. Acceptance testing is done at the end of development, and at the level of the user. This level of testing checks "system behaviours against the customer's requirements"(224). Usability testing should utilise

representative participants, representative tasks and a representative test environment with participant activities monitored by one or more observers(225).

In this chapter, we deal exclusively with acceptance level testing. We aim to discover if the TARDIS:REFLUX app meets the user requirements. System acceptability in turn has many dimensions including cost, utility, reliability and usability. Nielsen(226) provides a useful model for understanding the dimensions of system acceptability.

Figure 29: A model of system acceptability. Reproduced with permission from Nielsen and Landauer (1993)



Source: Nielsen (1993)

Utility can be defined as the ability of a system to perform specific tasks. The TARDIS:REFLUX app was designed to perform the task of tracking symptoms. A simple metric would be the demonstrating the ability of users to track symptoms using the TARDIS:REFLUX app. Another measure of utility would be the application the app a symptom diary for other conditions.

Usability can be defined simply as the ease of use(227). The dimensions that make an interface easy to use include efficiency, ease with which it is learnt or remembered. An interface is also more usable when it is aesthetically pleasing and works with few errors. Usability as a quality of an interface is measured by surveying the intended audience and asking them to rate an interface by ease of use (12). An interface that is difficult to user will lead to low adoption and high attrition of an interface.

It is important to define the level of measurement of a usability test. Testing can be summative or formative. Summative or task-level testing is specific. For example, a user may be asked to complete a task e.g. “Log into this system.” Usability can also be formative i.e. the user is asked to use the system and the researcher observes for problem discovery.

The tool we shall use to measure usability is the Systems Usability Scale (SUS) for summative assessment (See Section II appendix items). The SUS scale was developed as a response to the ISO 9241-11 standard. The SUS scale was developed as a “ready reckoner” scale and first published by John Brookes in 1996[10]. Since inception, this scale has lent itself to assessment of web-based products due to its brevity and ease of application. Data from its application has been generated leading to availability of comparable and perhaps normative data.

The SUS's author envisioned a one-dimensional scale. "The SUS yields a single number representing a composite measure of the overall usability of the system being studied". However, others have suggested that items in this scale (items 7 and 10) are a measure of the learnability of a system(228).

The SUS score has ten items rated on a Likert scale from 1 to 5. Each item is given a score depending on the position on the scale. For items 1, 3, 5, 7, and 9, which are positive questions, the score contribution is the scale position minus 1. For items 2,4,6,8 and 10, where the questions are negative, the contribution is 5 minus the scale position. This will result in positive scores ranging from 0 to 4 for each item. The scores are then summed and multiplied by 2.5, giving a composite measure ranging from 0 to 100.

In the next sections, we describe three experiments established to test dimensions of system acceptability. These are:

1. The TARDIS Pilot Study: the app was released to parents and caregivers of children with GOR. They were asked to compare usability of the app against a standard method (paper diary). The pilot study primarily focussed on the usability of the app.

2. The Reflux UX focus group: a focus group of target users was convened. The app was tested in this environment and users were asked to opine on usability, utility, efficiency and subjective appeal.

3. Dissemination: The app will be made available on general release on the iTunes app store and published on the NHS health apps library. Data will be collected on social acceptability e.g. download rates and patterns, and demographic characteristics of app users and subjects. The uptake of the tool by the public will also provide a measure of utility.

THE TARDIS PILOT STUDY

In paediatric gastroenterology and surgery clinics, the established method of recording symptom data over time is the symptom diary. Parents are offered a paper symptom diary on which they record symptoms and symptom frequency. The app is a tool that utilises the smartphone to deliver a symptom diary.

Therefore, the aim of the pilot study was to compare utility of a new questionnaire interface (the app) against the standard method i.e. paper diary. Of the comparable utility parameters, we focused on usability i.e. ease of use.

We identified a well-validated and standardized tool to measure usability i.e. the Systems Usability Scale (g SUS). The SUS was developed by John Brooke and first published in 1986. Since publication it widely applied for the assessment of hardware, software, website and other interactive technologies[3]. In a review of 2324 questionnaire responses Bangor, Kortum, and Miller [3] found the SUS to have high reliability over a wide range of assessed interfaces (Cronbach's alpha coefficient 0.91/Excellent). It has also been used to compare user experience of web-based versus paper diaries [4].

The SUS is a simple 10-question assessment rating user experience on a Likert-type 5-point scale. User ratings are scored, giving a composite score ranging from 1 to 100. Higher scores indicate higher user satisfaction and higher usability (13).

A full study protocol is available in the Section II appendix items. In the sections to follow, we describe the pilot study design, conduct and results.

Design

The null hypothesis was stated as:

“In tracking symptoms of disease, the usability of a smartphone symptom questionnaire (app) is the same as that of a paper symptom questionnaire .”

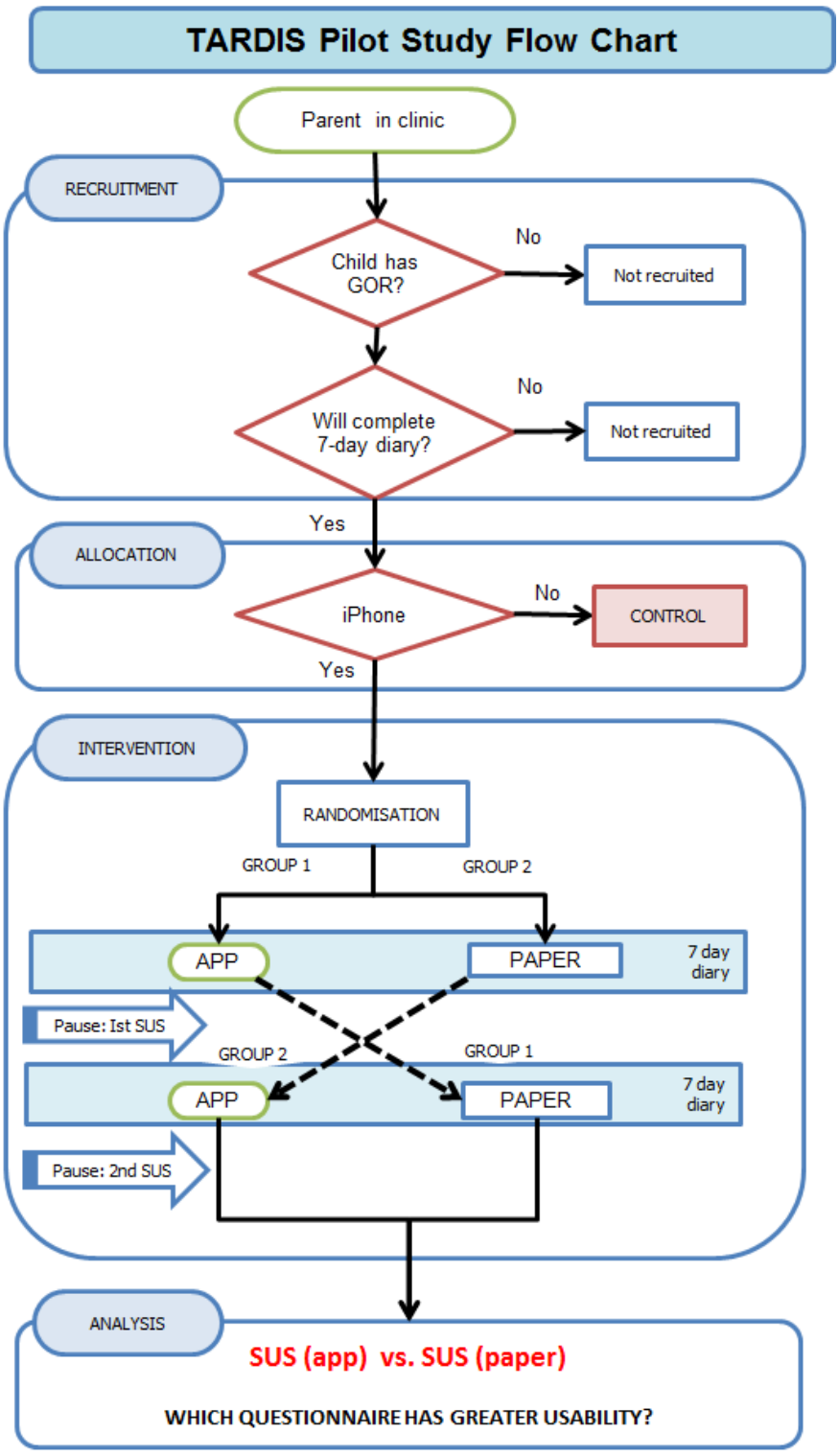
This was a randomised, crossover trial. We administered the symptom questionnaire in two formats, paper and app, over successive 7-day periods. Allocation to randomisation groups is pragmatic. The app is only available on the iPhone platform. Therefore, only participants with iPhones can be allocated to test both the app and the paper diary.

Participants with iPhones completed both the paper and the app in consecutive 7-day diary episodes. The order of use was allocated randomly. As a result of the cross-over, there were two sub-groups within the randomisation group i.e. those who use the app first (A) and those who use the paper diary first (B). At the end of each period participants were asked to rate each symptom diary using the Systems Usability Score (SUS). This crossover design generated paired data for each participant, allowing a direct comparison of user experience for each diary format.

Participants without an iPhone were asked to complete the paper diary and provide an SUS assessment of it after 7 days. They will comprise the control group. They allow detection of any effects attributable to iPhone ownership.

The study algorithm below details the phases of the study i.e. recruitment, allocation and investigation and analysis of outcomes.

Figure 30: The TARDIS Pilot Study algorithm



OUTCOME MEASURES

The primary objective was to measure usability. This was scored using the Systems Usability Scale (SUS).

The summary measures to be reported were:

1. Raw SUS score

2. **Median SUS score for each study group**: Previous studies(228) suggest that SUS scores are not normally distributed. Therefore, a Wilcoxon paired comparison was be used to compare SUS scores between groups.

Figure 31: SUS questionnaire for assessment of TARDIS:REFLUX

Great Ormond Street Hospital for Children NHS Foundation Trust

UCL

Thank you for participating in the Reflux UX Workshop.

Evaluation using the System Usability Scale©

Participant	

Please tick to rate.

	Strongly disagree				Strongly agree
	1	2	3	4	5
1. I think that I would like to use this system frequently					
2. I found the system unnecessarily complex					
3. I thought the system was easy to use					
4. I think that I would need the support of a technical person to be able to use this system					
5. I found the various tasks in this system were well integrated					
6. I thought there was too much inconsistency in this system					
7. I would imagine that most people would learn to use this system very quickly					
8. I found the system very cumbersome to use					
9. I felt very confident using the system					
10. I needed to learn a lot of things before I could get going with this system					

Secondary objectives include the measurement of

1. Response rate: To analyse differences in response rate for paper and smartphone diaries.

2. Validity: To assess face and content validity and confirm the appropriateness and relevance of the selected GERD symptoms.

Face validity will be assessed by expert review. The diary will be reviewed by clinicians working in this field. Symptoms included and written definitions were reviewed to achieve consensus on face validity.

Content validity will be assessed through the pilot of the diary. Participants will have an opportunity to include an "other" symptom, should they find the symptom experienced is not included. Frequent use of "other" indicates that refinements on construct need to be made. Symptoms frequently offered as "other" symptoms will merit review for future inclusion. Ambiguities between "other" symptoms and symptoms included in the diary will be clarified.

3. Measurement range: Frequency of symptoms will take a minimum value of zero (no symptoms recorded). Symptoms are recorded for each 24-hour period in blocks of 15 minutes. This fifteen-minute segmentation is a practical adaptation to ensure that the paper diary is not too visually dense. Therefore, within a 24-hour period, the maximum number of times a single symptom can be recorded is 4 times every hour i.e. 96 times in 24 hours. Clustering of values towards either end of this range will suggest that the scale can be limited further in the revised diary.

SELECTION AND RECRUITMENT

Participants were selected from paediatric gastroenterology and surgery clinics. Potential participants were parents of children with GOR attending clinic. The diagnostic standard was the expert opinion of consultant paediatric surgeons at GOSH.

SCREENING

Participating clinicians will review patient-identifiable data to identify children with gastro-oesophageal reflux from the list of patients attending clinic. Following routine clinic attendance, participating clinicians will invite parents to receive information on the study from the researcher.

If the response is positive, the participants will then be referred to the researcher.

To minimise the potential risk of coercion,

- participating clinicians will only introduce the study and the researcher after the clinic appointment has come to an end
- clinicians will introduce the researcher, but will not offer a view on participation
- The researcher will see potential participants in a separate clinic consulting room, ensuring that the conversation between potential participants and the researcher is not observed by the clinical team.

INCLUSION CRITERIA

1. Participant is a parent of a child with GOR.
2. Participant is willing to complete two 7-day diary episodes.

EXCLUSION CRITERIA

1. The participant's child has no symptoms of GOR.
2. Participant is not willing to complete two 7-day diary episodes.

RECRUITMENT AND INFORMED CONSENT

The researcher will attend paediatric gastroenterology and surgery clinics and administer the recruitment of study participants. The researcher will take informed consent from each subject prior to participation in the study, following adequate explanation of the aims, methods, anticipated benefits and potential hazards of the study.

The researcher has the required GMC Good Clinical Practice (GCP) training with suitable qualifications and experience to undertake this task responsibly. No study data will be collected prior to obtaining consent from the participant.

A Participant Information Sheet (Section II appendix items) will be given to the participant. The researcher will explain the participants are under no obligation to enter the study and that they can withdraw at any time during the study, without having to give a reason. The Informed Consent Form (see appendix Section V) will be given to the participant. If a positive answer is given, participants will be asked to sign and return the consent form.

REGISTRATION

Following consent, participants details will be documented on the Case Report Form (see appendix). The participant will be registered onto the database of participants.

The information required for registration of participants:

1. Person-identifiable data (PID): The following items of PID will be collected and stored in a secure and encrypted document (Registration.xls) on a secure GOSH personal drive. This is described further in the data confidentiality section (5.1).

a. Participant name

b. Participant email address

c. GOSH Clinical Record Number (CRN) of the participants child, whose symptoms are being recorded.

Demographic factors i.e. gender of participant, age and smartphone ownership will also be collected.

PARTICIPANT PATHWAY

In summary, the study has 5 phases. These are:

1. Recruitment and allocation: This is a face-to-face encounter. Participants are recruited following informed and written consent. They are registered onto the database of participants. Those suitable for randomisation are allocated to group A or B. Those not suitable for randomisation are allocated to the control group.

Participants will be allocated to either the case or control group, based on iPhone ownership.

Participants with an iPhone will be allocated to complete both the paper and smartphone diary. The order of completion will be allocated randomly.

Table 22: Study groups

Category	**Group title**	**Allocation**	**Intervention**
Randomised	Group A	Random	App , Paper
	Group B	Random	Paper, App
Control	Control	Participants with another phone brand	Paper only

2. 1st diary episode: The participants have a 7-day period to complete a symptoms diary remotely. Participants in the control group complete their SUS assessment of the paper diary. On returning their paper questionnaires, they have completed the study.

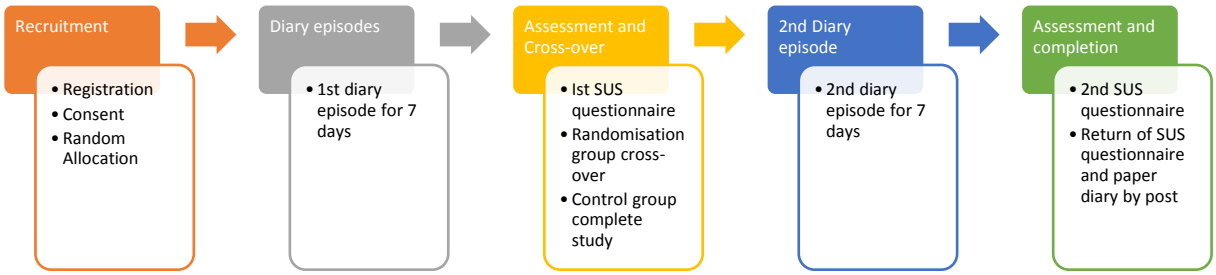
3. Assessment and Crossover: Participants in the randomisation groups complete an SUS assessment of the diary they have used in the 1st diary episode.

4. 2nd diary episode: Participants in the randomisation groups have another 7-day period in which they use the alternative diary.

5. Assessment and completion: Following completion of diary episodes, participants will complete their SUS assessment.

At the end of the study, all participants will return paper diaries and SUS questionnaires by post. Participants who have not completed the SUS for each diary are encouraged to do so. There will be email and telephone reminders to ensure prompt passage between study phases.

Figure 32: Participant pathway



DISCONTINUATION/WITHDRAWAL OF PARTICIPANTS

Participants were able to withdraw at any point. Participants will be classified as having withdrawn (a study dropout) if a SUS score is not submitted. At the end of each diary episode, an email requesting completion of the SUS will be sent. If no response is received, a further email reminder will be sent before this classification is made. The sample size calculation anticipates a 15% dropout rate.

STATISTICAL CONSIDERATIONS

We have an app that is only available for testing on the iPhone platform. The study design is pragmatic, with recognition given to this bias. A YouGov quarterly survey found that 86% of smartphone owners were under the age of 55 years. In a March 2011 survey(229), OfCom reported the incidence of smartphone ownership in the 25 -34 years age bracket as 42%. Amongst adults aged 35-54, the incidence of smartphone ownership was 29%. There are also discernible income differentials between smartphone users. In April 2012, a YouGov Poll identified that BlackBerry users were likely to be higher earners(230). 10% of BlackBerry users admitted to earning over £50,000 a year, compared to 7% of iPhone users and 5% of Android users. Therefore, our randomisation cohort is likely to be younger, wealthier and more likely to be male than the paper diary group.

To mitigate sampling bias, we will document age and gender and test for significant between-group differences, and also treat age and gender as confounding variables in any regression analysis. As we want to limit the person-identifiable and sensitive data collected, we will not gather income information. Sampling bias by income can be mitigated in future studies by randomising participants and providing smartphones to those who do not have one.

The primarily outcome is a comparison of paper diary versus app SUS scores. Therefore, the sample size estimation is based on prior SUS data.

SAMPLE SIZE CALCULATION

Bangor et al (20) analysed 2324 questionnaire responses and obtained a mean score of 70.1 with standard deviation 21.7 and standard error 0.45. The data were not normally distributed.

Table 23 summarises the sample size calculation, for 5% significance at 80% power. We assume an effect size of half a standard deviation represents a significant difference in usability. To detect a 10-point difference in paper versus smartphone SUS score, the study would need 39 participants. Allowing for a study dropout rate of 15%, we will aim to recruit 45 participants.

Formula= $n= [\sigma^2 * (Z_\beta+Z_\alpha)^2]/ \Delta^2$

Table 23: Sample size calculation parameters for TARDIS pilot study

Parameter	**Value**
α	5%
β	20%
Power (1-β)	80%
σ (standard deviation)	21.7
Zα (2-tailed)	1.96
Zβ for 0.8 power	0.84
Δ (effect size)	10
n (sample size for each group)	n=[σ2 * (Zβ+Zα)2]/ Δ2 n=(〖21.7〗^2.〖(0.84+1.96)〗^2)/〖10〗^2 n = 39

RECRUITMENT RATE

Based on previous recruitment experience (as research coordinator for the REMOS trial), 10% of patients attending paediatric surgery clinics will have symptoms of gastro-oesophageal reflux. In previous studies, 80% of parents approached agreed to participate. We estimate similar recruitment rates for this study. Based on the calculation below, we anticipate a recruitment rate of 28 participants/month and a 2 month recruitment period.

Table 24: Anticipated recruitment rate

Characteristic	n
Patients attending gastroenterology clinic	150
Patients attending surgery clinic	100
Weekly attendance (total)	250
Monthly attendance	1000
Percentage with GOR symptoms	10%
Potential number eligible for recruitment / month	100
iPhone possession (assumed at 35%)	35

Anticipated recruitment rate	80%
Recruitment rate / month	28
Time required to recruit 45	2 months

As study numbers are small, there will be no provision for interim analysis.

DATA MANAGEMENT AND QUALITY ASSURANCE

The study was reviewed and approved by an NHS Research and Ethics committee (NHS REC 12/NW/0837). Details regarding data storage and transmission are précised below. Full details of the research protocol and data collection and transmission tools developed for this project are available in the appendix (Study protocol).

CONFIDENTIALITY

All data will be handled in accordance with the UK Data Protection Act 1998. A systems level security policy (SLSP, see appendix) has been written and submitted as part of the NHS Research and Ethics application for this project.

Person identifiable data of participants will be stored on the GOSH Local Area Network, behind the NHS firewall. All other study data will be anonymized. The unique study number (USN) will be used for anonymization and identification. The data collection diaries will not bear the subject's name or other personal identifiable data. Regardless of format, the following data will be collected by the diary.

1. Unique Study Number
2. The symptom / event
3. Date and time of symptom or event

It will be the responsibility of the researcher to ensure the accuracy of all data entered in the diaries. The researcher will have sole responsibility for data collection and handling, and sole access to the study database.

DATA COLLECTION TOOLS

Standardised data collection tools and secure data storage tools have been developed. These are detailed in the table below and further described in the SLSP.

Table 25: Data collection and data storage tools

Element	Role	Description
REGISTRATION.XLS	Data storage tool	This is an Excel document containing the anonymising key, i.e. Unique Study Number. Contains person identifiable data of participants. Stored on the GOSH Local Area Network, behind the NHS firewall.
Smartphone / paper diary	Data collection tool	Contains diary responses indexed by Unique Study Number. Does not contain person identifiable data.
REST API	Data transmission tool	Secure transmission of diary data from smartphone to MYSQL database. Does not contain or transmit person identifiable data.
TARDIS MySQL database	Data storage tool	Contains diary responses indexed by Unique Study Number. Does not contain person-identifiable data. Stored behind the UCL firewall.

DATA STORAGE AND MANIPULATION.

Paper diary responses will be reviewed by the researcher. A copy of the paper diary will be made and stored in the corresponding patient's notes. All paper data will be transferred and keyed-in directly into an institutional MySQL database by the researcher.

Smartphone diary data will be transmitted securely via the REST API. The TARDIS MySQL database will be configured to receive these data. These data will be encrypted. This is further defined and described in the SLSP. The USN will used to differentiate participants. The MySQL database will be configured as an Open Database Connectivity (ODBC) data source. Once imported, data will be manipulated and analysed using R(231) software package, within the UCL WLAN environment.

Design and architecture of the SQL database and API are described in Section II Apppendix p.TARDIS:REFLUX database478.

Results

Assessment of Eligibility

Between May and October 2013, the researcher screened the clinic notes of patients attending the three source services to identify patients eligible for recruitment. Three sources of eligible patients receiving services at my institution (Great Ormond Street Hospital for Children) were identified.

1. General surgery clinic
2. Gastroenterology clinic
3. Gastroenterology investigations suite: Patients requiring oesophagoscopy, colonoscopy, pH impedance studies and other specialist gastroenterology related investigations attend this unit for investigations.

There were 432 sets of patient notes reviewed. Of these, 125 patients were identified as having a history of reflux. There were 16 patients who did not attend the clinic. There were 19 patients who were missed i.e. the researcher was unable to approach them for recruitment before they left. There were 25 patients whose parents were not approached because they were felt to be too complex (n=11), private patients (n=11) or because the consultant felt it was not appropriate for the patient to be involved (n=2).

Of the parents approached, 13 reported that the reflux had resolved. There were 11 parents who reported that reflux was not confirmed as a diagnosis. There were 2 participants who we were unable to recruit because of a language barrier. There were 4 potential participants who declined to participate in the study. A total of 61 participants were recruited to this study to anticipate dropout attrition.

Recruitment rate

In planning this study, we anticipated a recruitment rate of 28 participants per month and a recruitment period of 2 months. Between 21st May and 23rd September 2013, we recruited 61 participants to the TARDIS Pilot Study. Therefore, the actual recruitment rate was 15 participants per month.

Summary of participants

The characteristics of pilot study participants are summarised in the table below.

Table 26: Demographic characteristics of patients and parents participating in the TARDIS Pilot Study

Characteristics		**n**
Male /Female		32/29
*Patient age at recruitment (years)		3.13
*Age of main carer (years)		33
Parents attending clinic	Mother	45
	Father	3
	Both	13
*Data are median (interquartile range)		

Most parents recruited (74%) owned a smartphone. Among smartphone owners, the iPhone was the most commonly owned phone (78%). Details of smartphone ownership are detailed in the table below.

Table 27: Distribution of smartphone ownership

Type of phone		**n**
Non-smartphone		16
Smartphone		45
Type of smartphone owned	iPhone*	35
	Samsung Galaxy	8
	Blackberry	2
*2 participants had an iPhone but, due to screen damage or operational damage, were unable to use their iPhone to participate in the study.		

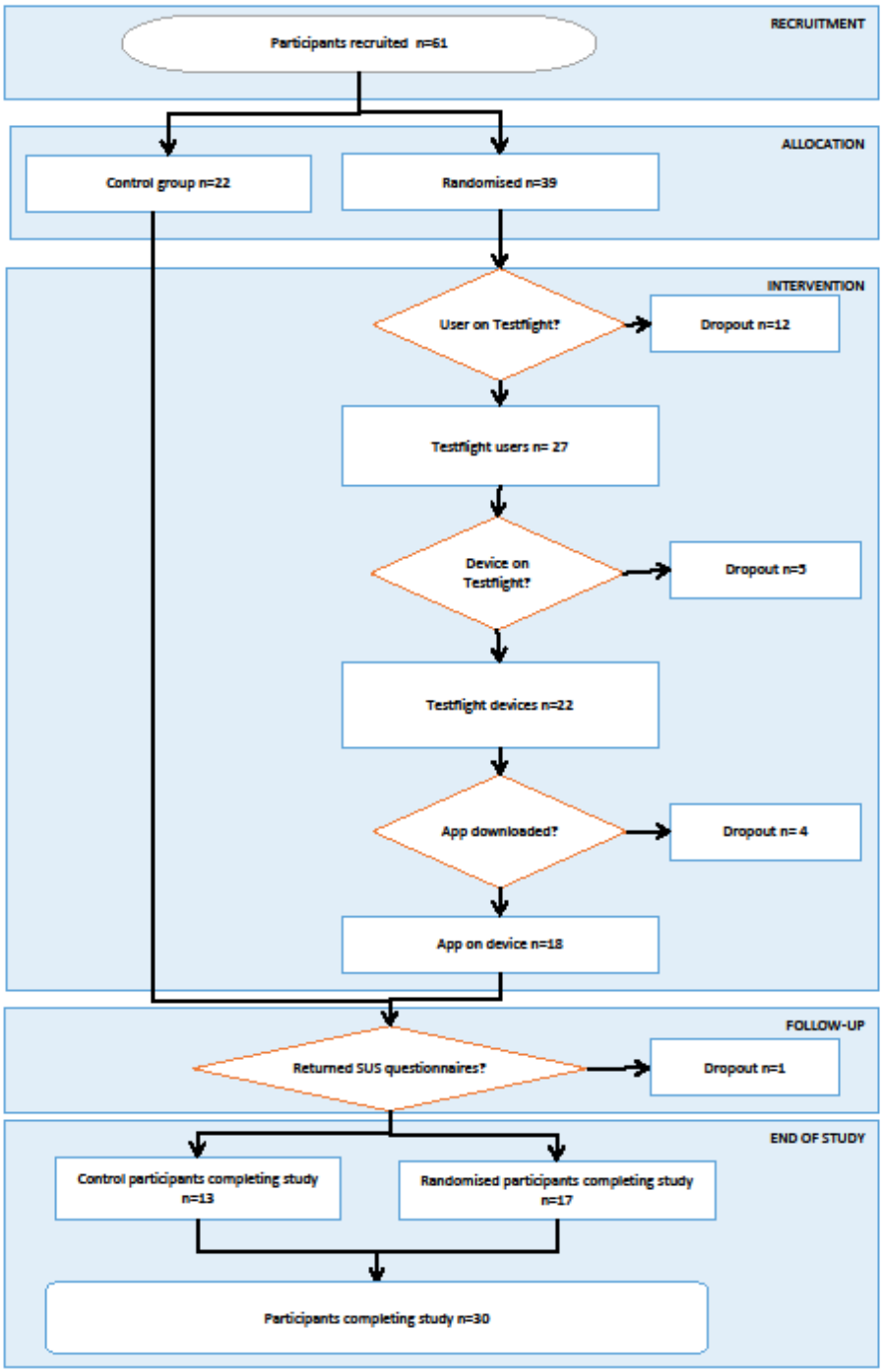
Some participants wanted to take part in the randomised arm of the study suing their iPad device. As the app was applicable to both the iPhone and the iPad, we recruited iPad –owning participants too. At recruitment 33 participants recruited preferred to their iPhone while 6 participants preferred to use their iPad for the study.

Allocation and completion

Patients were allocated into three groups depending on smartphone ownership. Attrition rates were high in both the control and randomisation groups. Only 13 of the 22 participants in the control group

completed the study (59%). Only 17 of 39 participants in the randomisation group completed the study (43% completion).

Figure 33: Patient participation, attrition and completion



Primary outcome: Usability

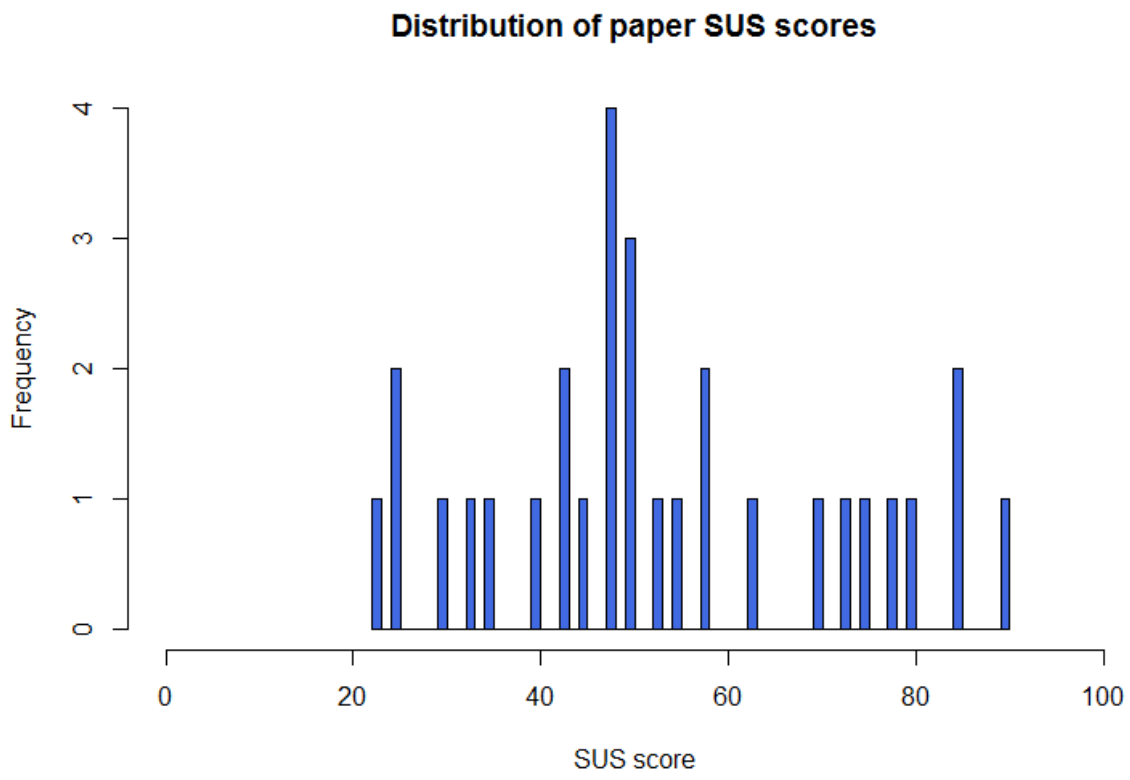
Our usability metric was the SUS score achieved by each symptom diary modality. We discuss the SUS scores for the paper diary and the app and then report on the comparison of scores for each modality.

Usability scores for the paper symptom diary

Participants in both the randomised and control arm of the study assessed the paper diary using the SUS tool. There were 61 participants polled. 30 returned SUS questionnaires rating the paper symptom diary. Therefore, the overall response rate was 49 %. Our attrition rate was therefore 51%. In the control group, 13 of 22 participants returned their SUS assessment (59% response rate). In the randomisation arm, 17 of 39 participants returned their SUS assessment (44% response rate).

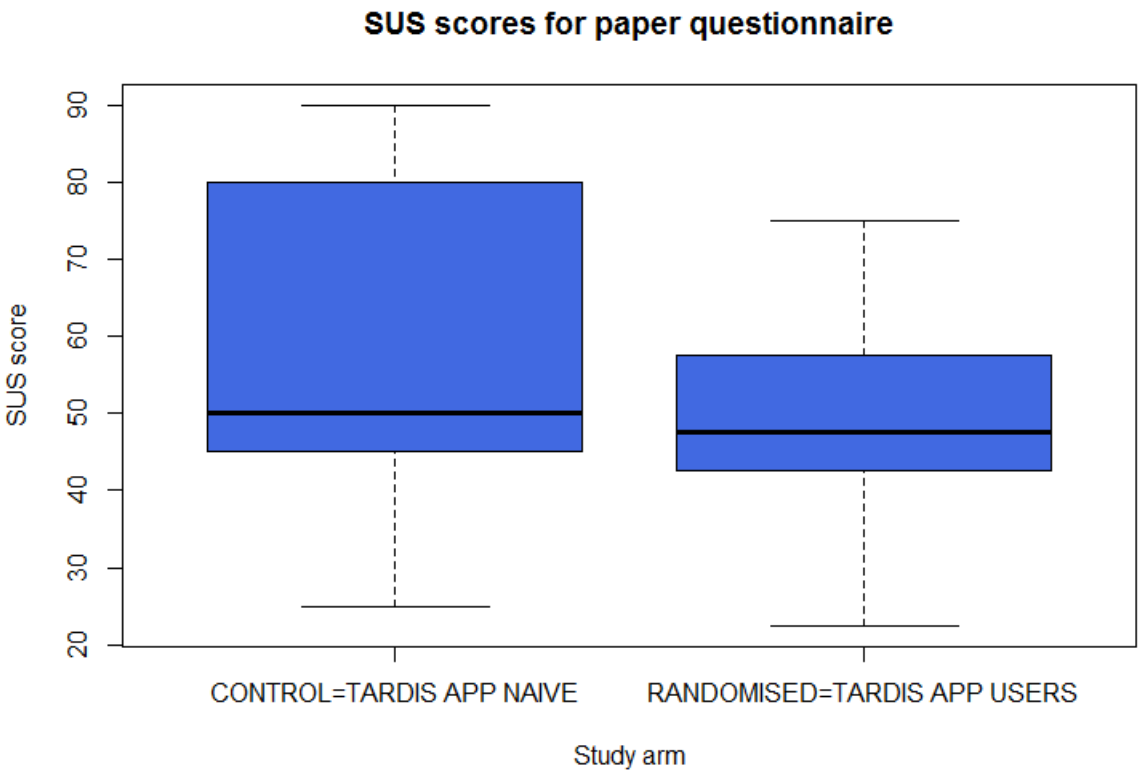
SUS scores range from 0 to 100, with higher scores reflecting greater usability. The histogram below (**Figure 34**) demonstrates the distribution of SUS scores for the paper symptom diary. The median SUS score was 50 (IQR 42.5-68.1). The mean SUS score was 53.3 (SD 18.9).

Figure 34: SUS scores for the paper symptom diary



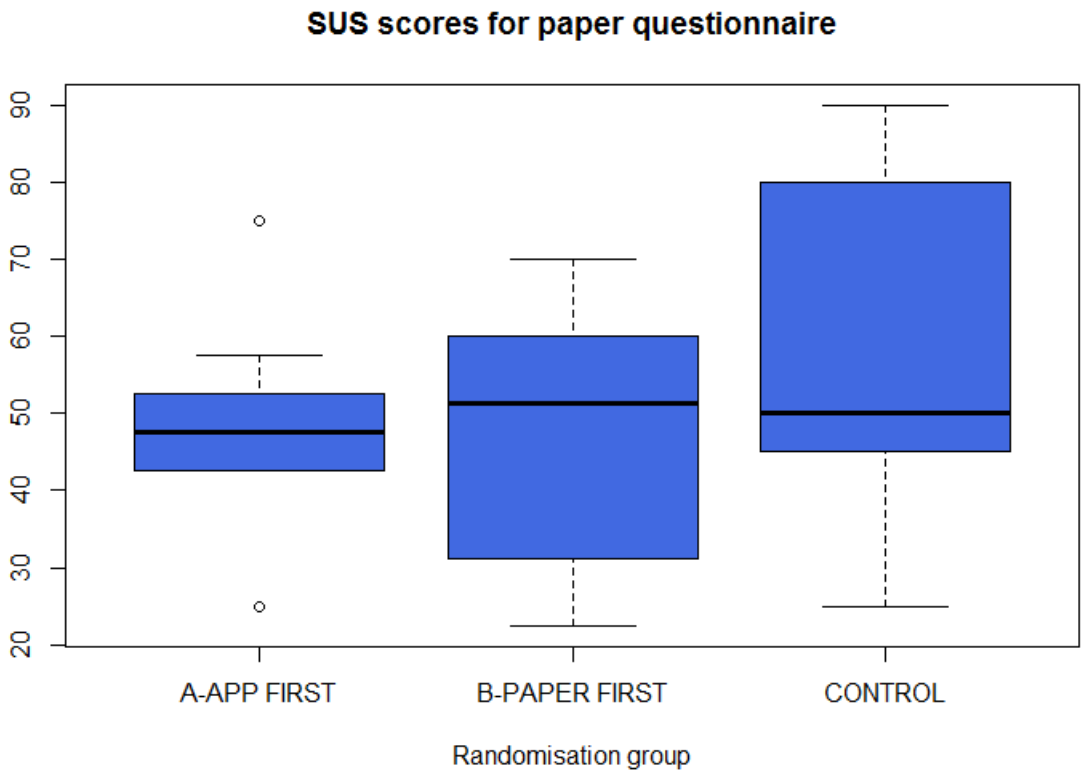
We compared the SUS scoring of paper diaries by participants who had not used the TARDIS:REFLUX app (Control, app naïve), to the randomised participants who had used the app. Participants not exposed to the TARDIS app (CONTROL) assigned a median score of 61.3 (IQR 46.9-81.3) to the paper diary. Participants who were exposed to the TARDIS app gave lower scores to the paper diary (median=47.5, IQR 42.5-57.5). The usability scores given by app-naïve participants were non-significantly higher (Mann Whitney U test, W=68, p=0.09) than those given by participants exposed to the app.

Figure 35: Comparing SUS scores for each study arm



We compared scores from control participants (CONTROL), randomised participants who used the app before the paper diary (A-App first), and randomised participants who used the paper diary before the app (B-Paper first). Participants who used the app first rated the paper diary at a median of 47.5 (IQR 42.5-52.5). Participants who used the paper diary first rated it at a median of 51.25 (IQR 31.9- 58.8). Control participants who only used the paper diary rated it a median of 50 (IQR 45-80).

Figure 36: Comparison of median SUS scores for the paper diary, categorised by randomisation group



Although participants who used the app first gave lower scores compared to participants who used the paper diary first, there was no significant difference in scores assigned by the three groups (Kruskal Wallis rank sum, chi-squared = 1.86, df = 2, p-value = 0.39).

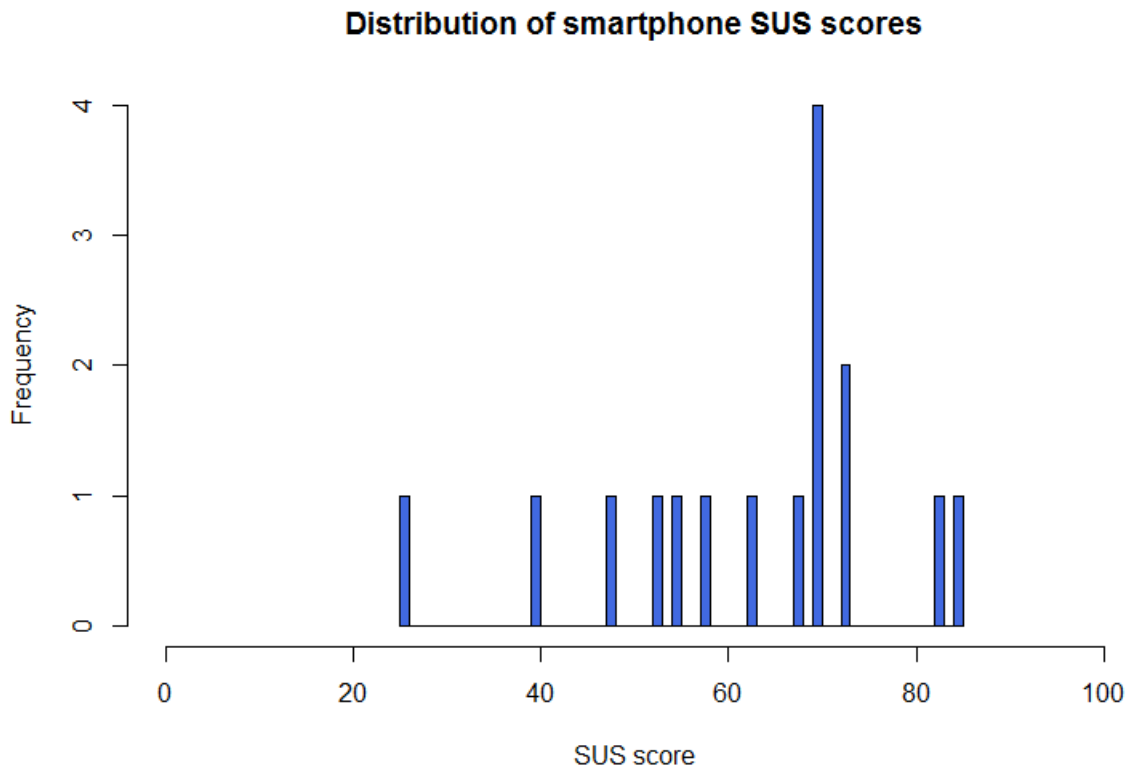
Usability scores for the TARDIS:REFLUX app

Participants in the randomised arm of the study assessed the app using the SUS tool. There were 39 participants polled. Of these, 17 returned questionnaires, resulting in a response rate of 35%. One smartphone SUS questionnaire was returned but was completely blank with none of the 10 items scored. This questionnaire was excluded from further analysis.

In the group A (app first), 9 of 20 participants returned their SUS assessment (45% response rate). In group B, 7 of 19 participants returned their SUS assessment (37% response rate).

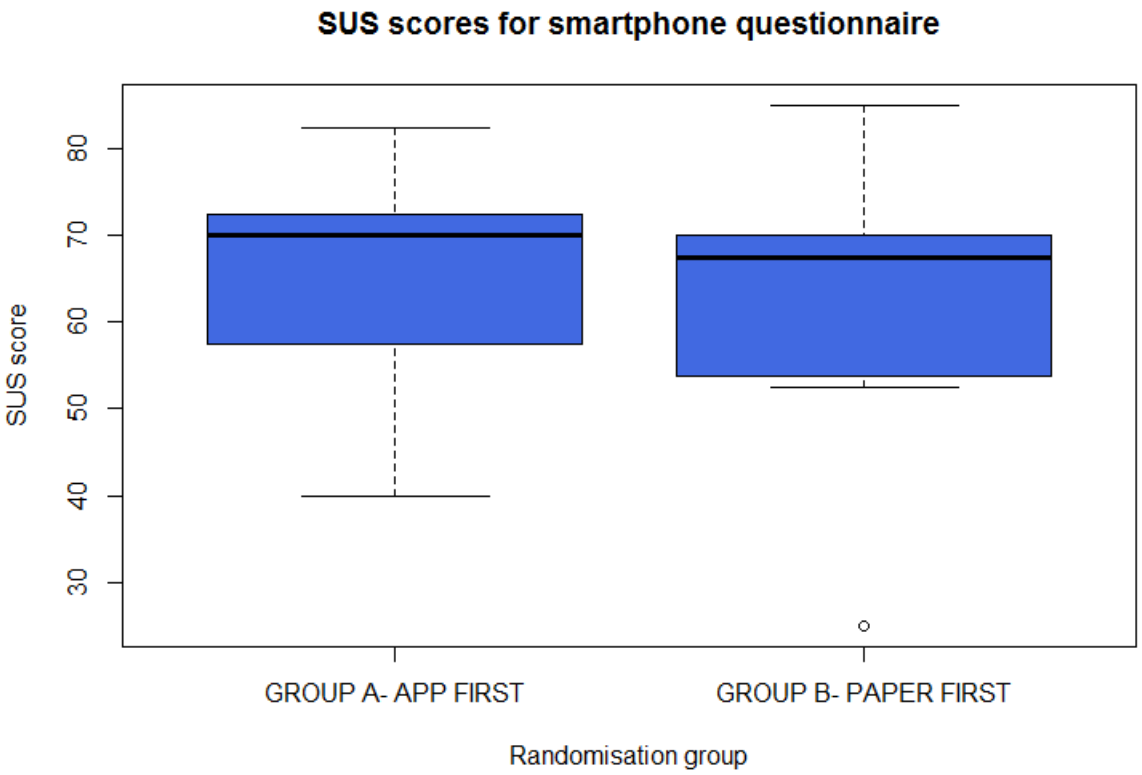
In Figure 37 below, we demonstrate the distribution of SUS scores for the app. For all respondents, the median SUS score was 68.8 (IQR 54.4-70.6). The mean SUS score was 62.5 (SD 15.7).

Figure 37: Distribution of SUS scores for the app



We compared the SUS scoring of apps by randomisation group (Figure 38). Participants who used the app first (Group A) gave a median score of 70 (IQR55-85). Participants who used the paper diary first (Group B) gave a median score of 67.5 (IQR 51.2-83.7). There was no significant difference between the median scores given by participants in groups A and B (Wilcoxon sum rank test, W=36, p=0.7).

Figure 38: participants in group A and B awarded similar SUS scores to the app.

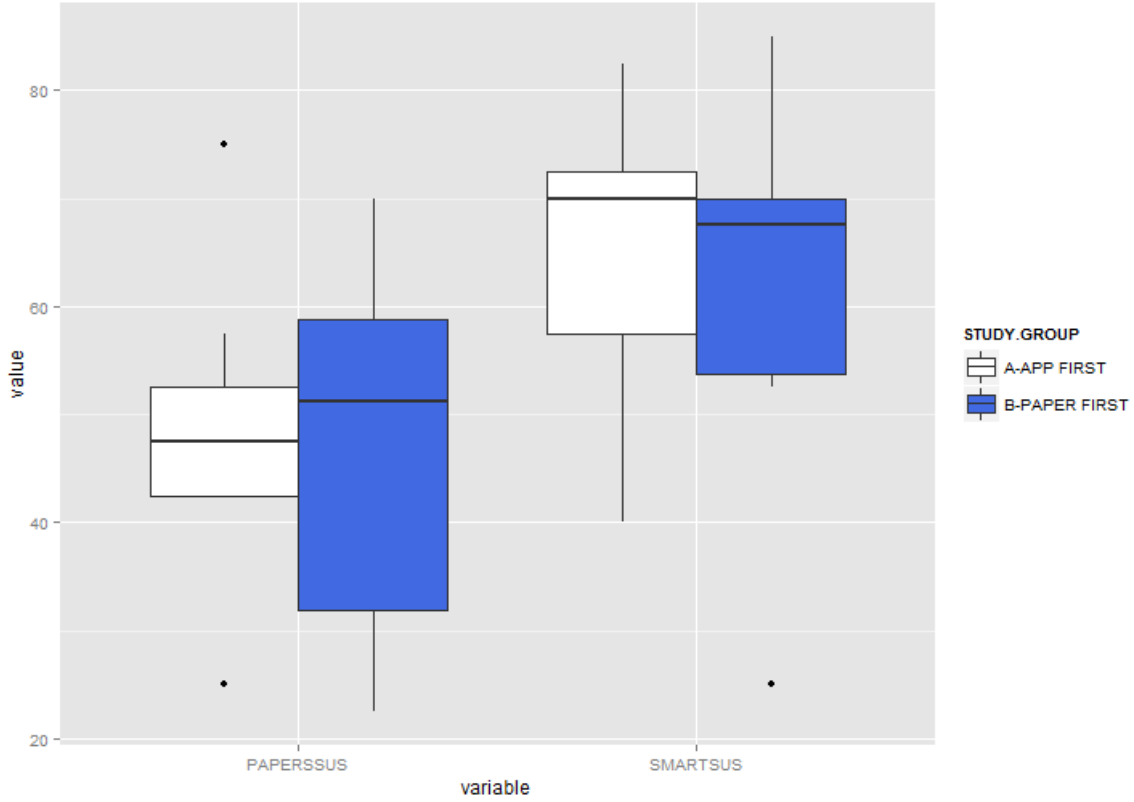


Crossover study: within group comparison

Participants in the randomisation arm had the opportunity to use both the paper diary and the app. We compared SUS scores assigned to each symptom diary by the same participant. The app was found to have higher usability when compared to the paper diary (Wilcoxon signed rank test V=99.5, p=0.03).

Participants who experienced the TARDIS:REFLUX app first gave lower scores to the paper diary (Mann Whitney U test, v=61, p=0.069). Participants who experienced the paper diary first gave higher scores to the paper diary (Mann Whitney U test, v=39.5, p=0.201).

Figure 39: Comparison of paper diary and app SUS scores depending on which diary was used first



From this assessment we can conclude that there was a sensitisation effect, where participants exposed to the app first rated the paper diary less highly.

Discussion

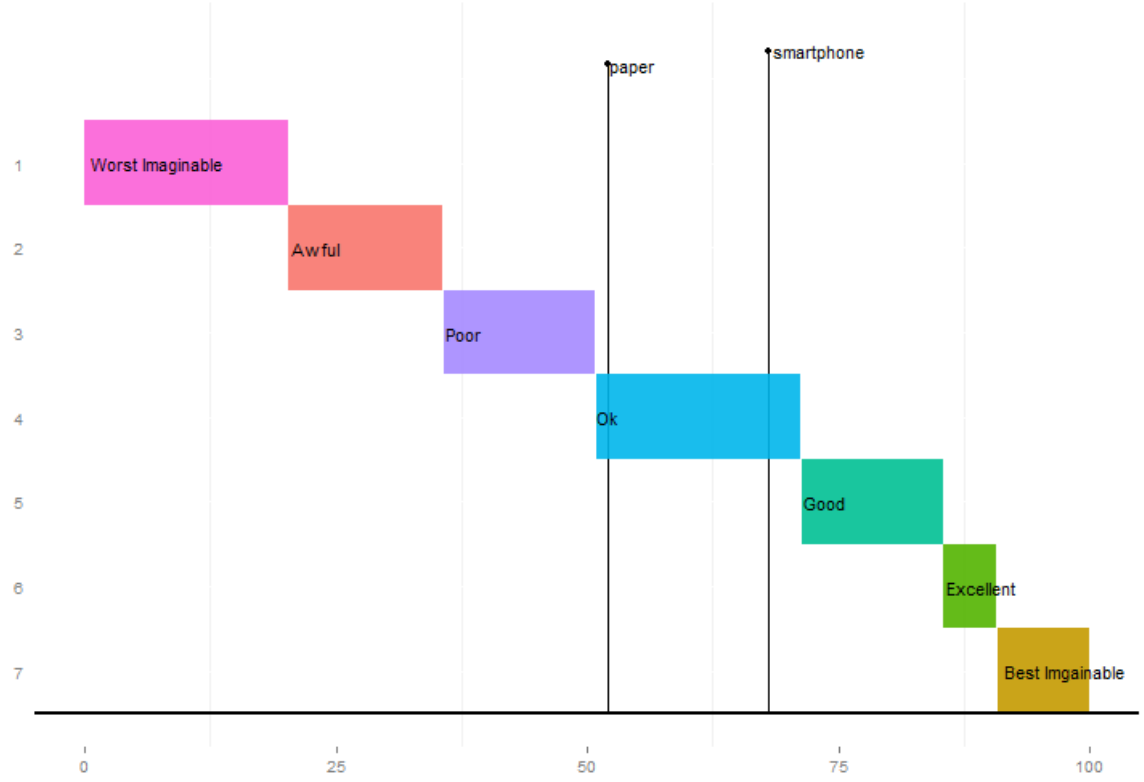
Interpreting the Pilot Study results

The 100-point SUS scale is useful for relative judgements. In the context of this study, we know that participants exposed to both the app and the paper diary had an overall preference for the app. However, the median scores do not tell us whether the app was merely the lesser of two evils.

Bangor et al(228) developed and validated an adjective rating to correspond to mean SUS scores. After completing and SUS assessment of a variety of interfaces, participants were asked to select an adjective or phrase that corresponded to their overall impression of the interface assessed. There were seven options i.e. “Worst imaginable”, “Awful”, “Poor”, “Ok”, “Good”, “Excellent” and” Best imaginable”. There was a high and significant (r=0.82, alpha <0.01) correlation between SUS scores and adjective rating for each participant.

In the context of this study, the both mean scores for the paper diary (53.3) and the app (62.5) lead to an adjective correlation of “Ok”. However, as demonstrated in the figure below, the smartphone score approaches good, while the paper score approaches “Poor”.

Figure 40: Adjective rating for SUS scores



Bangor et al(228) also reported on average SUS scores for multiple interfaces. From this analysis, we have an understanding on how the app and performs as an interface. The average interface score is 69.5. Therefore, we can cautiously conclude that the app (mean score = 62.5) meets average expectation.

Table 28: Summary of SUS scores by user interface type (from Bangor et al)

Type	**Number of SUS assessments**	**Mean SUS Score**
Web	1433 (41%)	68.2
Cell phones	593 (17%)	65.9
Interactive voice response	573 (17%)	72.4
GUI	250 (7%)	76.2
Hardware	237 (7%)	71.8
TV	185 (5%)	67.8
Total	3463	69.5

LIMITATIONS OF PILOT STUDY

SAMPLING BIAS

In this pilot study, we found that 16 out of 61 participants recruited did not carry a smartphone.

Mobile phone ownership is not universal. Annual UK national statistics[5] reveal growth in mobile phone ownership in the past decade. In 2001/2002, 65% of the population owned a mobile phone. In 2009, this proportion had risen to 81%. Mobile phone ownership varies with household income. In 2009, ownership of mobile phones was 67% in households representing the lowest 10% by income. In the highest 10% of households by income, mobile phone ownership was reported as 92%[6].

Mobile phone ownership varies with age. In March 2011, a survey by Ofcom, the communications industry regulator, found the incidence of mobile phone ownership in Great Britain to be 89% [5]. Between the ages of 16- 54, mobile phone ownership was over 95%.

Smartphone ownership is neither ubiquitous nor evenly distributed. According to a recent government survey (YouGov Smartphone Mobile Internet Experience Study 2011), only 35% of adults owed a smartphone in 2010 [7]. Although smartphones typically cost more than a standard mobile phone, income is not a key driver of smartphone ownership. YouGov's quarterly survey of mobile phone ownership has identified the key drivers of smartphone ownership as gender and age. Notably, 58% of smartphone owners are men.

Smartphone ownership varies with age. The YouGov quarterly survey found that 86% of smartphone owners were under the age of 55 years. OfCom also identified variations in smartphone ownership with age. In the March 2011 survey, OfCom reported the incidence of smartphone ownership in the 25 -34 years age bracket as 42%. Amongst adults aged 35-54, the incidence of smartphone ownership was 29%.

Due to cost limitations, we have developed a smartphone app for the iPhone only. There are discernible income differentials between smartphone users. In April 2012, a YouGov Poll identified that BlackBerry users were likely to be higher earners[8]. 10% of BlackBerry users admitted to earning over £50,000 a year, compared to 7% of iPhone users and 5% of Android users.

Our study sample was therefore likely to be biased towards younger, wealthier participants who carry iPhones and smartphones. This limits the generalizability of the study.

We also found that majority of our respondents were female. This finding is unlikely to represent sampling bias. Instead it is likely to reflect the circumstances and choices of individual families', where the female parent is most often the primary care giver.

SAMPLE SIZE

The pilot study applied summative or task specific testing. The user was asked to use a symptom diary for 7 days and then complete a comparative assessment. The outcome measure was a comparison of mean SUS scores. Therefore, knowing the variance of the SUS scores, we defined the effect size and set the degree of power and statistical confidence required. We then estimated the sample size.

When calculating the required sample size for the study, the response rate was taken into consideration. We estimated an attrition rate of 15%. We have observed an actual attrition rate of 51%!

If we re-calculate the sample size based on an attrition rate of 51%, the study would need to recruit 59 participants (39 *1.51) in the randomisation study arm. In addition to this, at least 30 recruits would be required for the control arm. Therefore, a total of 90 participants would need to be recruited.

Attrition results in missing data and is a source of bias. We contacted the parents of subjects who did not submit the questionnaire to identify reasons for non-compliance with monitoring task. Reasons for non-compliance mentioned:

1. Task too difficult
2. No time to complete task
3. No direct benefit to patient/child

Completion rates were observed to be lower in the randomised crossover group (43%) compared to the control group (59%). The randomised study arm required a greater commitment i.e. completion of two diary episodes, each 7 day long. For busy parents with an unwell child, maintaining such a commitment may be difficult.

The cross-over design may also have been a source of attrition. There were a few days between recruitment and commencement of diaries. Where possible, we contacted participants to remind them to progress to the next phase of the trial. In some cases, parents commented on having misplaced the symptom diaries since clinic review.

There was no compensation for parents participating in the study. As there was no direct applicability of diary findings to the clinical care of their child, parents may have required greater incentive to complete the study. In the absence of clinical motivation, the pilot study was essentially a software testing scenario.

In software testing, inducements or compensation is often offered to testers. As Al-Awar et al(223) observed:

"We have paid all our subjects, and have paid them enough to entice them into the laboratory."

As the pilot study was established as clinical trial, offering inducements may not have met ethical standards. However, offering parents compensation for their time and inconvenience should certainly be considered for future studies, as this may reduce attrition.

PUSH NOTIFICATIONS

Push notification services allow app vendors to communicate remotely with app users. An internet protocol (IP) connection is established between a server hosting the app and the smartphone. Messages and alerts can be sent to users from the app host. These services are offered to app vendors by app stores e.g. iTunes, GooglePlay.

Mobile health apps can use push notifications as a form of enhanced monitoring. For example, Buddyapp for therapy sends a daily text message asking users how they feel.

The pilot study was implemented using a development platform (Testflight). At the time of the study, push notifications were not available to non-enterprise users. The cost of enlisting this further service was prohibitive. However, such a service would have been useful for keeping in touch with participants

and implementing the study protocol. For example, push notification can be used to remind users to input daily data.

BLANK ITEMS IN THE SUS QUESTIONNAIRE:

There were two paper SUS questionnaires and one smartphone SUS questionnaire returned un-rated. Where a SUS score is left blank, three approaches have been suggested by Sauro[9]. One approach is to exclude the questionnaires with unscored items. This option would be unsuitable in the context of this study due to the low response rate. Furthermore, other questionnaire items were completed, excluding these questionnaires would lead to unnecessary loss of data. A second approach would be to re-weight the score items to minimise to reduce the overall contribution of blank items to the final score. This approach is recommended for questionnaires where there are up to 2 missing values(228). From the literature provided by the author who suggests this approach, choice of multiplier appears arbitrary. The effect of changing the measurement dimensions for selected responses but not the whole cohort is not quantified or elucidated.

A third would be to impute the missing data by assigning a "neutral" score of 3 to the unscored items. This was the approach was chosen for blank questionnaire items.

SUMMARY

The TARDIS Pilot study was designed to test system acceptability in the target group of users i.e. parents of children with gastroesophageal reflux disease. Our ability to make firm conclusions from this small study is limited by high attrition rates, sampling bias and methodological flaws. However, some observations can be made. These observations will be detailed and discussed in Chapter 6.

USER EXPERIENCE TESTING

The TARDIS Pilot Study described above captured some elements of acceptability e.g. utility and usability. However, as this was a product test, data on user experience and aesthetic dimensions required exploration. Elements of interest include e.g. subjective appeal, flow, efficiency etc. Not all dimensions are amenable to quantitative measurement.

In market research, convening a focus group allows a researcher to explore specific issues with a small group of people. This method of study is suitable where personal insights are of value. This medium can also allow areas of interest to be plumbed in great depth. Compared to one-to-one interviews, focus group environments "allow" participants to pause for reflection which, in a one-to-one interview situation, might be experienced as awkward silences[11].Enabling individuals to reveal their opinions within the safety of a group may lead to more open and honest revelations of their opinions on the product[11].

Reflux UX Workshop

We developed and convened a focus group to assess the TARDIS:REFLUX user interface with specific focus on usability, utility, efficiency and subjective appeal.

Grant funding for this workshop was received from a UCL Public Engagement grant. The workshop was titled 'REFLUX UX'. UX is a commonly adopted acronym representing User Experience.

Objective

The objective of the focus group was to determine how easy it is for participants to download and use the TARDIS:REFLUX app. Therefore, the focus was be on the system acceptability dimension of usability.

Specifically, we would asked assess whether the TARDIS:REFLUX app is:

- Easy to learn
- Efficient to use
- Used with few errors
- Subjectively pleasing

We also asked participants to comment on the utility of the app i.e. whether they might use it to track disease, and if there are other conditions that an app format may be useful for.

Outcome measures

The outcome measures were both quantitative and qualitative. The outcomes measures are defined as:

1. Observation: We observed the ease of download and activation of the app

a. Measures app download and activation time for each participant

b. Count and document problems encountered

2. Interviews: Participants were asked questions about their attitudes to apps and current usage patterns. Participants perception of the user interface was recorded and analysed, focusing on utility and usability. Below is a list of open and simple questions designed to initiate a discussion?

a. What apps do you use on your iPhone?

b. Do you think an app for gastroesophageal reflux is a useful thing?

c. What do you think about this app?

d. What would you change about this app?

e. What do you think of the colours used?

f. Let's review the symptoms you are able to enter. Are they all clear?

g. What would you change about the TARDIS:REFLUX app?

3. Quantitative survey: Focus group participants were asked to complete the SUS questionnaire.

Group composition

A focus group is comprised of the group leaders or animators and the group members. In constructing a focus group, a researcher aims for balanced composition in group membership. Balance in demographic factors e.g. age, sex and ethnic status is desirable. A typical focus group might contain 6 to 12 members.

For the REFLUX UX Workshop, researcher was the group leader (EWM). The researcher was assisted by a fellow trainee paediatric surgeon undertaking a post-graduate degree (Miss Rashmi Roshan Singh (RRS)) whose contribution is defined and acknowledged in the opening chapter of this thesis. Miss Singh was co-leader for the group sessions.

Advertising the focus group

Focus group attendees should, ideally, resemble the targeted users. The ideal composition of a focus group would be parents of children with gastroesophageal reflux disease. However, by the context of the study. To recruit based on patient-specific parameters would require registration of the focus group in the context of a clinical trial. Taking a pragmatic approach, we chose to recruit participants with some experience of chronic disease.

We were granted access to some 12 iPad devices (courtesy of UCL Advances). This allowed us to recruit participants who did not have an iPhone. We invited parents of children with chronic illness to attend the workshop at the UCL Institute of Child Health.

We advertised for participants in the following forums:

1. Parent forums and websites

a. Targeted advertisement on Facebook

b. Eventbrite

2. Mailing lists

a. UCL announce

b. ICH announce

c. GOSH announce

REGISTRATION OF PARTICIPANTS

Once participants were recruited, they were invited via email to join the Test team for the TARDIS: Reflux app on the Test flight platform. Those with iPhone were able to download the app prior to the focus group session. Registration as testers was completed prior to the workshop.

Workshop methods

We planned to run 2 sessions 2 weeks apart. Each workshop session aimed to recruit 6 participants and was scheduled to last 3 hours. Feedback from the first session would be used to modify the app. This updated product would then be presented to the second focus group. The feedback from the second session was used to modify the app before final release on the app distribution platforms. By taking a staged approach, we used focus group methodology as a tool to improve the user-interface in an iterative way. A schedule for the workshop comprising three main sessions was devised (**Table 30**). Each user was asked to complete a SUS questionnaire as part of the workshop follow-up.

Recording and consent

As the sessions were recorded by using videography (UCL Video Services) , consent from participants was required in keeping with UCL guidelines.

Data collection and analysis

Following the sessions, recorded interviews were transcribed. Transcribed interviews were used as source documents for analysis. The content within source documents was coded by themes (NVivo Version 10, QSR International Pty Ltd). SUS questionnaire data was collated (Microsoft Excel) and analysed (R).

Results

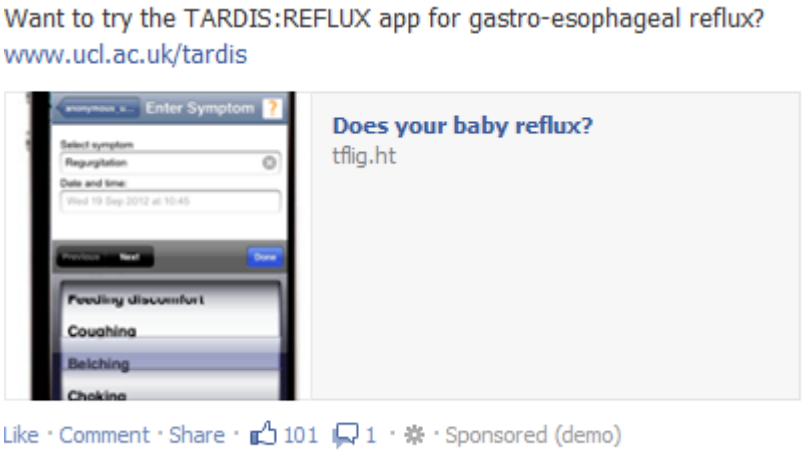
Recruitment

We initially sent out an advertisement for the focus group on mailing lists with a view to holding the first and second focus groups one month later. Despite announcing the focus group on multiple lists (UCL Announce, ICH all users email, GOSH all users mail), we received no responses. Therefore, after consultation, the workshop was postponed. Furthermore, due to early indications of sub-optimal recruitment, we decided to limit the workshop to one session and reschedule the workshop.

The workshop was re-advertised and updated to include a reward for participation. Participants were offered £30 in Marks and Spencer's vouchers to attend the focus group. In addition to the mailing lists already targeted, the re-advertisement was also distributed via further methods:

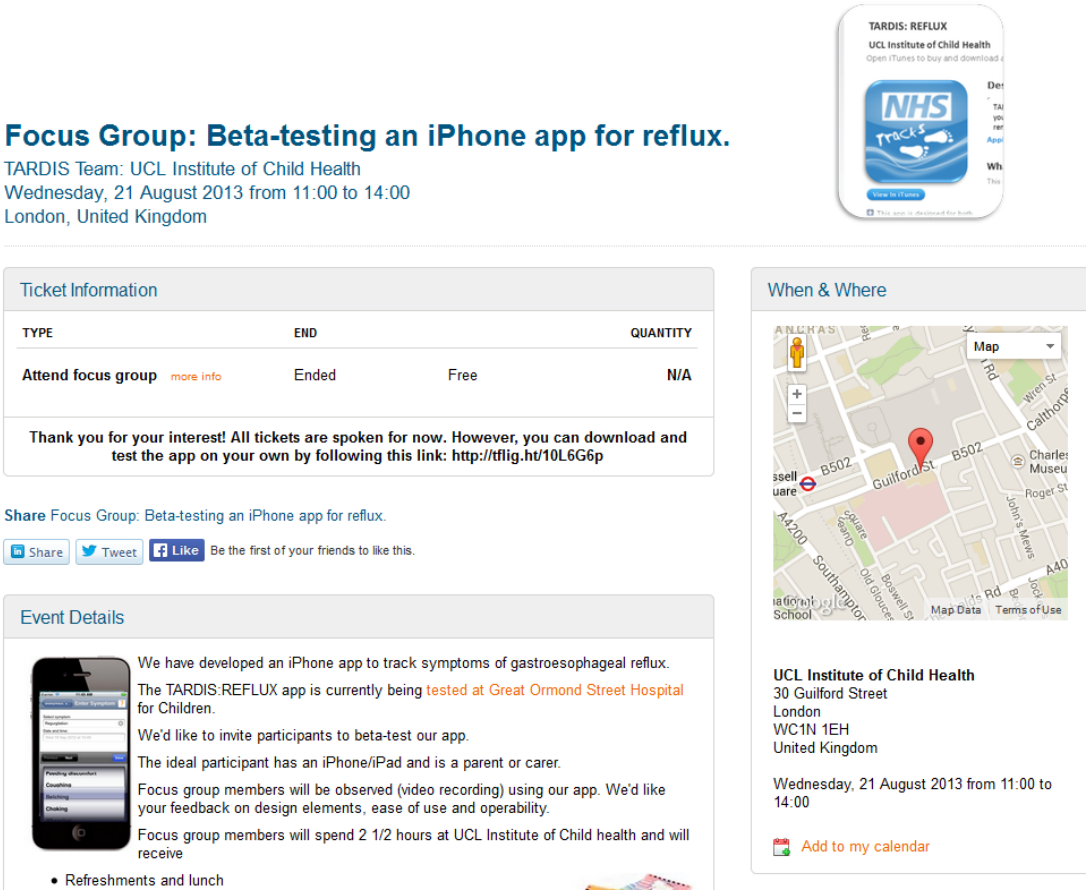
1. Facebook: This campaign ran for one month on www.facebook.com

Figure 41: Facebook campaign to find participants for the REFLUX UX workshop



2. Eventbrite: we used www.eventbrite.co.uk to advertise and distribute tickets to the reflux UX workshop.

Figure 42: Advertising campaign on www.eventbrite.co.uk



Participants

There were 16 tickets available for interested participants. Of these, we received 12 registrants. We sent an email to participants requesting confirmation 2 days prior to the focus group. Of the 12 initial registrants, 11 confirmed attendance. On the day of the workshop, 8 participants attended the TARIDS:REFLUX focus group. Demographic characteristics of participants and iOS device ownership are detailed in the table below:

Table 29: Summary of REFLUX UX Workshop participant characteristics

Participant	Gender	Age	Profession	Parent	Chronic illness	Device
1	FEMALE	26	Study Coordinator	NO	NO	IPHONE
2	FEMALE	25	Receptionist	YES	NO	IPHONE
3	FEMALE	45	Paediatric surgeon	YES	NO	IPHONE + IPAD
4	FEMALE	31	Primary school teacher	YES	YES	IPHONE
5	MALE	28	Student	NO	YES	IPHONE
6	MALE	29	Research Assistant	NO	YES	IPAD
7	FEMALE	35	Pharmacist	YES	NO	IPHONE + IPAD
8	MALE	25	Developmental biologist	NO	NO	IPHONE

Workshop

The REFLUX UX focus group was held in a seminar room at the UCL institute of Child Health. We used a large room with comfortable chairs. Tea, coffee soft drinks and snacks were provided.

The workshop was held on 21st August 2013. The program on the day is detailed in **Table 30** below.

Table 30: Workshop program

Title	**Reflux UX focus group**
Date	**21st August 2013**
Location	**UCL institute of Child Health , Lower ground floor Seminar room 2**
11am – 1130 am	Welcome and Registration
1130 am - 1215pm	App download and activation: This is an observed session, where the researcher notes problems experienced with during download and activation. Users who did not download the app prior during registration were observed performing this procedure. Each user is given the opportunity to try out the app.
1230 - 130pm	App Look and feel: • Each participant will have a face-to-face feedback session with a group leader. • Café-style feedback session on visual and design elements of the app. This session will be held over lunch.
130pm to 2 pm	Next steps: • First impressions are captured on the SUS assessment questionnaire. • We explain how to use the app over the next seven days
215pm	Workshop ends
Follow-up (7 days later)	Each participant uses the app for 7 days. At the end of this period, each user is asked to complete the SUS assessment sent via email.

Outcome measures

Observation

Download and Activation

Participants were observed downloading and activating the app. Seven participants successfully downloaded the app with no crashes during the session. One participant was unable to download the app onto their iPhone. This issue was temporised by giving the participant an iPad brought along by the group leaders (courtesy of UCL Advances). We were able to resolve the app download issue for this participant by establishing a Wi-Fi connection. The initial download was not successful because an incorrect Wi-Fi service was selected and, on clicking the invitation to download the app, the user was directed to a captive portal rather than the download link.

Table 31: Timing of Observed events

	Task completion time(seconds)			
Participant	**Login**	**Record Symptom**	**Record Event**	**Generate Report**
1	3	8	50	40
2	10	7	8	25
3	10	4	30	30
4	5	6	40	8
5	8	4	45	25
6	4	4	15	6
7	4	10	15	6
8	6	12	20	7
Mean	**6.3**	**6.9**	**27.9**	**18.4**
Standard deviation	2.8	3.0	15.7	13.3
Median	5.5	6.5	25.0	16.5

Login

Participants were observed when completing the Login task. Participants were required to tap on the app icon on their device and insert a username and password. To standardise the task, participants were asked to use

Username: tardis

Password: tardis

Participants required an average of 6.8±2.8 seconds to log in. Once logged in, all participants were at the home screen which presents three options: ‘Enter Symptom’, ‘Enter Event’, ‘View Report’.

Symptom

We requested all participants to record the symptom ‘Vomiting’. Participants were not given instructions on how to complete the task. The mean time to task completion was 6.9±3.0 seconds.

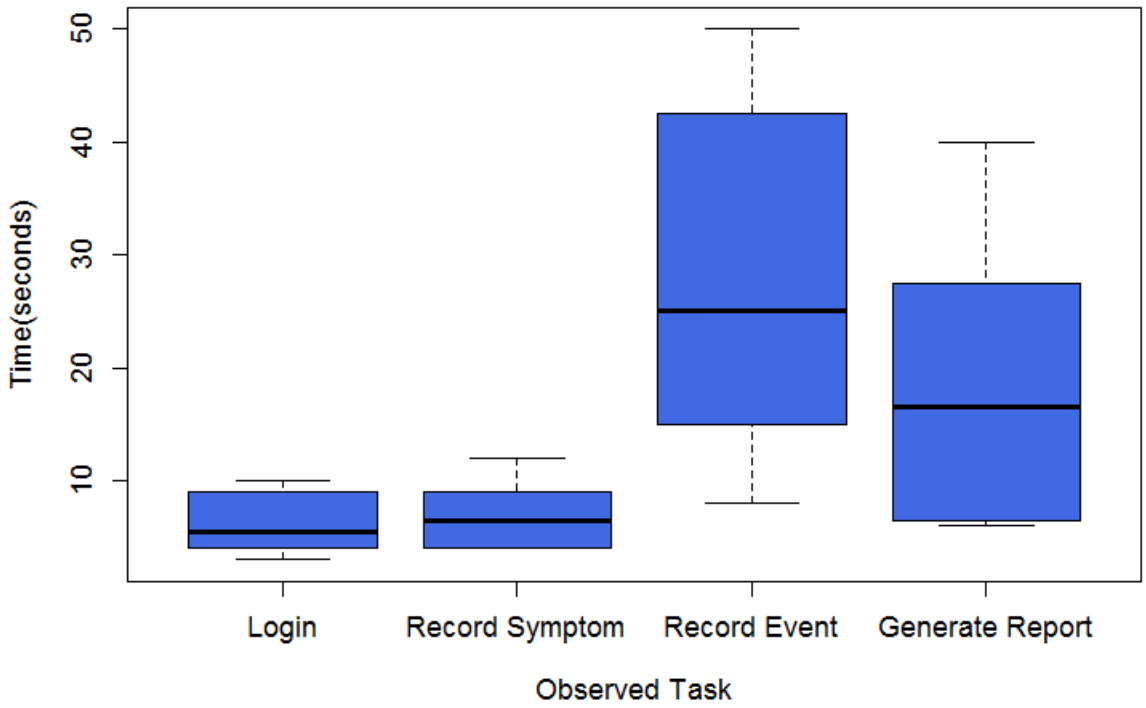
Event

We requested all participants to record the event 'Weight'. Participants were not given instructions on how to complete the task. The mean time to task completion was 27±16 seconds.

Report

We asked participants to record a few more symptoms of their own choosing. Participants were then asked to generate the report 'Symptom distribution'. Again, participants were not given instructions on how to complete the task. The mean time to task completion was 18±13 seconds.

Figure 43: Time taken to complete observed tasks



Interviews:

The focus group was video and audio recorded. The videographer (PM) deleted non-interview segments leaving only video taken when interviews were occurring. The running time of the resulting video was 39 minutes and 30 seconds. The researcher watched and listened to these recordings and transcribed them verbatim. All interviews were transcribed into the source document “Interviews” (see appendix Section II).

Content analysis

The transcripts were then analysed using the constant comparison method. In this method, themes emerging from the participants contributions are identified and compared with themes and contributions from subsequent interviews[11]. Emerging themes from each interview were used as reference codes (NVivo Version 10, QSR International Pty Ltd).

The process was repeated with subsequent interviews until no new coding references are generated. A typological framework for the reference codes was then established. A cross-referencing approach was taken, i.e. it was possible for a data unit from an interview to be coded under two different themes if the context allows. **Table 32** below lists and describes briefly the themes emerging from the focus group interviews. Coverage refers to the percentage of text coded in relation to the theme.

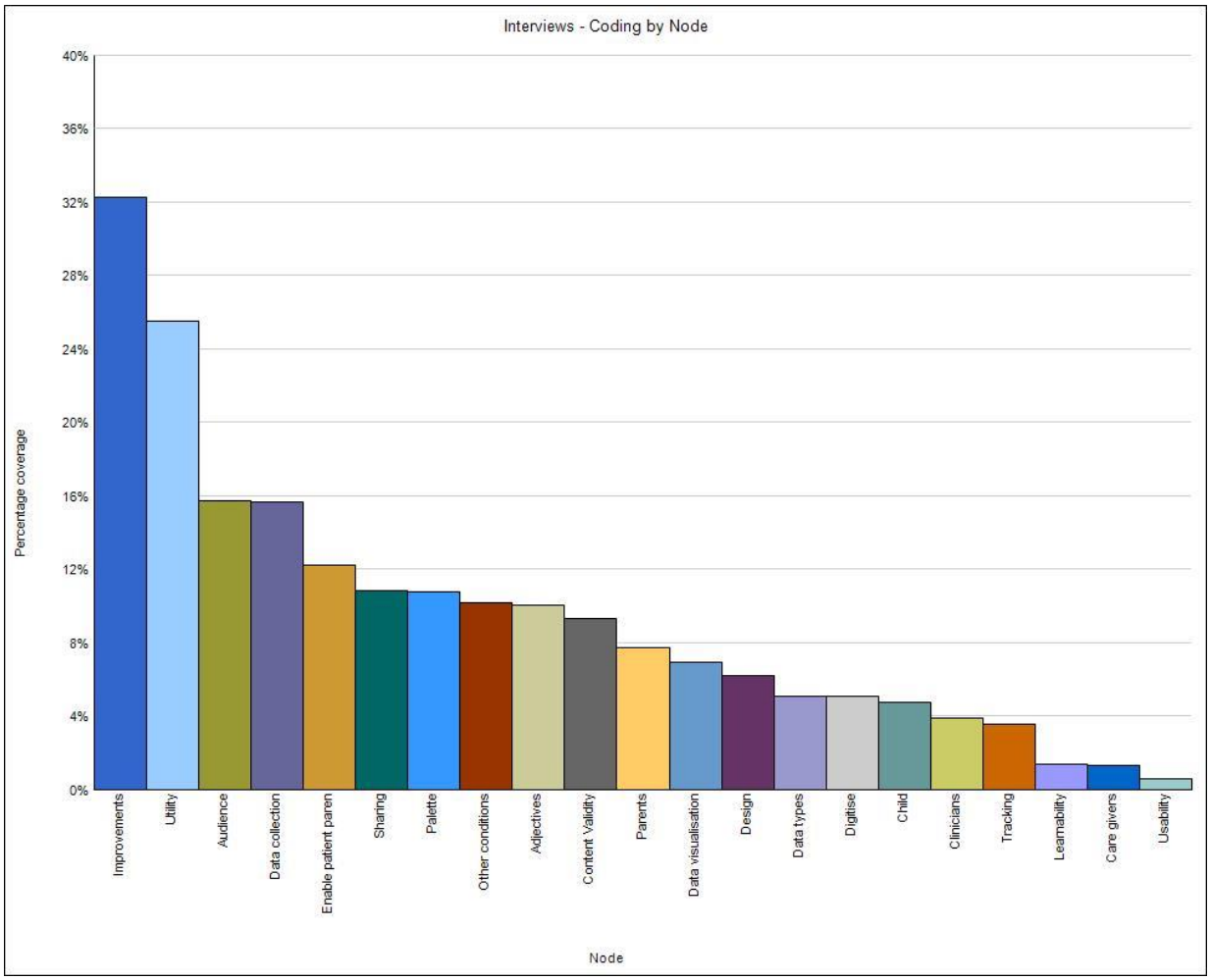
Details of the usability interviews and excerpts from participant responses are found in Section II Appendix p. 478.

Table 32: Summary of themes identified from interviewing participants at the REFLUX UX workshop

Themes	Coverage	Key points identified
Usability	37%	• Design efficiency: participants made suggestions on the app design that make it more efficient to use • Subjectively pleasing: participants assessed the subjective appeal of the app design and palette. • Learnability: the app because more accessible with repeated use • Content validity: participants observed how words and phrases used app captured their intended meaning
Utility	26%	• Tracking: apps can be useful for tracking conditions. • Data collection: An app can be used to collate and store data. It is used to digitise data which in turn enables sharing. • Sharing: apps can be used to share data from patients to healthcare providers. They can also be used to share data between health professionals. • Other conditions: Participants suggested other conditions for which the use of an app might apply. • Data visualisation: Apps may be used to manipulate data into views that make the information easier to understand • Clinical application
Audience	16%	• The app should be targeted at parents • The app should be targeted at children • The app could have features that appeal to children • The app should have some features that enable parents to engage children

The themes emerging, in order of frequency, are visualised in the chart below. Design improvements were by far the most commonly discussed topic (32% coverage). Utility of the app was the second most covered topic (26% coverage).

Figure 44: Frequency of themes (nodes)



Problem discovery

We reviewed the usability problems reported by each participant. Participants reported 66 problems in total. However, some participants reported on problems already reported by other participants. We reviewed focus group interviews to identify the problems reported by more than one applicant. There were 29 unique problems and 13 problems reported by more than one participant. Therefore, there were 42 discrete problems reported. The mean number of discrete problems identified was 8. The mean rate of problem identification for the group overall was 0.2.

Table 33: Problem discovery rate for each participant

Participant	**Number of problems identified**	**Unique problems identified**	**Rate of problems identified (42 discrete problems)**
1	6	3	0.14
2	8	4	0.19
3	7	3	0.17
4	6	2	0.14
5	14	6	0.33
6	8	5	0.19
7	6	2	0.14
8	11	4	0.26
Total	66	29	
Mean		4	0.2

Assuming that the total number of discoverable problems was identified during the REFLUX UX workshop, we defined the probability of problem identification for each participant as:

$$Probability[problem\ identification] = \frac{problems\ identified\ by\ participant}{total\ discrete\ problems}$$

SUS ANALYSIS

All participants (n=8) completed the SUS assessment of the TARDIS:REFLUX app. This is the same questionnaire utilised in the cross-over randomised beta-testing of the TARDIS app.

Table 34: Participant rating of the TARDIS:REFLUX app using the SUS questionnaire

Participant	SUS score
1	72.5
2	87.5
3	82.5
4	95
5	67.5
6	80
7	97.5
8	90

Applying Bangor's analogous adjectives(228), the app received a "good" mean usability rating of 84 out of 100 (Systems Usability Scale®).

DISCUSSION

This focus group discussion explored general views on smartphone apps in research and healthcare, as well as specific views on the TARDIS:REFLUX app. Our participants demonstrated enthusiasm for the use of apps as a research tool. Some specific insights have emerged. These are discussed here, and shall also be used to inform the next development cycle.

Insights:

Audience engagement

To engage with an audience, developers of a product must understand who the audience is and what their motivations may be. In the paediatric population, defining the target audience is not straight forward. However, in pre-verbal children, or those with cognitive deficits, accuracy of information can only be guaranteed by primarily polling parents.

The researcher designed this app with parents as the audience in mind. Tracking can be encouraged by goal-setting and receiving of feedback on performance(232). This is referred to as 'gamification'. For example, a popular fitness app HeiaHeia(233) awards users with 'badges' for milestones achieved e.g. achieving a weight loss goal. Participants felt that parents would be motivated to record the symptoms of their child. However, there are no other rewards or inducements for parents.

When developing an app, a researcher may be more focused on the measurement properties rather than aesthetics. However, the researcher must understand that the end user is not app-naïve. Users will be influenced by prior exposure to well-funded, high-value commercial apps e.g. Nike+. Enthusiasm and uptake may be limited if the researcher's app appears poorly designed or has little aesthetic appeal.

The researcher did not anticipate the cost of user interface polish from the outset. Achieving a sleek aesthetic to match commercial standards may be beyond the budget of research apps. Nonetheless, the aesthetic qualities of apps targeted towards a similar audience should be studied.

'System-focus'

In software projects, development is often focused on the machine or system, rather than the end user. Rubin and Chisnell[13] have identified this 'system-focus' as a factor that limits usability. This was certainly true in the TARDIS: REFLUX project. Due to the technical challenges of producing an app-based tool to measure GOR, the focus of the project was certainly on the app rather than its reception. This may have been mitigated by involving stakeholders from the outset i.e. at inception.

Piece-meal development

Developing an app for research and symptom tracking is a complex process. By necessity, aspects of the system and interface were designed piecemeal by different members of the team. The questionnaire items and decisions about measurement scales were developed by the clinical and research team. Responsibility for implementation of these decisions lay with developers. The integration of these combined visions may have appeared satisfactory to the development team. However, for an impartial observer the app may not completely satisfy either as a research tool or a lifestyle app.

Continuous versus discrete monitoring episodes

Continuous monitoring of a health condition may not be desirable. The app user may be made more aware of their chronic disease. Context-driven tracking, where users track symptoms for a pre-defined period with a pre-defined goal is more acceptable. Tracking with the involvement of clinicians is perceived more valuable than user-led tracking in isolation.

Passive monitoring

Interacting with an app should be as minimal and unobtrusive as possible. Technology should be harnessed to allow users to 'track-without-trying'. A contemporaneous example is digital scales linked to a software application. The transmission of weight data to the software application is automated. The wider application of smartphones as body sensors should be considered.

Sharing is desirable

App users are motivated to share data especially where it accelerates the patient pathway and helps them communicate with healthcare professionals. App tracking is perceived as a digital transformation of symptoms experienced by an individual into consumable health data with value. Users want this health data to be of value to them e.g. improve their own treatment, promote research of their condition. Privacy concerns arise due to the perception that health data may be of value to other players in health markets e.g. employers and insurers.

Limitations of the REFLUX UX workshop

Recruitment

Systems usability should utilise participants who are representative of the end user. Therefore, when we first advertised the workshop, we sought participants who were, ideally, parents of children with gastroesophageal reflux.

Despite advertising on multiple sites (UCL internal mail, Facebook, Google ad), we were unable to recruit parents of children with GOR to attend the focus group. Reasons for this failure include:

1. Narrow specification i.e. parents of children with GOR. Whilst these parents are over-represented at Great Ormond Street Hospital, they probably comprise a very small proportion of the London population.
2. Study was advertised on a Tuesday morning, which may be a difficult time for working parents
3. Focus group was located in central London
4. Not enough time between advertising event and focus group date (3 weeks)

The first attempt to launch the REFLUX UX workshop was also marred by a lack of reward. As observed by Al-Awar et al(223), usability testing often requires offering inducements to testers.

We changed strategy and decided to target the focus group more generally. Notably, the second series of advertisements, in which the £30 Marks and Spencer's vouchers were prominently featured, resulted in immediate responses and better uptake. We recruited 12 participants in 3 days.

Sampling bias

The app was only available for iPhone users. Although we did not limit attendance to iPhone users, the advertising mentioned that the app was only valuable for the iPhone. As we had access to only 3 additional iOS devices for the focus group, we hoped that most participants would bring and iPhone or iPad.

iOS devices are popular due to the high quality and wide availability of apps. Therefore, iPhone users are already self-selected to be enthusiastic about app usage. Accepting this bias, we deliberately avoided or truncated discussions about phone platforms. The discussion on patterns of app usage were kept brief.

The composition of the group was mostly health young professionals involved in research or students. Their app usage was high, but their burden of disease was low. Only one participant had a health condition that he tracked using an app. Only 3 participants were parents. Therefore, this group of people may not capture real-life insights on app usage when the target audience is parents of sick children.

Were we to repeat this study, we would recruit our focus group directly from GOSH outpatient clinics. That would demand more time and preparation, and an ethical approval process.

Sample size analysis

It is important to assess whether out sample size was sufficiently large to identify app design problems. In usability testing circles, sample size estimation is a perennial issue that has been addressed by

multiple authors. This question has been addressed by observations from experiments and mathematical modelling. The answer has ranged from “5-6 testers”(234)(226) to “at least 12”(225).

A system developer has a system they wish to test. They start off with one tester and subsequently add another, going on till they have a group of n testers. Initially, we expect that ever more problems will be found by more testers. Yet we also expect that there will comes a point where additional subjects identify problems already identified by other users. At this point, the probability of discovering new problems no longer increases. To determine the optimal number of testers (n), a researcher needs to quantify the likelihood of problem discovery and the number of problems to be identified.

Using binomial theorem to estimate sample size.

Lewis(235) used binomial probability theory to estimate the number of testers required. In their analysis, each tester represented a trial (n). The assumption made is that testers are trying to identify only 1 problem (k=1) problem. It is assumed that, a single observation of a problem was sufficient to trigger action to solve the problem. Therefore, the probability of problem discovery is estimated at 0.5 i.e. a single problem sufficiently obvious so as to confuse 50% of the population.

$$f(x) = \binom{n}{k} p^k q^{n-k}$$

Where

x = an integer denoting the number or problems identified

n = number of trials i.e. the number of testers

κ = number of successes

$n - \kappa$ = number of failures

p = probability of success in one trial

$q = (1 - p)$, probability of failure in one trial

Based on these assumptions, Lewis(235) recommended that 6 testers be could be used to detect a problem that would confuse 50% of the population. Considering the cost of usability testing, the suggestion that a system can be sufficiently tested with 6 users would be welcomed by most experimenters. However, limitation of the binomial model and its application to the REFLUX UX workshop should be appreciated before declaring satisfaction with our sample size.

Reviewing data from the REFLUX UX workshop, we summarised the number and proportion of problems discovered by each tester. In our study, we observed 42 discrete problems. The mean probability of problem discovery (p) was 0.2. Assuming a binomial distribution, we generated a participant-problem matrix for the REFUX workshop. The probability reported is the cumulative probability of discovering *at least* κ problems.

Table 35: Participant-problem matrix for the REFLUX UX workshop. Binomial distribution is assumed, p =0.2

Problems / Participants	1	2	3	4	5	6	7	8	9	10
1	0.00									
2	0.04	0.00								
3	0.10	0.01	0.00							
4	0.18	0.03	0.00	0.00						
5	0.26	0.06	0.01	0.00	0.00					
6	0.34	0.10	0.02	0.00	0.00	0.00				
7	0.42	0.15	0.03	0.00	0.00	0.00	0.00			
8	0.50	0.20	0.06	0.01	0.00	0.00	0.00	0.00		
9	0.56	0.26	0.09	0.02	0.00	0.00	0.00	0.00	0.00	
10	0.62	0.32	0.12	0.03	0.01	0.00	0.00	0.00	0.00	0.00
11	0.68	0.38	0.16	0.05	0.01	0.00	0.00	0.00	0.00	0.00
12	0.73	0.44	0.21	0.07	0.02	0.00	0.00	0.00	0.00	0.00
13	0.77	0.50	0.25	0.10	0.03	0.01	0.00	0.00	0.00	0.00
14	0.80	0.55	0.30	0.13	0.04	0.01	0.00	0.00	0.00	0.00
15	0.83	0.60	0.35	0.16	0.06	0.02	0.00	0.00	0.00	0.00
16	0.86	0.65	0.40	0.20	0.08	0.03	0.01	0.00	0.00	0.00
17	0.88	0.69	0.45	0.24	0.11	0.04	0.01	0.00	0.00	0.00
18	0.90	0.73	0.50	0.28	0.13	0.05	0.02	0.00	0.00	0.00
19	0.92	0.76	0.54	0.33	0.16	0.07	0.02	0.01	0.00	0.00
20	0.93	0.79	0.59	0.37	0.20	0.09	0.03	0.01	0.00	0.00

Some observations can be made from the participant-problem matrix above. Firstly, binomial distribution is only predictive where n ≤k i.e. where the number of testers if fewer than the number of problems requiring identification. Secondly, where the number of problems is fixed, the probability of problem detection increases with each participant and tends to 1. Therefore, the REFLUX UX workshop had only a 50% probability of discovering at least 1 problem despite having 8 participants. Compared to Lewis' findings (235), the workshop required 2 more testers to achieve the same problem detection probability. The workshop also performed less well when compared to the findings of Landauer and Nielsen (226). In a review of 11 usability studies, thy found the mean observed problems to be 42 and the mean problem identification rate to be 0.33. Compared to these data, the workshop observed number of problems is comparable, but the problem discovery rate for each participant is low.

There are some flaws to the assumption of binomial distribution in problem discovery. Firstly, independence of trials is a requirement for binomial probability. In problem discovery, each tester is treated as an independent trial. However, as demonstrated by Virzi(234), the nature of usability problem discovery is not random. More severe and obvious problems are detected by more than one tester, and detected prior to less severe problems. In observed data, Virzi(234) found that first 4-5 users find 80%

of problems in a 3 usability tests reviewed. Therefore, testers are cannot be treated as independent trials, especially in the context of a focus group

Secondly, using the average as a summary statistic for the probability of problem discovery does not sufficiently account for variability in the test environment and tester performance. In 1990, a published review of usability studies by Nielsen and Molich(236) observed the average user's problem discovery rate of 4 systems as ranging from 0.2 – 0.5. In a later (1993) review of usability studies, Nielsen(236) observed rates of 16-61% across 11 different systems. In 2006, Lewis(235), in a review of usability studies, found this probability ranged from 0.16 to 0.42. Estimates of problem discovery rates are also flawed by the pooling of data from evaluation of a variety of systems e.g. software, websites, computer graphical user interfaces. Although we expect some variation between users observing the same system (evaluator effect), data suggest great variance.

Alternative estimation using a Poisson model

In a 1993 publication, Nielsen and Landauer(226) acknowledged the limitations of the binomial model for describing usability problem discovery. Taking a Poisson distribution approach, they revisited the issue of estimating the tester sample size.

Poisson distribution describes the discrete probability of an event, where the event occurs randomly within a discrete interval of space or time. In usability terms, a Poisson model can be used to estimate the probability of discovering a problem during a usability evaluation exercise.

In a review of 11 usability studies, Nielsen and Landauer's assumed that the number of problems identified was the total possible number of problems identifiable (N). Based on this assumption, they calculated the mean rate of problem identification (33%).

The probability of a user identifying a problem was estimated using the simplified formula:

$$Found(i) = N(1 - (1 - \lambda)^{i})$$

Where

i = number of evaluators

N= number of problems to be found

λ = average probability of finding a problem during one test

1-λ = the probability of a problem remaining unfound, given that it has not been found already

As the number of evaluators (i) and the number of problems found (N) was known, they were able to estimate lambda (λ=0.31).

Figure 45: Copied with permission Nielsen and Landauer. The points in the chart are observed data from 11 usability studies. The lines represent fitting of the study data to a Poisson model.

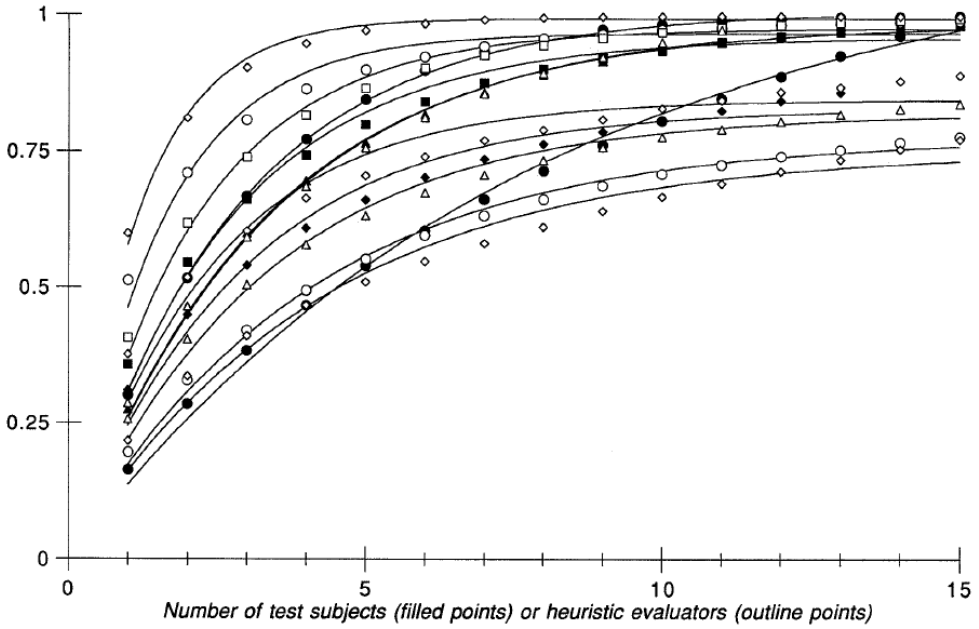


Figure 1 *Proportion of usability problems found with increasing numbers of subjects or evaluators for the interfaces in Table 1. The markers indicate the actual values from the studies and the lines indicate the fitted curves according to (EQ 1). The values from the various studies have been normalized to proportions rather than absolute number of problems to allow comparisons in a single figure.*

The figure above illustrates two observations. Firstly, the proportion of problems detected increases with the number of testers utilised. Secondly, data modelled by a Poisson process (lines) closely fitted observed data (points). This finding is confirmed by the high R^2 model fit values (0.97-1) for each study.

Problem discovery in usability testing appears to be better modelled by a Poisson distribution for a number of reasons. Firstly, defining the number of problems to be identified by testers cannot be defined prior to the test. In contrast to a binomial distribution (which requires the number of problems to be at least <1 less than the number of testers), an unlimited number of outcomes is possible.

Secondly, an infinite number of trials is assumed and the rate of problem discovery in a unit of time. Given the variability observed in performance of testers, the average rate of problem discovery during a usability test interval offers a better summary statistic than average tester performance.

One limitation of the Poisson model is the assumption that event occurrences in one interval are independent of others. The corollary in usability testing is the assumption of no linkage between the discovery of one problem and subsequent discovery of another. In reality, usability problems are often linked. They may be linked due to the functionality of a system e.g. finding a broken link on a website is linked to the inability to navigate to the page it leads to.

We used a Poisson model to estimate the probability of problem discovery during the REFLUX UX workshop.

The probability of problem discovery within usability test interval is described by the equation

$$f(\kappa;\ \lambda) = P(X = \kappa) = \frac{\lambda^{\kappa} e^{-\kappa}}{\kappa!}$$

where

each user represents a usability test interval

λ = 0.2 (rate of problem discovery)

κ is the number of problems to be identified

e is Euler's number i.e. 2.71828

Assuming a Poisson process, we designed a participant-problem matrix. The probability reported is the cumulative probability of discovering *at least* κ problems.

Table 36: Participant-problem matrix, assuming that problem discovery during the REFLUX-UX workshop is described by a Poisson distribution

Problem / Participant	1	2	3	4	5	6	7	8	9	10
1	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.06	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.12	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.19	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.26	0.08	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.34	0.12	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00
7	0.41	0.17	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00
8	0.48	0.22	0.08	0.02	0.01	0.00	0.00	0.00	0.00	0.00
9	0.54	0.27	0.11	0.04	0.01	0.00	0.00	0.00	0.00	0.00
10	0.59	0.32	0.14	0.05	0.02	0.00	0.00	0.00	0.00	0.00
11	0.65	0.38	0.18	0.07	0.02	0.01	0.00	0.00	0.00	0.00
12	0.69	0.43	0.22	0.10	0.04	0.01	0.00	0.00	0.00	0.00
13	0.73	0.48	0.26	0.12	0.05	0.02	0.01	0.00	0.00	0.00
14	0.77	0.53	0.31	0.15	0.07	0.02	0.01	0.00	0.00	0.00
15	0.80	0.58	0.35	0.18	0.08	0.03	0.01	0.00	0.00	0.00
16	0.96	0.62	0.40	0.22	0.11	0.04	0.02	0.01	0.00	0.00
17	0.97	0.66	0.44	0.26	0.13	0.06	0.02	0.01	0.00	0.00
18	0.97	0.70	0.48	0.29	0.16	0.07	0.03	0.01	0.00	0.00
19	0.98	0.73	0.53	0.33	0.18	0.09	0.04	0.02	0.01	0.00
20	0.98	0.76	0.57	0.37	0.21	0.11	0.05	0.02	0.01	0.00

As demonstrated in the table above, there is a 48% probability of discovering at least 1 problem during

the REFLUX UX workshop. To achieve a 95% cumulative probability of observing at least 1 problem, we would require at least 15 participants.

Iterative testing

The number of problems discovered increases with the number of testers deployed. However, the number of testers used in a single usability experimented is limited by the experimenters' ability to observe them. Furthermore, not all problems are immediately apparent. Virzi(234) found that early testers discover common and severe problems. Therefore, simply having the maximum number of testers one can observe (or can afford) is not viable. To discover as many problems as possible, an iterative testing process is recommended.

In this approach, the problems identified in the first test are identified and resolved. This resolution then sets up a new test environment. New users are introduced with the idea that new problems will be identified. This process was observed and described succinctly by Al–Awar et al(223) in 1981. When observing testers using a program for the first time in a usability seminar they noticed: "During the initial tests, the experimenter is very busy…As the programme became more accommodating, the number of interventions decreased to zero. "

When does the test-evaluate re-rewrite cycle come to an end? Al-Awar et al(223) offered a rule of thumb: Ninety five percent of users should be able to complete the tasks without any significant difficulties. An experimenter will know they had found that their final group of subjects when a group of testers complete a test without any significant difficulties. However, this approach limits the experimenter's ability to estimate the number of cycles, the cost and the time required to complete usability testing.

A further application of the Poisson model is the estimation of the number of iterative tests required. If an experimenter completes a usability tests with two participants, an empirical value for λ (rate of problem discovery) can be obtained. Revisiting our participant –problem matrix, we assume that each REFLUX UX workshop has 8 participants and hence has a problem discovery rate of 48%. Assuming that each workshop represents a Poisson process, we can see that 10 testing cycles would deliver a 95% probability of discovering at least 1 problem.

Table 37: Participant-problem matrix for problem discovery for hypothetical each RELFUX-UX workshop

Workshops	1	2	3	4	5	6	7	8	9	10
1	0.08	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.25	0.07	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.42	0.17	0.06	0.02	0.00	0.00	0.00	0.00	0.00	0.00
4	0.57	0.30	0.13	0.04	0.01	0.00	0.00	0.00	0.00	0.00
5	0.69	0.42	0.22	0.09	0.03	0.01	0.00	0.00	0.00	0.00
6	0.78	0.54	0.32	0.16	0.07	0.03	0.01	0.00	0.00	0.00
7	0.84	0.65	0.43	0.24	0.12	0.05	0.02	0.01	0.00	0.00
8	0.89	0.73	0.53	0.33	0.18	0.09	0.04	0.02	0.01	0.00
9	0.93	0.80	0.62	0.42	0.26	0.14	0.07	0.03	0.01	0.00
10	0.95	0.85	0.70	0.51	0.34	0.20	0.11	0.05	0.02	0.01
11	0.97	0.89	0.77	0.60	0.42	0.27	0.16	0.08	0.04	0.02
12	0.98	0.92	0.82	0.67	0.51	0.35	0.22	0.12	0.06	0.03
13	0.99	0.95	0.86	0.74	0.58	0.42	0.28	0.17	0.10	0.05
14	0.99	0.96	0.90	0.79	0.65	0.50	0.35	0.23	0.14	0.08
15	0.99	0.97	0.92	0.84	0.72	0.57	0.42	0.29	0.18	0.11
16	1.00	0.98	0.94	0.88	0.77	0.64	0.49	0.35	0.24	0.15
17	1.00	0.99	0.96	0.90	0.82	0.70	0.56	0.42	0.29	0.19
18	1.00	0.99	0.97	0.93	0.85	0.75	0.62	0.48	0.35	0.24
19	1.00	0.99	0.98	0.95	0.89	0.80	0.68	0.55	0.42	0.30
20	1.00	1.00	0.99	0.96	0.91	0.84	0.73	0.61	0.48	0.35

We were only able to achieve one cycle of focus group testing. A key reason was recruitment difficulties. A second reason was cost.

Cost

The TARDIS:REFLUX app was generously supported by a grant from UCL Grand Challenges Small Projects (£5000) and Train and Engage Fund (£750). The app was produced at a cost of £6308. Compared to estimates of cost given by commercial companies during the scoping exercise (£15000-£35000), the app was developed and deployed at below-market cost.

Table 38: Summary of app development costs

Development phase	Cost	UCL Grand Challenges	Train and Engage fund	Personal funds
Specification	Balsamiq software			120
Design and implementation	iOS Developer	3000		
	XCode SDK developer account (yearly subscription)			210
Testing	Materials (posters, envelopes, printed questionnaires, stamps)			60
	Advertising focus group-Facebook			240
	Developer review	1000		
	Vouchers for focus group attendees		240	
	Focus group catering		140	
	Focus group- videography		370	
Dissemination	iOS Developer	720		
	Advertising-Facebook			208
	Total	4720	750	838

In this study, the cost of app testing was not factored into the original funding application (UCL Grand Challenges). Testing was made possible by applying for a second grant (Train and Engage Fund).

Costs include equipment for testers, audio-visual recording of usability tests, expenses claimed by testers and inducements offered to testers. The cost of usability testing should be factored into the cost of developing a research app.

Summary

The focus group study design was modified in several major aspects due to recruitment issues. Firstly, we reduced the number of workshops from two to one. We postponed the date because of poor subscription to the initial date. We widened the target audience once the difficulty of exclusively recruiting parents of children with reflux in public forums became apparent. Despite these changes, the focus group participants offered surprising and valuable insights.

A focus group is versatile tool and effective tool. Qualitative data is obtained from the impressions and opinions of participants. However, quantitative and technical information that is unavailable during alpha testing can be obtained e.g. time taken to submit a symptom report. The focus group also generated some interesting and general insights about the utility of apps in symptom tracking and research. We obtained invaluable feedback on specific elements of the app. These will be used to inform the next development cycle.

CHAPTER 6: DISSEMINATION

The development of the app was envisioned as a proof-of-concept project. The culmination of this project is the public release and application of the tool. The public response to this app is a measure of its social acceptability.

LAUNCH

We submitted the app for review on the iOS iTunes platform on 5th January 2014. The app publication parameters were follows:

Table 39: iTunes app store submission parameters

Version	1
Copyright	2014, Dr. E. Macharia
Primary Category	Medical
Secondary Category	Health & Fitness
Age rating	4+ (Apps in this category contain no objectionable material)
Default language	UK English
Keywords	Gastroesophageal, reflux, baby, children, symptoms, vomiting, acid, heartburn, GERD, GORD, Nissen
Support URL	http://www.ucl.ac.uk/tardis
Privacy Policy URL	http://www.ucl.ac.uk/tardis/Privacy

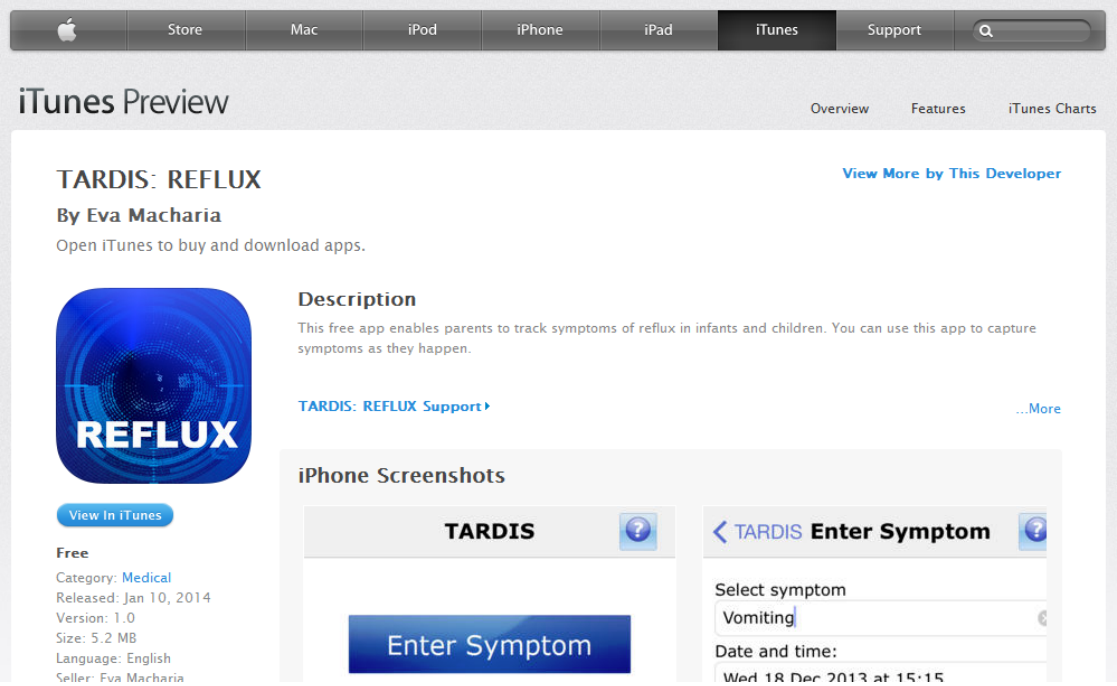
The app was accepted for publication on the iTunes Store on first review. It was published the TARDIS:REFLUX app for the iPhone on the iTunes store on 10th January 2014. The uniform resource locator (URL) assigned to the product on the store is:

https://itunes.apple.com/us/app/tardis-reflux/id792659458?mt=8

The project URL is www.ucl.ac.uk/tardis

Initially, the app was launched without any promotion. Over the first month, we monitored downloads and feedback to ensure there were no technical issues. Within this first month, we received no crash reports, negative or positive reviews.

Figure 46: Screenshot of listing of TARDIS:REFLUX app on the iTunes app store. https://itunes.apple.com/us/app/tardis-reflux/id792659458?mt=8



After this period, the app was promoted on the internet to encourage targeted user downloads.

Promotion

Facebook

We designed an advertising campaign for Facebook. Between the 8th and 15th of March 2015, an advertisement with the following target audience was displayed.

Table 40: Targeted advertising on Facebook

Campaign	TARDIS: REFLUX - App Installs
Start Date	20/03/2014
End Date	18/04/2014
Devices	Mobile, iOS
Country	USA United Kingdom India Canada Nigeria
Demographic	+18 years
Interests	Parenting, Babies, Family, Paediatrics

The app ran for a week and had a wide reach and average click through rate. The advertising resulted in a spike in app downloads from the iTunes store.

Table 41: Summary of performance of Facebook advertising campaign

Reach	37384
Frequency	1.3
Impressions	47925
Clicks	371
Unique Clicks	344
Click-Through Rate	0.8
Unique Click-Through Rate	0.9
Cost Per Click	0.3

NHS apps library

The app was also submitted for review to the NHS apps library (24th February 2014). It was accepted on first review and was published on the NHS health apps library on the 25th of March 2014. The URL assigned is: https://apps.nhs.uk/app/tardisreflux/

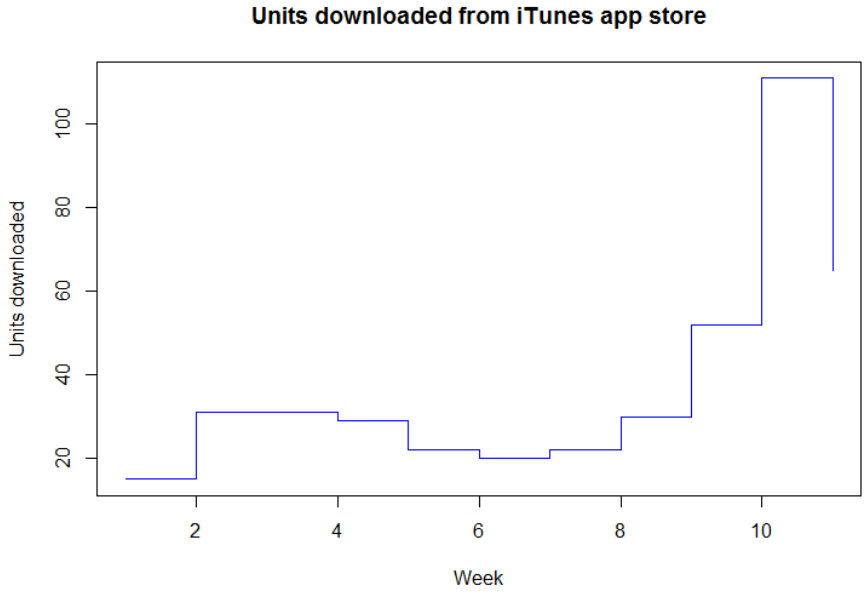
Figure 47: Screenshot of the TARDIS:REFLUX app on the NHS apps library

UPTAKE

iTunes Downloads

Between 10th January 2014 and 23rd March 2013, there were 428 units downloaded from the iTunes app store. The initial rate of downloads per week was low, with an average of 28 downloads per week. However, after the Facebook promotion in week 10, we observed a rapid increase in units downloaded a week. The weekly download rate rose to an average of 88 per week.

Figure 48: Number of TARDIS:REFLUX app units downloaded from the iTunes app store per week.



Most users downloaded the app onto an iPhone device. Notably, 71 users (16%) apparently downloaded the app onto an iPod touch. To our knowledge, the app does not function on this generation of iOS devices.

Table 42: Devices used to download the iTunes TARDIS:REFLUX app

Platform	Total
iPhone	315
iPod touch	71
iPad	37
Desktop	5
Total	428

The device was downloaded most frequently by users in the USA and Canada. The second largest group of users came from the Africa, Middle East and India territory with users in India contributing most unit downloads to this group.

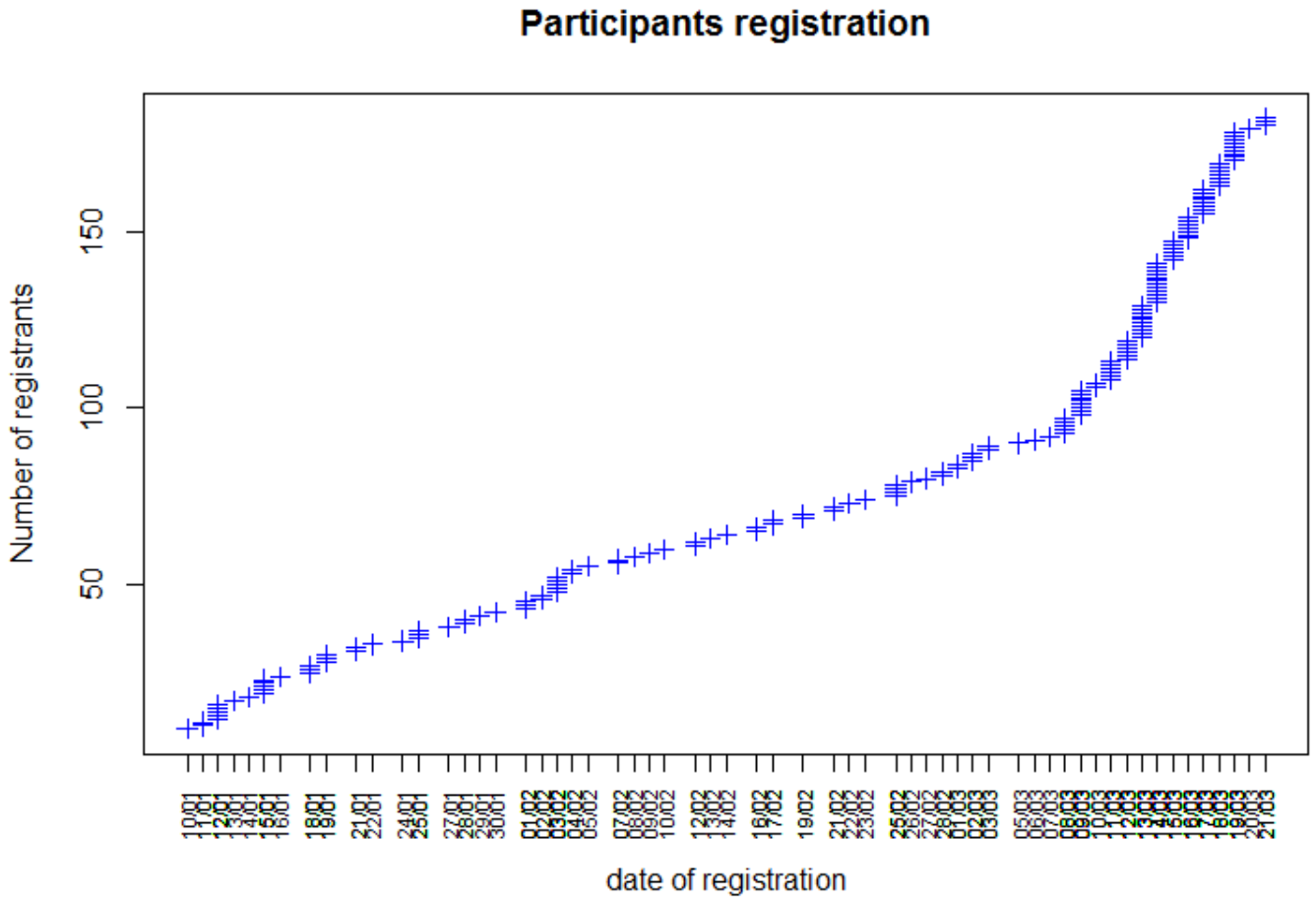
Table 43: Unit downloads per iTunes sales territory

Super Territory	Units
USA and Canada	200
Africa, The Middle East, and India	108
Europe	87
Asia Pacific	26
Latin America and The Caribbean	7
Total	428

Data submitted to TARDIS database

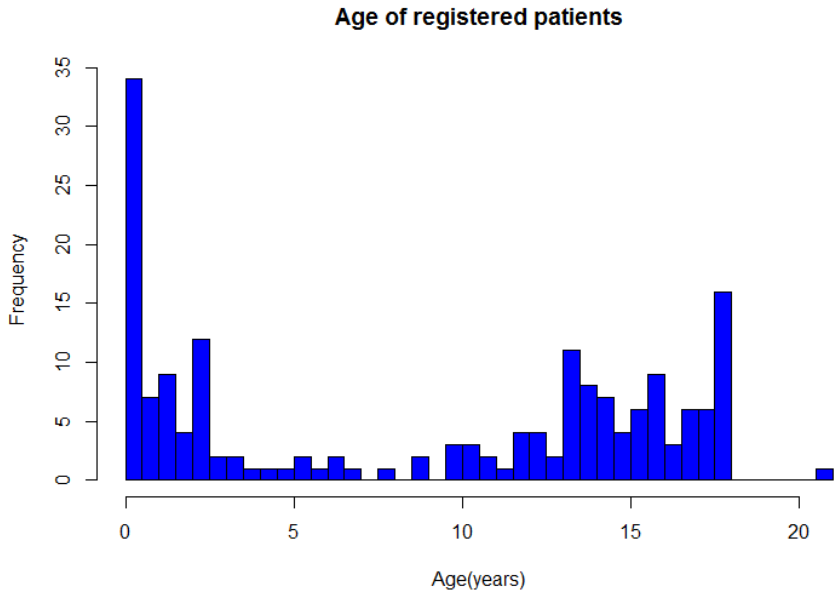
After download, logging into the app would result in a user registration on the TARDIS database. The first user registered on the 10th of January 2014 when the app was first launched on the app store. On the 8th of March 2014, a Facebook promotion was launched to increase the reach of the app. This ad campaign led to an increase in the number of user registrations. By the 26th of March 2014, there were 181 app registrants.

Figure 49: Number of users who registered and logged in on the app



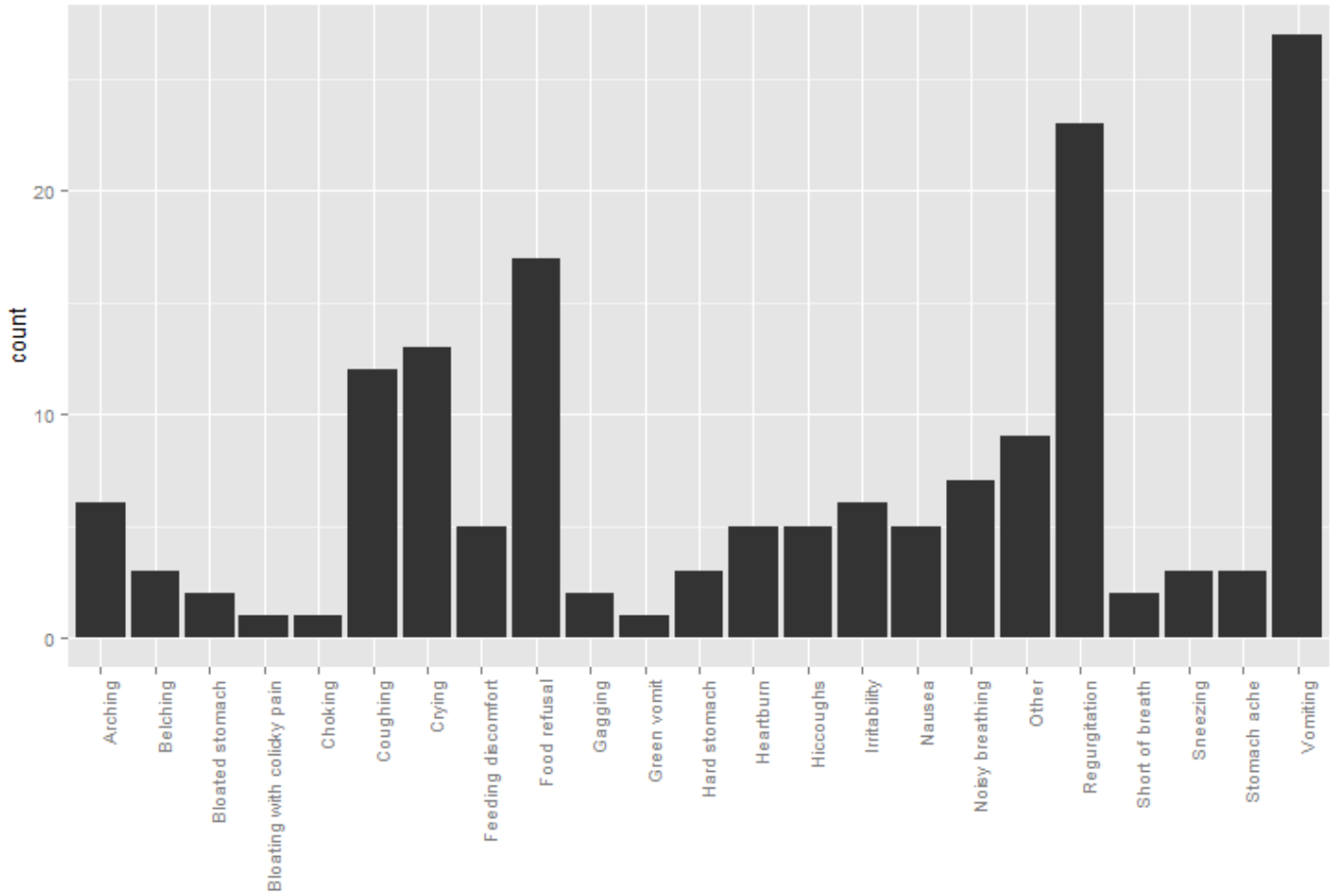
The mean age of registered patients was 8.8 years (standard deviation 6.9 years). The median age of registered patients was 10.6 years (IQR 1.2-15.2).

Figure 50: Age characteristics of subjects tracked



Not all registered users went on to track symptoms or events. Of the 188 registered users, 97 recorded and submitted symptom data. There 161 symptom episodes submitted. Within these, there were 23 symptoms reported.

Figure 51: Reported symptoms



Vomiting was the most commonly reported symptom (27 episodes). Regurgitation (23 episodes) and food refusal (17) were the next most common symptoms.The median number of symptoms submitted per subject was 1 (range 1-4). The mean number of symptoms submitted was 1.6 (Standard deviation).

There were 45 events reported by 35 registrants. Medication events were the most common category of event reported (12, 27%).

Table 44: Events reported by registered users on the TARDIS: REFLUX app on iTunes

Event category	Event	Episodes	Total
Medication	Started Omeprazole	5	14
	Started Ranitidine	2	
	Stopped Ranitidine	2	
	Increased Dose Omeprazole	1	
	Reduced dose Omeprazole/Lansoprazole	1	
	Started Domperidone	1	
	Stopped Domperidone	1	
	Stopped Omeprazole/Lansoprazole	1	
Weight	Weight	12	12
Procedure	Nasojejunal tube	4	8
	Gastrostomy Removal	2	
	Nasogastric tube	2	
Diagnosis	Constipation	3	11
	Diarrhoea	3	
	Nose Bleed	2	
	Productive cough	2	
	Pneumonia	1	
Total			45

Summary

The launch of the TARDIS: REFLUX app on a commercial and public platform is the final demonstration of the feasibility of this TARDIS tools. The reality of the app being downloaded and used by people naïve to the research study suggests that there is both utility and demand for this tool. It demonstrates the exciting possibilities of directly 'crowdsourcing' symptom level data for research.

The age of children being tracked is surprising. As gastroesophageal reflux is a condition of infancy, we expected most users to record symptoms of children under two. Instead, we observe a bimodal distribution with peaks at 6 months and 15 years.

The frequency of symptoms reported by the public is also insightful. In keeping with published data, vomiting and regurgitation are the most common symptom reported. Surprisingly, food refusal is the next most common symptom.

It is observed that app usage is not sustained. Most users report 1-4 symptoms, then give up using the app. This reinforces the suggestion arising from focus group testing i.e. symptom tracking is most

effective when driven by a contextual need. There should be a pre-defined goal e.g. assessing impact of an intervention on symptoms. Tracking should also be limited to brief and discrete episodes.

The implications of these findings and applications for future study are discussed further in Section V: Discussion.

SECTION III: STRATIFYING RISK OF FUNDOPLICATION IN CHILDREN WITH GORD: RETROSPECTIVE DATABASE STUDY METHODOLOGY

CHAPTER 1: INTRODUCTION

In Section 1 the epidemiology of GORD was reviewed. Gaps in the knowledge of this condition were identified. Key emerging themes were the paucity of symptom data and the diagnostic challenge presented by association with multiple comorbidities. These comorbidities are not mutually exclusive. Indeed, they may occur in the same patient. Therefore, understanding comorbidities as independent, inter-dependent or confounding variables is important.

In order to understand the effects of co-morbidities, it is necessary to study a large cohort of patients with GOR. Ideally, this would be done using a prospective study design with a uniform pro-forma data collection instrument. The disadvantage of this approach is that few patients could be included. Furthermore, a study of this nature would take several years as the treatment , treatment effect and outcomes emerge on such a scale.

Instead, a pragmatic assessment of data already collected and available was made. Indeed, our institution has treated children with GORD for more than 30 years, suggesting a large retrospective cohort available for study. Systematic review of data already collected may yield valuable insights.

The overall aim of this study is to address risk stratification in the surgical management of patients with GOR. In a clinical encounter, a parent may ask a surgeon: “Will fundoplication improve my child’s symptoms?”. The surgeon answering the question will weigh risks against benefits. Benefit can be defined in several ways e.g. improvement of symptoms, reduced medication requirement, reduced requirement for fundoplication, reducing risk of morbidity and mortality. From a surgical perspective, therefore, the key outcomes of interest are:

1. Risk of fundoplication
2. Risk of failure of fundoplication
3. Risk of mortality

METHODOLOGY AND RATIONALE

A data mining approach was chosen for the collection, collation and analysis of the retrospective GOR cohort.

A classic, experimental approach is hypothesis-driven. Firstly, the hypothesis is stated. Then, the null hypothesis is then defined. An experiment is designed to disprove the null hypothesis. Then follows the hypothesis-drives data acquisition, collation and analysis.

In contrast data mining is a process of knowledge discovery. The data are collated, described and analysed. Input-output dependencies are sought. A hypothesis may be reached **after** analysis. Therefore, the data drives the hypothesis.

The paradigm shift from classical experimental approaches to data-driven approaches has occurred in the last 20 years(237) . Several factors have contributed to this shift. Firstly, the digital age has led created the capacity to passively collect large amounts of data. Data mining tools and methods are becoming more readily available to the average researcher. Access to statistical packages and cloud computing in academic institutions are making ‘big data’ accessible to the pedestrian academic(237). Thus, the power of large datasets to answer important questions is becoming clear.

Data mining of large databases i.e. database studies are, essentially, retrospective cohort studies. Database studies may be descriptive or comparative. In the hierarchy of evidence(238), they qualify as level III evidence.

The goals of data mining are two-fold i.e. description and prediction. Descriptive data mining is generation of new and non-trivial knowledge based on the available data set. The aim of predictive data mining is to produce a model to describe the system based on the data available. Both these goals are achieved through standard statistical and computational methods.

Data mining methodology is appealing for the study of GOR for following reasons:

- *Complex systems*: Data mining is particularly suited complex and unknown biological systems where interdependencies between variables, and the relationships between the variables and outcomes not fully understood *a priori*. GOR is certainly a condition where many demographic and comorbid variables require consideration.
- *Untapped resources*: Institutions reflexively gather large amounts of data. The data gathered on patients with GOR at GOSH as a largely untapped resource with the potential to reveal insights on GOR in children.
- *Agnosia*: Data mining requires an agnostic stance. The researcher abandons pre-determined ideas and assumptions. In the study of disease, where clinical traditions and assumptions often bear greater weight than empirical data, such a naïve approach may lead to novel insights.
- *Pragmatism*: A prospective observational study of a cohort of patients with GOR would be an ideal methodology to describe the epidemiology of this condition. However, as described in Section 1, the symptoms associated with a diagnosis of GORD are multiple and sometimes poorly defined. A prospective clinical observational study involving blanket recruitment of patients with symptoms ranging from gurgling to hiccoughs would be impractical. Ethical approval for inclusion of a large group of patients with a poorly defined symptomatology would be difficult to obtain. Such a study would also be expensive to implement. The outcomes measured in GOR e.g. medication-independence, fundoplication, are latent. Given the frequency of this condition in the population, it is estimated that one would need 3-8 years of follow-up to obtain outcome data. For the purposes of a doctorate, such an approach would be impractical.
- *Real-world approach:* In clinical management of conditions where symptomatology is complex and tests are inconclusive, the physician's impression is very influential. A retrospective cohort will reflect medical decision-making and utilisation. This may give a truer account of the management of a condition in the population than the clinical stage management of a clinical trial.
- *Paucity of knowledge*: In the literature review, it became clear that the symptoms and comorbidities associated with GORD are multiple and poorly defined. Therefore, a study identifying frequency and distribution of comorbidities in a large cohort of patients with GORD would add value to current knowledge.

Retrospective database studies have inherent limitations. Selection bias is a primary concern. Database composition may be compromised by injudicious or inconsistent inclusion of study subjects.

Selective sampling is inevitable because none of the databases queried were explicitly designed for the purpose of gathering a cohort of patients with GOR. Moreover, no single database available at GOSH contained all the data required on each patient. The final collation, is therefore an amalgam of data gathered from multiple databases. Data are supplied into the databases by clinicians and administrators, for many different purposes. This provenance may affect data quality. To mitigate, real effort has been taken to describe the commission and function of each data source. Where possible, search parameters have been standardised across databases.

Another methodological problem with retrospective data is measurement bias. In describing the epidemiology of a condition, an attempt is made to establish an association between exposures (premorbid and comorbid conditions) and outcomes. However, inconsistent or incomplete measurement of variables may lead to an inaccurate description of their associations.

Symptoms are particularly prone to measurement bias. Clinicians, parents or patients are the measurement tool. Observations can be subjective. For example, what one observer may record as 'vomiting' may be recorded as 'regurgitation' by another.

Comorbid variables are less prone to measurement bias. Some objectivity is required to achieve a diagnostic standard e.g. congenital diaphragmatic hernia is a radiological and intra-operative diagnosis.

In summary, the quality of knowledge discovery in data mining is determined by the quality of the data collection. Robust methodology requires judicious use of data tools, statistical and computing methods. In this chapter, we describe the data collection approach taken.

DATA COLLECTION PLAN

Data mining may be considered a stylised application of standard data analysis tools(237) . Commonly, the steps followed are(237):

1. Defining the system: The system is the context of the study, akin to the population. The context of the problem is understood and the ecosystem of variables is explored.

2. Collect data: Data collection can be observational or experimental. This study is observational i.e. data are harvested from pre-existing data sources. The strategy for searching, inclusion and exclusion is defined.

3. Pre-process data: Data anomalies e.g. errors, missing points, outliers are identified. Data may also be encoded, for example converting disease status into a binary outcome variable. The feature selection strategy, in which salient variables are retained and extraneous variables are excluded, is defined.

4. Estimate the model: Descriptive tools are used to demonstrate the distribution of variables within the cohort. Standard descriptive statistics e.g. frequency, proportion, mean and medians are useful. Should descriptive tools demonstrate any patterns, predictive modelling tools such as linear and logistic regression can be used to further investigate the relationships between variables. Thus the data are used to generate a hypothesis model of how the system might work.

5. Interpret the model and draw conclusions: In this final step, we address the model in the context of the system. Does the model give us any novel or useful insights into how the system works?

In this section we describe the first three steps. In next section, we model the data and draw insights about GORD.

To enforce rigour, these data mining routines will be governed by principles of the ISPOR(239) checklist for retrospective database studies. This checklist(239) was developed to enable assessment of the quality of retrospective database studies and endorsed by the EQUATOR network(240). This checklist was used to design and structure both the search strategy and reporting of findings. It was useful for identifying data gaps and methodological problems. The resulting data collection plan is described in the table below.

Table 45: IPSOR checklist for reporting of retrospective database studies

Section	Characteristic	Reporting
Data Sources	Search Strategy	Rationale for database selection, strategy for patient identification within each database described.
	Relevance and Reliability	All data sources are described in their provenance and reliability. Data quality checks and cleaning procedures are described for each data source.
	Linkages	Where patient records are held on multiple databases, the strategy to verify linkages and ensure referential integrity is described.
Methods	Design selection	Systematic data search is guided by a priori research design. Hypothesis testing is described.
	Design limitations	Anticipated and unexpected limitations of design are described.
	Data analysis plan	Data mining approach described a priori.
	Treatment effect	Treatments / interventions of interest described.
	Study population and variable description	Cohort of interest defined. Meta-variables e.g. demography and comorbidities defined.
	Eligibility and censoring	Inclusion and exclusion criteria for data sources and subset analysis described.
	Operational definition and validity	Key variables defined e.g. eligibility for member status in a comorbid cohort.
	Timing of outcome, event capture	Window of inclusion for interventions and follow-up described. Limitations in event capture e.g. event occurring at another hospital, described where relevant.
Statistics	Control variables	For comparison analyses, case and controls described
	Relevant variables	Measured and unmeasured variables described.
	Influential cases	Role of influential cases assessed and factored into statistical analysis
	Statistical model	A priori hypothesis testing adhered to, informing model selection
	Testing assumptions	Assumptions e.g. normality, data independence assessed where relevant
	Multiple tests	Data mining approach to identify best model fit for data, multiple tests for each dataset
	Model predictions	Predictive power of each model assessed, verified and compared
Discussion	Theoretical basis	Relevance to theory discussed
	Practical vs. statistical significance	Clinical implication discussed
	Generalizability	Utility of data and findings in informing other clinical scenarios assessed.

CHAPTER 2: DEFINING THE SYSTEM

The cohort of patients, their variables e.g. comorbidities and their outcomes e.g. mortality can be defined in data mining terms as an ecosystem of data. The exploration of this data system is retrospective. Conceptually, the experiment / hypothesis test has already taken place. The role of the data miner is to observe and report the interactions of the subjects, variables and outcomes. Therefore, systematic exploration of data begins with clear definitions of the subjects, variables and outcomes.

The subjects arise from the population of children treated at GOSH. They are the subset of children who had symptoms and signs of GORD within a predefined capture period. The variables affecting these outcomes, as described in Section 1, are the demographics, comorbidities, symptoms and investigative results of patients with symptoms of GOR. We have defined the key outcomes of interest as :

1. Risk of fundoplication
2. Risk of failure of fundoplication
3. Risk of mortality

Rephrasing this in data mining terms, the target system is defined as:

$$y = f(x)$$

where:

- x= input variables i.e. demographics, comorbidities, symptoms and investigative results
- y= outcomes i.e. risk of fundoplication or re-do, or mortality.
  - Observed outcomes (y) are what actually happened to the patient in real life.
  - Expected outcomes (E(y)) are experimental i.e. predicted outcomes after applying a formula or function.
- f= function. This is the yet unidentified function that describes the relationship between x and y i.e. the model.
- The 'goodness of fit' of the model / function will be defined by its ability to minimise y-E(y) i.e. the difference between observed outcomes (y) and expected outcomes (E(y)).

It is important to note that not all input variables are measurable. For example, we have captured the gestation at birth of included subjects. However, we have not captured weight at birth. It may be that this is an important factor influencing mortality outcomes. Unmeasured factors can contribute to the difference between observed and expected outcomes. Therefore, the input variable (x) is a product of measured and unmeasured factors. Therefore, the system may be more fully described as:

$$y = f(m, u)$$

Where;

m = measured variables

u = unmeasured variables

The goal of the mining task is to identify the function f i.e. the model that describes the relationship between input variables and outcomes.

DEFINING THE POPULATION

We describe the sociodemographic and health profile of the sample population i.e. patients attending Great Ormond Street Hospital (GOSH). The organisation of services and its effect on this demographic is also described.

Great Ormond Street Hospital

GOSH is a tertiary care hospital i.e. one which offers specialist consultative care based on referral from primary (e.g. GP) or secondary (e.g. district hospital) care services. Therefore, GOSH patients are a subset of the patients attending local hospitals, community paediatricians and GPs. Referral patterns are also based on geographical catchment. The southern border of this catchment is the River Thames. The northern, eastern and western boundaries are marked boundary is the M25 motorway. The eastern and western boundaries are marked by the Circular course of the M25 motorway. The local hospitals linked to GOSH detail below: Local services are provided to the following hospitals:

Table 46: Local hospitals within the geographical catchment area of GOSH

- Barnet
- Basildon
- St Mary's (Paddington)
- Central Middlesex
- Chase Farm
- Royal London
- Chelsea and Westminster
- Ealing
- St John's (Chelmsford)
- Hemel Hempstead
- Hillingdon
- University College London
- Homerton
- Newham General
- West Middlesex
- North Middlesex
- Northwick Park
- Whittington
- Queen's Hospital (Romford)
- Princess Alexandra (Harlow)
- Watford
- Royal Free
- Southend
- Whipps Cross

Specialist care

GOSH does not have an Accident and Emergency department, and so patients are referred to GOSH for a particular specialist service. Rare diseases benefit from concentration of clinical experience. GOSH is the only centre in the UK for certain conditions e.g. epidermolysis bullosa, gene therapy, cloaca and bladder exstrophy. For these conditions, the services have a national catchment. Specialist services at GOSH and their catchment area are below (Table 47).

Table 47: Specialist services available at GOSH at the time of data collection

Specialist service	Catchment area
Audiological medicine	Local
Bone Marrow Transplant	Regional
Cardiology, cardiothoracic surgery and cardiac intensive care unit (CICU)	Regional
Cleft lip and palate and craniofacial	Regional
Clinical genetics	Regional and National
Cystic fibrosis	Regional
Dental and Maxillofacial Surgery	Local
Dermatology (including Epidermolysis Bullosa)	National
Ear Nose and Throat (ENT)	Local
Endocrinology	Local
Gastroenterology	Local
Haematology and Haemophilia	Regional
Heart and lung transplant	Regional
Immunology	Regional
Infectious Diseases	Regional
Metabolic Medicine	Regional
Nephrology	Regional
Neurology, Neuro-disability and Neuromuscular	Regional
Neurosurgery	Local
Oncology	Regional
Ophthalmology	Local
Orthopaedics	Local
Palliative Care	Local
Paediatric and Neonatal Intensive Care Units (PICU/NICU)	Regional
Plastic Surgery	Regional
Pulmonary Hypertension	Regional
Radiology (Imaging) and Interventional Radiology	Local
Respiratory Medicine	Local
Rheumatology	Regional
Specialist neonatal and paediatric surgery (SNAPS)	Local
Speech and language therapy	Regional
Tracheal service	National
Urology	Local

The implications of specialist provision on selection of our cohort must be considered.

- GORD severity: Patients are referred to GOSH for secondary opinions or for management of conditions that cannot be managed elsewhere. The initial management steps of GORD can be carried out in the community e.g. feeding modulations, ASM. Therefore, patients primarily referred for review of GORD are likely to have severe GORD.
- GORD requiring surgical management: Most of the referring hospitals in the catchment area have no paediatric surgical service. Therefore, patients referred primarily for review of GORD would be

reviewed for consideration for gastrostomy +/- fundoplication. This creates a bias in the population at GORD i.e. selection for patients whose GORD has failed medical management.

- Disease rarity: GOSH is a leading centre for rare diseases of childhood. GORD may not be the primary condition for which the patient is attending GOSH. It may be a comorbidity that may or may not be associated with their primary condition for referral e.g. epilepsy.

- Severity of comorbid condition: GOSH is one of a minority of centres providing super-specialist services e.g. extracorporeal membrane oxygenation (ECMO), epispadias. Therefore, our cohort may have a higher incidence of rare comorbidities than is observed in the population. These rare conditions are often associated with severe morbidity and higher mortality. This may skew morbidity and mortality outcomes in this population.

Private versus NHS patients

GOSH offers care for patients through the NHS or as private patients. There are two kinds of private patients seen at GOSH.

1. UK: These are UK-based patients seeking a private consultation. Reasons include bypassing NHS waiting lists, guaranteeing continuity of care with a particular doctor, or seeking a second opinion. These patients are included in the analysis.

2. International: The international private patient (IPP) facility serves patients referred from beyond the UK. Referral patterns are idiosyncratic and reflect the health capacity of referring countries rather than established NHS referral protocols. The risk of incomplete data and losing patients to follow-up is greater with the IPP service. Therefore, IPP records are excluded from analysis. The GOSH patient identifier (PatientID) for both IPP and UK private is a six-digit number prefaced by PX e.g. PX888777. However, UK patients have a 10-digit NHS number while IPP do not. Therefore, records with both an NHS and private PatientID are rationalised, verified by date of birth (DOB), merged and included.

In summary, the referral pathways and organisation of services at GOSH results in a caseload that does not, inherently, reflect the distribution of disease in the population. Instead, the caseload reflects NHS policy for specialist provision, specialisms available at GOSH, disease severity and rarity.

DATA HANDLING

In querying the databases, data quality considerations emerged.

1. Consistency: This is uniformity of definitions or descriptions. Checking for consistency is particularly important where there is a human and subjective element e.g. results of investigations. For example, following endoscopy does reported 'severe oesophagitis' apply to a pre-defined set of findings? Use of established protocols and grading systems is helpful.

2. Redundancy: Data are redundant when they can be omitted without loss of meaning or function. For example, querying the theatre database for 'fundoplication' would yield multiple records for the same patient during the same theatre episodes. This was because transfer to, admission into and discharge from theatre, within the same episode, were archived returned as separate records. Records were rationalised using PatientID and episode date. Redundant records can be deleted. However, this process risks introducing errors of deletion. Therefore, a rationale and methodology for addressing redundancy was customised for each data source.

3. Referential integrity: Multiple databases hold data on patients at GOSH. A key task when amalgamating data is to ensure accurate data linkage. To address this, we have used a 2-step verification using DOB and name.

4. Verification: The primary verification method is cross-referencing patient data from two separate databases holding similar data e.g. operation notes and theatre administrative records. Where there has been a change of the coding or reporting methods over time, this is reported.

5. Timeliness: This attribute of data quality describes how contemporaneous data are for the purpose at hand. In review of databases, it was clear that not all records were created contemporaneously. For example, mortality data was often updated with delay to the PIMS database. Where there was lag, this was noted and impact on data quality addressed.

6. Completeness: Missing values can arise from inadequate sampling, difference in clinical pathways for individual patients and asynchronous databases. As the database is retrospective, it is impractical to gather missing data. To estimate and mitigate bias, each data source was reviewed for missing elements and this is factored into analysis.

Collation

Data are collated in spreadsheets / flat files/ tables (Microsoft Excel). A flat file consists of data organised into rows and columns / fields. To enforce this format, data are saved in the 'comma separated values (.csv) format (Microsoft Excel). After, .csv tables are linked and merged into a single relational database (Microsoft Access) i.e. [RetrospectiveGOR.db]. .csv and .db are data formats have import and export functionality.

Transfer

Computational tools were used to secure and verify data transfer. The researcher authored macros for use in Microsoft Excel and Access packages. Macros are a set of statements/ instructions which can be stored and executed repeatedly within a computer program. They are particularly useful for data processing and cleaning tasks as common errors recur and require similar remedies. Macros were

authored in Structured Query Language (SQL). It was possible to build in data validation procedures to minimise copying / pasting errors. Audit was possible, with each data import/export procedure generating an error file where failed verification has resulted in no transfer. Therefore, data are securely transferred without loss of information.

Merger

Data type validation procedures were used in the transfer of data. The four meta-categories of data types across systems are ‘date’, ‘number’, ‘string’ and ‘text’. Within these categories are further subtypes e.g. ‘date’ can be further described as date (YYYY-MM-DD), datetime (YYYY-MM-DD, hh:mm:ss) and time (hh:mm:ss). For example, a time of 0220hrs will be imported into a string column as 220 and into a text column as 0220. Consistency of data type was particularly important when merging, importing and exporting data. Enforcement of data type across databases was utilised to highlight and prevent conversion errors.

In all GOSH databases, patient records are indexed by a unique patient identifier (PatientID) and date of birth (DOB). The PatientID is a 6-digit numerical identifier assigned to new patients registered at GOSH. Assignment is in a serial fashion i.e. on a certain day, 3 new admissions will be assigned *****2, *****3, ******4 etc. A PatientID is a unique identifier for each patient. Used to index all subsequent paper and electronic records created. Therefore, this parameter was chosen to index and identify in patients included in [Retrospective GOR.db] and ensured referential integrity across database tables.

DATA PROCESSING AND CLEANING PROCEDURES

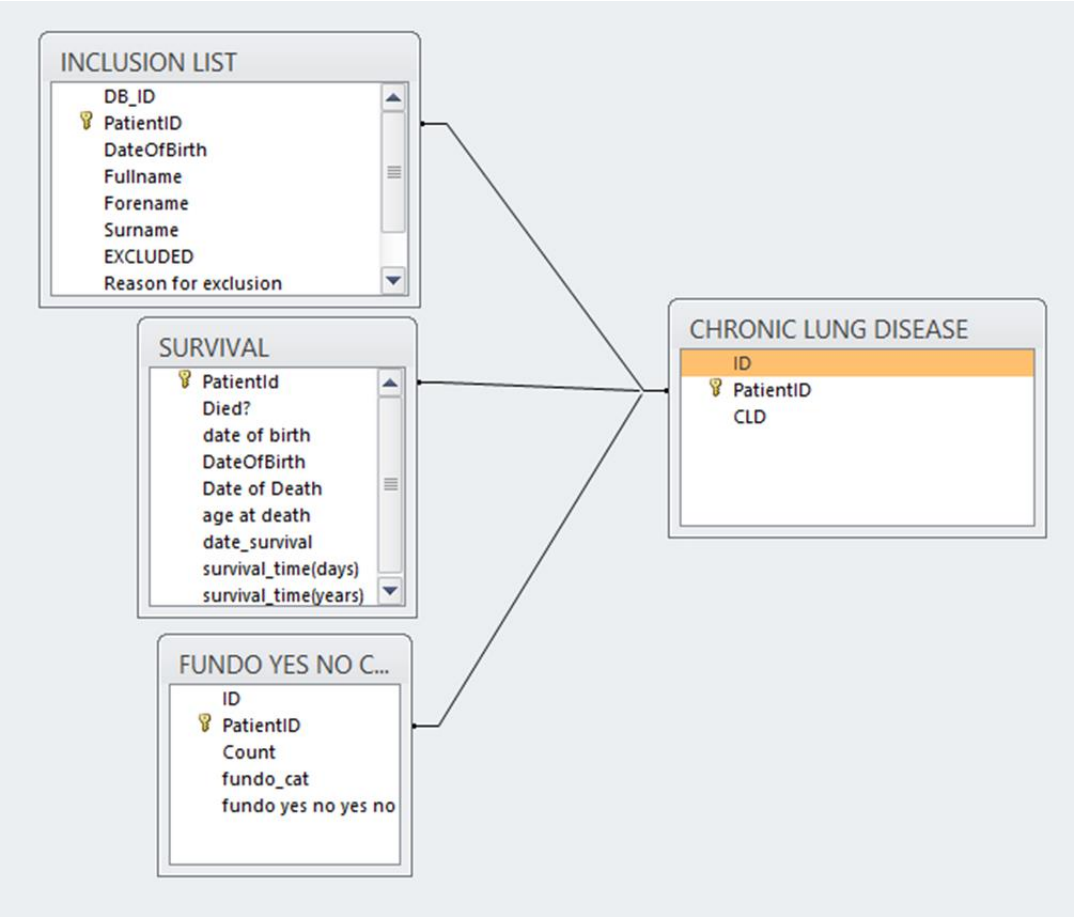
Data cleaning or pre-processing is defined as “detecting and removing errors and inconsistencies from data in order to improve the quality of the data.”(241). For all databases, errors and inconsistencies were anticipated. Data processing procedures common to all database searchers are described below.

Maintaining referential integrity

All patients at GOSH are registered with a unique PatientID. All subject records in the RetrospectiveGOR.db were indexed. A primary index table ‘INCLUSION LIST’ contained the PatientID of all patients included. In this table, PatientID was indexed with no duplicates. DOB and patient name were used for further verification. PatientID in the INCLUSION LIST was then used to link all records imported into RetrospectiveGOR.db. Therefore, PatientID was set as the primary index key and referential integrity between linked records established.

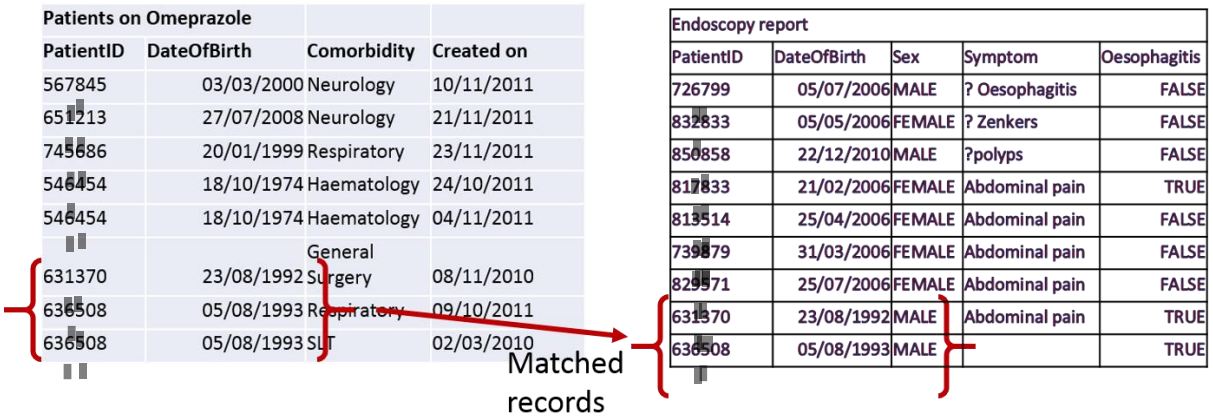
The relational database structure, linking the inclusion list, comorbidity data [CHRONIC LUNG DISEASE] and outcome data [SURVIVAL], [FUNDO] is illustrated in **Figure 52** below.

Figure 52: CHRONIC LUNG DISEASE is linked to INCLUSION LIST, SURVIVAL and FUNDO by the primary key i.e. PatientID



Data linkages were formed using SQL macros called joins. A join combines fields from two or more tables by matching values common to each. There are two main categories of joins i.e. inner and outer joins. Inner joins put together only matching data from tables. For example:

Figure 53: Two tables with matched and unmatched data before SQL merge. All PatientID and DateofBirth and PatientID couplets are fictitious for the purpose of anonymising the data.



An inner join matching is performed, matching Patient ID in the tables above. The following syntax is used.

SELECT Omeprazole.*PatientID, Omeprazole.DateofBirth, Omeprazole.Comorbidity, Endoscopy.Sex, Endoscopy.Oesophagitis*

FROM *Omeprazole*

INNER JOIN *Endoscopy* ON *Omeprazole.PatientId* = *Endoscopy.PatientID*;

The yield of this inner join is illustrated in Figure 54 below.

Figure 54: An inner join returns only matched data

Inner Join

PatientID	DateOfBirth	Comorbidity	Sex	Oesophagitis
631370	23/08/1992		MALE	TRUE
636508	05/08/1993		MALE	TRUE
631370	23/08/1992	General Surgery	MALE	TRUE
636508	05/08/1993	Respiratory	MALE	TRUE
636508	05/08/1993	SLT	MALE	TRUE

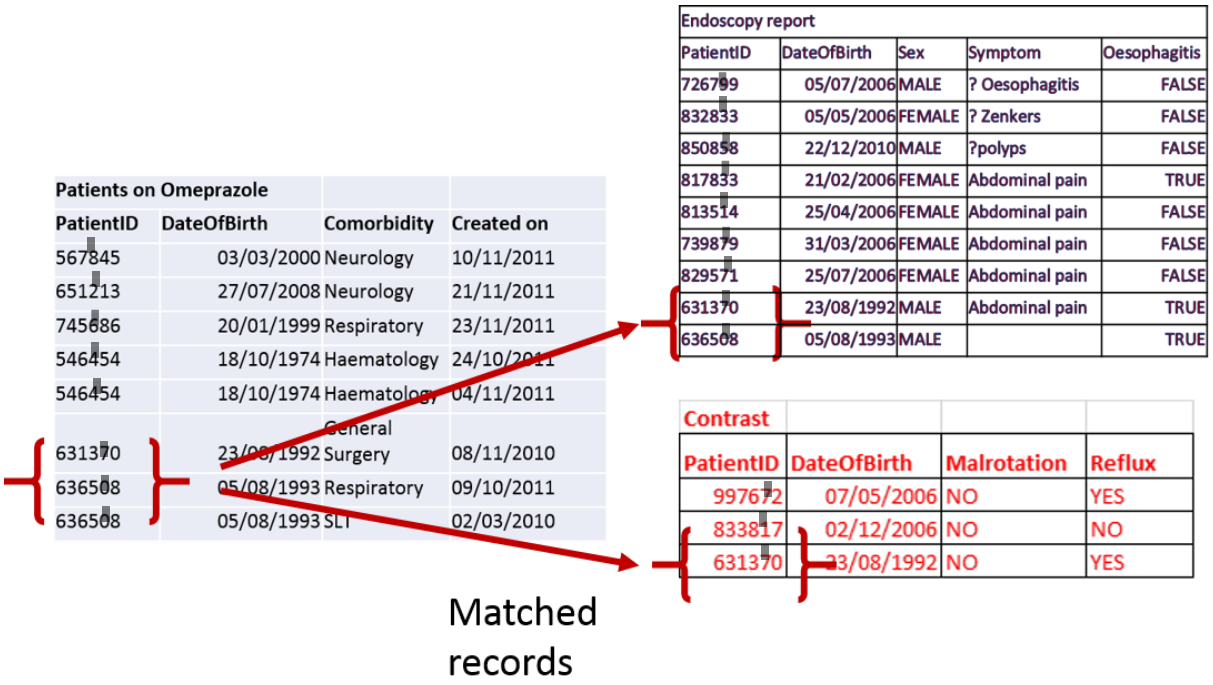
Outer joins put together matching data, but can also return unmatched data from tables.

Unmatched records are expected as individual patients have different clinical pathways. For example, in the example below (**Figure 55**), an inner join would only capture patients who had both endoscopy and contrast. An outer join captures those who have had either endoscopy or contrast or both.

Outer joins can be full, leftward or rightward. Full joins return all data, both matched and unmatched. .If a 'left join' is established, a query will always return data in the left table and matched data in the

right. The table containing primary index data are conventionally placed on the left. This convention illustrates one-to-one or one-to-many relationships between the primary index data and variables.

Figure 55: Three tables before an outer join is performed



SELECT *Omeprazole.PatientID, Omeprazole.DateOfBirth, Omeprazole.Comorbidity, Endoscopy.Oesophagitis, Contrast.Malrotation, Contrast.Reflux*

LEFT JOIN *Omeprazole_Endoscopy* ON *Omeprazole_Endoscopy.PatientID= Omeprazole.PatientID*

LEFT JOIN *Contrast* ON *Omeprazole_Endoscopy.PatientID = Contrast.PatientID*

Figure 56: Left outer join

PatientID	DateOfBirth	Comorbidity	Oesophagitis	Malrotation	Reflux
567845	03/03/2000	Neurology			
651213	27/07/2008	Neurology			
745686	20/01/1999	Respiratory			
546454	18/10/1974	Haematology			
546454	18/10/1974	Haematology			
631370	23/08/1992	General Surgery		NO	YES
636508	05/08/1993	Respiratory	TRUE		
636508	05/08/1993	SLT	TRUE		

If a 'right join' is established, a query will always return data in the right table, and matched data in the left. To create [RetrospectiveGOR.db] use full outer joins to preserved unmatched records as well as left outer joins to capture one to many relationships. This strategy led to some redundancy in merged data. This was addressed in data pre-processing and cleaning procedures (page181).

Duplicated records

A search of the CDD database was made for each duplicated PatientID. The correct DOB was identified by reviewing clinical records and finding the most consistent entry. This erroneous record was then deleted.

- Private and NHS PatientID: UK-based private patients may have both a private alias (GOSH ID prefixed by PX e.g. PX654321) and a standard GOSH ID e.g. 654123. Amalgamating these identifiers into the same column is not possible as a conflict of data type arises. The PX-ID is a string and the GOSH-ID is treated as a number in Microsoft Access. To resolve this conflict, PX-ID letter proxies were substituted for the number 9999. This converts PX-ID into a 10-digit numerical data type. PX-IDs could then be identified by selecting numbers exceeding 1000000.

The data is then sorted by DOB. Where DOB of private patients are matched to DOB of NHS patients, the records are joined. The merger is then validated by inspecting firstname and surname for

Figure 57: SQL query replacing private patient identifiers with GOSH PatientID

```
INSERT INTO [private px alias] (PatientID)

SELECT [DB REGISTRATION]. PatientID

FROM [DB REGISTRATION]

WHERE ((([DB REGISTRATION]. PatientID)>1000000));
```

correspondence. . Where a private and GOSHID correspond to the same patient, the PatientID and the private alias are linked.

Finally, we run an SQL query to replace the PX-ID with the patient's GOSH-ID. This resulted in a single

Figure 58: SQL query joining private and NHS records matched by DOB

```
SELECT [dob in private px alias]. PatientID, [dob in fundo px].Date of
Birth

FROM [dob in private px alias] LEFT JOIN [find date of birth] ON [dob in
private px alias]. [PatientID] = [find date of birth]. [ID2]

WHERE ((([find date of birth].ID2) Is Null));
```

identifier for each patient.

Errors and inaccuracies:

We expect misspelling, misreporting, duplication and contradictory values. This is especially true when data are keyed in manually. Such errors can be identified by finding outliers i.e. values that are unusual compared the rest of the population e.g. a weight of 152kg in a 2-year-old child is more likely to be 15.2 kg. However, we did identify errors which, if undetected, would impair the data verification and referential integrity.

1. PatientID errors: PatientID was established as the unique identifier for the RetrospectiveGOR.db. However, we did detect likely keystroke errors in this field.

- E.g. PatientID 624xx2, DOB 12/02/1994, FirstName, Surname
- PatientID 924xx2, DOB 12/02/1994, FirstName, Surname

In this example, DOB as well as Firstname and Surname are used for validation and the records are then merged.

2. DOB errors. We identified instances where DOB keystroke error results in a PatientID with two different corresponding DOB e.g. example, a patient with the PatientID 60xx01 may have the following two records e.g. :

Table 48: PatientID corresponding to 3 different DOB

PatientID	DOB	Full name
60xx01	11/12/1988	Same Name
60xx01	11/12/1998	Same Name
60xx01	11/12/1999	Same Name

These errors can be identified by ordering records by PatientID, then reviewing DOB in duplicated records. The SQL query below illustrates the cleaning procedure on the table [DB REGISTRATION].

Figure 59: SQL query to identify and duplications in PatientID and errors in DOB

```
SELECT [DB REGISTRATION]. [PatientID], [DB REGISTRATION]. [Date
of Birth], [DB REGISTRATION]. [Full name], [DB REGISTRATION].
[Forename], [DB REGISTRATION]. [Surname]
FROM [DB REGISTRATION]
WHERE ((([DB REGISTRATION]. [PatientID]) In (SELECT [PatientID] FROM [DB
REGISTRATION] As Tmp GROUP BY [PatientID] HAVING Count(*)>1 ))) ORDER BY [DB
REGISTRATION].[Date of Birth];
```

3. Name errors: Patients with different names were indexed against the same hospital number. Typically, this is where patients have been admitted as neonates. Neonates are often registered using their mother's surname. Indeed, they may be admitted before parents have chosen a name. Therefore, the first name on record is 'Baby'. Once the birth has been recorded with the registrar of births, hospital records are amended to record the child's legal name. Another source of name error is misspelling e.g.

- 99xx99, Shanae Drian
- 99xx99 Shane-Drane

We verify the matched records using DOB. The most frequently used version of the name is utilised, and duplicates are deleted.

DATA SOURCES

There are multiple GOSH databases holding information on patient's demographics, referral and clinic letters, investigations and interventions. We describe the databases, their structure and purpose. The databases are listed as:

1. Clinical documents database (CDD)
2. Patient information management system (PIMS)
3. Radiology database (PACS)
4. Histology databases (PATHLAB)
5. pH study and pH impedance reports (document folder on PC)
6. Theatre database (PIMS-OR)

Search strategies, data processing and cleaning procedures are described for each data source in Section III appendix p.

SUMMARY OF RECORDS

In total, 17607 records were extracted from the various database searches. We identified 13902 unique records of patients with symptoms and signs of GOR seen at GOSH between 2006 and 2010

. Most records identified came from keyword searches of CDD for acid suppression medication.

Figure 60: Summary of search yields, inclusions and exclusions resulting in the final Retrospective GOR.db cohort

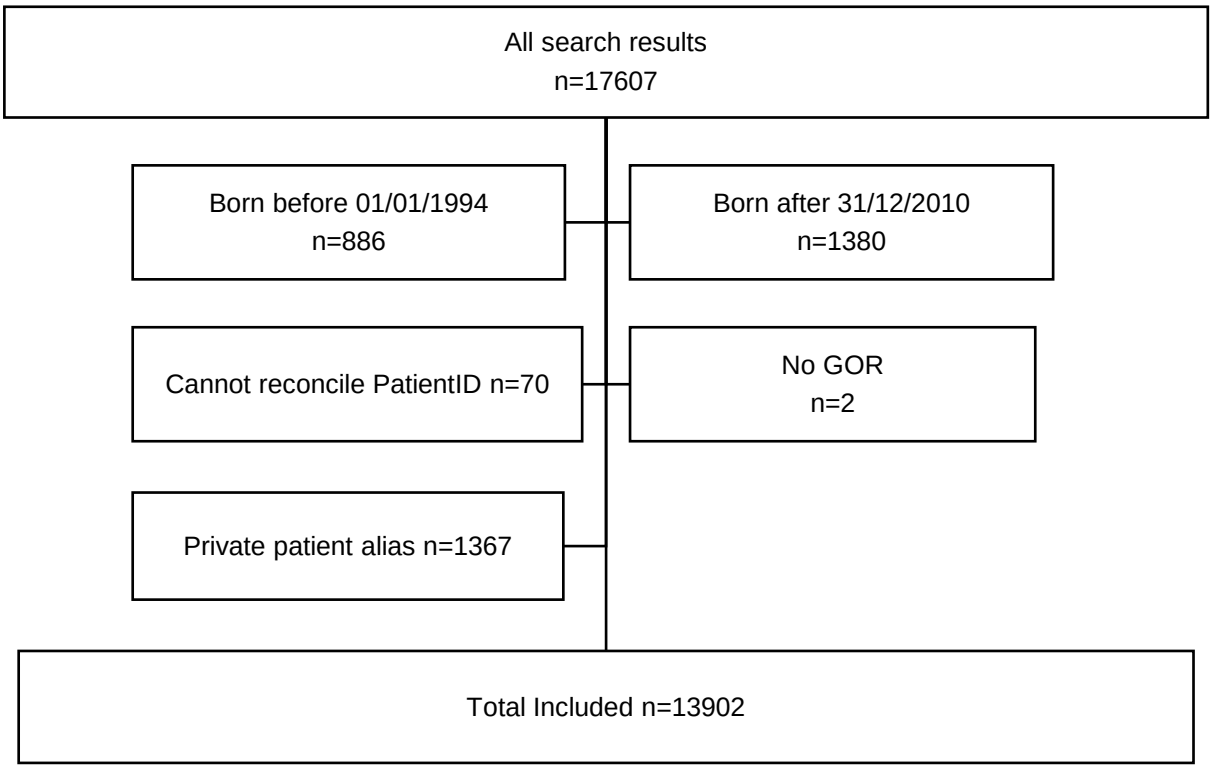


Table 49: Total numbers included from each search

Source	Total included
CDD: GORD	2900
CDD: Acid suppression medication	12117
CDD: Fundoplication	598
CDD: Oesophageal atresia	592
CDD: Achalasia	40
CDD: Prematurity	1417
Radiology database	2103
Gastroenterology investigations: pH impedance	345
Gastroenterology investigations: OGD and biopsy	1915
Theatre database search	611
PIMS mortality data	925
CDD / Palliative care mortality data	39

We were able to gather data on multiple demographic and comorbid variables. Outcome data on fundoplication and mortality was also collated. The variables are summarised in the table below:

Category	Table 50: Summary of input and outcome variables Measured variables		
Demographics (n=2)	• DOB • Sex		
Comorbidities (n=36)	• Tracheal and laryngeal anomalies • Cleft and craniofacial anomalies • Tracheostomy • Sleep apnoea • Aspiration • Chronic Lung Disease • CF • Congenital diaphragmatic hernia • Asthma • Malrotation • Achalasia • Metabolic disease • Prematurity	• Cardiac comorbidities • Cardiothoracic surgery • Neurological impairment • Genetic and chromosomal anomalies • Oesophageal atresia with or without tracheoesophageal fistula • Feeding and swallowing disorders • Constipation • Exomphalos • Chronic renal failure • Endocrine disease	• Immune disorders • Bone marrow transplant • Oncological disease • Epidermolysis Bullosa • Renal dysplasia • Skeletal anomalies • Dental disease • Consanguinity • Iron deficiency anaemia • Gastroschisis • Hirschsprung's disease • Intestinal atresia • Anorectal malformation
Investigation (n=4)	UGI– level of reflux pH impedance- reflux index pH impedance- reflux episodes/hour OGD and biopsy		
Interventions (n=3)	Fundoplication Number of fundoplications Mortality		

RELATIONAL DATABASE

When data collection was complete, flat file data were amalgamated into the Retrospective GOR.db. The data structure of this database is described in **Table 51** below. Fields in the table represented all the dimensions of data collected. Data types were enforced to ensure accuracy of data import and transformation tasks. An integer was used to represent binary variables: 1 represents Yes; 0 represents No.

Table 51: Data structure of Retrospective GOR.db

Category	Data Dimension	Data Type	SQL DATA TYPE	Description
Primary Key	PatientID	NUMBER	STRING	Fixed length n
Demographics	DOB	DATE	DATE	Format: YYYY-MM-DD
	Male/Female	TEXT	ENUMERATED(MALE/FEMALE)	Male/Female
Comorbidities	e.g. NI, CLD, OATOF	Yes/No	INTEGER	1 or 0
Investigations	Level of reflux on UGIC	NUMBER	INTEGER	As per levels of reflux described in Table 68
	pH impedance: reflux index	NUMBER	DECIMAL	Maximum number of digits is 3, 1 decimal place
	pH impedance: refluxes per hour	NUMBER	DECIMAL	Maximum number of digits is 3, 1 decimal place
	OGD and Biopsy: Yes /No	NUMBER	INTEGER	1 or 0
Interventions	Fundoplication: Yes /No	NUMBER	INTEGER	1 or 0
	Fundoplication: Count	NUMBER	INTEGER	1 or 0
Outcomes	Fundoplication: Yes /No	NUMBER	INTEGER	1 or 0
	Re-do fundoplication	Text	None, One, Greater than One	Categorical
	Mortality	NUMBER	INTEGER	1 or 0
	Survival Age	NUMBER	DECIMAL	Maximum number of digits is 3, 1 decimal place

Multiple linkages were established between tables using PatientID. The schema below demonstrates the relational structure of the database.

Figure 61: Database schema demonstrating data linkage using primary key PatientID

Summary

In this section, the search strategy, data cleaning and processing methods utilised to collate the Retrospective GOR.db have been described. In the next section, we explore and describe the cohort, focusing on demographic factors, comorbidities, investigations and outcomes. We will present a narrative of the descriptive data mining operations and their results.

In the subsequent chapter, we will analyse the strength and direction of association between variables and outcomes. We will describe the predictive data mining operations carried out and their results.

CHAPTER 3: DESCRIPTION OF THE GORD COHORT

In the previous chapters, we described the methodology used to identify a cohort of patients with GORD. We identified 13902 unique records of children with symptoms and signs of GORD seen at GOSH between 2006 and 2010.

In this chapter the cohort is further described. We investigated the incidence and prevalence of GORD. We focused on patient variables i.e. demographics, comorbidities and investigations. Outcome measures – fundoplication, RF and mortality- were described. Lastly, we investigated the strength and direction of association between variables and outcomes. This step is in preparation for the modelling exercise described in the next chapter.

The exploration of the data is a feature selection exercise (242). When constructing a model to explain a system, the data may have multiple variables or dimensions. Not all variables may be relevant or contribute to predictive performance. Equally, as is seen with genetic studies, there may be thousands of variables (e.g. gene loci) to consider. This introduces the 'curse of dimensionality'. As the number of variables increases, the dimensional 'space' they occupy increases and the data points are rendered sparse.

High-dimension data increases demand on processing time, data storage and computational memory. For these reasons, researchers may aim to filter variables and include only relevant ones in modelling tasks. Identifying which variables will prove useful to the model requires evidence-based approach, heuristics and pragmatism. Understanding which features are useful, relevant or redundant is key (242). A detailed discussion of feature selection approaches is beyond the scope of this thesis. I have, however, described the approach to feature selection taken for this data set.

STATISTICAL METHODS

For each variable, standard descriptive statistical methods were applied. The distribution of variables, central tendency and dispersion was described. Where data were tested for normality tests used e.g. Shapiro-Wilks test, a W statistic and p-value are stated. Normally distributed data were tested for independence using paired or un-paired t-tests as relevant, with mean, standard deviation, t statistic and p-value reported. For skewed continuous data, Wilcoxon rank sum test was applied, and the Z statistic and p-value reported. Where no significant difference is identified between comparison groups, the results are marked 'n.s.'. For unpaired non-parametric data, the Chi-squared and Fisher's independence tests was applied. McNemar's test was used for paired non-parametric data. The chi-squared statistic and p-value were reported.

Both linear and logistic regression were also used to assess the relationship between variables and outcomes. Odds ratios and confidence intervals is reported. Survival analysis using log-rank comparison was used to assess survival times within subgroups.

Variables were assessed individually to understand their influence on outcomes. Variables which individually influence risk of fundoplication and/ or risk of mortality were selected for inclusion in subsequent model building. Linear predictors e.g. regression coefficients, odds ratios, etc were used as the weighting factors in most cases. Variables were also assessed for interdependence. Pruning of

redundant features was necessary for some modelling procedures. Where performed, the methodology and rationale for identifying redundant features was described.

For analysis of mortality data, logistic regression was used to generate comparative coefficients for mortality risk. Survival age was also estimated. Median survival estimated is the time at which the survivorship equals 50% i.e. cohort half-life. In subsets where mortality is low or latent i.e. 50% of the subset survival longer than the observation period, median survival age is inestimable. Therefore, restricted mean was used to compare survival age.

For consistency, a strategy for number precision is defined(243) i.e. number reported in 2-3 effective digits, in keeping with European Association of Science editors' guidelines. Risk ratios were reported to allow two significant digits for standard deviation.

DEMOGRAPHICS

There were 7913 male patients (57%) and 5989 female patients (43%).

To restrict the study to paediatric patients i.e. <16 years, we restricted the range of patients' DOB i.e. patients born between 01/01/1994 to 31/12/2010. This is the capture period.

Age at presentation with GORD was investigated. As GOSH is a tertiary referral centre for GORD, patients are likely to have GORD prior to referral. Identifying age at first onset of symptoms would require access to records in primary and secondary care which was not possible. Alternatively, one could poll caregivers, which for a cohort of 13902 patients, was also not considered feasible. We estimated, instead, age at first referral to GOSH. It was assumed that, at first referral, the patient had GORD. This assumption is safe as paediatric GORD is a usually a condition that begins in infancy and persists in a subset of patients, although there may be some patients in which GORD may first occur at an older age.

A "PatientID" is assigned to a patient when they are first seen at GOSH. As PatientID are assigned serially, the year of registration can be identified based on PatientID. We obtained the first PatientID assigned in every year from 2006 to 2012 from a PIMS administrator. Data before 2006 were held on a separate and archived database and registration data could not be extracted.

Table 52: PatientID assigned at the start of every year from 2006 to 2012.

Year	Number assigned	PatientIDs assigned /year
01/01/2006 02:53:16	82xx04	19,540
01/01/2007 20:00:08	84xx44	20,853
01/01/2008 01:10:13	86xx97	23,427
01/01/2009 06:26	89xx24	22,516
01/01/2010 06:29	91xx40	23,528
01/01/2011 13:42	93xx68	26,352
01/01/2012 00:30	96xx20	

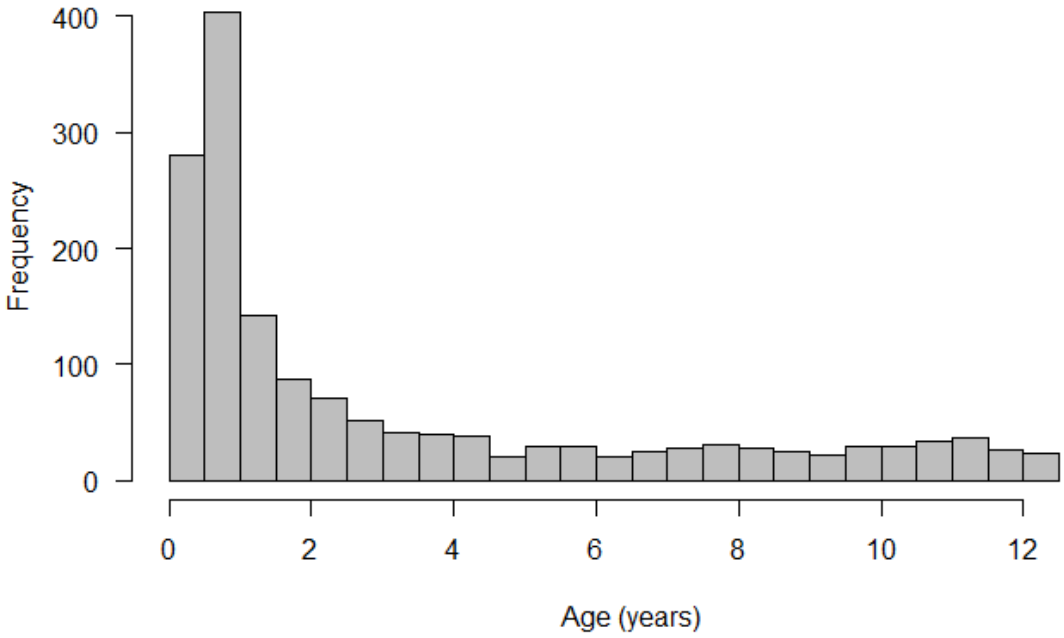
Therefore, a patient with a PatientID of 899999 would have been registered between the 01/01/2009 and the 31/12/2009. For this and other patients, the mid-year point i.e. 2/7/2009 was taken as the date of registration.

Age at referral was estimated as:

$$Age\ at\ referral\ (days) = Date\ of\ registration\ \ (Mid\ year\ point) - Date\ (Birth)$$

Age at referral could be estimated for 8897 patients (64%). The mean age of presentation for the cohort was 3.3 ± 3 years. The median age at presentation 1.4 years (1 month -12.3 years). We were unable to estimate date of registration for 36% of patients, many of whom were registered prior to 01/01/2006. Therefore, age at referral data are incomplete. The difference between mean and median demonstrates that the age data are skewed, with patients being diagnosed under the age of three (Figure 62).

Figure 62: Histogram of age at presentation



For patients who died during the capture period (n=964), we calculated the survival age as difference between DOB and date of death (DOD). Mean age of death was 5.2 ± 4.9 years. The median age at death was 3.2 (0.05 – 18) years.

We calculated survival age as difference between DOB and 31/12/2010. For patients who survived to the end of the follow-up period, mean survival age was 9.4±4.7 years. Median survival age was 8.6 (2-19).

INCIDENCE

Incidence rate describes the risk of developing a new disease or condition within a defined time frame. We identified the number of newly registered patients at GOSH from 2006 -2010. We then counted the number of patients with GORD who were newly registered. Incidence of GORD could be estimated for each year using:

$$Incidence\ GORD = \frac{new\ registrations\ with\ GORD}{new\ registrations}$$

Incidence of GORD was stable between 7-9% over 5 years. The average incidence of GORD between 2006 and 2010 was 8.1%.

Table 53: Registration of new patients at GOSH, and incidence of GORD between 2006 – 2010

Year	New patients registered	Number registered with GORD	Percentage (%)
2006	19,540	1596	8.2
2007	20,853	1617	7.8
2008	23,427	1910	8.2
2009	22,516	1984	8.8
2010	23,528	1790	7.6
Total	**109864**	**8897**	**8.1**

COMORBIDITIES

There were 6395 patients with GORD and no other identified comorbidities. There were 7507 patients with GORD and other conditions. We identified 35 premorbid and comorbid conditions were identified. The median number of comorbidities was 1 (range 1-8) .

Comorbidities were further categorised into 7 anomaly categories in keeping with established A, B, C, D approach to patient assessment:

- **A**irway anomalies
- Disorders of **B**reathing
- Disorders of **C**irculation
- Neurological **D**isability and Prematurity
- Gastro-intestinal disorders
- Miscellaneous anomalies

These categories are further described in Table 61 below.

Table 54: Comorbidities conditions observed in patients with GOR included in the cohort.

Category	Comorbidity	n	% (n=13902)
Airway	Tracheal and laryngeal anomalies	472	3.3
	Cleft and craniofacial anomalies	149	1.1
	Tracheostomy	29	0.21
	Sleep apnoea	27	0.19
	Aspiration	26	0.19
Breathing	Chronic Lung Disease	517	3.7
	Cystic Fibrosis	113	0.81
	Congenital diaphragmatic hernia	134	0.34
	Asthma	28	0.20
Circulation	Cardiac comorbidities	2593	19
	Cardiothoracic surgery	117	0.84
Disability	Prematurity	1417	10
	Neurological impairment	1940	14
Gastrointestinal	Oesophageal atresia with or without tracheoesophageal fistula	302	2.2
	Feeding and swallowing disorders	205	1.6
	Achalasia	40	0.3
Miscellaneous	Chronic renal disease	621	4.5
	Endocrine disease	289	2.1
	Metabolic disease	254	1.8
	Immune disorders	246	1.8
	Bone marrow transplant	205	1.3
	Oncological disease	190	1.4
	Epidermolysis Bullosa	98	0.7
	Renal tract structural anomalies	70	0.5
	Skeletal anomalies	60	0.4
	Dental disease	28	0.2
	Consanguinity	18	0.13
	Genetic and chromosomal anomalies	19	0.14

Airway anomalies

The comorbidities related to airway are discussed below in order of frequency of occurrence. Comparisons were made between patients with stated comorbidities and the rest of the cohort at large.

Binary logistic regression was used to estimate if each comorbidity independently predicted fundoplication or mortality risk. Odds ratios were obtained by exponentiation of the regression coefficient and the 95% confidence interval (CI). Within some subsets e.g. children with tracheostomy, all patients with a particular comorbidity had a fundoplication rate = 100%. Therefore, having a tracheostomy was a 'perfect predictor' (p.p.) of fundoplication. The odds and confidence of odds for perfect predictors cannot be calculated using simple logistic regression. Various statistical approaches to perfect predictor variables have been suggested and will be explored in detail in the next chapter on modelling GORD.

Table 55: Risk of fundoplication in patients with airway anomalies.

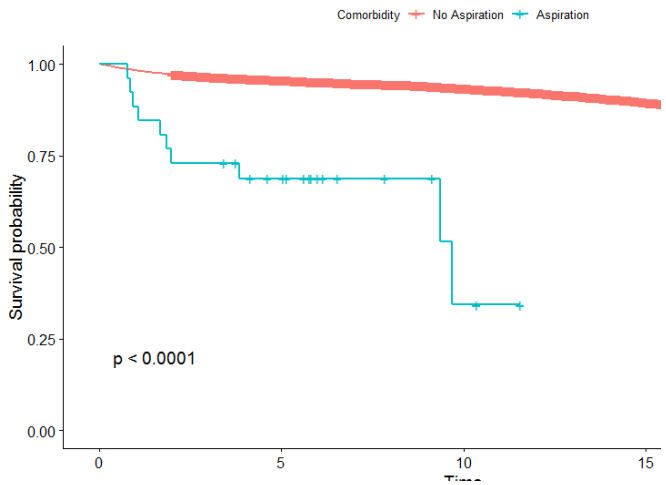
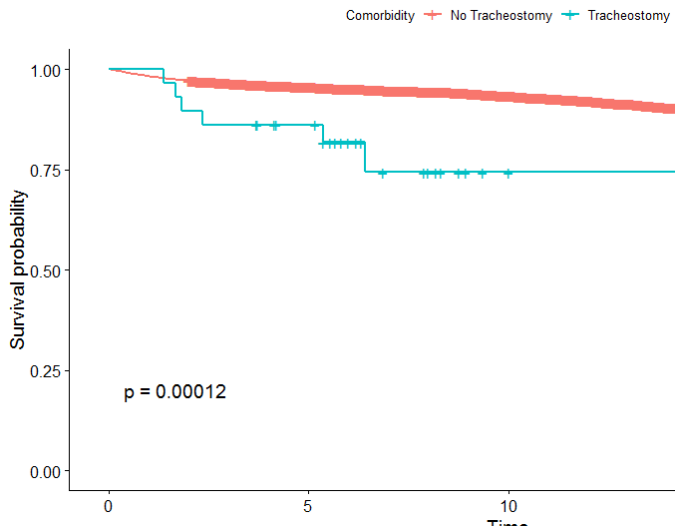
Anomaly	**Total**	**Fundoplication risk (OR)**				
		n	**%**	**OR**	**CI**	**P***
Tracheal and laryngeal	472	61	13	1.8	1.6-2.4	<0.05
Cleft and craniofacial anomalies	149	42	28	4.8	3.3-6.9	<0.05
Tracheostomy	29	29	100	Perfect predictor (p.p.)		
Sleep apnoea	27	27	100	p.p.		
Aspiration	26	26	100	p.p.		

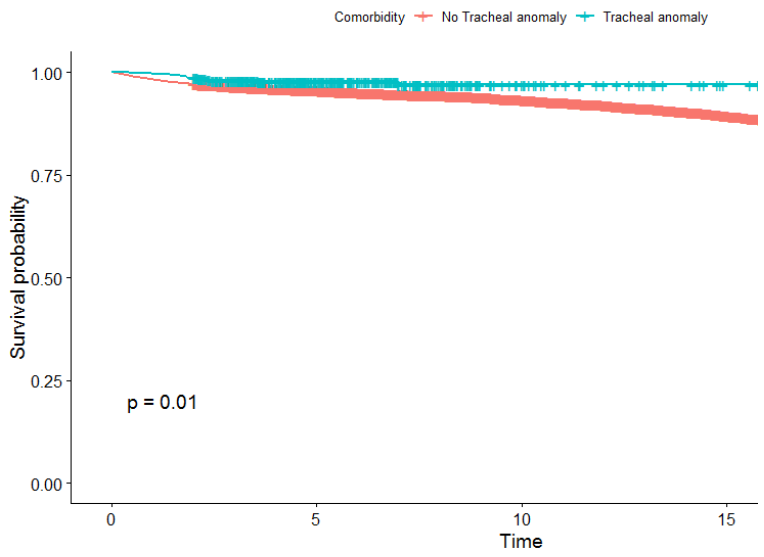
In the table below mortality associated with airway anomalies is reported. Table 56: Mortality risk and survival in patients with airway anomalies.

Anomaly	**Total**	**Mortality risk**				**Survival**	
		n	**%**	**OR**	**CI**	**Mean age (SE) yrs.**	**p***
Aspiration	26	10	39	8.4	3.7-18	9.4(1.3)	<0.0001
Tracheostomy	29	6	21	3.5	1.3-8.	13.4 (1.2)	<0.001
Sleep apnoea	27	4	15	2.3	0.7 – 6.1	16.2 (1.2)	n.s.
Cleft and craniofacial anomalies	149	7	4.7	0.7	0.3- 1.3	15.9 (0.2)	n.s.
Tracheal and laryngeal	472	11	2.3	0.31	0.16-0.54	18.4 (0.1)	<0.05
*Log rank test p-value							

Median survival estimated is the time at which the survivorship equals 50% i.e. cohort half-life. In the illustration below, patients with a history of aspiration had 50% survivorship at a median age of 9.8 (0.8-11.6) years.

Figure 63: Mean survival age is significantly lower in children with a history of aspiration, tracheostomy and tracheal anomalies.

Comorbidity	Kaplan Meier curves. Red = Comorbidity absent . Green = Comorbidity present.
History of aspiration	
History of tracheostomy	

Tracheal anomalies	

Disorders of breathing

There were five comorbidities related to the respiratory system identified as prevalent in the cohort of patients with GOR. These are:

- Chronic lung disease CLD)
- Congenital diaphragmatic hernia (CDH)
- Cystic fibrosis (CF)
- Asthma
- Acute respiratory failure (ARF)

Fundoplication rates were particularly high in patients with CDH, asthma and respiratory failure.

Table 57: Fundoplication rates and odds in patients with respiratory conditions

Anomaly	**Total**	**Fundoplication risk (OR)**			
		n	**%**	**OR**	**CI**
CLD	517	125	24	4.2	3.3-5.1
CDH	134	54	40	8.4	5.9 – 12
CF	113	8	7.1	0.9	0.4 – 1.7
Asthma	27	27	100	p.p.	
ARF	25	25	100	p.p.	

Mortality risk was markedly elevated in patient with a history of ARF.

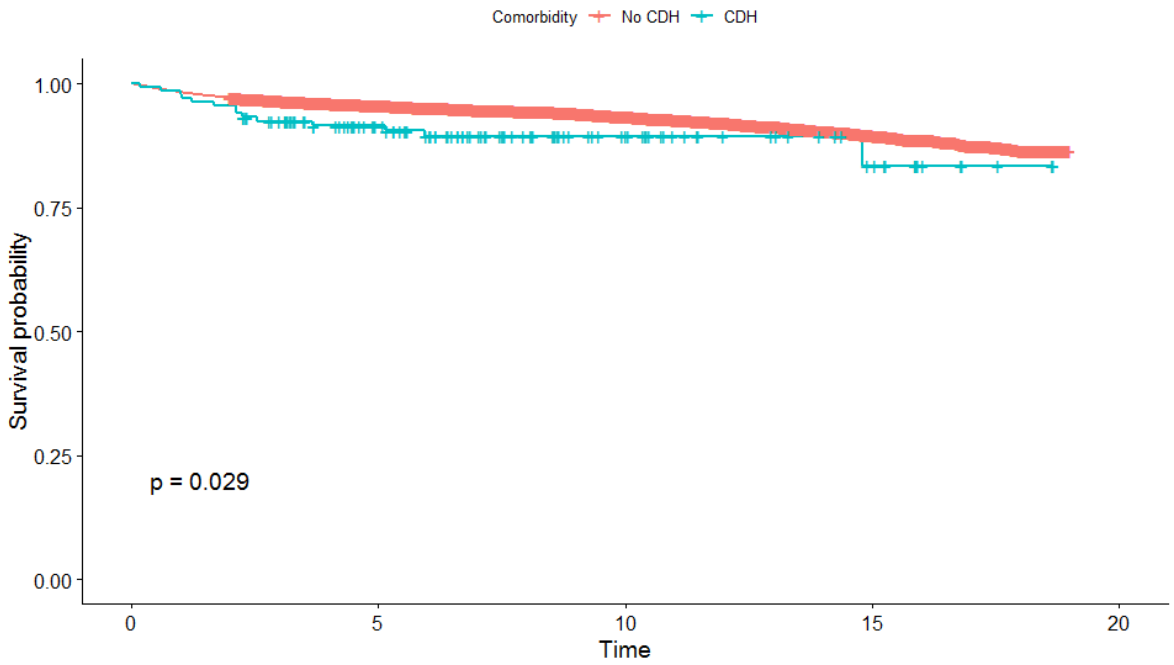
Table 58: Mortality risk and survival in patients with respiratory disorders.

Anomaly	**Total**	**Mortality risk**				**Survival**	
		n	**%**	**OR**	**CI**	**Mean age (S.E) yrs.**	**P***

CLD	517	35	6.8	0.97	0.7 -1.4	17.6 (0.2)	n.s.
CDH	134	14	10	1.6	0.9 – 2.7	16.9 (0.5)	0.03
CF	113	3	2.7	0.4	0.1 – 0.97	18.5 (0.3)	n.s.
Asthma	27	2	7.4	1.1	0.2 – 3.6	17.6 (0.6)	n.s.
ARF	25	10	40	9	3.9- 20	9.6 (2)	<0.01
*Log rank test p-value							

Survival age was significantly different in patients with ARF and CDH.

Figure 64: Mean survival was significantly lower in children with a history of diaphragmatic hernia.



Disorders of circulation

There were two categories of comorbidities related to the circulatory system that were prevalent in this cohort.

- Cardiac comorbidities: These are patients seen by the cardiologists for a spectrum of physiological cardiac disease not requiring surgery
- Cardiac surgery: These are patients seen by the cardiac surgeons who had surgery for structural cardiac anomalies.

Fundoplication risk was increased in patients with cardiac disease, but not for patients with structural anomalies.

Table 59: Risk of fundoplication in patients with physiological and structural cardiac anomalies.

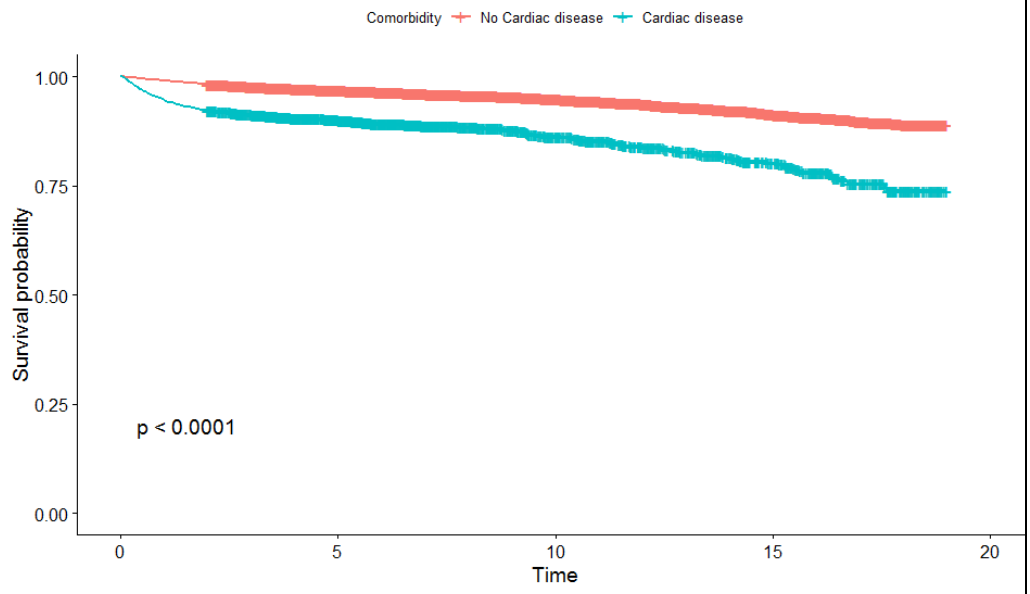
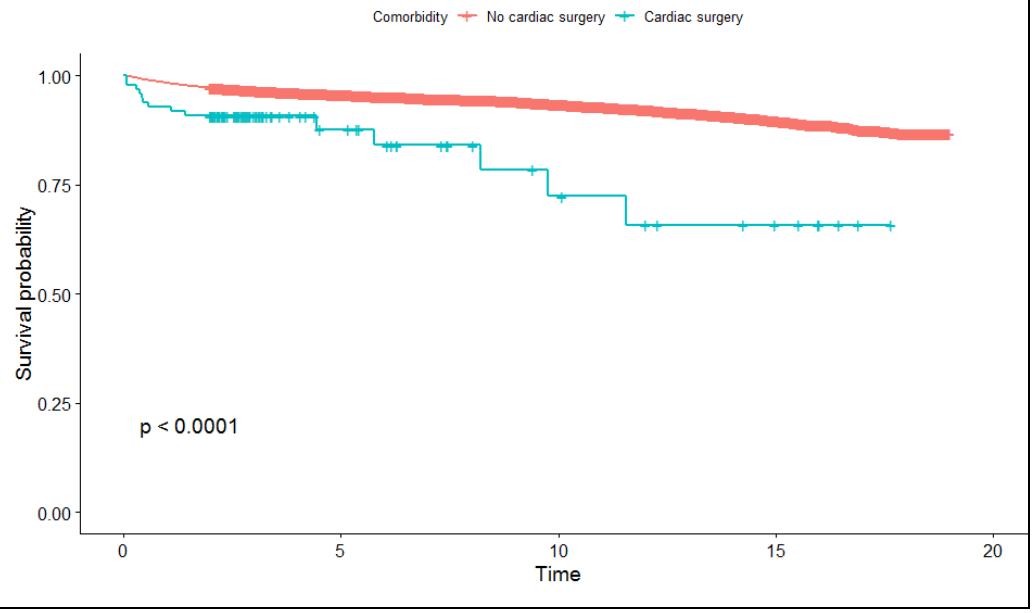
Anomaly	**Total**	**Fundoplication risk (OR)**			
		n	**%**	**OR**	**CI**
Cardiac disease	2593	291	11	1.7	1.5 – 1.9
Cardiac surgery	99	8	8	1	0.5-2

Mortality risk was higher in patients with cardiac disease and structural anomalies when compared the rest of the cohort. Mean age of survival was also significantly lower for both groups.

Table 60: Mortality risk and survival in patients with physiological and structural cardiac anomalies.

Anomaly	**Total**	**Mortality risk**				**Survival**	
		n	**%**	**OR**	**CI**	**Mean age (SE) yrs.**	**P***
Cardiac disease	2593	338	13	2.6	2.2-2.9	16.3 (0.1)	p<0.0001
Cardiac surgery	99	14	14	2.2	1.2-3.8	14.3 (1.0)	p<0.0001
*Log rank test p-value							

Figure 65: Mean survival was significantly lower in children with cardiac diseases and a history of cardiac surgery.

Comorbidity	Kaplan Meier curves. Red = Comorbidity absent . Green = Comorbidity present.
Cardiac disease	
History of cardiac surgery	

Neurological disability and prematurity

Patients with neurological impairment made up the largest comorbidity subgroup undergoing fundoplication.

Table 61: Fundoplication rates in patients with neurological disability and prematurity.

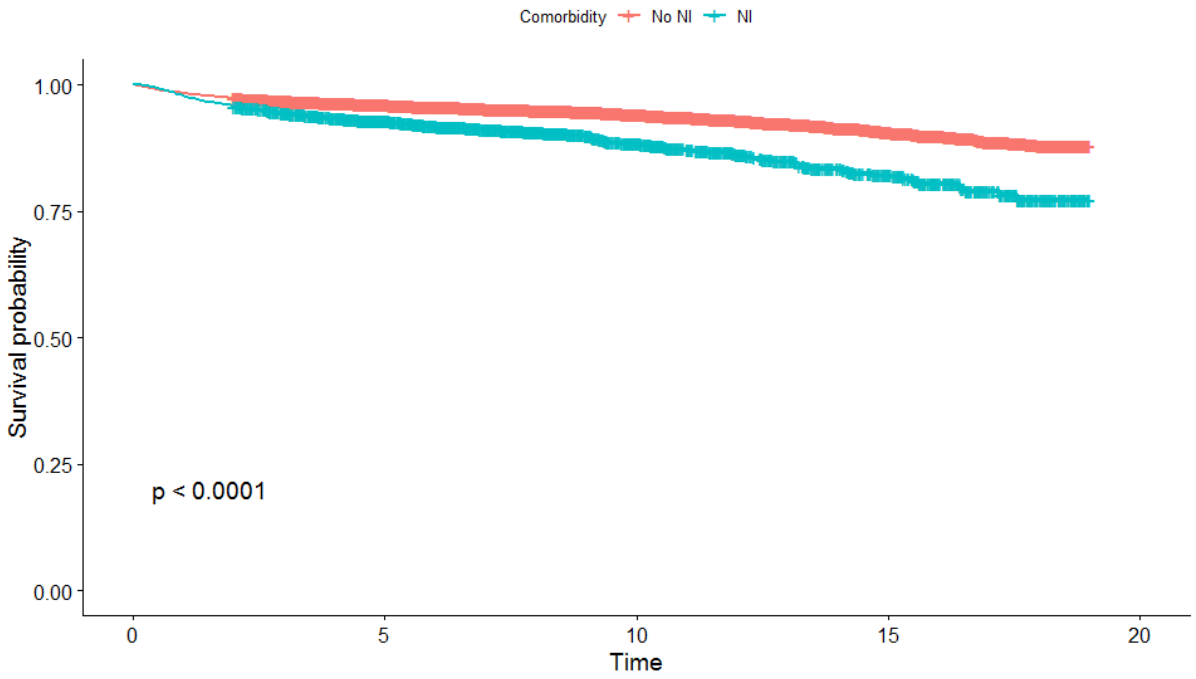
Anomaly	Total	Fundoplication risk (OR)			
		n	%	OR	CI
Neurological impairment	1940	611	57	11	9.9-13
Prematurity	1417	207	15	2.3	1.9 – 2.7

Mortality risk was significantly elevated and survival reduced by neurological impairment.

Table 62: Mortality risk and survival in patients with neurological disability and prematurity.

Anomaly	Total	Mortality risk				Survival	
		n	%	OR	CI	Mean age(SE) yrs.	P*
Neurological impairment	1940	224	23	2	1.7- 2.3	13.6 (0.1)	P<0.001
Prematurity	1417	101	7.1	1	0.8-1.3	17.3 (0.2)	n.s.
*Log rank test p-value							

Figure 66: Mean survival was significantly lower in children with NI.



Gastrointestinal anomalies

We identified gastrointestinal (GI) anomalies prevalent in patients with GORD. These are

- Oesophageal atresia with or without tracheoesophageal fistula (OATOF)
- Feeding and swallowing disorder
- Achalasia

Table 63: Risk of fundoplication in patients with gastrointestinal anomalies.

Anomaly	Total	Fundoplication risk (OR)			
		n	%	OR	CI
OATOF	302	75	25	4.1,	3.1 – 5.4
Feeding and swallowing disorders	205	94	46	11	8.2 – 14
Achalasia	40	11	28	4.5	2.2 – 8.8

Mortality risk was comparable to the GORD cohort in patients with gastrointestinal anomalies.

Table 64: Mortality risk in patients with gastrointestinal anomalies

Anomaly	Total	Mortality risk			
		n	%	OR	CI
OATOF	302	15	5	0.7	0.4 – 1.1
Feeding and swallowing disorders	205	15	7.3	1.1	0.6 – 1.8
Achalasia	40	0	0	p.p.	
*Log rank test p-value					

Other GI comorbidities

Constipation was identified in 12 patients in the cohort of patients with GOR. No patients with constipation underwent fundoplication. Mortality in this subgroup was reported in 1 of 12 patients (8.3%). Eight patients with a history of malrotation were identified. Fundoplication was performed in one of these patients.

Two patients were identified with gastroschisis; two had Hirschsprung disease. One had exomphalos and another had an anorectal malformation. One patient with intestinal atresia was also identified.

There are no known syndromic associations between these comorbidities and GORD. Furthermore, the frequency of these incidental comorbidities is low. Therefore, we did not proceed to a systematic search for other patients with both GORD and these comorbidities. These comorbidities are excluded from further consideration.

Miscellaneous anomalies

In this subgroup are comorbidities identified in the cohort that which do not fit into categories above. These are:

- Metabolic disorders
- Haematological disease
- Immune disorders
- Oncological disorders
- Chronic renal failure
- Bone marrow transplant
- Epidermolysis Bullosa
- Skeletal anomalies
- Dental disease
- Endocrine disease
- Genetic and chromosomal anomalies
- Consanguinity

Table 65: Fundoplication in patients with miscellaneous co-morbidities co-incident with GORD.

Anomaly	**Total**	**Fundoplication risk (OR)**			
		n	**%**	**OR**	**CI**
Metabolic disorders	254	45	18	2.6	1.9 – 3.6
Haematological disease	455	5	1.1	0.1	0.05 – 0.3
Immune disorders	246	22	8.9	1.2	0.7 – 1.8
Oncological disorders	190	3	1.6	0.2	0.05 -0.5
Chronic renal disease	621	45	7.2	0.9	0.7-1.2
Epidermolysis Bullosa	98	2	2	0.25	0.04 – 0.8
Skeletal anomalies	60	37	32	20	12-34
Dental disease	28	11	39	7.8	3.5 – 16
Endocrine disease	289	34	12	1.6	1.1 – 2.3
Bone marrow transplant	193	2	1	0.1	0.02-0.04
Renal tract structural anomalies	70	38	54	15	9.1 -24
Genetic and chromosomal anomalies	19	19	100	2.5×10^7	p.p.
Consanguinity	18	18	100		

Mortality risk was for each comorbid condition was compared to the rest of the unaffected cohort.

Table 66: Mortality risk associated with miscellaneous comorbidities

Anomaly	Total	Mortality risk				Survival	
		n	%	OR	95%CI	Mean age (S.E) yrs.	P*
Oncological disorders	190	51	27	5.1	3.7 – 7.1	13.9 (0.5)	p<0.0001
Consanguinity	18	4	22	3.8	1.1-11	12 (2.0)	p<0.01
Renal tract structural anomalies	70	15	21	3.7	2-6.4	16.4 (0.6)	p<0.001
Bone marrow transplant	193	37	2	3.3	2.3 – 4.7	15 (0.5)	p<0.0001
Metabolic disorders	254	41	16	2.7	1.9 – 3.7	15.6 (0.5)	p<0.0001
Skeletal anomalies	60	10	17	2.7	1.3 – 5.1	16.2 (0.8)	p<0.01
Genetic and chromosomal anomalies	19	3	16	2.5	0.6 – 7.6	13.6 (1.5)	p<0.05
Epidermolysis Bullosa	98	14	14	2.3	1.2 – 3.9	16.5 (0.6)	p<0.01
Haematological disease	455	59	13	2.1	1.5 – 2.7	16.6 (0.3)	p<0.0001
Immune disorders	246	30,	12	1.9	1.3 – 2.7	16.5 (0.4)	p<0.001
Dental disease	28	3	11	1.6	0.4 – 4.6	16.1 (0.7)	n.s.
Endocrine disease	289	23	8	1.6	0.7- 1.7	17.4 (0.3)	n.s.
Chronic renal disease	621	42	6.8	0.97	0.7 – 1.3	17.9 (0.2)	n.s.

INVESTIGATIONS

UGI Contrast

There were 2103 UGIC records included in the Retrospective GOR.db. Reports were manually reviewed to identify the level of reflux identified.

Finding of reflux	n	%
Inadequate study	35	1.7%
No reflux	1115	53.0%
Proximal oesophagus	347	16.5%
Mid Oesophagus	239	11.4%
Distal Oesophagus	346	16.5%
Oropharynx	20	1.0%
Aspiration	1	0.05%
Total	**2103**	**100.0%**

The majority of patients had an UGIC and no fundoplication (n=1668, 79%). A smaller number (n=400, 19%) had an UGIC as well as fundoplication. There were 115 patients who had an UGIC both before and after fundoplication. Thirty-five patients (1.7%) had UGIC and no fundoplication.

Table 67: Level of reflux detected on UGIC pre- and post-fundoplication

Oesophageal level or reflux	Pre -fundoplication	%	Post-fundoplication	%	Total
No reflux	783	(70)	332	(29.8)	1115
Proximal	328	(95)	19	(5.5)	347
Mid	229	(96)	10	(4.2)	239
Distal	312	(90)	34	(9.8)	346
Oropharynx	16	(80)	4	(20.0)	20
Aspiration	0	(0)	1	(100)	1
	1668		400		2068

Some patients had and UGIC performed after fundoplication. We report on reflux status on UGIC on this subset in the table below. There were also UGIC contrast studies performed after patients had fundoplication. As the pre-and post- fundoplications were not systematically requested in paired / unpaired patients, no direct comparison can be made. However, it is notable that the reflux was less commonly identified in patients post-fundoplication.

Table 68: Reflux status on UGIC in patients with or without fundoplication

Reflux	No fundoplication		Had fundoplication		Total
No reflux	783	(47%)	332	(83%)	1115
Reflux	885	(53%)	689	(17%)	953
Total	1668		400		2068

The relationship between the level of reflux and fundoplication was assessed with logistic regression analysis (Table 69) in available studies (n=1742). Reflux to the oropharynx significantly reduced the risk of fundoplication.

Table 69: Relationship between the level of reflux and the risk of fundoplication

Level of Reflux	OR	95% C.I. for EXP(B)					p
None	0.81	(	0.23	-	2.1	)	n.s.
Distal	0.65	(	0.18	-	1.8	)	n.s.
Mid-oesophagus	0.41	(	0.12	-	1.1	)	n.s.
Proximal	0.35	(	0.11	-	0.92	)	n.s.
Oropharynx	0.19	(	0.04	-	0.77	)	<0.01

Logistic regression identified oropharyngeal reflux as an independent predictor of fundoplication. Therefore, UGIC level of reflux was included as a factor in the GORD risk stratification model included in the next chapter.

We analysed the subset of patients who had UGIC and fundoplication(n=115). The UGIC in this subset could be considered paired i.e. pre- and post-fundoplication. both pre- and post-fundoplication Pre-fundoplication, 72% of patients had reflux on UGIC. Post-fundoplication, 13% of patients had reflux. The difference was found to be significant ($p<0.001$) when data were analysed using McNemar's test for independence.

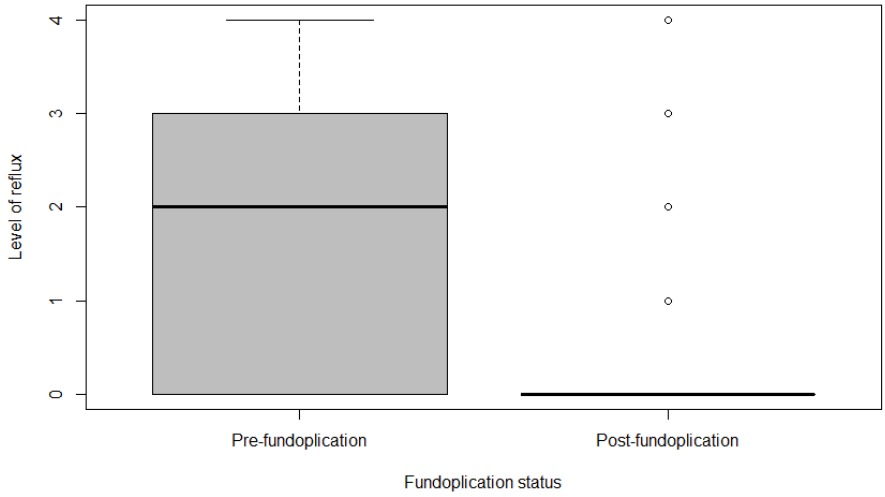
The level of reflux in this paired subset was examined. The level of reflux was re-coded as an ordinal category where no reflux is assigned a zero and aspiration is assigned a five. Inadequate studies were left blank.

Table 70: Level of reflux is re-coded as an ordinal category 0-5

Oesophageal level of reflux	Integer
None	0
Distal	1
Mid	2
Proximal	3
Oropharynx	4
Aspiration	5
Inadequate study	

Pre-fundoplication, the median level of reflux was 2 (0-4) i.e. mid-oesophageal. After fundoplication the level was 0 (0-4) i.e. no reflux. Post-fundoplication level of reflux was significantly lower (Wilcoxon rank sum, V=2990, p<0.0001).

Figure 67: Pre- and post-fundoplication median level of reflux in paired sample analysis



It would also be possible that a high level of reflux increased risk of mortality. Using logistic regression, we found no association between level of reflux on pre-fundoplication UGIC and risk of mortality.

Table 71: Mortality risk in patients with UGIC who have fundoplication

Level of reflux	**Mortality risk**								
	n	**%**	**OR**	**(95%CI)**					**p**
None	781	46	0.8	(	0.2	-	2.1	)	n.s.
Distal	314	19	0.7	(	0.2	-	1.8	)	n.s.
Mid	229	14	0.4	(	0.1	-	1.9	)	n.s.
Proximal	328	19	0.3	(	0.1	-	0.9	)	n.s.
Oropharynx	16	1	0.2	(	0.04	-	0.08	)	<0.05
	1699	100							

In summary, only a small subset undergoing fundoplication had UGIC data available. Oropharyngeal reflux was associated with reduced risk of fundoplication and reduced risk of mortality. There was no clear ordinal relationship between mortality data and level of reflux on UGIC. Although the direction of association between UGIC and outcomes is not clear, these data suggest that UGIC should be incorporated as a variable in the GORD model.

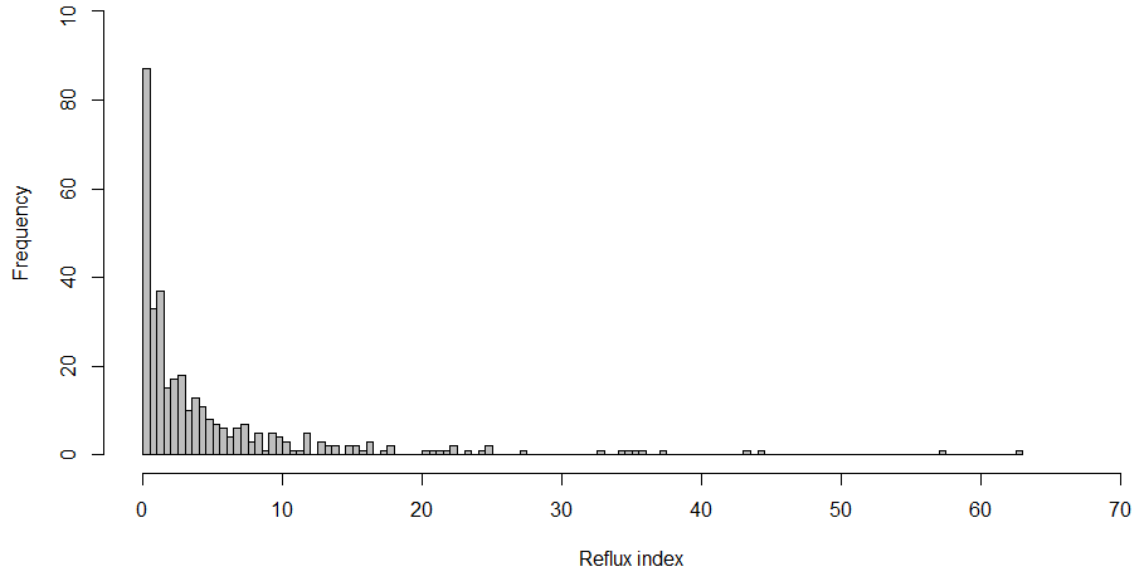
pH impedance

We included 345 pH impedance reports. The median age at pH impedance testing was 4.0 years (IQR 1.8-8.1) years. Key variables of interest were the reflux index and the reflux episodes per hour.

Reflux index (RI): This is the percentage of time that pH is less than <4. Most patients had a reflux index of <5%. RI values were not normally distributed (W=0.6, p <0.05, Shapiro-Wilks test).

RI values had a left-ward skew. Median RI was 2.1 (IQR 0.5 -5.8).

Figure 68: Histogram demonstrating the distribution of reflux index in the investigated population.

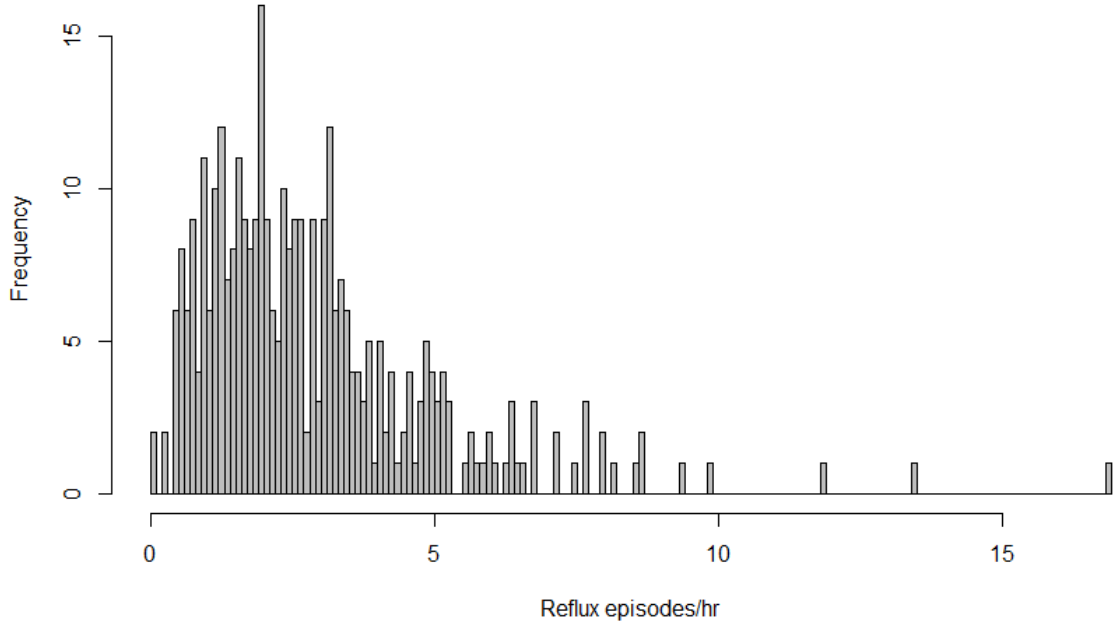


Reflux episodes: The total number of reflux episodes recorded during the pH impedance monitoring period is counted. The recommended study duration is 24 hours. However, we found variation in the study duration. The median study period was 20.6hrs (range 2.5 to 48 hours).

It was necessary therefore to standardise for variations in duration. The number of reflux episodes per hour was calculated for each patient. Most patients had 0-5 reflux episodes per hour.

Data were not normally distributed (W=0.8 , p<0.05, Shapiro-Wilks test). Data had a leftward skew. The median reflux episodes/hour was 2.4 (IQR 1.4-3.6).

Figure 69: A histogram demonstrating the distribution of reflux episodes per hour.

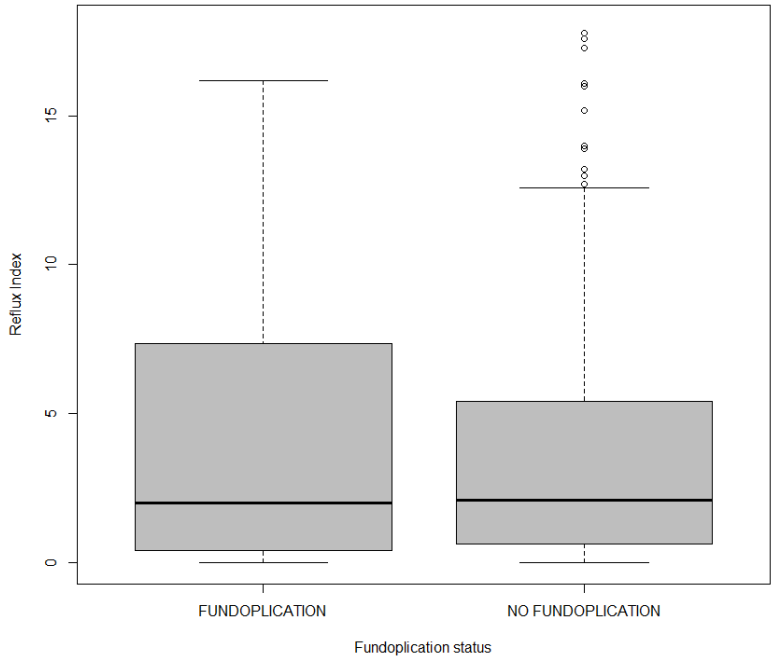


Relationship between pH Impedance findings and fundoplication

We considered whether MII findings influenced the decision to perform fundoplication. There were 76 patients (22%) who had a pH impedance study who went on to have a fundoplication. Having a pH impedance study significantly increased risk of having fundoplication (OR = 3.5, 2.7- 4.8 95%CI).

Of the specific test parameters, we focused on reflux index and reflux episodes per hour to assess whether either was predictive of outcome. Median reflux index did not vary significantly in patients who had fundoplication (2.1, IQR 0.6-5.4) compared to those who did not (2, IQR 0.4-7.4).

Figure 70: Reflux index in patients with / without fundoplication

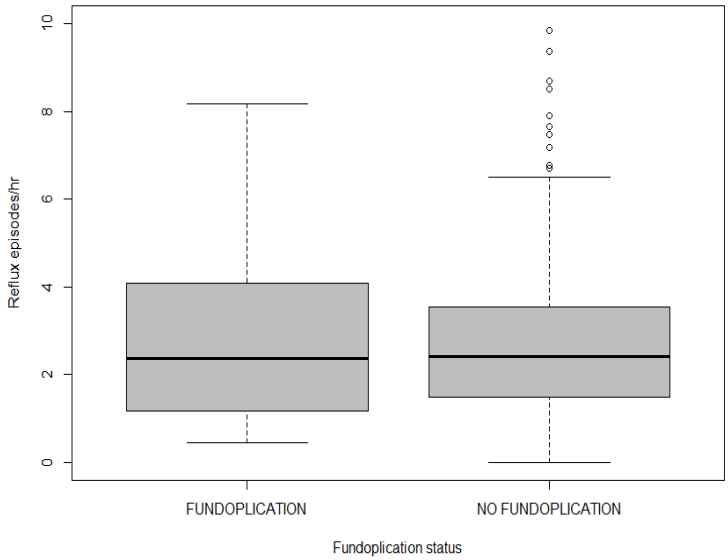


	Fundoplication	No fundoplication
Median	2.0	2.1
IQR	6.9	4.8

Mann Whitney U test , W=10220, p=0.92

There was no significant difference reflux episodes/hr in patients with (2.4, IQR 1.2-4.1) and without fundoplication (2.4, IQR 1.4-3.5).

Figure 71: Reflux episodes / hr in patients with / without fundoplication



	Fundoplication	No fundoplication
Median	2.3	2.1
IQR	2.2	1.2

Mann Whitney U test , W=10148, p=.88

Relationship between pH Impedance findings and Mortality

There were 17 patients (4.9%) who had a pH impedance study who died within the study period. Having a pH impedance study did not alter risk of mortality (OR = 0.7, 0.4- 1.1 95%CI). There was no significant relationship between reflux index and risk of mortality (OR = 0.99, 0.93- 1.06 95%CI). There was no significant relationship between reflux episodes/hour and risk of mortality (OR = 0.997, 0.79-1.26 95%CI). No association was identified between pH impedance investigative findings and fundoplication status or mortality.

In summary, pH impedance data was incomplete for majority of patients undergoing fundoplication. Within the small subset of data available, no association with fundoplication or mortality was identified. Therefore, pH impedance findings were not carried forward as a modelling variable.

OUTCOME MEASURES: FUNDOPLICATION

As described in Chapter 2, data on fundoplication were extracted from two sources i.e. PIMS-OR and CDD. PIMS-OR was launched in January 2006. Therefore, CDD records were instrumental for identifying patients who had fundoplication before 2006. THE PIMS-OR data, however, is valuable for its granularity. From these data it is possible to observe patterns of practice e.g. operative approach, trend towards laparoscopy etc. Therefore, the PIMS-OR data are described below in some detail.

Fundoplication data from PIMS-OR

Between January 2006 and December 2010, 611 anti-reflux procedures were conducted on 544 patients at GOSH hospital. The majority of patients (n= 493, 81%) received one fundoplication. A minority (19%) had one or more revision fundoplications. One patient underwent a third revision within this 5-year period.

Figure 72: Serial fundoplication carried out at GOSH between January 2006 and December 2012

Fundoplication	n	%
First	493	80.7
Second	108	17.7
Third	9	1.5
Fourth	1	0.2
Total	**611**	**100**

Primary fundoplication

In patients having first / primary fundoplication, a laparoscopic approach was most commonly taken.

Table 72: Approaches used for primary fundoplication

Primary fundoplication approach	Number	Percentage (%)
Laparoscopic approach	283	57.4
Laparoscopic converted to open	7	1.4
Open abdominal approach	203	41.2
Total	493	100.0

Fundoplication was most commonly performed with creation of gastrostomy. We identified 36 patients who had gastrostomy prior to fundoplication.

Table 73: Concomitant creation or revision of gastrostomy in patients having primary fundoplication

Primary fundoplication	n	%
Fundoplication + creation of gastrostomy	291	60
Fundoplication only	166	34
Fundoplication + revision of gastrostomy	36	7.3
Total	**493**	**100**

We identified four patients (0.8%) who had a pyloroplasty in addition to the fundoplication.

Revision of Fundoplication

We identified 108 patients who had a second fundoplication. Of these, 51 had a revision of a primary procedure that had been conducted at GOSH subsequent to 2006. The rest (n=57) had their primary procedure performed either elsewhere, or before 2006. Open surgery was frequently used for revision fundoplication.

Table 74: Approaches used for first revision of fundoplication

Approach	n	%
Open abdominal	67	62
Laparoscopic	37	34
Laparoscopic-converted-to-open	4	4
Total	**108**	**100**

Five patients (4.6%) had a pyloroplasty in addition to the second fundoplication. We identified 9 patients who underwent a third fundoplication. Of these, laparoscopic revision was possible in 4 (44%) cases. No pyloroplasties were performed. One patient had a fourth fundoplication with no pyloroplasty.

Fundoplication data from CDD records

We searched CDD to identify patients who had fundoplication. Those already identified on the PIMS-OR search were excluded. The search yielded a total of 469 records.

Cross-referencing with PIMS-OR, we found that the majority of patients had primary fundoplication at GOSH (n=360, 77%) prior to January 2006. A minority of patients received fundoplication elsewhere (n=109, 23%).

Majority of patients (85%) received one anti-reflux procedure (range 1-4).

Merged fundoplication data

We merged the records of patients known to have fundoplication from CDD (n=469) and PIMS-OR (n=611). In total, we identified 1080 patients who had a fundoplication within the capture period. The total number of fundoplications carried out was 1166. The majority of patients had one fundoplication

(n = 1005, 93%). Considering the Retrospective GOR.db cohort, the proportion of patients who had one or more fundoplication was 7.8% (Table 75).

Table 75: Fundoplication distribution in the Retrospective GOR.db cohort

Number of fundoplications	n	%
0	12822	92.2
1	1005	7.2
2	66	0.47
3	7	0.05
4	2	0.01
Total	13902	100

The relationship between fundoplication and comorbidities

Fundoplication rates varied with comorbid conditions. The highest rate of fundoplication was found in patients with cleft and craniofacial anomalies. In the table below, the comorbid conditions which, on univariate analysis, are associated with increased risk of fundoplication, are summarised (Table 76).

Several comorbidities have very large OR with some >100. We observe that these comorbidities have a fundoplication rate of 100%. This indicates that there is complete separation i.e. the variable is a 'perfect predictor'. Complete separation occurs where all observations of a variable have one value. For example, there were 26 patients with history of aspiration. All had fundoplication. This may be a reflect clinical practice. Aspiration can be life threatening. Best practice dictates that all patients with a history of GORD-related aspiration receive fundoplication.

However, complete separation in this case may also reflect the rarity of the event. This outcome is rarely observed. There may be patients with a history of aspiration who *did not* have fundoplication. However, as aspiration itself is a rare event, this combination was not observed within the sampled cohort.

When modelling data, a model cannot be fit where 'perfect predictor' variables are included in the formulation. In the next chapter, the strategy for dealing with complete separation is described. We utilise the selected features to construct, test and validate a model of GORD.

Table 76: Fundoplication rates and odds ratios for each comorbidity

		Fundoplication						
Comorbidity	**Total**	**n**	**%**	**OR**	**CI**			
Skeletal anomalies	60	37	62	20	(	12	- 34	)
Renal tract structural anomalies	70	38	54	15	(	9.1	- 24	)
Neurological impairment	1940	611	57	11	(	9.9	- 13	)
Feeding and swallowing disorders	205	94	46	11	(	8.2	- 14	)
Congenital diaphragmatic hernia	134	54	40	8.4	(	5.9	- 12	)
Dental disease	28	11	39	7.8	(	3.5	- 16	)
Cleft and craniofacial anomalies	149	42	28	4.8	(	3.3	- 6.9	)
Achalasia	40	11	28	4.5	(	2.2	- 8.8	)
Chronic Lung Disease	517	125	24	4.2	(	3.3	- 5.1	)
OATOF	302	75	25	4.1	(	3.1	- 5.4	)
Metabolic disease	254	45	18	2.6	(	1.9	- 3.6	)
Prematurity	1417	207	15	2.3	(	1.9	- 2.7	)
Tracheal and laryngeal anomalies	472	61	13	1.8	(	1.6	- 2.4	)
Cardiac comorbidities	2593	291	11	1.7	(	1.5	- 1.9	)
Endocrine disease	289	34	12	1.6	(	1.1	- 2.3	)
Immune disorders	246	22	8.9	1.2	(	0.7	- 1.8	)
Cardiothoracic surgery	117	8	8	1	(	0.5	- 2	)
Chronic renal disease	621	45	7.2	0.9	(	0.7	- 1.2	)
CF	113	8	7.1	0.9	(	0.4	- 1.7	)
Epidermolysis Bullosa	98	2	2	0.25	(	0.04	- 0.8	)
Oncological disease	190	3	1.6	0.2	(	0.05	- 0.5	)
Haematological disease	455	5	1.1	0.1	(	0.05	- 0.3	)
Bone marrow transplant	205	2	1	0.1	(	0.02	- 0.4	)
Tracheostomy	29	29	100					
Sleep apnoea	27	26	100					
Aspiration	26	26	100					
Asthma	27	27	100					
ARF	25	25	100					
Consanguinity	18	18	100					
Genetic and chromosomal anomalies	19	19	100					

OUTCOME MEASURES: MORTALITY

Associations between co-morbidities, results of investigations and interventions and mortality were investigated. Child mortality statistics are as defined by the Office of National Statistics (ONS). Mortality statistics were compared to published government statistics available (ONS 2010, ONS 2015). Mortality rates are expressed as number of deaths per 1000 per population per year.

Between 01/01/1994 and 31/12/2012, 964 patients (6.9%) in the cohort died, a mortality rate of 3.8 per 1000 per year.

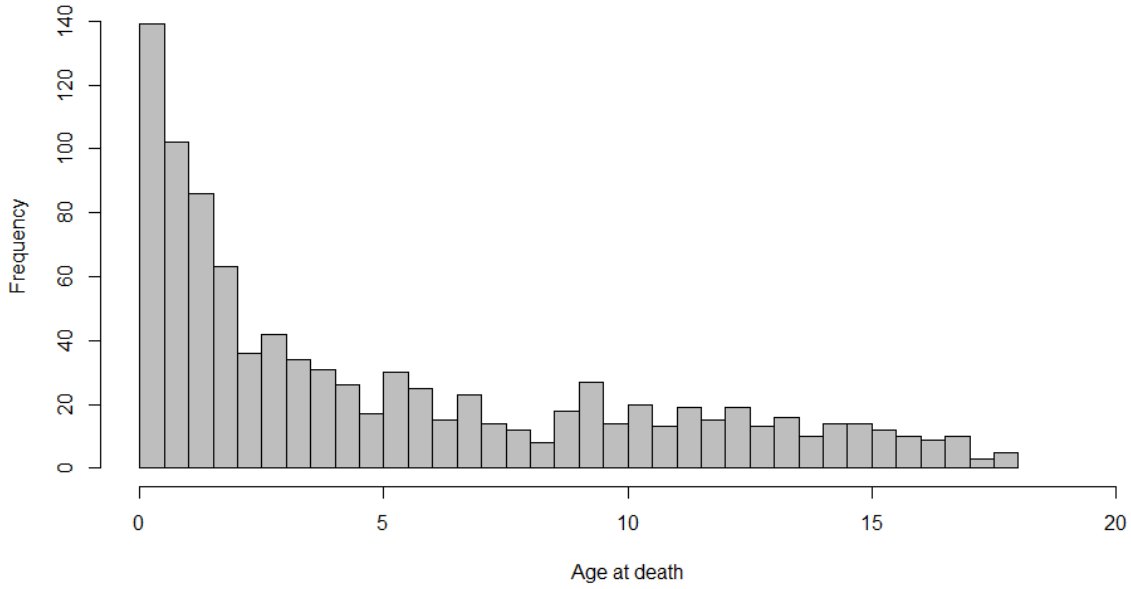
Data were stratified into neonatal, infant and childhood mortality and compared to ONS data averaged for the years 1994-2012 (Table 77).

Table 77: Comparison of mortality rates in GORD cohort compared with ONS data. Data are mean (s.d.)

Mortality rate	Description	Retrospective GOR.db	ONS
Neonatal	Death under 28 days	0.14 (0.15)	3.5 (0.4)
Infant (>28 days – 1 year)	Death under 1 year	1.6 (1.3)	5.2 (0.7)
Childhood	Death between 1 year and 18 years	5.4 (3.4)	1.4 (0.3)

As demonstrated in the histogram below, distribution of age at death was non-normal (W=0.86, p<0.001) and skewed left. The median age at death was 3.2 (IQR 8.1) years.

Figure 73: Frequency distribution - age at death.



This distribution is unsurprising and reflects the bimodal distribution of all-cause mortality in the population, with a peak in infancy and a second peak in the 8th decade of life. As the data are have an upper bound due to the inclusion of only paediatric patients into the cohort. Therefore, the second peak is not seen.

Five-year survival of patients with GORD was found to be 93%. These data are further described in the life tables below. Probability of mortality is highest in infancy. It gradually falls then ranges between 20-60 patients/ year between the ages of 4-18 years.

Table 78: Life tables describing survival in a cohort of patients with GOR.

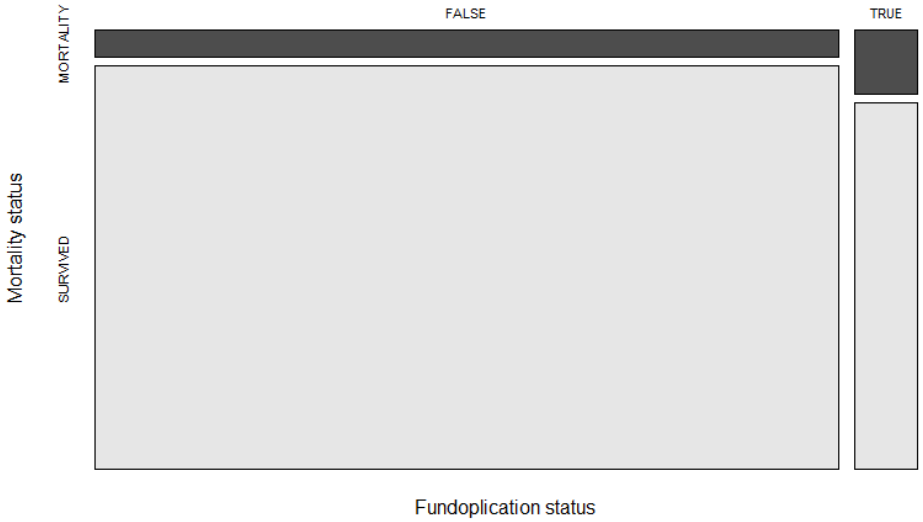
Age	Number Entering Interval	Number Withdrawing during Interval	Number Exposed to Risk	Number of Terminal Events	Proportion Terminating	Proportion Surviving	Cumulative Proportion Surviving at End of Interval	Probability Density
0	13902	0	13902	237	0.02	0.98	0.98	0.017
1	13665	0	13665	152	0.01	0.99	0.97	0.011
2	13513	792	13117	79	0.01	0.99	0.97	0.006
3	12642	947	12169	67	0.01	0.99	0.96	0.005
4	11628	1018	11119	43	0	1	0.96	0.004
5	10567	1063	10036	54	0.01	0.99	0.95	0.005
6	9450	1102	8899	38	0	1	0.95	0.004
7	8310	1032	7794	25	0	1	0.94	0.003
8	7253	883	6811.5	28	0	1	0.94	0.004
9	6342	745	5969.5	41	0.01	0.99	0.93	0.006
10	5556	686	5213	35	0.01	0.99	0.93	0.006
11	4835	630	4520	33	0.01	0.99	0.92	0.007
12	4172	596	3874	32	0.01	0.99	0.91	0.008
13	3544	594	3247	30	0.01	0.99	0.91	0.008
14	2920	592	2624	27	0.01	0.99	0.9	0.009
15	2301	606	1998	20	0.01	0.99	0.89	0.009
16	1675	577	1386.5	16	0.01	0.99	0.88	0.01
17	1082	579	792.5	7	0.01	0.99	0.87	0.008
18	496	493	249.5	0	0	1	0.87	0
19	3	3	1.5	0	0	1	0.87	0

The relationship between fundoplication and mortality.

There were 1080 patients who underwent fundoplication within the capture period. Of these, 161 died (15%). In patients who did not have fundoplication (n=12822), mortality was significantly lower at 6% (χ^2=115, p<0.001, df= 1).

Comparing mortality risk in both groups, the gross figures would seem higher in the no fundoplication subset. However, as only a minority of patients underwent fundoplication, it appears clear that the mortality burden was greater with fundoplication, as demonstrated in the mosaic chart below.

Figure 74: Mortality in patients with and without fundoplication



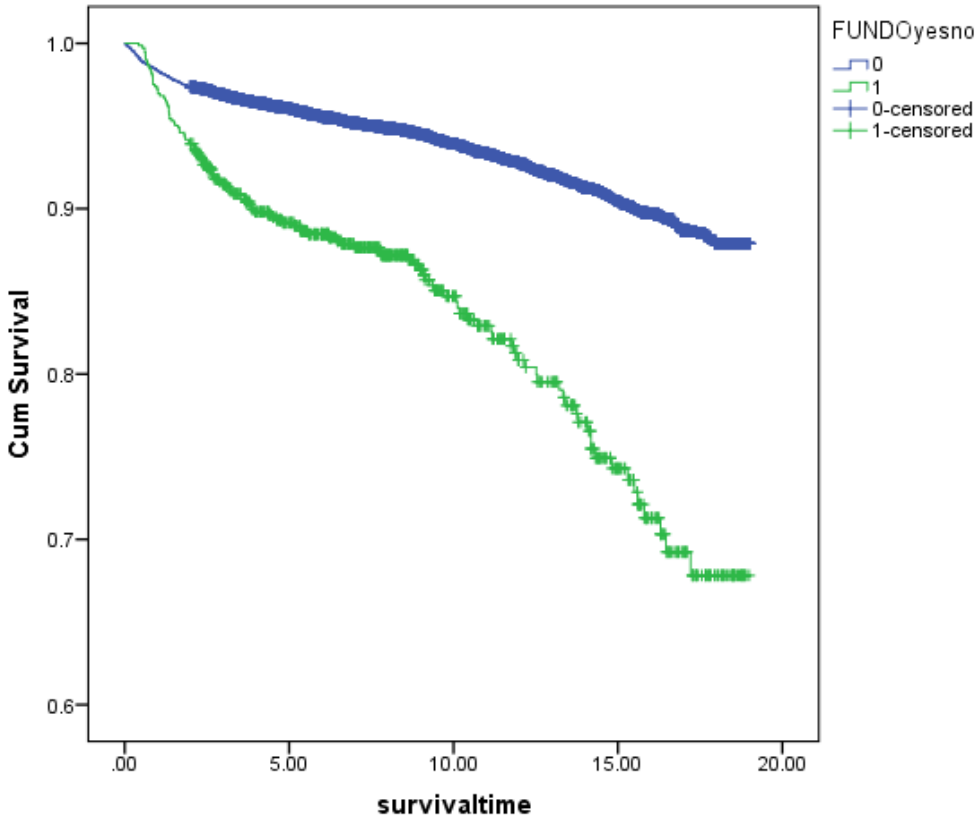
Using logistic regression, we estimated the fundoplication increased risk of mortality almost three-fold (OR 2.6 (95% CI 2.1 -3.1)). Mortality risk was significantly reduced in patients who did not have fundoplication .

Table 79: Mortality risk in children with or without fundoplication

fundoplication status	**Total**	**Mortality risk**				**Survival**	
		n	**%**	**OR**	**CI**	**Mean age (SE) yrs.**	**P***
Fundoplication	1080	161	15	2.3	(2.3 -3.1)	16.2 (0.2)	<0.001
No fundoplication	12822	803	6	0.06	(0.06-0.07)	17.8 (0.04)	<0.001
*Log rank test p-value							

Although mortality was higher in patients who had fundoplication, mean survival age was comparable (Wilcoxon rank sum test, W=442140 , p>0.05). Mean survival in patients who had fundoplication 16.2 (SE 0.2) years.

Figure 75: Comparing survival of patients with and without fundoplication



Mortality data presented here do not contain information on cause of death. Therefore, we cannot conclude from this that fundoplication causes or contributes to patient mortality, only that fundoplication is associated with mortality. Indeed, patients have other comorbidities that may also be associated with increased mortality risk. Therefore, in the modelling exercise, we will look for interactions between comorbidities and fundoplication in increasing mortality rate.

Comorbidities predicting mortality

Logistic regression was used to assess the relationship between comorbidities and mortality. The table below contains a summary of comorbidities that were found to independently increase mortality risk of mortality.

Table 80: Summary of variables independently and significantly predictive of mortality

Variable	OR	95% CI			
ARF	9.0	(	4	-	20)
Aspiration	8.4	(	3.8	-	18)
Tracheostomy	3.5	(	1.4	-	8.7)
Tracheal and laryngeal anomalies	3.2	(	1.75	-	5.8)
Bone marrow transplant	3.2	(	2.3	-	4.7)
Skeletal anomalies	2.7	(	1.4	-	5.3)
Epidermolysis Bullosa	2.3	(	1.3	-	4)
Renal tract structural anomalies	3.7	(	2.1	-	6.6)
Oncological disease	5.1	(	3.7	-	7.1)
Immune disorders	1.9	(	1.3	-	2.8)
Metabolic disorders	2.7	(	1.9	-	3.7)
Haematological disease	2.1	(	1.6	-	2.7)
Consanguinity	3.8	(	1.3	-	11.7)
Neurological impairment	2	(	1.7	-	2.3)
Cardiac surgery	2.2	(	1.3	-	3.9)
Cardiac comorbidities	2.6	(	2.2	-	2.9)
CF	0.4	(	0.1	-	0.97)

Although we can conclude that patients with these comorbidities died, we cannot conclude that patients died directly from these comorbidities. It is also possible that fundoplication together with these morbidities is more strongly associated with mortality.

In the next chapter on modelling, we examine interactions between fundoplication, comorbidities and mortality. We also estimate Cox proportional hazards of comorbidities and compare the relative risk of mortality attributable to each comorbidity.

Did fundoplication interacting with these comorbidities reduce risk of mortality? Using logistic regression, we searched for interactions between the variables above and fundoplication. We identified one interaction i.e. patients with metabolic disorder and fundoplication had an decreased risk of mortality (OR 0.18, 95%CI05-0.89).

We can summarise that fundoplication is independently associated with mortality. Contrary to expectation, survival analysis demonstrated that fundoplication does not confer survival benefit in patients with GORD. This is presumably because those factors that are associated with fundoplication are also associated with mortality. In order to test whether fundoplication leads to mortality, a randomised controlled trial would be needed, controlling for morbidities. It is doubtful whether such a trial would be practical or even ethical.

Summary

Data mining has both descriptive and predictive elements. In this chapter, we reported on the descriptive data mining tasks. We described the data set or cohort, with its general and special features. We explored the cohort and identified demographic characteristics.

The aim of descriptive data mining was feature selection i.e. useful variables for building a statistical model of GORD. We have examined and identified several comorbidities alter risk of fundoplication. Whilst UGIC findings appeared to be a factor in risk of fundoplication, pH impedance was non-contributory. This completes the descriptive element of the data mining task.

In the next chapter, we report on the predictive component of this data mining exercise. Features selected above will be contributory variables in various data models predicting fundoplication risk in children with GORD.

SECTION IV: DATA MODELS: USING COMORBIDITIES TO ESTIMATE FUNDOPLICATION RISK IN PATIENTS WITH GORD

What are the risks faced by patients with GORD?

Firstly, there is the risk of non- or mis-diagnosis. If GORD remains undiagnosed, patient symptoms may remain untreated. This may result in ongoing feeding discomfort and failure to thrive(246). On the severe end of the spectrum, there is a risk of acute life-threatening events (ALTE) related to aspiration(247).

Once diagnosed, patients are treated with acid suppression medications (ASM). The main risk here is that the medications do not work, so that feeding discomfort, failure to thrive, aspiration pneumonia and ALTEs continue. For this subgroup surgery is an option. However, surgery is a non-trivial undertaking. Fundoplication is associated with a 4-10 day stay in hospital, and the risks of wound infection, pneumonia(17) and post-operative pain(248). Specific complications of fundoplication include bleeding and bowel injury(17). There is also the inherent risk of selecting the wrong patients for fundoplication. The complications of surgery are heavily borne if the surgery was not indicated in the first place. Therefore, the second risk to consider is that of unnecessary surgery.

For patients who undergo surgery for GORD, there is again the risk of unsuccessful treatment. In published reports, it is estimated that 5%(182)-18%(159) of infants and children undergo revision of fundoplication. Revision surgery again exposes the patient to the inherent morbidity.

Finally, there is a risk that interventions do not prolong survival. Post-operative mortality data are often limited by the collection of 30-day mortality outcomes. Where longer follow-up is conducted, mortality rates are high. For example, Wockenforth et al(200) identified 20% mortality in a cohort of 255 children who underwent gastrostomy. The median follow-up time was 2·8 years (range 0·5-11·2). Although most of this mortality is presumably due to the underlying condition, can the surgeon proposing an operation be sure that surgery itself does not increase the mortality risk?

These questions come to the fore when counselling parents in clinic. Understanding and quantifying these risks is vital to facilitating the process of informed consent. In paediatric practice, informed consent can be defined as the process through which parents with capacity voluntarily give consent for a treatment after receiving full information about the risks, benefits and alternatives to the treatment. The clinic questions that parents might ask are:

1. My child has vomiting, severe scoliosis and epilepsy. Is fundoplication indicated for her?
2. Given my child's complex needs, are they likely to require fundoplication?
3. Will fundoplication work? Will it help relieve my child's symptoms?
4. My child has been really unwell following aspiration and chest infection. What is the risk of death with GORD? Will fundoplication prevent or delay this?

RESEARCH QUESTION

These specific parental concerns can be summarised into research questions:

1. What is the effect of fundoplication on symptoms?
2. What comorbid factors influence the risk of fundoplication?
3. What comorbid factors predict risk of failed fundoplication?

4. What are the factors that influence survival in patients with GORD?

The first research question cannot be addressed in this study. As discussed in Section II, symptom data were not systematically collected for patients in our cohort. This appears to be a blind spot for many workers in this field. As identified by Martin et al(71), only 36% of papers included in a systematic review of paediatric fundoplication literature described the symptoms that led to surgery. This is a major limitation of this study as well as GORD research in general. Addressing this limitation, we have devised a method to improve the systematic collection of patient symptom data (Section II). Accepting this limitation, this chapter will focus on the remaining research questions.

We have data available to answer question 4 i.e. survival in patients with fundoplication. However, achieving a focused piece of work necessitates sacrificing breath and scope in favour of depth. Some aspects of GORD mortality have been discussed in the previous chapter. No further exploration of this topic will be undertaken. The focus instead was on research questions 2 and 3 i.e. understanding the risks or primary and revision fundoplication.

Clinicians often rely on their personal experience to answer these questions. However, a simple memory exercise is prone to recall and interpretation bias. Here, a systematic review of retrospective data is presented. Any associations or correlations identified will reflect the experience of surgeons working at a single institution. Analysis of findings will therefore include a comparison of our findings to those reported in the literature.

The ultimate aim of this exercise is to extract information from our database that can be used to support clinical decision making. Returning to the clinic scenario, the generalizability and application of our findings can be assessed by asking: “Does this finding help parents and clinicians make a decision about the care of a child with GORD?”

MODELLING APPROACH

A machine learning approach has been used. In computer science, a program is said to learn when the performance of a task improves with experience. The program is exposed to a subset of the data on which to train. This training data are synonymous with the experience. Once the program has learned the rules to describe the training data, a modelling task is undertaken. For this task, the program is exposed to a fresh subset of data i.e. the test set. The program is now tested on whether it can predict outputs based on inputs using the rules learnt from the training data. The program is assessed on its predictive performance. If performance is wanting, the algorithm is modified and the program again exposed to training and test data. This recursive process continues till the program is optimised.

When the program is optimised, it is then applied to a third set data - the validation data. The performance of the program is measured against its ability to correctly predict outcomes in the validation set. Validation data are often obtained from a data set that is different to the training and test sets e.g. data from another laboratory or hospital. The process of validation is ideally carried out by experimenters different to those who trained and tuned the program.

A model is comparable to the computer program learning the rules of data. A model is a postulate. It is a proposed set of rules that describes the relationship between inputs and outputs, describing the data.

The learner can be said to be supervised or unsupervised(249). The supervised learner has a goal i.e. to use the inputs to predict specific defined outputs. The supervised learner tests her predictions against reality to understand how correct they are. If the model is found lacking, adjustments are made. The terms model and algorithm will be used synonymously to represent our hypothetical learner.

Supervised learning approaches can be further divided into broad categories determined by the type of output to be predicted(249). Regression methods apply to the prediction of quantitative outputs e.g. length of stay (days) or fundoplication. Examples of regression approaches include linear regression and regression trees. Classification is used for categorical outcomes e.g. Survived vs. Died. Examples of modelling approaches for classification tasks include decision trees and nearest neighbour classification. There are hybrid methods that can apply linear methods to classification tasks. For example, the logit function is a link function that enables application of linear methods to classification of categorical outcomes. Neural networks and support vector mechanisms are other examples of modelling approaches for both numeric prediction and classification.

Unsupervised learning approaches simply describe patterns in data. No attempt is made to use inputs to predict outcomes. These methods are useful for detecting patterns and clusters in data. Examples of algorithms include association rules and k-means clustering.

Table 81: Examples of model algorithms for supervised and unsupervised machine learning approaches

Machine learning approach	Task	Example of algorithm
Supervised	Numeric prediction	Linear regression
		Regression trees
	Classification	Decision trees
		Nearest neighbour classification
	Dual use	Neural networks
		Support Vector Mechanisms
Unsupervised	Pattern detection	Association rules
	Clustering	k-means Clusters

The data in Retrospective GOR.db are complex. While it is possible to identify which models might apply, it is not possible to specify *a priori* which model best describes the data. Therefore, for each research question we will explore several applicable modelling approaches. Both supervised and unsupervised algorithms will be utilised. Subsequently, each model explored will be evaluated and models with the greatest predictive power will be identified.

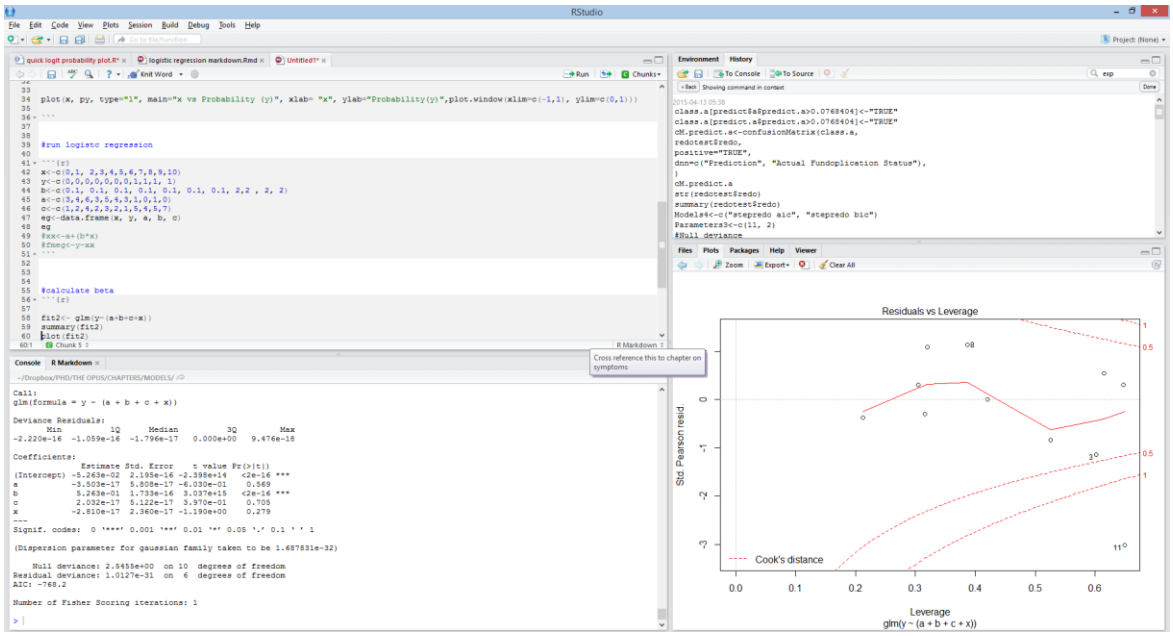
Data storage and statistical analysis

As shall be described in Section V on search strategy (p194), data were collated in a series of .csv spreadsheets or "flat files" (Microsoft Excel 2010). Flat files contained rows and columns of data. Each row represents a single record or participant. Each column represents observation of a specific parameter e.g. age. Observations were indexed by a unique patient identifier i.e. PatientID. Multiple flat files were collated into a single relational database (Microsoft Access 2010) using PatientID for data linkage. For data manipulation and statistical analysis, R was used. (231).

R is a non-proprietary software that accepts open source contributions. This means that individuals can write applications / "packages" to automate specific tasks. The open source community have contributed to hundreds of packages for automating complex and esoteric tasks(231). Formal standards for developing and submitting packages are established, thus assuring package quality. Packages are curated and archived in the Comprehensive R Archive Network (CRAN)(250).

An open source implementation of R was used i.e. RStudio (Version 0.98.953 – © 2009-2013 RStudio, Inc.). RStudio is described as "an integrated suite of software facilities for data manipulation, calculation and graphical display". It is a bundle of R components that allows ready visualisation of the console, data, history and graphical environments.

Figure 76: RStudio environment. Tiles contain (clockwise) the working directory document, the history and environment panes, a graphical visualisation window and the command console



Coding inputs and outputs

Software algorithms handle data based on predefined assumptions. They are agnostic to data input errors. An algorithm presented with a list of numbers 1 through 10 will handle the data differently if the data are coded as scalar (quantitative) or ordinal (qualitative). Therefore, each modelling task must begin with careful definition of the class of data being handled. Imported data must be inspected to ensure that input and output variables are correctly classified.

The terms inputs, independent variables, features and predictors are used synonymously. Input data can be quantitative or qualitative. Qualitative data can be categorical or ordinal. With ordinal data, there is ordering between the categories e.g. mild, moderate and severe. Nominal categories are applied where it is not possible to enumerate the difference between categories e.g. Chelsea, Arsenal, Manchester United football clubs.

The terms output, dependent variables, responses and outcomes are also used synonymously. Outputs are also broadly divided into qualitative and quantitative.

Columns containing continuous data will be described as numeric vectors. Columns containing categorical data are described as factors with each category corresponding to a 'level'. The number of levels corresponds to the number of categories in the column.

Table 82: R Coding of continuous and categorical variables

Data type	Variable type	Example	R coding
Quantitative/ Continuous	Numeric	e.g. 0.5, 0.6, 0.9 1.1, 1.7	Numeric
		e.g. 1,2,3,4	Integer
Qualitative/ Categorical	Nominal/ String	e.g. "juicy", "tangy", "bitter"	Character
		e.g. Present, Absent	Factor (with two levels)

		e.g. 1, 2A, 2B, 2C, 3, D	Factor (with four levels)
		e.g. True/False	Logical
	Ordinal	e.g. Small, Medium, Large	Ordered

For each vector or factor, we can describe the sample size, frequency of observations (e.g. proportions, percentiles) of interest, the central tendency (e.g. mean/ median) and the variation (variance). Various illustrative tools are used i.e. tables and figures.

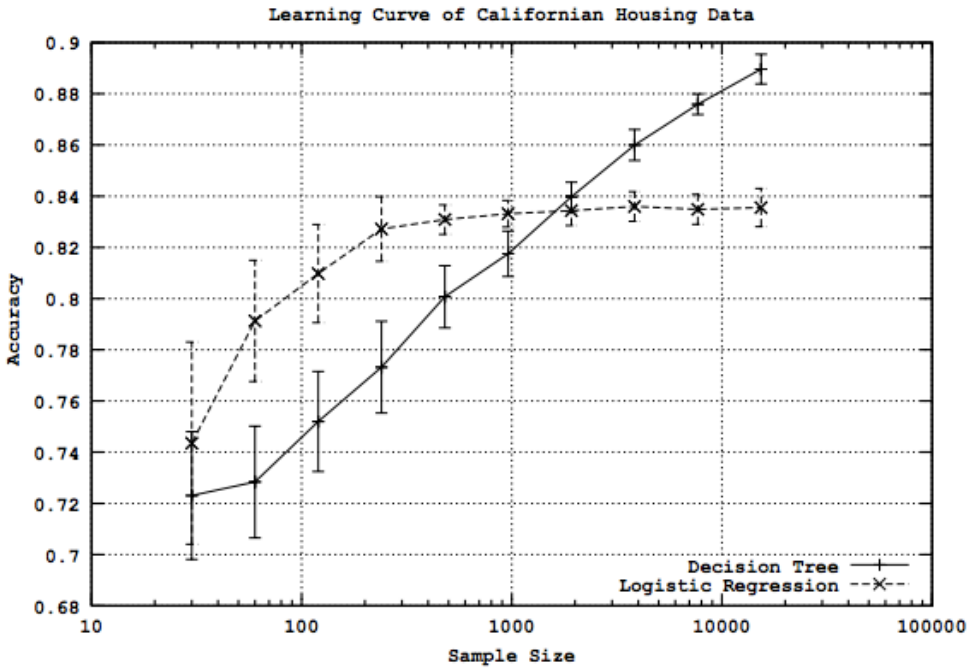
Data sampling and partitioning

As discussed above, we have chosen a machine learning approach to explore the data. This involves recursively exposing the learning algorithm to a subset of data- the training set- to identify patterns. The pattern is then applied to a different subset of data – the test set and the performance of the algorithm is measured. The model may also be tested on a third set of data for validation.

There are sampling principles to consider when partitioning the data. The ideal training set would be the whole population. However, researchers rarely have access to whole population data. Therefore, researchers try to sequester largest possible training subset, keeping in mind that a validation subset will be required. This is because, like human learners, machine learners have a learning curve. The accuracy of the learner's predictions increases with the volume of data available.

A good example of the machine learning curve is taken from Perlich et al(251). They compared decision trees and logistic regression models applied to the same data and demonstrate the relationship between accuracy and training data size. For both model types, accuracy increased as sample size increased from 10-1000.

Figure 77: Learning curves of decision tree and logistic regression models.

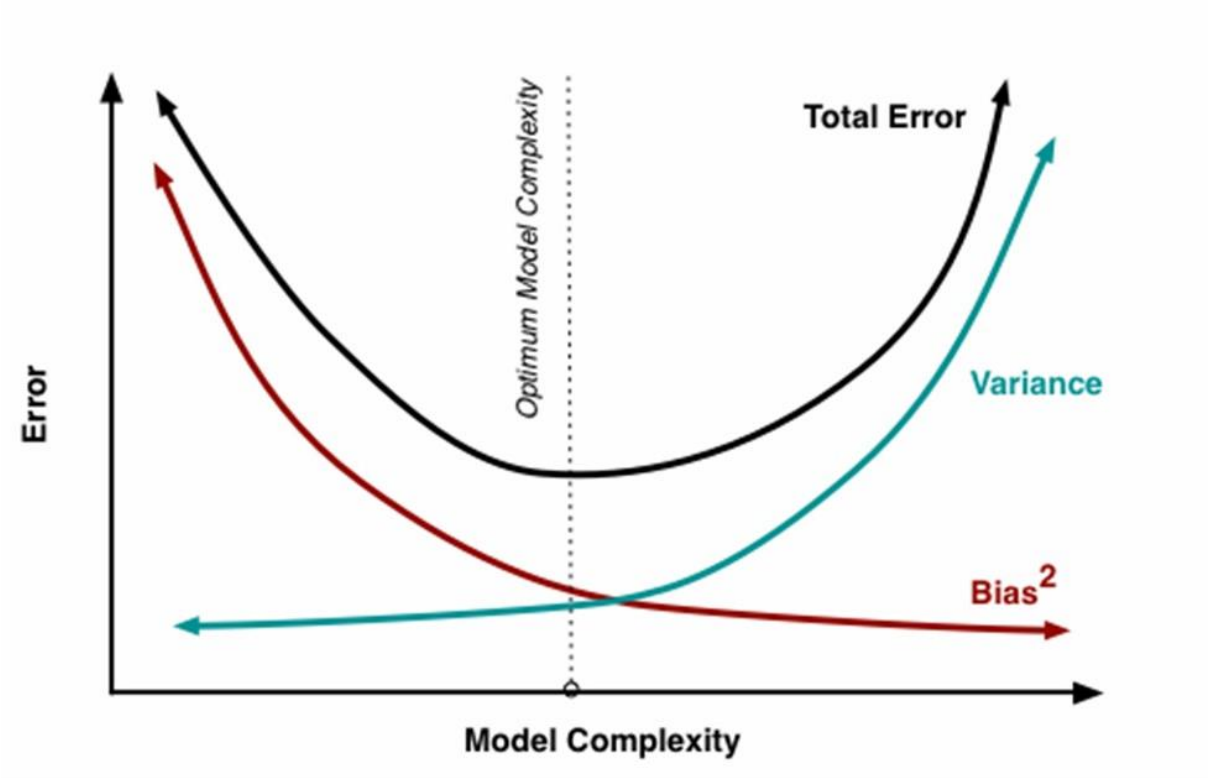


Reproduced with permission from Perlich et al. Tree induction vs. logistic regression: a learning-curve analysis. C Perlich, F Provost, JS Simonoff . The Journal of Machine Learning 2003, Volume 4, pages 211-255

No matter how large, developing the model on a single training set has limitations. Primarily, there is the risk of overfitting, where accuracy becomes sensitive to a particular training set. Overfitting results in a model that closely fits a particular data set but is poorly applicable to alternative sets. Therefore, where possible, having multiple training data sets is ideal.

Increasing the variation in data available to the model reduces the error. This problem is characterised as the balance-variance trade-off(249). As a model increases in complexity, the ability of individual variables to bias the data decreases. This reduces error i.e. the probability the model makes inaccurate predictions. However, as the model becomes more complex, it fits more closely to variability of each input. Error associated with variance is seen to increase. Therefore, a model with high bias risks under-fitting while a model with high variance runs the risk of over-fitting. The balance-variance relationship is illustrated as a trade-off because both are components of error that work in different directions as model complexity increase (Figure 78).

Figure 78: Bias variance trade-off as model complexity increases. Both bias and variance contribute to total error. Reproduced with permission from Hastie, T.; Tibshirani, R. & Friedman, J. (2001), The Elements of Statistical Learning, Springer New York Inc. , New York, NY, USA .



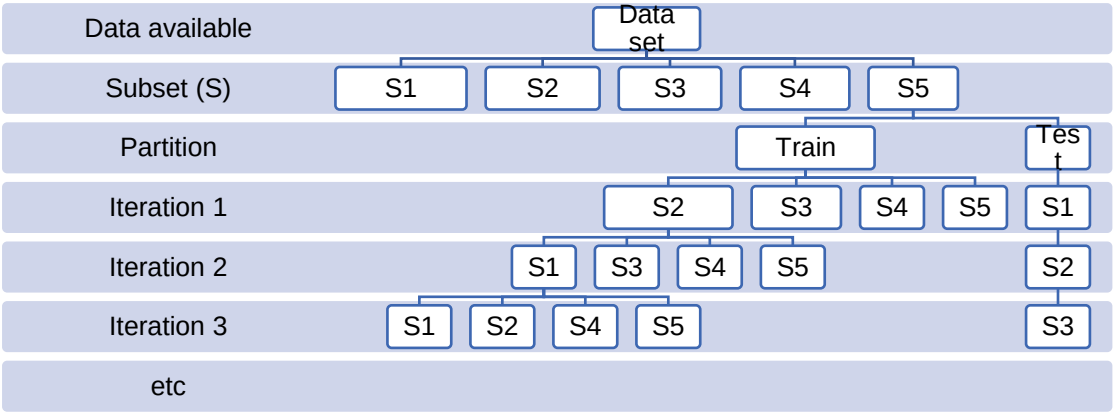
The ideal model is one where the complexity balances and minimises bias and variance. For each modelling approach, methods for minimising both bias and variance will be explored.

Data sampling approaches e.g. multiple training sets, can be used to address the balance-variance trade-off. An alternative to multiple training data sets is random re-sampling. With each re-sampling, a new dataset that is non-identical to the previous data set is generated. Bootstrapping is an example of random sampling.

With bootstrapping, a sample of size N is randomly sampled with replacement from the original data sample (of size ≥N)(252). Each sample or bootstrap is used to estimate the parameter of interest e.g. sample mean. The sampling is repeated a large number of times e.g. *m* times. This leads to *m* estimations of the mean and in turn, an estimation of the variability of the mean. Bootstrap methods are primarily used for describing the characteristics of data samples. However, statistical inference methods can be applied to the bootstrap sample in some circumstances.

Another variation on bootstrap resampling is cross-validation(252). The data set is divided into subsets (n=K, hence K-fold cross-validation). The model task is run iteratively. In each task, a subset is held back for testing. For each task, the subsets used for training and testing are different. This has the effect of running both training and testing on different data sets. Ten-fold cross-validation is commonly applied.

Figure 79: Cross-validation approach to sampling training and testing data sets



We might also consider stratified sampling. This is particularly useful for features or outcomes that are rare. The sampling procedure is specified to ensure that features in the test and validation sets are proportional. Indeed, stratification of data can be combined with cross-validation to achieve stratified cross-validation.

What is the optimal size of the test set? Like the training set, having a large test set is ideal. The larger the population, the lower the variance of the measure of performance. For example, the estimate of accuracy is more reliable with large, low variance samples. As training and test sets arise from the same

finite population, there is inherent tension between optimising the training set learning curve and accuracy of test set measures.

There are several proposed methods for calculating the best training:test splits peculiar to a data set(253). However, this discussion goes beyond the scope of this thesis. Therefore, we will adhere to convention and apply a 70/30 ratio of training to test data.

We do not have access to a validation set. However, a comparison of key findings to those in literature will serve as the both validation and generalizability.

Formulation

Mathematical formulae are useful when illustrating underlying statistical assumptions. By convention, input variables are represented by the symbol X. Where there is more than one input component in a set of input, the input is represented as a vector by adding a subscript $_j$ i.e. X_j. Where the variable is discussed, uppercase is used e.g. X_j. Where the value of the variable is observed, lowercase characters are used e.g. x_j.

The subscript of $_i$ is used to represent and individual observation e.g. the $_i$*th* observation in the set of x_j inputs is represented as x_i.

Qualitative data can also be described as a finite set of data e.g. when describing operative intervention, we can describe possible outcomes as group G where:

G={Fundoplication, No fundoplication}

Categorical data are often recoded into numerals e.g. 'Fundoplication' – the event- will be represented by the digit 1. 'No fundoplication' is recoded 0.

Outcome will be represented by Y. When we model the data, we make predictions of values of Y. This is denoted $\hat{Y}$ or "Y-hat". If the observed values i.e. Y are found in a range R or in a group of categorical outputs G, then we expect that $\hat{Y}$ should also take values in range R or set G. Hypothesis testing, discussed below, will explore the relationship between Y- outcome and $\hat{Y}$- predicted outcome.

Model selection

Having used several approaches to model the data, the researcher then needs to decide which of the models fits the data best and which model performs the best when challenged with predicting outcomes.

Model selection is a broad and contentious topic in statistical analysis. Here, we limit discussion to the methods utilised in this chapter.

Goodness of fit

There are two broad approaches to model evaluation. The first is statistical inference and the second is residual deviance analysis.

Statistical inference

When we perform a modelling task we make predictions of values of Y (denoted $\hat{Y}$ or " Y-hat"), given values of X.

- If, observed values Y are found in a range R or in a group of categorical outputs G
- Then, $\hat{Y}$ should also take values in range R or set G.

Statistical inference is the process through which we test whether the predicted values fit the expected distribution. To measure this, we will take a classical or frequentist approach.

The basis of frequentist inference is repeated sampling from a population. We ask, if there was repeated sampling of data sets similar to the one at hand, what is the likelihood of making the same prediction. Common frequentist inferential measures include p values and 95% confidence intervals.

Analysis of Deviance

The second element of model evaluation is assessing the difference between observed and expected values. A model is said to fit the data well when there is little difference between observed and fitted or predicted values. This difference is known as a residual. The deviance residual is the contribution a residual of each observation makes to the overall fit.

For example, linear regression utilises ordinary least squares methodology to enumerate the difference between observed and expected values. The best model minimises the residual sum of squares. This is known as the least squares method.

Logistic regression uses a similar parameter, residual deviance. We optimise logistic regression models

Equation 1: Deviance

$$Deviance = -2 \log Likelihood$$

by minimising residual deviance. There are various methods used to estimate residuals. For this opus, deviance is calculated:

To understand this deviance parameter, we need to first understand the likelihood term.

Maximum likehood estimation is a method used to estimate the probability of making an observation, given an understanding of the probability distribution. It is a heuristic method introduced by R. A. Fisher.

Maximum likelihood estimation can be applied to any probability distribution. For example, for a Poisson distribution, the parameter of interest θ will be solved for λ. For a binomial distribution, the parameter of interest θ is solved for the probability of the outcome p. As logistic regression will be used in this analysis, maximum likehood estimation for logit functions is utilised. The best fit for the value of $\hat{p}$ can be estimated given x.

The derivation of maximum likelihood is beyond the scope of this thesis.

Null versus residual deviance

The smaller the deviance, the better the model.

A null model has no parameters and only has an intercept i.e. $y = \beta$

A saturated model has set of parameters x. $\hat{y} = \beta x$

The fit of a saturated model can be tested against that of a null model. The residual deviance of the saturated model with parameters should be better i.e. smaller than the null deviance.

We can test the difference more formally. This test statistic is the difference between null and residual deviance. The test statistic follows a chi-squared distribution with degrees of freedom equal to the difference in degrees of freedom for the null and predicted model. The probability p of obtaining the difference in the two models can be estimated. By convention, a threshold of 0.05 is set. Therefore, when the difference in deviance p is <0.05 we conclude that the saturated model fit is poor.

Difference of deviance

The residual deviances of different saturated models can also be compared using various methods. If models are nested i.e. one is a subset of another, analysis of variance (ANOVA) using the test statistic can be performed.

Information criteria

There is a trade-off between model complexity and loss of information. The more complex the model, the closer it fits the data. However, the closer the fit, the higher the model bias. This is an illustration of the bias-variance trade-off described prior. Information criteria are a way of estimating where the balance between bias and variance lies.

We shall make frequent use of the Akaike information criterion (AIC). The AIC is calculated as follows:

Equation 2: Calculating Akaike information criterion (AIC)

$$AIC = -2\log L + 2k$$

Where,

k is the number of parameters in the model

L is the maximum likelihood estimate for the model

The deviance term (-2log likehood) is coupled with a penalty factor for parameters included (2k). Generally, the smaller the AIC, the better the model. The more the parameters used, the greater the AIC. Therefore, it is responsive to over-fitting of data.

In isolation, AIC is not useful. However, the AIC can be used to compare model performance.

Model performance

For two-class problems (i.e. there are only two possible outcomes), several tools can be used to assess the predictive power of a model.

Classification performance

Data is organised to cross-tabulate observations versus predictions. This tabulation is known as a confusion matrix. From this tabulation, performance metrics can be estimated. These are: accuracy, miscalculation rate, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), likelihood ratios.

		Observed	
		Positive	Negative
Predicted	Positive	True positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

$$Sensitivity\ (true\ positive/recall\ rate) = \frac{TP}{TP + FN}$$

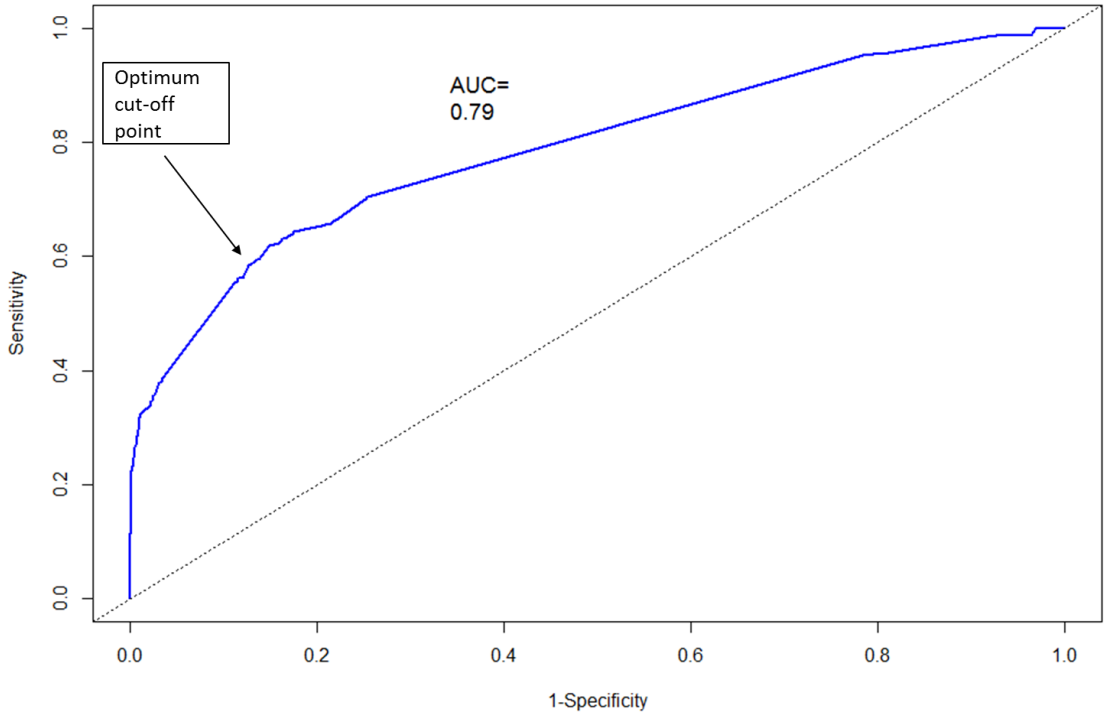
$$Specificity\left(false\frac{positive}{fall}out\ rate\right) = \frac{FP}{FP + TN}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Receiver operating characteristics

The ROC curve is used to visualise the relationship between sensitivity and specificity. A test or model that delivers the highest sensitivity and highest specificity is ideal. Sensitivity is plotted on the vertical axis. The horizontal axis has values for 1-speficicity.

Figure 80: Example of a receiver operator characteristic (ROC) profile



A perfect test would have 100% specificity and 100% sensitivity. In reality, however, there are no perfect tests and there is often a trade-off between sensitivity and specificity.

On a ROC curve, one can see the point of highest sensitivity and highest specificity. This is the optimum cut-off point. It is located by finding the highest point on the vertical axis and most leftward point on the horizontal axis. This point is therefore located in the upper left corner of the plot.

Another ROC metric is the area under the curve. A test with no discriminatory power will have a sensitivity of 50% and a specificity of 50%. Such a test would result in a straight line plot with an intercept of 0 and a gradient of 1. This plot would have an area under the curve of 0.5. A perfect test would have 100% specificity and 100% sensitivity and therefore, an AUC of 1. The academic system of grades is popularly used to interpret AUC values.

Table 83: ROC 'academic' grading system

AUC	Grade	Description
0.9-1	A	Excellent
0.8-0.9	B	Good
0.7-0.8	C	Fair
0.6-0.7	D	Poor
0.5-0.6	F	Fail

The probability at the optimal cut-off can be estimated for most modelling methods. This probability is a useful boundary diagnostic boundary. For example, in the illustration above, disease can be predicted with 60% sensitivity and 85% specificity when the predicted probability is 45%.

Clinical interpretation and generalizability.

For each model, we will assess clinical implications and applicability. Generalizability here refers to the prediction capability of a model on independent test data(249). This can also be considered to be a data validation step.

In the next we compare our findings to those available in published literature. It tests the relationship between reality suggested by the model and the facts on the ground.

Summary

In summary, we have defined the research questions. These are:

1. What is the effect of fundoplication on symptoms?
2. What comorbid factors influence the risk of fundoplication?
3. What comorbid factors predict risk of failed fundoplication?
4. What are the factors that influence survival in patients with GORD?

We focus on answering questions 2 and 3 using the retrospective data we have gathered. Data storage, manipulation and statistical analysis has been described. We have described the modelling approach i.e. machine learning using both supervised and unsupervised algorithms. Lastly, we have defined our strategy for model evaluation, selection and interpretation.

In the next section, we report on the progress and findings of the data modelling procedures.

CHAPTER 2: RISK OF FUNDOPLICATION

In this chapter, the risk of de novo fundoplication is modelled. We utilised various methods. The models presented here are those that demonstrated both useful fit and clinically relevant inference. Modelling methods applied are logistic regression, decision trees and association rules.

LOGISTIC REGRESSION

We define Y as a finite set of outcomes, where:

$$Y = \{0, 1\}$$

0= No fundoplication

1= Fundoplication

When y is a continuous variable, we can describe the relationship between X and Y using the formula

$$y = \alpha + \beta x$$

Where

α is the intercept

β is the slope

However, when Y is categorical, the formula no longer describes the relationship. Identifying values of y given values of x can no longer be computed using ordinary least squares regression.

When y is categorical {0,1} the probability of y given a subset of comorbidities of x is a continuous value that is limited to the range 0 and 1.

As p ranges between 0 and 1, the relationship between p(y) and α+βx will be non-linear. We therefore require a link function that links p(y) and α+βx .

We can introduce a non-linear function F, where

$$F(-\infty) = 0$$

$$F(\infty) = 1$$

The most commonly applied link function is the logit function (Δ) and is the basis of logistic regression. Other link functions for binomial regression are probit and complementary log functions. Here, we will focus on the logit function.

The logit is the log of the odds.

$$\Delta = logit(p) = log\frac{p}{1-p}$$

We can re-state the relationship between the outcomes and the inputs α+βx:

$$F(-\infty) = \Delta(\alpha + \beta x) = 0$$

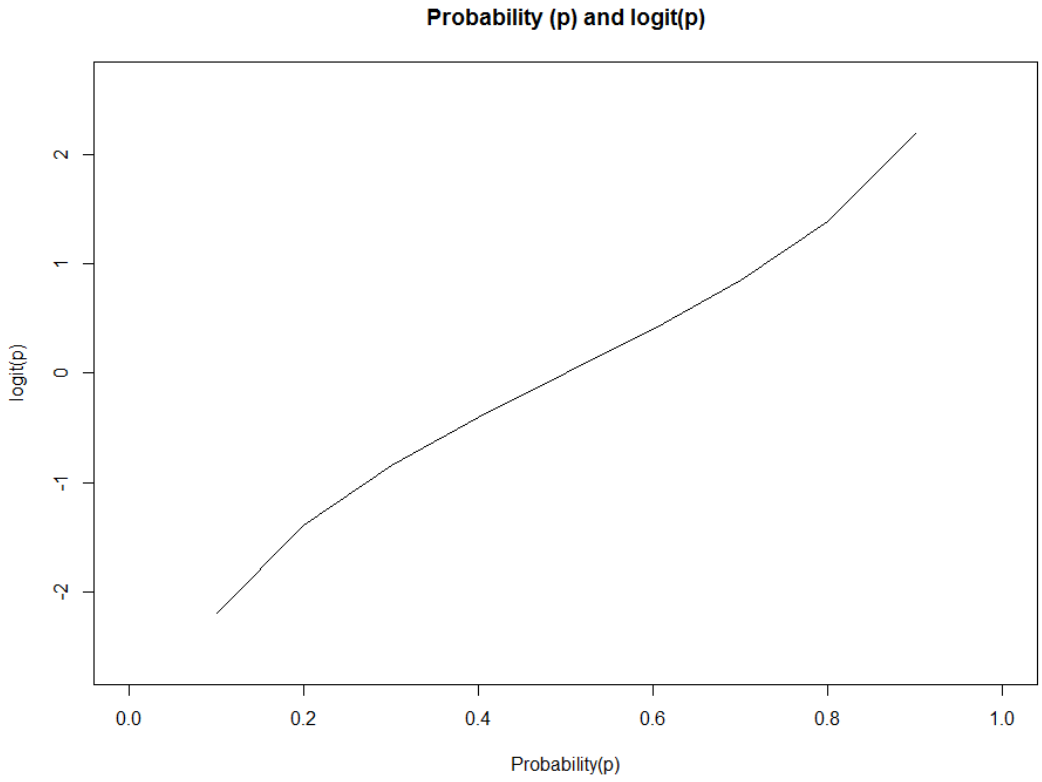
$$F(\infty) = \Delta(\alpha + \beta X) = 1$$

We have used the logit transformation to link categorical y into a continuous variable logit(p).

When $p(y = 0|x), logit(p) = -\infty$

When $p(y = 1|x),\ logit(p) = \infty$

As probability ranges from 0 to 1, logit(p) ranges from –infinity to +infinity.



Given known values of x (comorbidities), we would like estimate to the risk of y i.e. fundoplication for each individual patient.

Where

$$y = \alpha + \beta x$$

The relationship is stated using the logit link function:

$$log\frac{p}{1-p} = \alpha + \beta x$$

We can obtain the probability of y by exponentiation:

$$P(y) = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$$

where

e is the base for natural logarithm and is approximately equal to 2.72.

When x and y are known, we can use logistic regression to estimate α and β. α is also known as β0. It is the intercept i.e. the value of logit (p) when the p is 0. It is therefore the log odds for the reference group i.e. when outcome is 0.

Log odds are difficult to interpret. However, the exponential of the log odds is the odds ratio. Therefore, by exponentiation of α we obtain the odds ratio for the reference group i.e. when the outcome is 0.

Equally, exponentiation of β coefficients gives us the odds ratio for y, given x. The odds increase multiplicatively. The exponent of β gives the change in the odds of Y given one unit change in x.

We can also construct confidence intervals (CI) around our β values. We can also exponentiate these confidence intervals for the beta coefficient, yielding the 95%CI for the odds ratio.

Goodness of fit

The beta coefficients in logistic regression observe a t-distribution. Statistical inference is achieved using a frequentist approach. We expect:

- If, observed values Y observe a t distribution
- Then, $\hat{Y}$ should also observe a t distribution

We can estimate the p value i.e. the chance that, given sample size s, difference in coefficients β would be observed by random chance. This is the Student's t-test. We use this fact estimate standard errors for the sampling distribution and hence infer statistical significance.

$$t\ score = \frac{\beta_1 - \beta_0}{s_{b1}}$$

Where

t is a t distribution

β is the beta coefficient

s is the standard error

The t-distribution is well described. Therefore, T score values corresponding to the difference in coefficients can be read from a Student's T-test table (See Section IV appendix items, p517).

A number of R packages will be used to automate calculations and visualise data. These are described in Section IV appendix items, p514).

Pre-processing and data sampling

The database has 13902 observations of patients. Each column or vector contains information describing demographic, comorbid or interventional characteristics of each patient. The outcome variable of interest is fundoplication (column= hadfundo) This is coded as a logical variable i.e. 1/0. Columns 18 to 46 containing the input vectors describing comorbidities. They are coded as factors each with 2 levels i.e. PRESENT/ABSENT.

For this analysis, we partition the complete dataset into training and test samples, observing a 70 to 30 ratio. Data are randomly selected without replacement. A seed (observation at which sampling begins) is used to make sampling reproducible.

We now have a new dataset which will be used to train the model i.e. "train". The training set therefore contains 9734 observations with 31 features. Of these features, 29 are comorbidities, PatientID is a character vector containing the unique identifier for each record. 'hadfundo' is a logical vector containing the fundoplication status i.e. 1/0. There are no missing data. In the training data set, we observe that 92% of patients did not have fundoplication and 8% had fundoplication. The test data were a remnant subset of the 'base' data set after the training data was selected without replacement. 'Test' contained

4168 observations with 31 features including 29 comorbidities and no missing data. The fundoplication rate was 8%.

Variable selection

There are several variables with a fundoplication probability estimate of either 0 or 1. These variables are ‘perfect predictors’ of outcome as discussed elsewhere. To proceed, we have to address the problem of complete separation, as these variables will cause the log-likelihood function not to converge or reach a maximum

Table 84: Variables where only one outcome is identified

	Fundoplication	
Comorbidity PRESENT	FALSE	TRUE
Acute respiratory failure	0	18
Sleep apnoea	0	17
Tracheostomy	0	20
Asthma	0	22
Aspiration	0	18
Chromosomal anomalies	0	13
Consanguinity	0	11
Total	0	119

There are several ways of dealing with the linear separation problem. Firstly, we could leave these variables in the formulation, accepting that the standard errors and Wald estimates may be inaccurate.

Secondly, we could fit the model with and without the 'problem variables' and observe the differences. The patients with these comorbidities comprise a small but significant percentage of the total of patients who had fundoplication (119 of 757=15%).

Lastly, we could delete these 'problem variables' However, these are influential variables. Although removing the variables of concern will remove the linear separation problem, we will also lose information about the relationship between these comorbidities and the dependent variable. Furthermore, we may achieve biased estimates for the remaining variables.

We chose to take the second approach i.e. fitting the model with and without the problem variables. Below, we review results of the saturated model first, looking at inference, goodness of fit and predictive performance.

Modelling

Step 1: Running the model

The saturated model formula (logit1) is stated as:
Equation 3: Formula for logit1 model

```
fm<- hadfundo~tracheal+cleft+sleep+trachy+aspiration+cld+cdh+asthma+cardiac+ni+oatof+
achalasia+swallow+chrom+consang+prem+haem+endocrine+renal+skeletal+bone+dental+
oncology+metabolic+aresp+immune+cardsurg+eb+simplegord
```

Step 2: Interpreting the output

Reviewing the output, we find that intercept is first reported. Next, we are presented with information about the variables. In the second column we find the estimates i.e. the beta coefficients for each variable of x. This is the value by which the log odds of the outcome would change if the particular variable was increased by 1 unit. Several coefficients e.g. chronic lung disease (CLD), congenital diaphragmatic hernia (CDH) are predictive of outcome at the <0.001 significance level.

In the third column we find the standard error associated with beta coefficient. The standard error can be used to control for variance of the estimates. Each beta coefficient is divided by its standard error. The resulting quotient is the z value. Given a large enough sample, the z value is assumed to be normally distributed. Probability for z values can be obtained with reference to a standard normal table (Section IV appendix p.515). This p value for this normality test is found in the Pr(>|z|) column. The last column is a visual aid to identify the p values that are significant. The key for the significance coding is found at the end of the table of coefficients.

Table 85: Logit1 model coefficients

Coefficients:					
	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**	
(Intercept)	-4.0745	0.1391	-29.29	< 2e-16	***
achalasiaPRESENT	1.494	0.55	2.72	0.0066	**
arespPRESENT	15.7492	766.6422	0.02	0.98361	
aspirationPRESENT	17.3467	755.1046	0.02	0.98167	
asthmaPRESENT	17.3959	693.1176	0.03	0.97998	
bonePRESENT	-0.8497	0.7454	-1.14	0.2543	
cardiacPRESENT	0.4111	0.1341	3.07	0.00217	**
cardsurgPRESENT	0.5914	0.498	1.19	0.23502	
cdhPRESENT	1.956	0.2917	6.71	2.00E-11	***
chromPRESENT	17.2512	811.9148	0.02	0.98305	
cldPRESENT	1.1141	0.183	6.09	1.10E-09	***
cleftPRESENT	0.509	0.3841	1.33	0.18512	
consangPRESENT	16.7671	888.6696	0.02	0.98495	
dentalPRESENT	1.6118	0.6397	2.52	0.01174	*
ebPRESENT	-0.2213	1.0164	-0.22	0.8276	
endocrinePRESENT	0.6197	0.2707	2.29	0.02209	*
haemPRESENT	-0.8762	0.5483	-1.6	0.11008	
immunePRESENT	0.8588	0.3582	2.4	0.01652	*
metabolicPRESENT	0.8198	0.2581	3.18	0.00149	**
niPRESENT	2.4307	0.1319	18.43	< 2e-16	***
oatofPRESENT	1.7295	0.227	7.62	2.60E-14	***
oncologyPRESENT	-0.9527	0.7657	-1.24	0.21341	
premPRESENT	0.8415	0.13	6.47	9.70E-11	***
renalPRESENT	1.7772	0.474	3.75	0.00018	***
simplegordPRESENT	0.8377	0.1579	5.3	1.10E-07	***
skeletalPRESENT	1.6719	0.4317	3.87	0.00011	***
sleepPRESENT	16.9734	741.4309	0.02	0.98174	
swallowPRESENT	1.5794	0.2479	6.37	1.90E-10	***
trachealPRESENT	0.0443	0.2807	0.16	0.87449	
trachyPRESENT	16.2249	725.9785	0.02	0.98217	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 5320.3 on 9733 degrees of freedom

Residual deviance: 3949.9 on 9704 degrees of freedom

AIC: 4010

Number of Fisher Scoring iterations: 16

The next set of values is the null and residual deviance. A comparison of null versus residual deviance is used assess the goodness of fit of the model. The difference in degrees of freedom between the models is 29. This corresponds to the fact that the saturated model has 29 variables, while the null model has an intercept but no variables. The residual deviance is smaller than the null deviance, which tells us that the saturated model is a better fit than the null model.

The next value to be reported is the AIC. This is of no interest in isolation, but can be used to compare goodness of fit of subsequent models.

Lastly, the summary reports the Fisher scoring iterations. As described before, the model was estimated based on maximum likelihood estimation (MLE). However, multiple trials at fitting the model were attempted to identify the MLEs that generate the best fit. The Fisher scoring iterations value tells us how many attempts at fitting the model were made before the to achieve the output.

Returning to the coefficients, we calculate the odds ratios for the output. Odds ratios are more intuitive; therefore the coefficients and their confidence intervals are exponentiated. The data are presented as odds ratios with 95% confidence intervals. The variables that do not fall below the 5% significance threshold are excluded from the summary below.

Table 86: Odds ratios for significant predictors of fundoplication

	OR	2.5%	97.5%
achalasiaPRESENT	4.4548	1.41479	12.41586
cardiacPRESENT	1.5085	1.15891	1.9611
cdhPRESENT	7.0712	3.95563	12.43363
cldPRESENT	3.0467	2.11694	4.34169
cleftPRESENT	1.6637	0.75183	3.41263
dentalPRESENT	5.0116	1.37063	17.22679
haemPRESENT	0.4164	0.12083	1.08002
immunePRESENT	2.3602	1.11679	4.58949
niPRESENT	11.3669	8.80598	14.77095
oatofPRESENT	5.6378	3.57698	8.72507
premPRESENT	2.3198	1.79481	2.98937
renalPRESENT	5.9134	2.29898	14.75288
simplegordPRESENT	2.3112	1.70073	3.15978
skeletalPRESENT	5.3224	2.27758	12.49089
swallowPRESENT	4.8518	2.97244	7.86818

These data can be interpreted as follows: having a haematological disease reduces the risk of fundoplication. However, having the other comorbidities reported above increases the risk of fundoplication. Neurological impairment appears to be the greatest risk factor and leads to an 11-fold increase in risk of fundoplication.

Step 3: Goodness of fit

Statistical inference

The deviance residual for the logit1 model ranges from -2 to 3 with a median of -0.278. Already, we have some indication about the fit of this model. There are observations with a deviance residual >2 suggesting poor fit.

Table 87: Deviance residual statistics for logit1 model

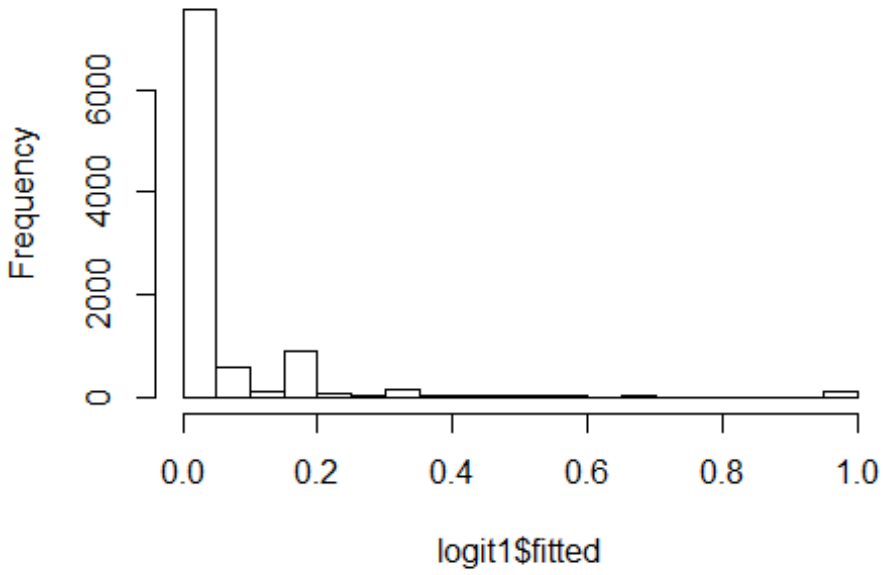
Deviance Residuals				
Min	IQ	Median	3Q	Max
-2.217	-0.278	-0.278	-0.225	3.017

For each model, we examine goodness of fit using various illustrators and measures.

Review of fitted values

The odds of fundoplication are plotted on a histogram. The plot is bimodal with a peak between 0-0.1 and another between 0.9-1.

Figure 81: Fitted values for the logit1 model



The plot is heavily skewed to the left. This corresponds to the observed event rate i.e. 92% of patients did not require fundoplication. There are >6000 patients with a fundoplication probability of <0.1.

Null vs Residual deviance

We compare null and residual deviance. The null deviance is 5320 and the residual deviance is 3950. As residual deviance is lower, this suggests it fits data better than a null model. We can calculate whether this difference meets statistical significance. This difference between null and residual

deviance observes a chi-squared distribution. Using a chi-squared test, we can calculate the probability of observing this difference in deviance.

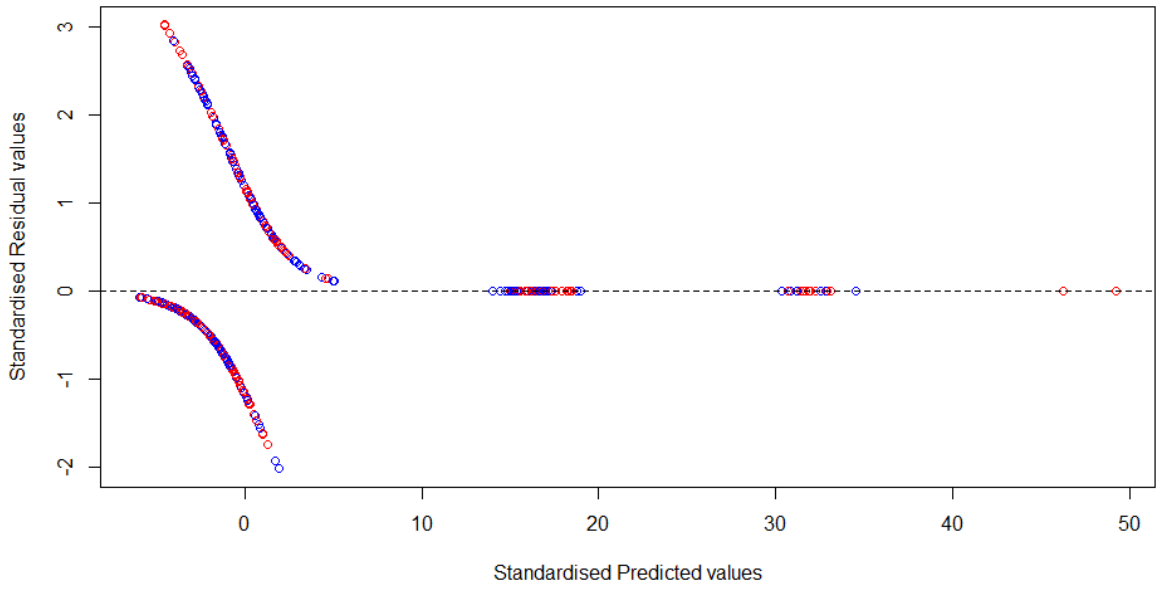
The p-value is <0.001 (chi squared =1416, df=29); i.e. < 5% and therefore not attributable to chance. We can therefore infer that the logit1 model performs better than a null model.

Residual vs. fitted distribution

We plot the residuals to assess the goodness of fit i.e. the distance of the observed data points from the fitted data points. Pearson residuals are calculated and reported here.

A perfect model would have a residual deviance and predicted deviance that matched perfectly. The values. A well- fitted model has points evenly distributed as a horizontal band around the x axis. Reviewing our residual vs. fitted / predicted values plot, we can say that values close to the horizontal line are well fitted. Where points lie above the horizontal line, the model has under-predicted the fitted values. Where the points lie below the line, the model has over-predicted the fitted values.

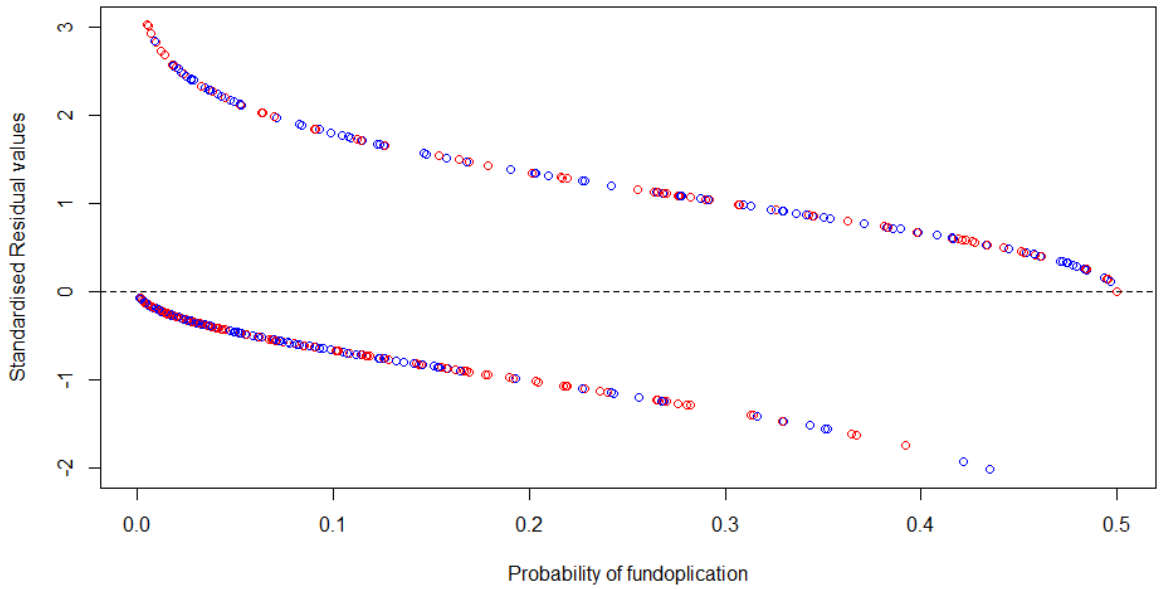
Figure 82: Standardised predicted values plotted against standardised residual values



The residual vs. fitted plot can also be used to assess the homogeneity of variance. This can be assessed by inspection i.e. looking at the vertical scatter of two points across the x axis. Here, we observe that the residuals are strongly patterned when the predicted values range from 0 to 10 (log odds). However, as the x axis becomes >10, there is a sub-population of data points that have different variance to the rest of the population.

This distribution is difficult to interpret. This plot is not intuitive as the x axis presents the log odds of fundoplication. We re-plot using probability of fundoplication against the residual values. The probability of fundoplication is obtained through the inverse logit function.

Figure 83: Probability is plotted against standardised residual values



Again, we observe that the data are highly symmetrical and patterned around the horizontal. However, there is no clustering around the horizontal, as you would observe with a well fitted model. The model does not fit the data well and may be mis-specified. There may be one or more covariates that do not influence the log-odds of success in a linear fashion.

Information criteria

The AIC for the logit1 model is 4010. The BIC is 4225. These values cannot be interpreted in isolation. However, they will be used for comparison with subsequent models.

Step 4: Predictive performance

Next, we assess the logit1 model's classification performance. This is done by applying the model on the test set of data.

Test data set contains 4168 observations of 29 comorbidities. We apply the logit1 model and generate predictions of the probability of fundoplication.

The result is a series of predicted probabilities. Predicted values are reviewed using a histogram. We observe that the distribution of predictions on test data is similar to training data. Most patients are predicted to have a fundoplication probability of <0.1.

Figure 84: predicted probabilities for observations in test data generated using logit1 model

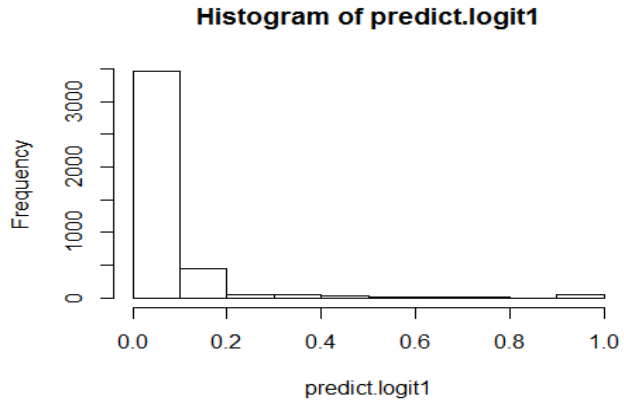
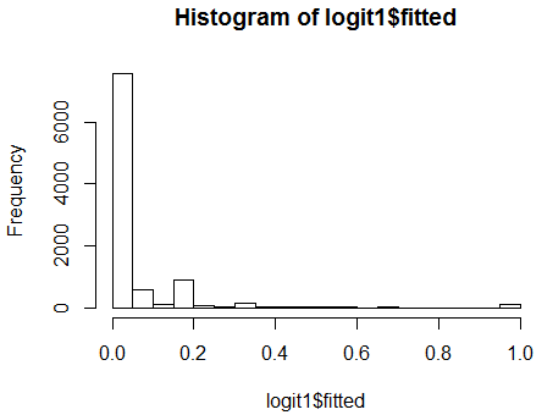


Figure 85: Fitted probabilities for the training data arising from the logit1 model



Classification performance

For the initial analysis, we assume that the model classifies probabilities of > 0.5 to a status of TRUE i.e. fundoplication predicted. We create a new vector containing predicted class. We then compare the predicted classification with the observed classification in a confusion matrix and calculate the classifications error.

Table 88: Confusion matrix and statistics for logit1 model, probability cut-off 0.5

		Actual fundoplication status	
		True	False
Predicted	True	82	14
	False	241	3831

Statistics	
Accuracy	0.94
95% CI	0.93-0.95
Sensitivity	0.25
Specificity	0.99
PPV	0.85
Negative predictive value (NPV)	0.94

With a probability cut-off of 0.5, the sensitivity of the logit1 model is low (25%). However, we have observed in our histogram of fitted and predicted probabilities most patients have a probability of <0.1. Therefore, a probability cut-off of 0.5 is far too high for the data at hand.

To illustrate the effect of probability on sensitivity, we estimate sensitivity for probability cut-offs ranging from 0.05 to 0.95.

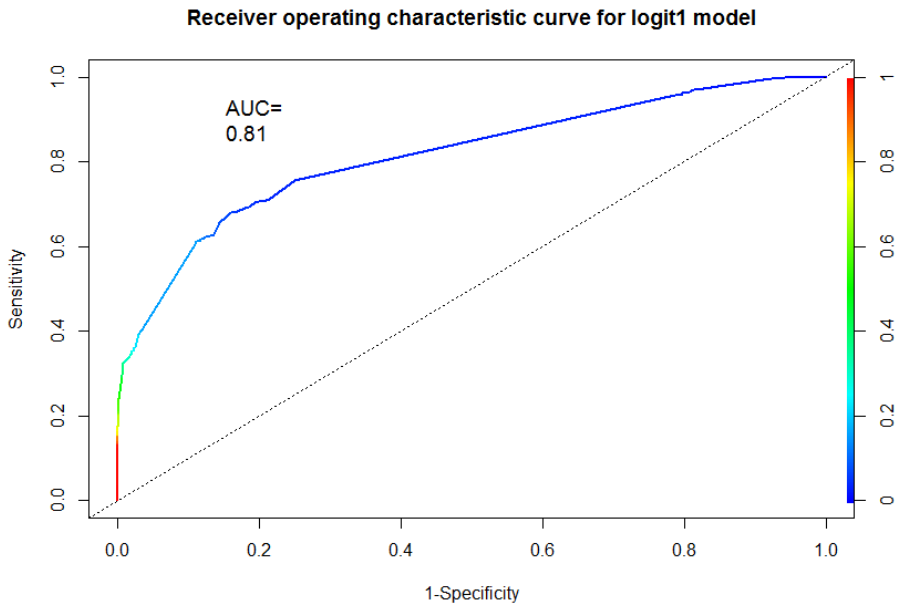
Table 89: Sensitivity - specificity trade-off

	Sensitivity	Specificity	False positive rate
Probability cut-off	% for fundoplication	% No fundoplication	1 % wrongly selected for fundoplication
0.05	69	81	19
0.1	63	87	13
0.2	40	97	3
0.3	35	98	2
0.4	31	99	0.8
0.5	25	100	0.4
0.6	21	100	0.2
0.7	18	100	0.1
0.8	15	100	0
0.9	14	100	0
0.95	13	100	0

Sensitivity decreases as the probability cut-off is increased. Specificity varies in the opposite direction. We can better visualise and automate the estimation of optimal probability using a receiver operating characteristics (ROC) analysis.

Receiver operator characteristics

Figure 86: ROC curve for the logit1 model



On inspection, the uppermost and most leftward point in the curve corresponds to a probability cut-off of 0.2. However, visual inspection is simply not reliable. Differentiation is used to identify the point of inflection of the curve. This is the basis of the Youden method. The optimum probability cut-off is 0.04. The performance metrics at this cut-off are presented in the table below.

Table 90: Confusion matrix and statistics for logit1 model, probabity cut-off 0.04

		Actual fundoplication status	
		True	False
Predicted	True	224	718
	False	99	3127

Statistics	
Accuracy	0.80
95% CI	0.75-0.83
Sensitivity	0.69
Specificity	0.81
PPV	0.24
Negative predictive value (NPV)	0.97

The area under the ROC curve (AUC) is a measure of the model's discriminatory power. The AUC for logit1 is 0.81. This value is of limited utility in isolation. However, as the AUC is greater than 0.5, this

indicates that the logit1 models a better predictor than chance (which would yield an AUC =0.5). Using the academic scale for interpretation, the logit1 model has a 'good' AUC.

Step 5: Clinical implications - logit1

With a probability cut-off of 0.4, we observe a sensitivity of 69%. The clinical implication is that 7 of 10 patients who require fundoplication can be identified using the logit1 model. The false positive rate (1-specificity) is 20%. Clinically, reducing risks of unnecessary surgery (high specificity) must be balanced against the risk of not offering fundoplication to a patient who needs it (high sensitivity). No studies can tell us which scenario is to be preferred. However, my clinical instinct would favour model heuristics that avoid unnecessary surgery and all its attendant risks. Therefore, further optimisation will favour high specificity over sensitivity.

The logit1 model is not accurate (80%). Considering the no-fundoplication rate is 92% in the training data, the error rate is high at 25%. The logit1 model requires further optimisation. This can be done by:

1. Addressing variables associated with complete linear separation
2. looking for the effects of interactions between the co-variates
3. reducing noise arising from uninformative parameters through stepwise regression
4. addressing the bias associated with the rarity of the fundoplication event.

Improving the model: addressing linear separation

We have chosen to address linear separation by examining the model with and without the perfect predictor variables. The saturated model, logit1, has been fit above. Below we, fit logit.sep i.e. a model without the perfect predictor variables i.e. sleep apnoea, tracheostomy, aspiration, asthma, chromosomal anomalies, consanguinity and acute respiratory failure. The new model applying the reduced formula is named logit.sep.

Equation 4: Formula for logit.sep model

```
fm.sep<-hadfundo ~ tracheal+cleft+cld+cdh+cardiac+ni+oatof+achalasia+swallow+prem+haem+
endocrine+renal+skeletal+bone+dental+oncology+metabolic+immune+cardsurg+eb+simplegord
```

Statistical inference

Compared to the logit1 model, haematological conditions do not appear to significantly reduce the risk of fundoplication. Furthermore, endocrine and metabolic conditions are associated with an increased risk of fundoplication in the logit.sep model (6). Coefficients with a p value >0.05 for one model are left blank. Coefficients with a p-value >0.05 for both models are excluded from the table.

Table 91: Comparing odds ratios for logit.sep and logit1 models

	Logit1			Logit.sep		
	OR	2.5%	97.5%	OR	2.5%	97.5%
niPRESENT	11.4	8.8	14.8	15.6	12.2	20.1
cdhPRESENT	7.1	4.0	12.4	8.0	4.6	14.0
renalPRESENT	5.9	2.3	14.8	7.0	2.8	17.2
oatofPRESENT	5.6	3.6	8.7	6.3	4.0	9.7
skeletalPRESENT	5.3	2.3	12.5	5.7	2.6	13.3
swallowPRESENT	4.9	3.0	7.9	5.2	3.3	8.3
achalasiaPRESENT	4.5	1.4	12.4	4.8	1.5	13.8
dentalPRESENT	5.0	1.4	17.2	4.7	1.3	16.9
cldPRESENT	3.0	2.1	4.3	3.0	2.1	4.2
simplegordPRESENT	2.3	1.7	3.2	2.9	2.1	3.9
immunePRESENT	2.4	1.1	4.6	2.9	1.4	5.5
cleftPRESENT	1.7	0.8	3.4	2.5	1.3	4.6
premPRESENT	2.3	1.8	3.0	2.4	1.9	3.1
metabolicPRESENT				2.4	1.4	3.9
cardiacPRESENT	1.5	1.2	2.0	1.9	1.5	2.4
endocrinePRESENT				1.8	1.1	3.1
haemPRESENT	0.4	0.1	1.1			

The OR generated by the two models are different, reflecting a difference in effect size. For example, in the logit.sep model, neurological impairment is associated with a 15-fold increased odds ratio of fundoplication, compared to the 11-fold increase suggested by the logit1 model. However, review of fitted probabilities yields frequencies that are identical on inspection. Most patients have a fitted probability of 0 -0.1, in keeping with the observed event rate.

Figure 87: Fitted probability arising from the logit.sep model

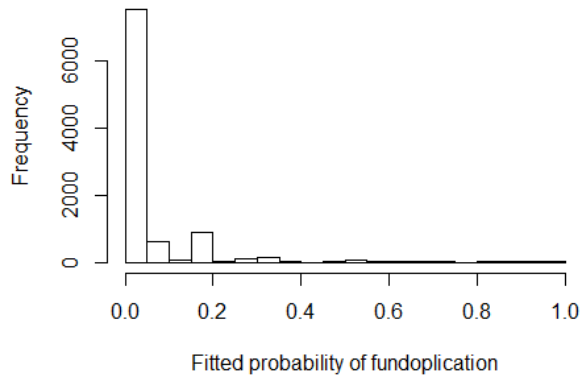
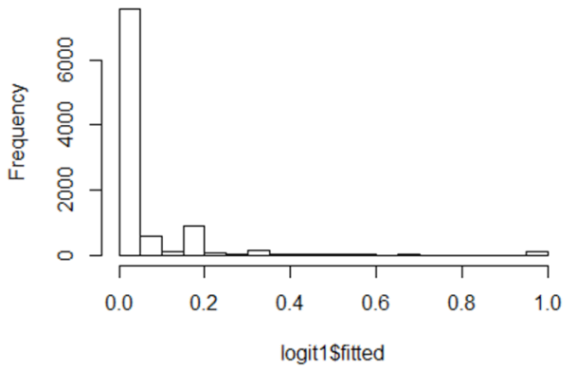


Figure 88: Fitted probability arising from the logit1 model

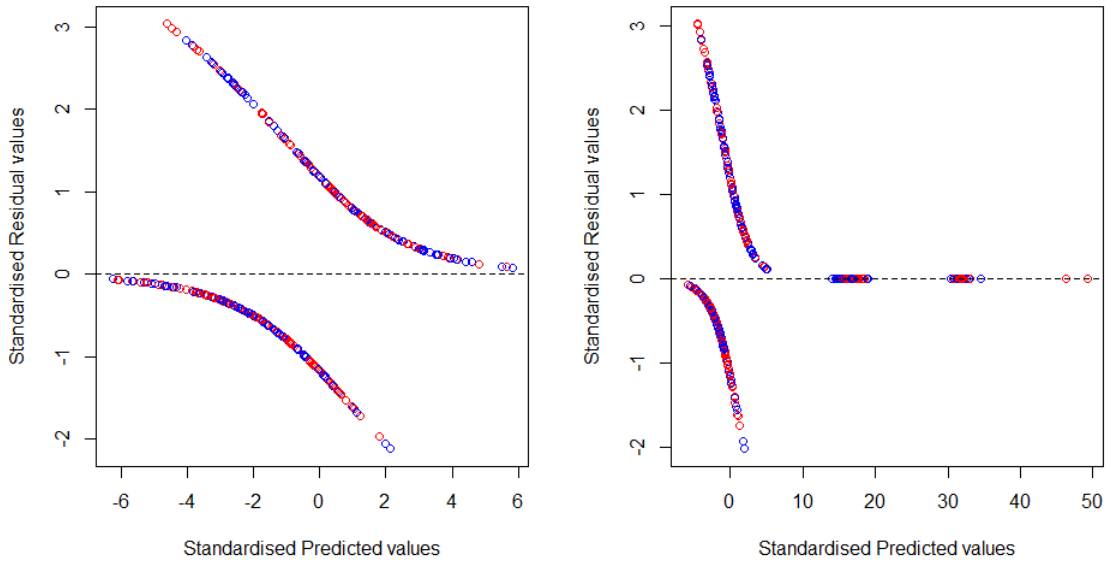


Goodness of fit

Deviance residuals range from -2.2 to 3 with a median of -0.278. Like logit1, there are observations with a deviance residual >2, indicating poor fit. Null deviance is significantly greater than residual deviance ($p < 0.001$ ($\chi^2=1249$, df=22), indicating that the logit.sep model fits the data better than a null model. Compared to the logit.sep model plot, the logit1 plot has a sub-population of standardised residuals

with a value of 0. This is illustrated in the residual vs. fitted distribution (Figure 89). These points lie on the horizontal and have homogeneity of variance. They arise from the perfect predictor variables. On the logit.sep plot, excluding perfect predictors has removed this sub-population.

Figure 89: Comparing residual vs fitted profiles of logit.sep and logit1 models



Both AIC and BIC for the logit.sep model are greater than those for the logit1 model. Focusing on the BIC, we can make some inferences about the model goodness of fit.

Table 92: Comparison of AIC and BIC of logit1 and logit.sep models

Model	Parameters	AIC	BIC
logit1	29	4010	4225
logit.sep	22	4117	4282

Logit.sep Logit1 has fewer parameters than and the same number of observations. Therefore, we would expect the penalty for parameters to be less for logit.sep. However, logit.sep has a higher BIC. This suggests that the likelihood term for BIC is greater for logit.sep, suggesting greater deviance. We confirm this by calculating the log-likelihood and deviance for each model.

This difference in deviance is statistically significant. This inference can be carried out using analysis of variance as logit.sep is a nested model of logit1.

Table 93: Comparing log likelihood and deviance for logit1 and logit.sep models

	logit1	logit.sep	p
Log-likelihood	-1975	-2036	
Deviance	3950	4071	<0.0001

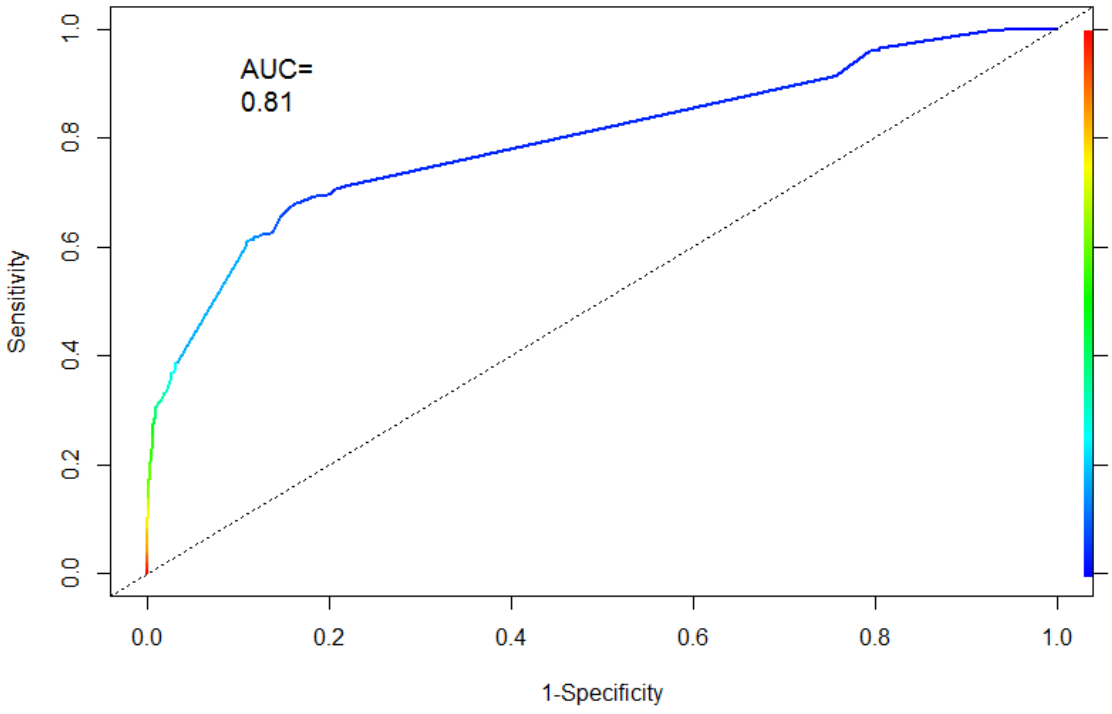
```
anova(logit1, logit.sep, test="Chisq")

Analysis of Deviance Table

Model 1: hadfundo ~ tracheal + cleft + sleep + trachy + aspiration + cld +
    cdh + asthma + cardiac + ni + oatof + achalasia + swallow +
    chrom + consang + prem + haem + endocrine + renal + skeletal +
    bone + dental + oncology + metabolic + aresp + immune + cardsurg +
    eb + simplegord
Model 2: hadfundo ~ tracheal + cleft + cld + cdh + cardiac + ni + oatof +
    achalasia + swallow + prem + haem + endocrine + renal + skeletal +
    bone + dental + oncology + metabolic + immune + cardsurg +
    eb + simplegord


  Resid.Df       Resid. Dev         Df      Deviance         Pr(>Chi)
1     9704            3950
2     9711            4071            7       121              <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Therefore, we can conclude that logit.sep fits the data less well than logit1. Removing perfect predictors from the logit1 model does not improve fit for the remaining variables.

Predictive performance

To visualise the optimal cut-off where sensitivity and specificity are maximised, we plot an ROC curve.

Figure 90: ROC curve for the logit.sep model



The optimal cut-off probability is estimated and found to be 0.06. The confusion matrix for this probability cut-off is generated.

Table 94: Confusion matrix and performance statistics for logit.sep model, probabity cut-off 0.06

		Actual fundoplication status	
		True	False
Predicted	True	220	629
	False	103	3216

Statistics	
Accuracy	0.82
Sensitivity	0.68
Specificity	0.84
PPV	0.26
Negative predictive value (NPV)	0.97

The AUC of the logit.sep model is 0.81. Although this suggests the model is a 'good' predictor, it is no better at classification than logit1 (AUC= 0.81).

Clinical implications

With a probability cut-off of 0.06, sensitivity is 68%. This is a similar sensitivity to that achieved using the logit1 model at a probability cut-off of 0.04. Therefore, the logit.sep model is a more sensitive predictor than the logit1 model.

The false positive rate for the logit.sep model is higher at 17%. Therefore logit.sep would incorrectly assign approximately four times more patients to fundoplication than logit1.

Although the AUC for the logit.sep and logit1 models are similar, the analysis of deviance and the BIC both suggest that the logit1 model fits the data better.

Clinically, the model associated with a lower risk of unnecessary fundoplication is preferred. On balance, therefore, logit1 is the better of the two models.

The high variance on the residual deviance suggests the model could be improved. Often, undetected interactions between variables can introduce bias into the model. Therefore, the next strategy will be searching for interacting variables.

Improving the model: addressing interactions

In patients with coincident comorbidities, there may be a potentiating effect of one comorbidity on the other. This phenomenon is seen clinically e.g. prematurity is a risk factor for neurological impairment. Statistically, this effect can be sought out by interrogating the data for interactions.

To illustrate the procedure, we search for interactions of various comorbidities with NI. We search for 2-way interactions between NI and sleep apnoea. The interaction term ni*sleep is included in the regression formula.

Coefficients					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-4.0745	0.1391	-29	< 2e-16	***
niPRESENT: sleepPRESENT	NA	NA	NA	NA	
niPRESENT	2.4307	0.1319	18	< 2e-16	***
sleepPRESENT	16.9734	741.4309	0	0.98174	
cldPRESENT	1.1141	0.183	6	1.10E-09	***
…	1.956	0.2917	7	2.00E-11	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

There was no coefficient calculable for the ni*sleep term. This is because there are no patients with the coincident variables of neurological impairment and sleep disorder in the training set.

Repeating the procedure for ni*cld, we are able to calculate a coefficient. Therefore, there are patients with both these conditions. However, the coefficient is not found to be significant.

	Estimate	Std. Error	z value	Pr(>|z|)	
niPRESENT: cldPRESENT	-0.0331	0.3695	-0.09	0.92852	

In contrast, the interaction between NI and tracheal anomalies disorders has a significant beta coefficient.

	Estimate	Std. Error	z value	Pr(>|z|)	
niPRESENT: trachealPRESENT	2.322	0.696	3.33	0.00085	***

Therefore, the interaction term NI*tracheal will be included in the regression formula.

In this manner, other interactions between comorbidities is carried out in a serial fashion. We identify the following interactions:

Table 95: Interaction terms that significantly predict risk of fundoplication

Coefficients					
Interaction	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**	
niPRESENT:trachealPRESENT	2.322	0.6964	3.334	0.000855	***
niPRESENT:cleftPRESENT	4.873129	1.488736	3.273	0.001063	**
niPRESENT:cardiacPRESENT	0.99384	0.26592	3.737	0.000186	***
niPRESENT:swallowPRESENT	2.75757	0.8243	3.345	0.000822	***
niPRESENT:premPRESENT	-0.95652	0.26416	-3.621	0.000294	***
cldPRESENT:metabolicPRESENT	1.83396	0.91469	2.005	0.044963	*
cardiacPRESENT:premPRESENT	-0.66776	0.28253	-2.364	0.018102	*
premPRESENT:achalasiaPRESENT	-3.85658	1.91899	-2.01	0.044463	*
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Apply logistic regression

We re-specify the formula to include the interactions terms identified above. The interacting model is called logit.int.

Equation 5: Formula for logit.int model

```
fm.int<-hadfundo~ni*tracheal+ ni*cleft+  ni*cardiac+  ni*swallow+   ni*prem+  cld*metabolic+
  cardiac*prem+ prem*achalasia+  tracheal+cleft+sleep+trachy+aspiration+cld+cdh+
asthma+cardiac+ni+oatof+achalasia+swallow+chrom+ consang+prem+haem+endocrine+renal+
skeletal+bone+dental+oncology+metabolic+aresp+immune+cardsurg+eb+simplegord
```

Logistic regression is carried out with the following result. The variables are sorted in order of beta coefficient / estimate.

Table 96: Summary of coefficients for logit.int model including interaction terms and perfect predictors

Coefficients:					
	Estimate	Std. Error	z value	Pr(>|z|)	
asthmaPRESENT	17.482	662.024	0	0.97893	
aspirationPRESENT	17.186	728.514	0	0.98118	
chromPRESENT	16.699	834.09	0	9.84E-01	
consangPRESENT	16.456	909.033	0	0.98556	
sleepPRESENT	16.301	727.619	0	0.98213	
trachyPRESENT	15.139	727.017	0	9.83E-01	
arespPRESENT	15.074	746.236	0	0.98388	
niPRESENT:cleftPRESENT	4.383	1.504	3	0.00357	**
niPRESENT:swallowPRESENT	2.422	0.839	3	0.0039	**
niPRESENT	2.206	0.207	11	< 2e-16	***
niPRESENT:trachealPRESENT	2.022	0.708	3	0.00431	**
cldPRESENT:metabolicPRESENT	1.997	0.913	2	0.02866	*
achalasiaPRESENT	1.953	0.548	4	0.00036	***
cdhPRESENT	1.92	0.304	6	2.50E-10	***
oatofPRESENT	1.705	0.237	7	6.90E-13	***
renalPRESENT	1.677	0.478	4	0.00045	***
skeletalPRESENT	1.655	0.449	4	0.00023	***
premPRESENT	1.494	0.224	7	2.30E-11	***
dentalPRESENT	1.395	0.646	2	0.03072	*
cldPRESENT	1.048	0.194	5	6.60E-08	***
niPRESENT:cardiacPRESENT	1.003	0.278	4	0.00032	***
immunePRESENT	0.769	0.357	2	0.03103	*
cardsurgPRESENT	0.758	0.502	2	0.13132	
simplegordPRESENT	0.709	0.213	3	9.00E-04	***
metabolicPRESENT	0.577	0.282	2	4.10E-02	*
endocrinePRESENT	0.478	0.277	2	0.08464	.
cardiacPRESENT	0.118	0.225	1	0.59965	
swallowPRESENT	-0.116	0.759	0	0.87884	
ebPRESENT	-0.347	1.025	0	0.73463	
cardiacPRESENT:premPRESENT	-0.662	0.295	-2	0.02498	*
trachealPRESENT	-0.925	0.522	-2	0.07616	.
oncologyPRESENT	-0.932	0.754	-1	0.21655	
niPRESENT:premPRESENT	-0.968	0.277	-4	0.00047	***
bonePRESENT	-1.143	0.756	-2	0.13026	
cleftPRESENT	-1.206	1.04	-1	0.24602	
haemPRESENT	-1.233	0.635	-2	0.05217	.
premPRESENT:achalasiaPRESENT	-3.323	1.825	-2	0.06861	.
(Intercept)	-3.945	0.2	-20	< 2e-16	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Null deviance: 5320.3 on 9733 degrees of freedom Residual deviance: 3863.9 on 9696 degrees of freedom AIC: 3940 Number of Fisher Scoring iterations: 16					

The largest coefficients with correspondingly large standard errors correspond to perfect predictor variables e.g. asthma and aspiration. For these, no maximum likelihood estimate can be achieved. The patients with these comorbidities comprise a small but significant percentage of the total of patients who had fundoplication (119 of 757=15%). Leaving in these variables in the formula may lead to inaccurate

standard errors and Wald estimates Furthermore, none of these variables are associated significantly with the outcome of fundoplication. Therefore, we re-specify the formula to exclude these 'problem variables'. This reduced model is called logit.sep.int.

Logistic regression is re-applied.

The estimate values for the remaining variables are marginally different. However, the values with a p value <0.05 remain the same. Therefore, the formula without the 'problem variables' will be used henceforth.

Table 97: Summary of coefficients for logit.sep.int model

Coefficients:					
	Estimate	Std. Error	z value	Pr(>|z|)	
niPRESENT:cleftPRESENT	4.687	1.478	3	0.00152	**
niPRESENT:swallowPRESENT	2.549	0.837	3	0.00232	**
niPRESENT:trachealPRESENT	2.379	0.672	4	0.0004	***
niPRESENT	2.307	0.203	11	< 2e-16	***
cdhPRESENT	2.039	0.29	7	2.00E-12	***
cldPRESENT:metabolicPRESENT	1.999	0.918	2	0.02944	*
achalasiaPRESENT	1.971	0.546	4	0.00031	***
renalPRESENT	1.78	0.462	4	0.00012	***
oatofPRESENT	1.763	0.232	8	3.20E-14	***
skeletalPRESENT	1.734	0.434	4	6.40E-05	***
premPRESENT	1.506	0.221	7	9.00E-12	***
dentalPRESENT	1.323	0.657	2	0.04395	*
niPRESENT:cardiacPRESENT	1.213	0.268	5	6.00E-06	***
cldPRESENT	1.031	0.192	5	7.40E-08	***
immunePRESENT	0.868	0.339	3	0.01053	*
cardsurgPRESENT	0.752	0.5	2	0.13288	
simplegordPRESENT	0.736	0.211	3	0.0005	***
metabolicPRESENT	0.595	0.275	2	3.07E-02	*
endocrinePRESENT	0.443	0.277	2	1.10E-01	
cardiacPRESENT	0.123	0.222	1	0.58031	
swallowPRESENT	-0.124	0.763	0	0.87063	
ebPRESENT	-0.32	1.025	0	0.75488	
cardiacPRESENT:premPRESENT	-0.646	0.283	-2	2.28E-02	*
niPRESENT:premPRESENT	-0.91	0.266	-3	0.00064	***
trachealPRESENT	-0.912	0.522	-2	0.08045	.
oncologyPRESENT	-1.006	0.766	-1	0.18905	
cleftPRESENT	-1.205	1.041	-1	0.24705	
bonePRESENT	-1.219	0.761	-2	1.09E-01	
haemPRESENT	-1.311	0.656	-2	0.04557	*
premPRESENT:achalasiaPRESENT	-3.72	1.905	-2	0.05077	.
(Intercept)	-3.972	0.198	-20	< 2e-16	***

Statistical inference

We re-express the beta coefficients as odds ratios and confidence intervals. The variables that do not fall below the 5% significance threshold are excluded from the summary.

Table 98: Odds ratios for significant coefficients of the logit.sep.int

Variable/ Interaction	OR	2.5%	97.5%
niPRESENT:cleftPRESENT	108.51	9.53	4242.43
niPRESENT:swallowPRESENT	12.80	3.00	91.39
niPRESENT:trachealPRESENT	10.79	3.16	45.72
niPRESENT	10.05	6.79	15.09
cdhPRESENT	7.69	4.32	13.50
cldPRESENT:metabolicPRESENT	7.38	1.15	47.16
achalasiaPRESENT	7.18	2.28	19.88
renalPRESENT	5.93	2.38	14.53
oatofPRESENT	5.83	3.67	9.13
skeletalPRESENT	5.66	2.44	13.43
premPRESENT	4.51	2.93	6.97
dentalPRESENT	3.76	0.99	13.31
niPRESENT:cardiacPRESENT	3.36	1.99	5.69
cldPRESENT	2.80	1.92	4.06
simplegordPRESENT	2.09	1.39	3.18
metabolicPRESENT	1.81	1.04	3.07
niPRESENT:premPRESENT	0.40	0.24	0.68
haemPRESENT	0.27	0.06	0.82
(Intercept)	0.02	0.01	0.03

This analysis yields some new insights e.g. having neurological impairment and a cleft lip/ palate is the most significant risk factor for fundoplication. Interestingly, neurological impairment combined with prematurity reduces the risk of fundoplication.

Goodness of fit

The null deviance (5320) is greater than residual deviance (3958). The difference in deviance is significant ($p<0.001$) confirming that the logit.sep.int model fits the data better than a null model.

The logit.sep.int AIC and BIC are both marginally greater than the logit1 AIC. Logit.sep has the highest AIC, BIC and deviance. This model will perform less well compared with logit1 and logit.sep.int and is therefore dropped from further consideration.

Table 99: Comparing the model diagnostic parameters of the logit1, logit.sep and the logit.sep.int models

Models	logit1	logit.sep	logit.sep.int
Parameters	29	22	30
AIC	4010	4117	4020
BIC	4225	4282	4243
Log-likelihood	-1974.94	-2036	-1979.21
Deviance	3949	4071	3958

The logit.sep.int model deviance is greater than the logit1 model. Using analysis of variance (ANOVA) difference in deviance between the two models is not significant.

Figure 91: ANOVA - difference in deviance between logit.sep.int and logit1

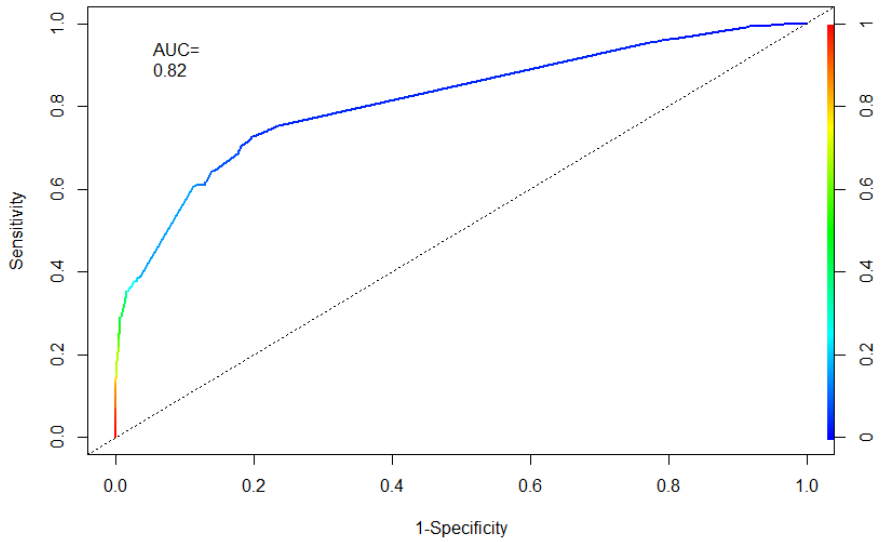
```
anova(logit1, logit.sep, test="Chisq")

Analysis of Deviance Table

Model 1: hadfundo ~ tracheal + cleft + sleep + trachy + aspiration + cld +
    cdh + asthma + cardiac + ni + oatof + achalasia + swallow +
    chrom + consang + prem + haem + endocrine + renal + skeletal +
    bone + dental + oncology + metabolic + aresp + immune + cardsurg +
    eb + simplegord
Model 2: hadfundo ~ tracheal + cleft + cld + cdh + cardiac + ni + oatof +
    achalasia + swallow + prem + haem + endocrine + renal + skeletal +
    bone + dental + oncology + metabolic + immune + cardsurg +
    eb + simplegord
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      9704       3950
2      9711       4071 -7     -121   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the difference in deviance between the models is not significant, we can conclude that logit.sep.int and logit1 models fit the data equally well.

Predictive performance

The AUC for the logit.sep.int model is 0.82. It is a better predictor than random change. However, it is negligibly better at prediction than the logit1 (AUC = 0.81).

Figure 92: ROC curve for logit.sep.int model



The optimal probability cut-off is identified at 0.05. This corresponds to a sensitivity of 73% and a specificity of 80%.

Table 100: Confusion matrix and performance statistics for logit.sep.int model, probability cut-off 0.05

		Actual fundoplication status	
		True	False
Predicted	True	235	759
	False	88	3086

Statistics	
Accuracy	0.80
Sensitivity	0.73
Specificity	0.80
PPV	0.24
Negative predictive value (NPV)	0.97

The clinical implication can be summarised as follows: If the optimal cut-off probability of fundoplication is estimated as 5%. This probability of 5% corresponds well to the real event rate i.e. only 8% of patients in the training sample had fundoplication. At this cut-off 73% of patients who need fundoplication can be identified. However, the false positive rate associated with this cut-off is 20%. This rate is unacceptably high and would result in 1 in 5 having an unnecessary operation. Further optimisation of this classifier is required.

We will try stepwise regression and bias reduction. To further optimise the model, we address influential variables using bias reduction techniques. We also address non-contributory parameters using stepwise regression.

Improving the model: Stepwise regression

Next, we optimize this model by identifying which contributing parameters to keep and which to drop. This is achieved through step-wise regression. Both AIC and BIC can be used as metrics to identify the information index of individual contributing parameters.

Stepwise regression using AIC

To recapitulate, stepwise modelling is logistic regression carried out in a serial fashion. Variables are

Equation 7: Calculating AIC

$$AIC = -2 \log L' + 2k$$

sequentially removed from the model and the effect on AIC is assessed. To recapitulate:

Where;

k = number of parameters included (comorbidities)

L' = Likelihood function

AIC is a model fitness parameter that is responsive to overfitting. The more parameters in a model, the higher the penalty. For stepwise regression, AIC is applied by using a k value of 2.

A null model will have the lowest AIC. A saturated model will have a high AIC. A variable which, when dropped from the saturated model, results in the least reduction in AIC can be dropped. The base

Equation 8: Formula for stepwise logistic regression model

```
fm.sep1<-hadfundo~ ni*tracheal+ ni*cleft+ ni*cardiac+ ni*swallow+ ni*prem+ cld*metabolic+
  cardiac*prem+ prem*achalasia+ tracheal+cleft+cld+cdh+ cardiac+ni+oatof+achalasia+swallow+
  prem+haem+endocrine+renal+skeletal+bone+ dental+oncology+metabolic+immune+cardsurg+
eb+simplegord
```

formula is the same as used for logit.sep.int model. It includes 8 interaction terms and 22 independent variables. The formula is restated to utilise a stepwise algorithm.

Equation 9: Logistic (logit.sep.int) and stepwise logistic (stepaic) regression operations

```
logit.sep.int<-                    glm                    (fm.sep1,
        family=binomial(logit),
        data=train)

stepaic<-step(logit.sep.int,
            direction="both",
            k=2)
```

```
Deviance                                                          Residuals:
    Min                      1Q       Median                      3Q                 Max
 -2.988   -0.278   -0.278   -0.209    3.187
```

Table 101: Coefficients for the stepaic model

Coefficients:					
	Estimate	Std. Error	z value	Pr(>|z|)	
niPRESENT	2.317	0.201	11.54	< 2e-16	***
oatofPRESENT	1.775	0.232	7.66	1.80E-14	***
cdhPRESENT	2.041	0.29	7.04	1.90E-12	***
premPRESENT	1.515	0.219	6.92	4.60E-12	***
cldPRESENT	1.029	0.191	5.38	7.40E-08	***
niPRESENT:cardiacPRESENT	1.179	0.266	4.43	9.20E-06	***
skeletalPRESENT	1.734	0.434	3.99	6.50E-05	***
renalPRESENT	1.778	0.462	3.85	0.00012	***
achalasiaPRESENT	1.973	0.546	3.61	0.0003	***
simplegordPRESENT	0.745	0.208	3.58	0.00034	***
niPRESENT:trachealPRESENT	2.381	0.671	3.55	0.00039	***
niPRESENT:premPRESENT	-0.918	0.265	-3.46	0.00054	***
niPRESENT:cleftPRESENT	4.66	1.479	3.15	0.00162	**
niPRESENT:swallowPRESENT	2.549	0.837	3.05	0.00232	**
immunePRESENT	0.859	0.339	2.53	0.01131	*
cardiacPRESENT:premPRESENT	-0.659	0.283	-2.33	0.01968	*
cldPRESENT:metabolicPRESENT	1.986	0.918	2.16	0.03048	*
metabolicPRESENT	0.596	0.275	2.16	0.03051	*
dentalPRESENT	1.319	0.656	2.01	0.0445	*
premPRESENT:achalasiaPRESENT	-3.72	1.905	-1.95	0.05076	.
haemPRESENT	-1.257	0.651	-1.93	0.05336	.
trachealPRESENT	-0.909	0.522	-1.74	0.0816	.
endocrinePRESENT	0.438	0.277	1.58	0.11351	
bonePRESENT	-1.143	0.75	-1.52	0.12738	
oncologyPRESENT	-1.007	0.766	-1.31	0.18874	
cleftPRESENT	-1.181	1.042	-1.13	0.25689	
cardiacPRESENT	0.167	0.219	0.77	0.44407	
swallowPRESENT	-0.127	0.763	-0.17	0.86824	
(Intercept)	-3.982	0.194	-20.51	< 2e-16	***

In the summary of the stepaic model result, we see there are 20 independent variables and 8 interaction terms remaining. The original formula contained 22 independent variables and 8 interaction terms. The results summary includes the results where there is a decrease in AIC observed. Two variables have been dropped i.e. ‘eb’ and ‘cardsurg’ are dropped.

Stepwise regression using BIC

Bayesian information criterion (BIC) is also responsive to overfitting by introducing a penalty the number of parameters. However, BIC penalizes the number of parameters more strongly than AIC and is therefore more sensitive to over-fitting(254).

Equation 10: Calculating BIC

$$BIC = -2\log L' + k\log(n)$$

BIC can be estimated by the formula:

Where;

k= number of parameters (comorbidities)

n=number of observations (number of patients)

L' = Likelihood function

We re-run the stepwise regression utilising BIC as the variable selection criterion. The base formula is that used for logit.sep.int model i.e. fm.sep1.

Equation 11: Logistic (logit.sep.int) and stepwise logistic (stepaic) regression operations

```
logit.sep.int<- glm(fm.sep1,
        family=binomial(logit),
        data=train)

stepbic<-step(logit.sep.int,
            direction="both",
            k=log(nrow(train)))
```

The BIC criterion results in more parameters being dropped. The base formula contained 22 independent variables and 8 interacting terms. The stepbic model summary has 12 independent variables and 4 interaction terms.

Table 102: Coefficients for the stepbic model

Coefficients:					
	Estimate	Std. Error	z value	Pr(>|z|)	
niPRESENT:cleftPRESENT	4.8145	1.4637	3.29	0.001	**
niPRESENT:swallowPRESENT	2.8162	0.8157	3.45	0.00056	***
niPRESENT:trachealPRESENT	2.7337	0.6614	4.13	3.60E-05	***
cdhPRESENT	1.8386	0.2651	6.93	4.10E-12	***
renalPRESENT	1.666	0.4164	4	6.30E-05	***
niPRESENT	1.6632	0.1016	16.37	< 2e-16	***
skeletalPRESENT	1.644	0.4117	3.99	6.50E-05	***
niPRESENT:cardiacPRESENT	1.5547	0.2295	6.77	1.20E-11	***
oatofPRESENT	1.4368	0.2056	6.99	2.80E-12	***
cldPRESENT	0.9766	0.1793	5.45	5.10E-08	***
premPRESENT	0.7209	0.1176	6.13	8.90E-10	***
swallowPRESENT	-0.2736	0.7408	-0.37	7.12E-01	
cardiacPRESENT	-0.3608	0.1501	-2.4	1.63E-02	*
trachealPRESENT	-1.2406	0.5103	-2.43	1.51E-02	*
cleftPRESENT	-1.2486	1.0189	-1.23	2.20E-01	
haemPRESENT	-1.8363	0.6397	-2.87	0.0041	**
(Intercept)	-3.2829	0.0667	-49.19	< 2e-16	***

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for binomial family take n to be1)
Null deviance: 5320.3 on 9733 degrees of freedom
Residual deviance: 4015.0 on 9717 degrees of freedom
Number of Fisher Scoring iterations: 7

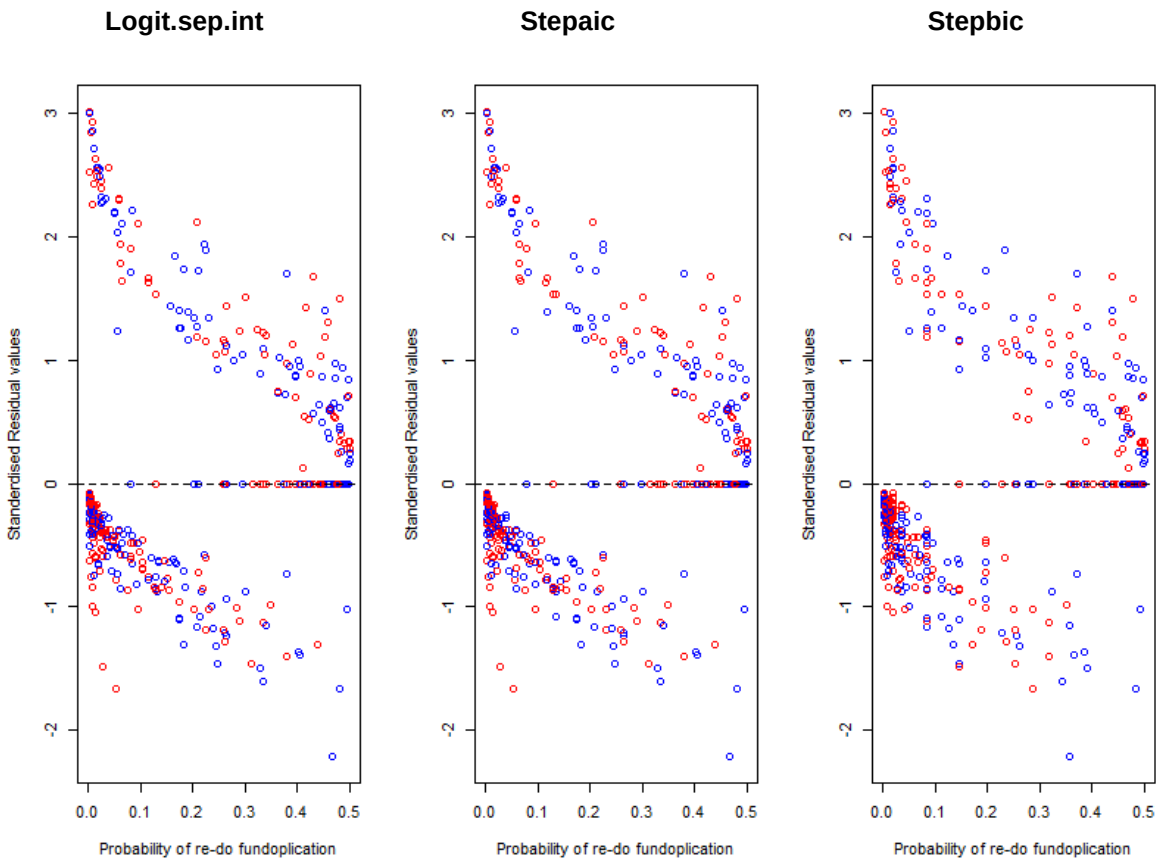
Goodness of fit

Residual vs fitted distributions

We review the residual vs. fitted plot for the stepaic, stepbic and logit.sep.int models. As with the logit.sep.int residual vs. fitted plot, we find that the residuals are patterned. There are two populations of points, suggesting predicted values with different variance. Firstly, there is a cluster of fitted values whose residual value is 0. These have been perfectly predicted and lie in the higher range of probabilities i.e. >0.1.

Secondly, there is a population of points scattered above and below the horizontal in a curved band. All three models demonstrate clustering below the horizontal when the probability ranges from 0 -0.1. The residuals are curved around the horizontal, suggesting a non-linear relationship between one or more variables and the outcome.

Figure 93: Fitted values vs. standardised residuals for the stepaic, stepbic and logit.sep.int models



This may indicate that an element of bias persists in the stepwise models. Having sought and identified interactions, residual bias may arise from unmeasured parameters. It may also be an indication that the model has been mis-specified and will require further optimisation.

Information criteria

Information criteria for the stepwise regression models are compared. The table below confirms that BIC is more responsive than the AIC to a change in the number of parameters.

Table 103 Comparison of AIC and BIC of the logit1, logit.sep.int, stepaic and stepbic models

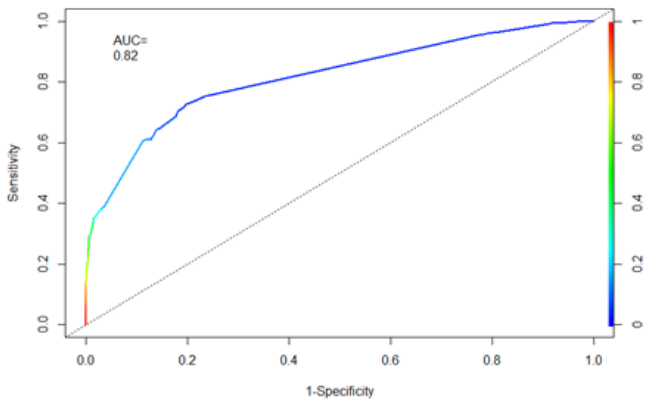
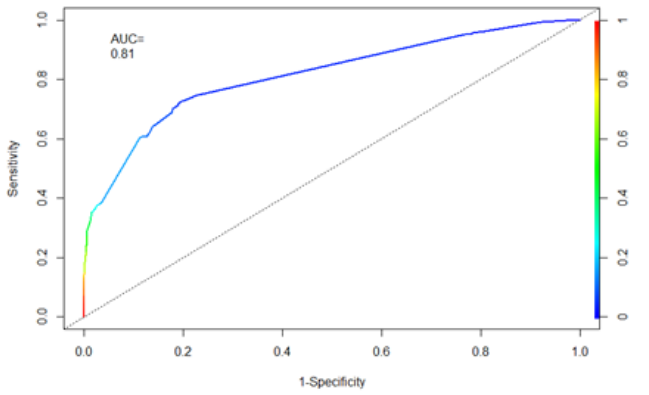
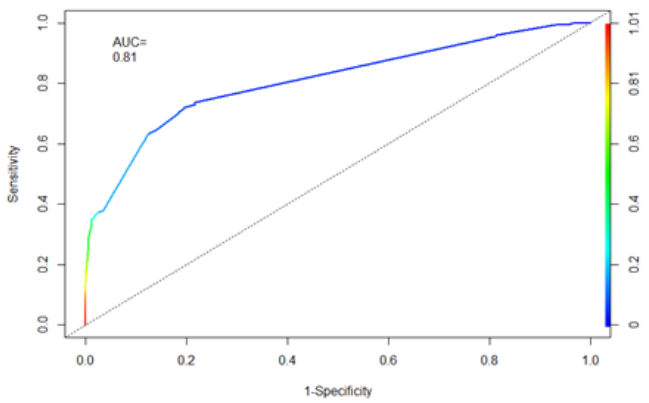
Model	Parameters	AIC	BIC
logit1	29	4010	4225
logit.sep.int	30	4020	4243
stepaic (stepwise regression using AIC)	28	4018	4227
stepbic (stepwise regression using BIC)	16	4049	4171

The BIC for the stepbic is the lowest. Therefore, we accept this model as the best fit for the data. Next, we review the predictive performance of the stepwise models.

Predictive performance

We now apply the stepwise model to the test data and assess model performance.

Figure 94: Comparison of ROC curves for the logit.sep.int, stepaic and stepbic models

Model	ROC	Performance
Logit.sep.int		AUC= 0.82 Cut-off =0.05 Sensitivity = 73% Specificity = 80%
Stepaic		AUC= 0.81 Cut-off =0.05 Sensitivity = 73% Specificity = 80%
Stepbic		AUC= 0.81 Cut-off =0.04 Sensitivity = 74% Specificity = 78%

The model performance metrics are very similar. Both AIC and BIC indicate that model fit is improved by stepwise regression. However, the predictive ability of the logit.sep.int model is not improved.

The stepbic model has a lower BIC and lower probability cut-off compared to the stepaic model. Therefore, we accept the stepbic as the better of the two stepwise model.

The optimal cut-off for stepbic (0.043) corresponds to a sensitivity of approximately 74%. The false positive rate is 22%. The AUC of the stepbic model is marginally less than that of the logit.sep.int model. Of the three models, Logit.sep.int marginally remains best for both goodness of fit and performance.

Clinical implications

The odds ratios of significant coefficients of the stepbic model are as follows:

Table 104: Odds ratios arising from the stepbic model

	OR	2.5%	97.5%
niPRESENT:cleftPRESENT	123.286	11.22045	4733.201
niPRESENT:swallowPRESENT	16.7131	4.11133	115.7255
niPRESENT:trachealPRESENT	15.3899	4.60992	64.1061
cdhPRESENT	6.2876	3.70045	10.4828
renalPRESENT	5.2909	2.30001	11.8039
niPRESENT	5.2762	4.32087	6.4361
skeletalPRESENT	5.1757	2.30617	11.6547
niPRESENT:cardiacPRESENT	4.7338	3.02995	7.4547
oatofPRESENT	4.2071	2.77597	6.2256
cldPRESENT	2.6554	1.85618	3.7503
premPRESENT	2.0564	1.62768	2.5821
cardiacPRESENT	0.6971	0.51518	0.9289
trachealPRESENT	0.2892	0.08848	0.6899
haemPRESENT	0.1594	0.03553	0.4624

We can summarise these odds ratios as follows: the 14 comorbidities listed alter the risk of fundoplication. The first 11 listed increase the risk of fundoplication. Of these factors, the comorbid presence of neurological impairment and cleft lip/palate carries the greatest increase in risk for fundoplication (120-fold). However, cardiac, tracheal and haematological anomalies reduce the risk of fundoplication compared with a child with no co-morbidities.

As the residual vs. fitted plot suggested, there remains some unexplained bias in the models above. This bias may arise from the fact that the event rate (fundoplication) is relatively low in the cohort (1080 of 13902 patients). Therefore, the next strategy will be to apply will be application of bias reduction techniques to counter the effect of rare outcomes.

Improving the model: Bias reduction

Rare events e.g. comet strikes, accidental death, etc., are difficult to study. This is because the predictors are strongly associated with the outcome that is observed far more frequently. The degree of bias is dependent on the absolute number of events in the less-likely category. The lower the number of cases, the greater the bias. Indeed, the proportional rarity of the event matters far less than the sparseness of rare outcome observations.

Analysis of rare events datasets has some inherent problems. Logistic regression in particular is applied to rare events data with some limitations in mind. Maximum likelihood estimation of the logistic model is known to suffer from small-sample bias. Typically, logistic regression underestimates the probability of rare events. A sample may therefore have an event frequency of 0.5%. However, logistic regression will be less accurate for 5 rare events in 1000 compared to 500 rare events in 100000 (16). Secondly, data on explanatory variables are often insufficient and bias may arise due to the practicalities of data collection. For example, if we are studying the outcome of a comet hitting the earth, we have more observations of comets not hitting the earth, than those that do. Therefore, when comets hit the earth, there would be fewer data on fewer variables recorded.

In our cohort, fundoplication is observed far less frequently than no fundoplication. More data on comorbidities are available for patients who have not had fundoplication.

Table 105: Observations on comorbidities are more frequent in patients without fundoplication

	Comorbidities	No comorbidities	
No Fundoplication	6797	6025	12822
Fundoplication	**829**	251	1080
Total	7626	7048	13902

A potential solution to rare events data analysis is two-fold. Firstly, data collection strategies could be adapted to ensure that there are parameter data for rare events. Sampling should take into account the probability that there is less information available for rare outcomes. Secondly, the statistical method should be adjusted for rare event bias in maximum likelihood estimation.

To address the bias in maximum likelihood estimation, Firth (255) suggested the introduction of a counteracting score function. As the sample size increases, the bias increases, but the counteracting term tends to zero. This method is known as maximising penalised likelihood. A full explanation of Firth's method is well beyond the scope of this thesis. However, it is the method applied to address the rare event bias in our data. We have automated the bias reduction process by using the brglm package for R(256–258) designed by Kosmides (2) which is an implementation Firth's penalized likelihood method.

Applying logistic regression with bias reduction

We will use the pruned formula generated from stepwise reduction as the base formula for bias reduction. Stepwise regression has demonstrated which comorbid factors have no predictive value. Therefore, including these 'noisy' but non-contributory variables adds no value to the fit of the model. Following stepwise regression 16 comorbidities were retained and 13 discarded. The formula is restated with the retained variables. We re-run the analysis based on this reduced formula, but using a bias reduction technique. The model is called 'brfit'.

Equation 12: Formula for brfit model

```
fm.reduced<-train$hadfundo~tracheal+cleft+cld+cdh+ni+oatof+swallow+cardiac+prem+haem+
renal+skeletal+ ni*tracheal+  ni*cleft+  ni*cardiac+  ni*swallow
```

Table 106: Coefficients for the brfit model

Call:

brfit<-brglm(fm.reduced, family=binomial(logit), data=train, method="brglm.fit")

summary(brfit)

Coefficients:					
	Estimate	Std. Error	z value	Pr(>|z|)	
ni+cleft	4.0434	1.2293	3.29	0.00101	**
ni+tracheal	2.5821	0.6352	4.06	4.80E-05	***
niPRESENT:swallowPRESENT	2.5686	0.7449	3.45	0.00056	***
cdhPRESENT	1.8389	0.2644	6.95	3.50E-12	***
renalPRESENT	1.6718	0.4154	4.02	5.70E-05	***
niPRESENT	1.6613	0.1014	16.38	< 2e-16	***
skeletalPRESENT	1.6369	0.4099	3.99	6.50E-05	***
niPRESENT:cardiacPRESENT	1.5343	0.2285	6.71	1.90E-11	***
oatofPRESENT	1.4402	0.2044	7.04	1.90E-12	***
cldPRESENT	0.9777	0.1783	5.48	4.20E-08	***
premPRESENT	0.7203	0.1171	6.15	7.80E-10	***
swallowPRESENT	-0.0684	0.6639	-0.1	9.18E-01	
cardiacPRESENT	-0.3521	0.1491	-2.36	1.82E-02	*
cleftPRESENT	-0.8547	0.8323	-1.03	3.04E-01	
trachealPRESENT	-1.1251	0.4795	-2.35	0.01894	*
haemPRESENT	-1.6757	0.5952	-2.82	4.87E-03	**
(Intercept)	-3.2781	0.0665	-49.27	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null deviance: 5253.7 on 9733 degrees of freedom

Residual deviance: 4015.6 on 9717 degrees of freedom

Penalized deviance: 3972

AIC: 4050

Statistical inference

Applying bias reduction, we find that all remaining parameters, except for cleft and swallow anomalies, are significantly correlated with the outcome of fundoplication. A reduction in both the magnitude of the odds ratio as well as the confidence interval is observed. For example, [neurological impairment+ cleft lip/palate] combined is associated an odds ratio of 62 (7-5208) decreased from 123 (11-4733). This narrowing of the confidence interval suggests better fit.

Table 107: Odds and CI for the brfit model

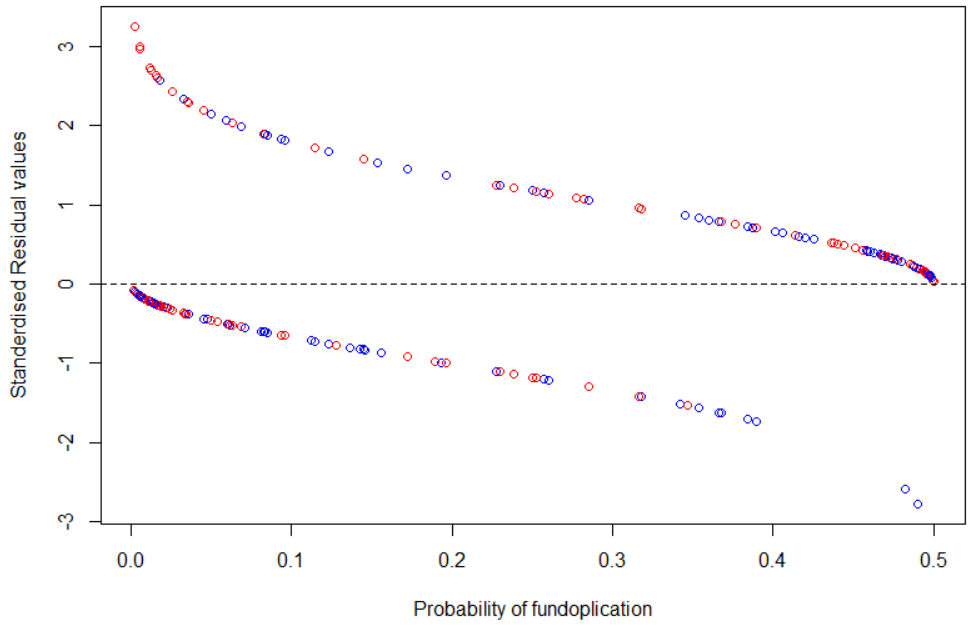
	OR	25%	95%
cleftPRESENT:niPRESENT	62.6	7.7	5208.4
niPRESENT:swallowPRESENT	23.5	6.0	226.4
trachealPRESENT:niPRESENT	13.7	4.2	68.4
cdhPRESENT	6.1	3.5	10.5
renalPRESENT	6.1	2.7	13.2
skeletalPRESENT	5.5	2.4	12.8
niPRESENT	5.4	4.4	6.6
niPRESENT:cardiacPRESENT	4.5	2.9	7.3
oatofPRESENT	4.2	2.8	6.3
cldPRESENT	2.6	1.8	3.7
premPRESENT	2.1	1.6	2.6
swallowPRESENT	0.8	0.1	2.5
cardiacPRESENT	0.7	0.5	1.0
cleftPRESENT	0.4	0.0	1.4
trachealPRESENT	0.3	0.1	0.8
haemPRESENT	0.2	0.0	0.5

Goodness of fit

Residual vs. Fitted plot

Reviewing the residual vs. fitted plot, we observe the absence of points along the horizontal. Compared with the residual plots from the logit.int1, stepglm and stepbic models, this model has a single data cluster rather than two. This is due to removal of perfect predictor variables.

Figure 95: Fitted values vs. standardised residuals for the brfit model



The data are distributed in a curved band about the horizontal across a probability range of 0-.5. There is curvature in the residuals suggesting a non-linear relationship between one or more variables and the outcome. It may also be an indication that there are unmeasured but influential variables.

Information criteria

The brfit model has 16 parameters. Therefore, we expect less of a parameter penalty and hence a smaller BIC compared to the full 29-parameter models. On comparing the models' information criteria, we find the brfit model has a BIC 50 units less than that of the logit1 and logit.int1 models. Therefore, brfit is more parsimonious i.e. greater explanatory power with fewer variables.

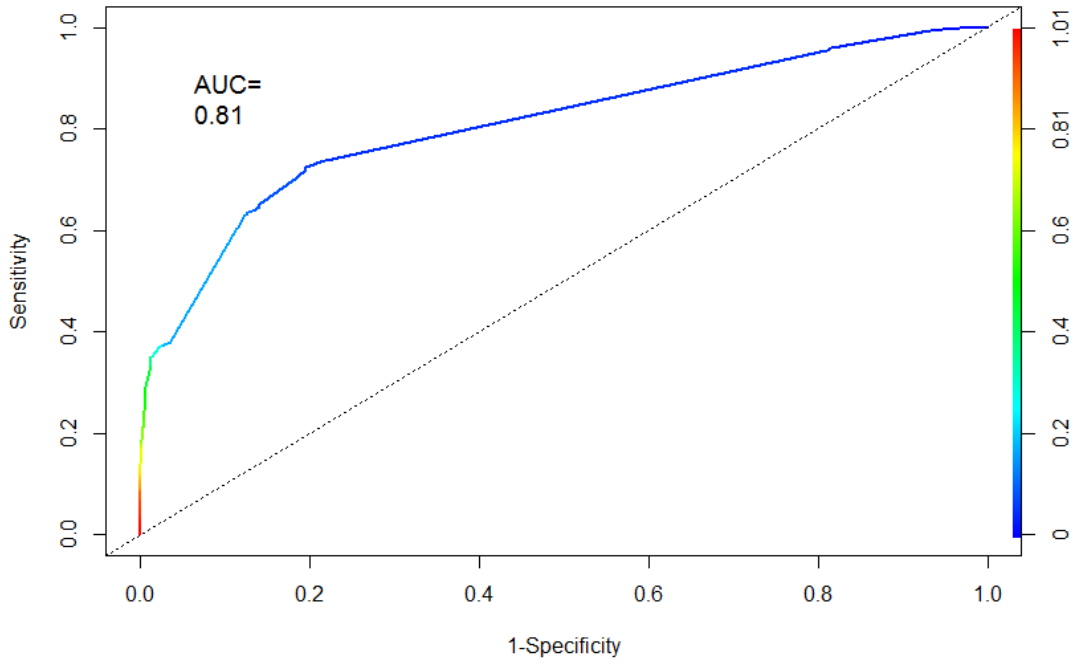
Table 108: Summary of AIC and BIC of logistic regression models

Models	Parameters	AIC	BIC
logit1	29	4010	4225
logit.int1	30	4117	4243
logit.sep.int	30	4020	4243
stepbic	16	4049	4171
brfit	16	4050	4172

Predictive performance

The brfit model has an AUC of 0.81. This indicates that its predictive performance is as good as the stepbic model which had an AUC of 0.82. Therefore, bias reduction does not improve the classification performance of the model.

Figure 96: ROC curve for the brfit model



Clinical implications

The probability cut-off that maximises sensitivity and specificity is 0.06. This corresponds with a sensitivity of 72% and a specificity of 80%. If the probability is >0.06, 72% of patients are correctly assigned to fundoplication. However, this corresponds to a false positive rate of 20%.

Table 109: Confusion matrix and performance statistics for brfit model, probability cut-off 0.06

		Actual fundoplication status	
		True	False
Predicted	True	234	752
	False	89	3093

Statistics	
Accuracy	0.80
95%CI	0.79-0.81
Sensitivity	0.72
Specificity	0.80
PPV	0.24
Negative predictive value (NPV)	0.97

Table 110: Summary of logistic regression models

Model	Description	Parameters	Predictive comorbidities	BIC	Sensitivity (probability cut-off)	False positive rate*	AUC
Logit1	Basic logistic regression	29	15	4225	69%(0.04)	4%	.81
Logit.sep	Excluding variables generating linear separation	22	16	4282	68% (0.06)	17%	.81
Logit.int1	Including interaction terms and excluding variables generating linear separation	30*	18	4243	73%(0.05)	20%	0.82
stepbic	Stepwise regression on reduced model with BIC	16	15	4171	74% (0.043)	22%	0.81
brfit	With bias reduction	16	14	4172	73% (0.06)	20%	0.81
*8 interaction terms, 22 independent variables							

Summarising logistic regression models

The results of the logistic regression are summarised in Table 110 above. For the analysis of predictive performance, a probability cut-off of 0.05 is used to estimate sensitivity and specificity for all of the models. The strategies used to improve the predictive performance of the model improved sensitivity from 69% in the original model to 74% in the final model. However, there was a commensurate increase in the false positive rate from 4% to 20%. Despite optimisation, model AUC remained static at 0.81-2.

The brfit model has the highest sensitivity for the highest probability when applied to the test data. Compared to the stepbic, there is less risk of failure of linear separation as the probability cut-off is higher. It shall be accepted as the best specification of variable interactions when logistic regression is applied.

Our various strategies have not improved the predictive performance of the models. However, we have managed to prune the regression statement and identify key comorbidities.

The models can also be summarised based on which comorbidities that consistently influence fundoplication risk. These are:

1. Neurological impairment
2. CDH
3. CLD
4. Prematurity
5. OATOF
6. Renal anomalies
7. Skeletal anomalies

We have also identified some key interactions:

1. NI and cleft anomalies
2. NI and tracheal anomalies
3. NI and swallow disorders
4. NI and cardiac disease

Lastly, we have identified some comorbidities that appear to reduce the risk of fundoplication i.e. haematological and cardiac disease and the presence of tracheal anomalies.

These findings can inform clinical practice. During history taking, the surgeon should identify whether these comorbidities are present and active.

The most sensitive model (brfit) delivered a false positive rate of 20%. Clinically, relying on this model would result in unnecessary fundoplication for 1 in every 5 patients receiving the procedure. Therefore, it is important to acknowledge that the logistic approach does not fit the data sufficiently well. This is probably due to insufficient specification of the model, i.e. some variance is due to variables that are not in the model.

In the next section, we apply alternative modelling approaches.

DECISION TREE MODELLING

A decision tree (DT) is the result of a learning algorithm used for classification tasks. Metaphorically, the complete data set is at the "root" of the "tree". The data are serially subdivided according to an established rule or principle. This creates the branches of the "tree". Eventually, the classification process defines branches, twigs then leaves that describe data of a particular class or outcome. This progressive splitting is known as recursive partitioning, continues until a stop criterion is reached.

Take a simple task e.g. separating oranges from a boxful of apples and oranges. The classification task begins by identifying features that best predict the target class i.e. oranges. Features include:

1. Colour- oranges are yellow, apples can be green, pink or red
2. Taste: oranges are tart, where apples are sweet

Some features are more distinguishing than others. For example, colour clearly distinguishes orange oranges from green, pink or red apples. However, both apples and oranges can take a round shape.

It is most efficient to begin classification using the most defining features. The application of classification features or rules is known as classifying heuristic. The successful heuristic should result in a pure selection of the target class. An efficient heuristic achieves this in the fewest subdivisions possible. Therefore, the performance of a classifier can be measured by its ability and efficiency in achieving this 'purity'.

Advantages and disadvantages of decision tree modelling

The advantages to utilising DT modelling are:

- Intuitive: This procedure mimics the decision-making steps that a surgeon might use. We start out with a large sea of information. We ask broad questions e.g. age of child, male or female. We then hone our questions e.g. what are the comorbidities, what are the daily symptoms? The algorithm-building process and output are more intuitive and applicable than formulae generated from a regression analysis.

DTs utilise only the most important features. This enforces a standard approach to standard problems. When a clinician encounters a patient who is not represented in the algorithm, they perceive that this patient is an outlier. Therefore, the clinician assigns the patient to a class (treatment/ no treatment), understanding that the outcomes seen for classifiable patients cannot be guaranteed for non-classifiable patients.

- Interpretability: The results of a DTs are presented as an algorithm. This is a simple way of transmitting learned information that does not require the statistical knowledge of the underlying rules. Algorithms are commonly used in clinical practice to communicate guidelines and pathways. They are a useful medium for evidence based medicine.
- Suitability to data: This approach is suited to data with minimal continuous and nominal characteristics. Classification tasks are often biased by features with numerous levels. Our GOR database contains comorbidity data and outcome data that are binary. Therefore, our data set is particularly suited to analysis by classification tree.

DT modelling offers a flexible approach to the data. Unlike linear or logistic regression, both continuous and categorical can be included in a single model. Furthermore, little data pre-processing is required.

- Heuristics: The model can be modified to assign cost and weight. In this way, real-world knowledge or intuition is introduced in the construction of the model.
- Performance metrics: Like the regression approach trialled previously, the performance of DT algorithms is measurable. Metrics include classification accuracy, sensitivity and false positive rates.

There are some disadvantages to using DTs.

- Trees are high-variance classifiers. Small variations in the data may result in a completely different DT. This is due to the hierarchical nature of the process. An incorrect split at the top of the tree carries forward and alters the structure of the rest of the tree. This can be mitigated through 'bagging' i.e. using an ensembles of trees to identify an 'average' tree structure. This method will be discussed later.
- Over-fitting: it is possible to create complex DTs that the fit majority of the data, but do not generalise the data well. This is mitigated by pruning procedures.
- The model is heavily dependent on the heuristic decisions made at each node. Again, this limitation can be mitigated using ensembles.
- Large trees lead to complex algorithms can be difficult to interpret and hence are poorly applicable.

Metrics of classifying heuristics

Entropy

Entropy is a measure of the degree of disorder in a system. Therefore, it is an inverse measure of purity. An entropy of 0 corresponds with purity. A sample with entropy of 1 contains the maximal amount of disruption. For example, if we have 2 classes (apples, oranges) and an equal number of apples and oranges, a 50:50 split represents the maximal amount of entropy(259).

Entropy is calculated using the following formula:

Equation 13: Entropy

$$Entropy\ (S) = \sum_{i=1}^{c} - p_i \log_2(p_i)$$

where

- S is a segment of data
- c is the number of different classes
- p is the proportion of values that fall into a target class

For our apples and oranges classification task, where the proportion of apples to oranges is equal, we can calculate the entropy of a 50:50 split.

$Entropy\ (S) = \sum_{i-1}^{c} - p_i \log_2(p_i)$

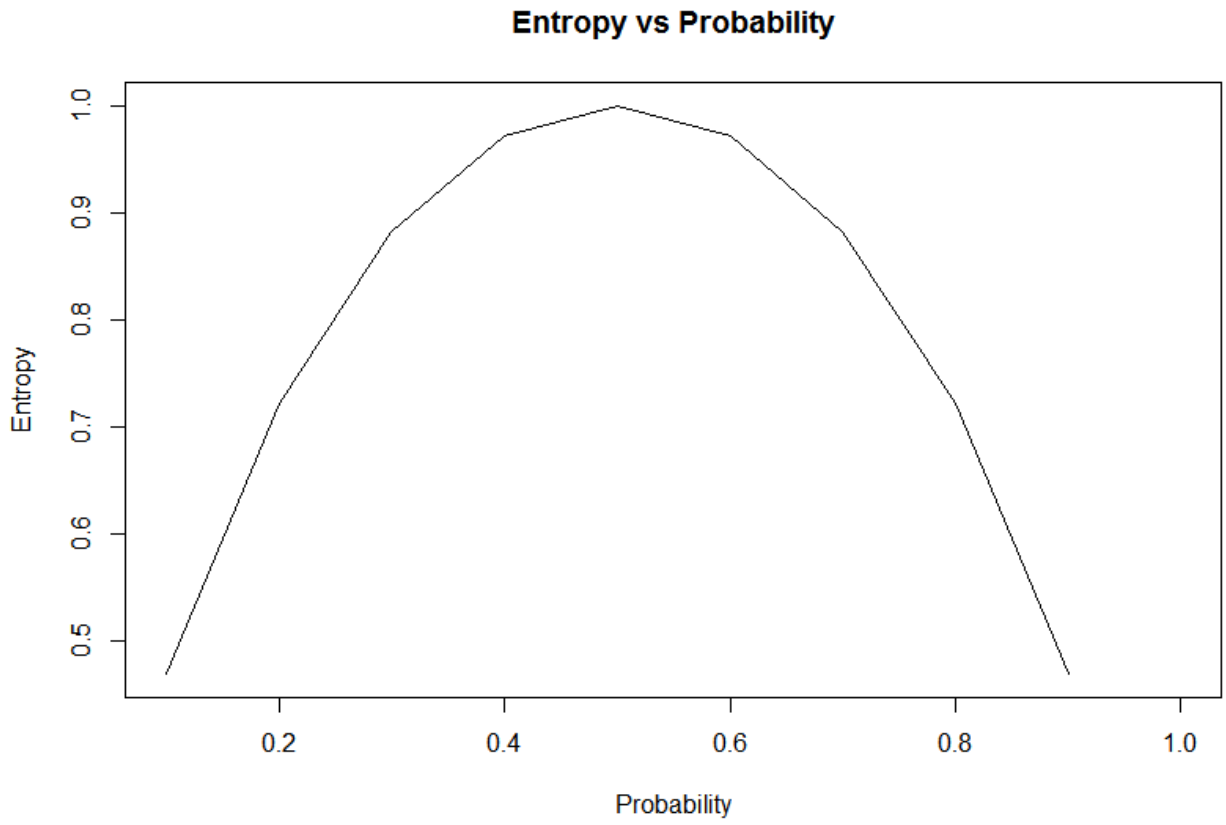
*Entropy (S)= -0.5*log2(0.5) -0.5*log2(0.5)*

Entropy (S)= -0.5(-1) –(0.5(-1))

Entropy (S)= 0.5 +0.5=1

Entropy is also a measure of certainty. How certain can we be that the target variable is homogenous in the subset generated after a split? This certainty can be expressed as probability. Purity (entropy=1) is equivalent to a sample containing all (probability=1) or none (probability = 0) of a particular class. Misclassification risk for pure samples is 0. A probability of 0.5 corresponds to maximal entropy(252) and maximal misclassification risk. Therefore, the entropy/probability function is symmetrical (Figure 97).

Figure 97: Entropy/probability function

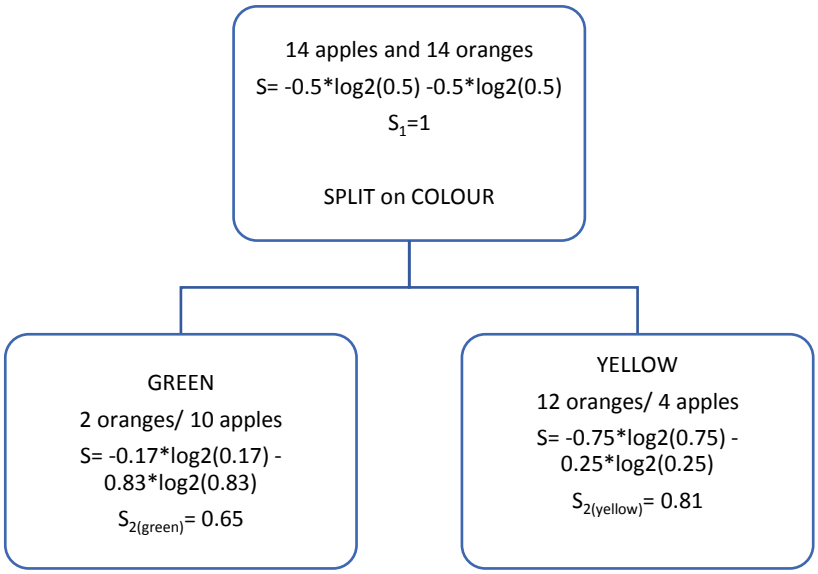


In classification tasks, the learner approaches the data with the aim of calculating the 'best' split. Therefore, decreasing the degree of disorder (i.e. entropy) with each split can be used as a split criterion. Splitting procedures based on entropy utilise the 'information gain' metric (254).

Information gain

In a series of classifications, a reduction in entropy with each split is used as a criterion for each forward step of the algorithm. For example, our boxful of fruit may be split by colour or by taste. First we split the box using colour:

Figure 98: Splitting a boxful of fruit on colour criteria



Each child node's entropy is weighted with the proportion of observations falling into that subset. The entropy after the split (S_2) is the average of entropy of the child nodes.

$$S_2 = (0.65 + 0.81)/2_{\cdot} = 0.73$$

Equation 14: Information gain

$$Information\ Gain\ = S_1 - S_2$$

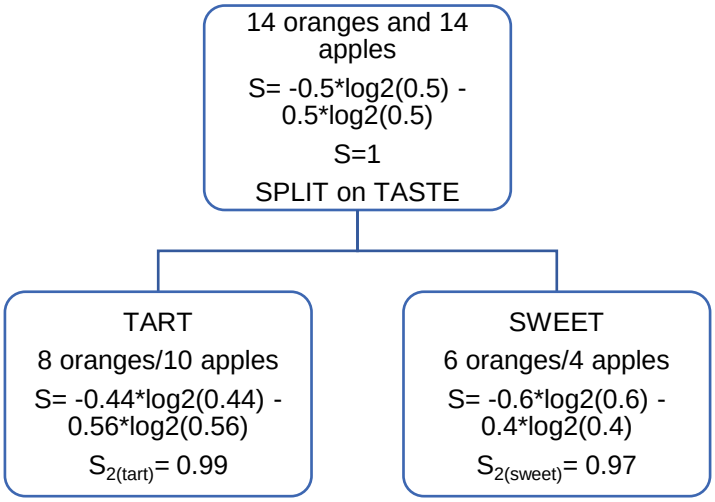
The difference in entropy before (S_1) and after (S_2) the split is known as information gain(260).

$$Information\ Gain\ = S_1 - S_2$$

Information gain (colour) =1-0.73=0.27

If instead we split on the taste of the fruit, we find:

Figure 99: Splitting a boxful of fruit using taste criterion



The information gain based on this split is:

Information gain (taste)=1-((0.99+0.97)/2)=0.02

In a series of partitions, the more the information gain at each split, the fewer the splits required. Therefore, the criterion with the highest information gain is the preferred (261). In this illustration, colour has higher information gain and is the better criterion.

LIMITATIONS OF INFORMATION GAIN

For 2-class problems there are limitations of portioning on information gain. Splitting can result in information gain ties. Ties arise where the information gain for either branch is the same. Breaking the tie may require real-world information or heuristics. For example, a tie between tart and sweet apples may be solved by other information e.g. tart apples have longer shelf life. For this thesis, ties will be broken randomly.

Furthermore, balance-variance trade-offs must be considered when dealing with multi-class (>2 target classes) problems. For example, take a model that resolves to single observations. Each terminal node is a singleton and singletons, by definition, are pure. Misclassification risk in each terminal node is zero. However, the resulting model is not generalizable. Such a model would perform poorly when exposed to test data. Therefore, information gain as a criterion can lead to overfitting i.e. give greater emphasis for minor features.

Gain ratio

To counter the limitations of information gain, we require a parameter that penalises small subsets. Such a parameter was described by Quinlan (241,260).

Gain ratio is derived from information gain. It is designed to reduce bias towards highly branching attributes. A parameter, intrinsic information, is used to standardise gain ratio. Intrinsic information is a measure of the number of splits. Given two splits with the same information gain, the gain ratio will be lower for the split generating more subsets.

$$Intrinsic\ information = -\sum \frac{S_i}{S} log2 \frac{S_i}{S}$$

$$Gain\ ratio = \frac{Information\ Gain}{Intrinsic\ Information}$$

In addition to information gain, there are two other commonly used node impurity measures. These are misclassification error rate and Gini index.

Gini impurity index

The Gini impurity index (GII) is a measure of error in the training data set. It is calculated by summing probability of an observation being correctly classified and multiplying this by the probability of misclassification.

where,

- k- is the class of interest
- p is the proportion of observations that have class k
- m is a terminal node belonging to the region R on the feature space
- N is the number of observations in a particular node

Equation 15: Gini impurity index

$$Gini = \sum_{k \neq k} p_{mk} * p_{mk}\,!$$

$$= \sum_{k \neq k} p_{mk}(1 - p_{mk})$$

In a 2-class problem, where a terminal node is pure and contains N observations of class k, the probability of y=k being correctly classified if assigned to this node is 1.

$$p(y = k) = 1$$

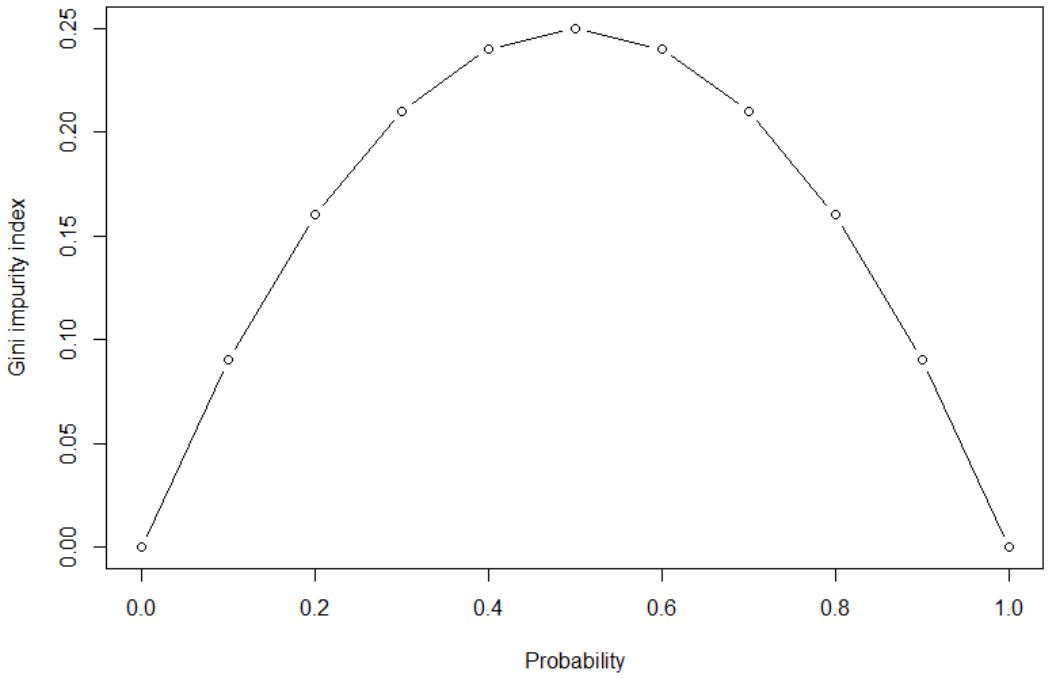
If the node does not contain any observations with class k, then

$$p(y \neq k) = 0$$

$$Gini_{y \in Nmk} = (0 * 1) = 0$$

Therefore, GII ranges between 0 and 1. It reaches its minimum 0 when purity is achieved. It is therefore a parabolic function similar to information gain.

Figure 100: Parabolic function of Gini impurity index



When a split is made on a variable, the GII for the two daughter nodes should be less than the GI for the parent node. The higher the difference, the greater the importance of the variable. Variables therefore can be ranked in order of importance based on GII.

GII and entropy change similarly with probability. Both relate to probability with a parabolic function. The GII, like entropy, is maximal with a 50:50 split i.e. the maximal amount of disorder in a 2-class system. However, GII is a measure of classification error whilst entropy is a metric describing sample distributions within a population. GII can therefore be used as a criterion to minimise classification error.

Misclassification error rate

Another useful parameter for classification tasks is the misclassification error rate. Again, we have a 2-class problem, where

Equation 16: Misclassification error

$$misclassification\ error = \frac{1}{N_m} \sum_{i \in R_m} I\,(y \neq k(m))$$

- k- is the class of interest, one of many in a feature space
- p is the proportion of observations that have class k
- m is a terminal node belonging a region in a feature space (R)
- N is the number of observations in a particular node
- R is the feature space / sampled population

Then misclassification error is the probability that an outcome y is wrongly classed as k

Unlike the Gini index and information gain, the misclassification rate is not sensitive to the probability of events in splitting nodes. Therefore, it is not useful in identifying split criteria. However, it is a useful metric for pruning trees. This is done using the 'complexity parameter' which will be described below.

Pruning

The complexity of the tree describes its size i.e. both the number of splits and number of terminal branches. How large should a tree be grown? A large tree will over-fit the data. However, a small tree may fail to capture important structure. To mitigate this, the modeller may choose to 'prune' the algorithm tree based on various heuristics. Strategies include:

1. Limiting the number of observations in terminal branches of the tree.
2. Limiting the number of splits or branches.
3. Allowing splits that meet a threshold e.g. entropy reduction, misclassification risk or information gain. This 'look forward' rule has limitations. There will be situations where a split criterion that does not meet the threshold 'set up' for a subsequent high value split.

Classification trees can be optimised to all three heuristics using 'complexity parameter' and a process of cross-validation.

We define

- C is the cost of growing a tree i.e. the larger the tree, the greater the cost.
- α- the number of observations in a terminal node
- T- a tree
- $|T|$- the number of observations in a terminal node

The misclassification risk of T is

$$R(T) = \sum_{i=1}^{k} P(T_i)R(T_i)$$

We define a complexity parameter (CP) α that lies between 0 and 1. When α is small, the tree has more numerous splits and terminal nodes. When α is large, the tree has fewer splits, and therefore fewer terminal nodes. Therefore α measures the 'cost' of adding a new variable or splitting criterion to the tree. Trees with a small α risk overfitting the data and trees with a large α risk under-fitting the data.

$$C_\alpha(T) = C(T) + \alpha T$$

$R(T_0)$ is the risk for zero split tree i.e. the root node. The CP is a measure of the change in misclassification risk when a node is split. A threshold for CP can be set. This threshold is effectively a stopping criterion. If the CP for a particular split is less than the threshold, a split is not attempted. Cross-validation i.e. running multiple trials using various splits, can be used to identify the misclassification risk with each split. Misclassification risk is then used to optimise the complexity parameter.

Basic decision tree modelling

Step 1: Importing and processing the data

The following R packages were utilised in this analysis are detailed in the appendix, section IV.

For this section ‘fundoplication’ will be denoted F . Conversely, ‘no fundoplication’ will be denoted NF.

Step 2: Building the classifier

The first DT model was based on a saturated formula including all the comorbid variables available. This model is called dt.raw.

Equation 17: Formula for the dt.raw decision tree model

```
dt.raw<-rpart (fundo~tracheal+cleft+sleep+trachy+aspiration+cld+cdh+asthma+cardiac+ni+oatof+
achalasia+swallow+chrom+consang+prem+haem+endocrine+renal+skeletal+bone+dental+
oncology+metabolic+aresp+immune+cardsurg+eb+ simplegord, data=train)
```

For this initial analysis dt.raw, we fit a DT model to the training data. By default, the Rpart algorithm utilises the GII to select the splitting criterion. We review the output as a DT output table summarising each split. The features of the tree are:

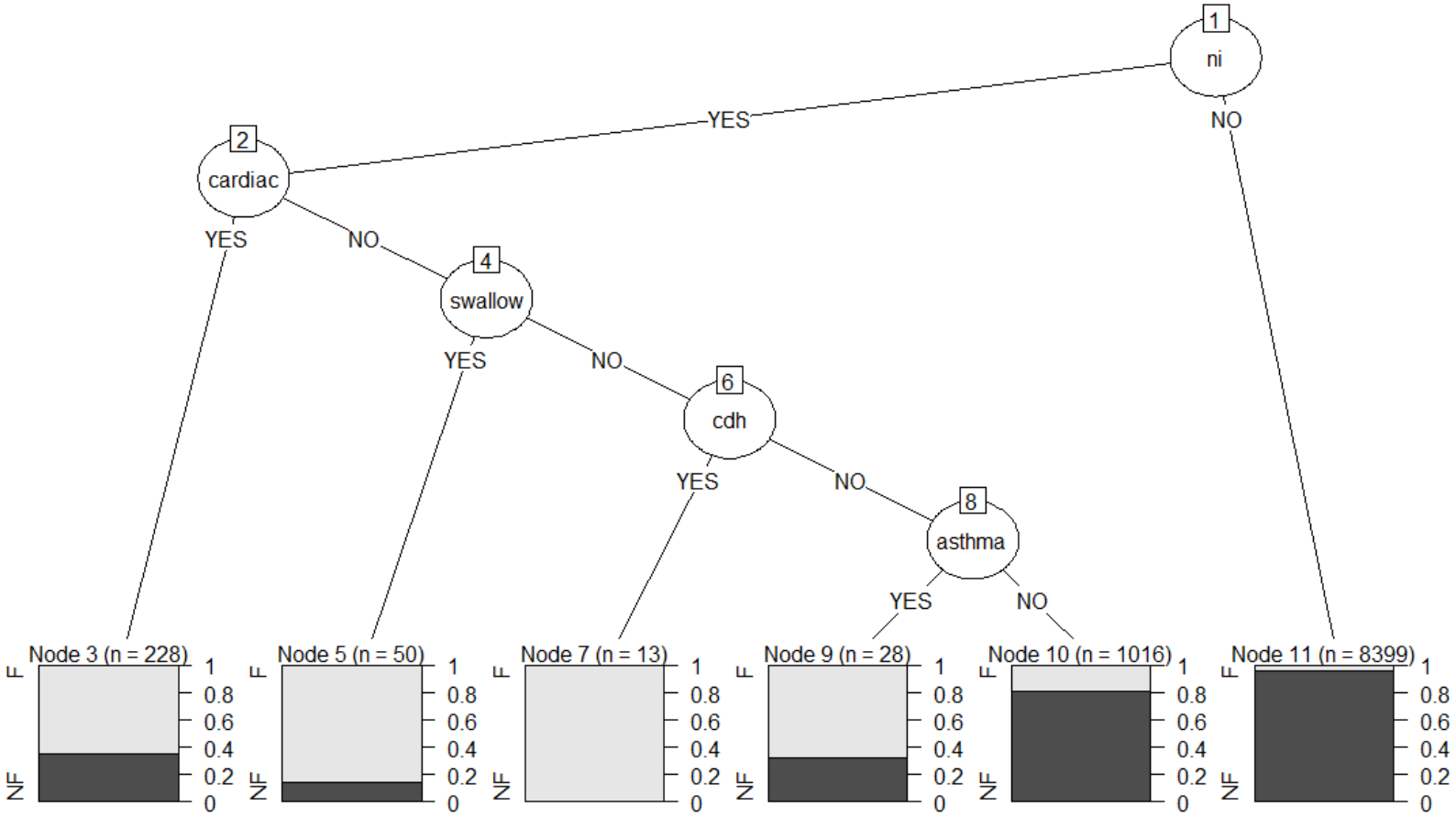
- A root with n=9734 observations
- Splits: the first split is row [2] and is based on the variable ni.
- Terminal nodes: row [3] is a terminal node. There are n=231 observations in this node. The error /misclassification rate is 35% i.e. 65% patients who had fundoplication [F] and 35% did not . Row [13] is also a terminal node containing n=8381 observations. In this node 96% of patients did not have fundoplication [NF].

Figure 101: Decision tree output table for dt.raw

```
Fitted party:
[1] root
|   [2] ni in PRESENT
|   |   [3] cardiac in PRESENT: F (n = 231, err = 35%)
|   |   [4] cardiac in ABSENT
|   |   |   [5] swallow in PRESENT: F (n = 58, err = 21%)
|   |   |   [6] swallow in ABSENT
|   |   |   |   [7] cdh in PRESENT: F (n = 16, err = 0%)
|   |   |   |   [8] cdh in ABSENT
|   |   |   |   |   [9] asthma in PRESENT: F (n = 11, err = 0%)
|   |   |   |   |   [10] asthma in ABSENT
|   |   |   |   |   |   [11] cleft in PRESENT: F (n = 8, err = 0%)
|   |   |   |   |   |   [12] cleft in ABSENT: NF (n = 1029, err = 18%)
|   [13] ni in ABSENT: NF (n = 8381, err = 4%)

Number of inner nodes:    6
Number of terminal nodes: 7
```

An algorithm chart is a more intuitive visualisation of the DT and shall be used henceforth.

Figure 102: Decision tree dt.raw



Key

F= Fundoplication NF= No fundoplication YES = comorbidity is present NO= comorbidity is absent

Step 3: Statistical inference

The key observation is the primacy of NI. If NI is absent, fundoplication is almost never indicated. If NI is present with cardiac disease, then fundoplication is often performed. NI in conjunction with disordered swallow, CDH, asthma and cleft palate also predispose to fundoplication. This initial analysis classified most examples correctly. There were 27 patients wrongly assigned to fundoplication, and 557 patients wrongly assigned to no fundoplication. This yields a misclassification rate of 6%. We measure the predictive performance of the model against test data.

Step 5: Predictive performance

We import the test data, a random selection of 4168 examples. We then use the dt.raw model to predict the *class for examples in the test data.*

Table 111: Confusion matrix and performance metrics for dt.raw DT model

		Actual fundoplication status	
		True	False
Predicted	True	90	233
	False	41	3804

Statistics	
Accuracy	0.93
Sensitivity	0.28
Specificity	0.99
PPV	0.69
NPV	0.94

The sensitivity of the dt.raw model is very low at 28%. The clinical implication is that only 28% of patients are correctly identified as requiring fundoplication. Specificity is high at 99%. This yields a false positive rate of 1% i.e. only 1 of 100 patients are wrongly assigned to fundoplication.

The error rate is only 6.6%. A model that assumes that no patients require a fundoplication would only misclassify 6% of observations. However, the event rate in the data is also low i.e. 6% . Therefore, a model assuming no fundoplication would only err by 6%.

Optimising the classifier through pruning

Increasing the number of splits can reduces error in the training set as the model fits the data exactly. However, when applied to test data, error increases with each split as the model is highly biased to the training data. This observation can be used to estimate the optimal number of splits. The model is trained to minimise test error. This is a process known as cross-validation.

The first step is to permute the training data randomly. This ensures that each subset will comprise a different population. Then, the data is split into 10 subsets.

In the next step, the classification task is carried out on the 9 of the 10 subsets. This results in 9 models. The tenth subset (Gi) is then used to cross-validate the resulting 9 models. We obtain a class prediction for each observation in Gi using each of the 9 models, resulting with nine predictions. These 9 predictions are averaged resulting in a probability of a correct prediction. This probability is an estimate of misclassification risk (beta). It is also known as the complexity parameter (CP). The smallest risk of error i.e. the smallest CP corresponds to the 'best' model. The number of splits corresponding to the smallest CP is optimal for minimising misclassification risk.

The misclassification error is dependent on the number of classes and the number of observations within each class. The root node error is the probability of being classified NF when your status is F before the first split.

In our dataset, the fundoplication rate is 6%. Therefore, the root node error for the whole data set is 0.06. There are two other error parameters of interest i.e. relative and cross-validation / absolute error. Relative error (rel error) is a factor of the absolute cross-validation error (xerror).

$$Relative\ cross-validation\ error * Root\ node\ error = Absolute\ cross-validation\ error$$

Visualising the relationship between CP and relative error results in a characteristic curve. As the CP decreases from infinity (x axis), the cross-validation (xval) error rate (y axis) initially drops from 1, plateaus at a minimum relative error before starting to rise again. This pattern reflects the change in error rate when moving from an under-fitted to an over-fitted tree. The best split is obtained by the tree whose error is the minimum error + 1 standard deviation (xstd). This value is known as the 'error level' and is plotted on the as a horizontal line on the CP vs x-val error rate axes.

$$Error\ level = minimum\ error + 1\ standard\ deviation$$

There may be several tree sizes represented below the error level and along a cross-validation error plateau. To break such a tie, we review the standard deviation of the tied trees. The tree with the smallest standard deviation is selected as the tree with optimal splits.

Pruning the tree using complexity factor

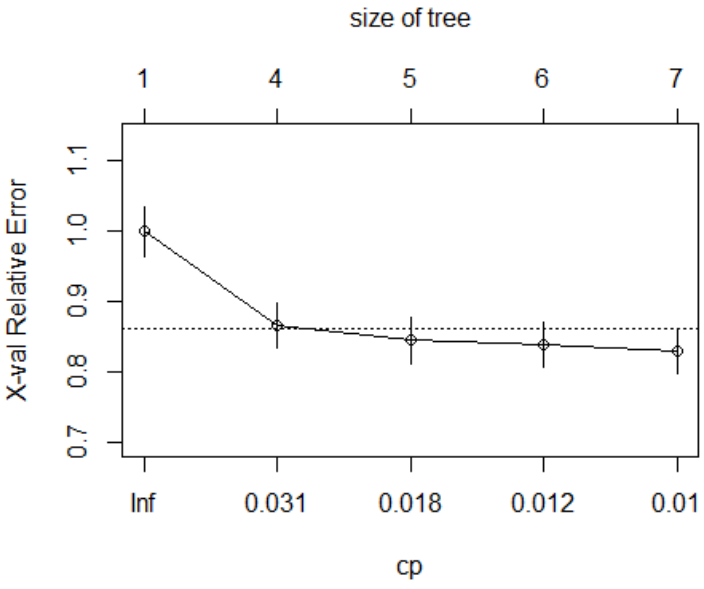
Using a training set of containing 9734 observations, we construct a cross-validation tree. The pruned tree is called dt.pruned.

Figure 103: Using cross-validation to identify optimal tree size

```
Variables actually used in tree construction:
[1] asthma  cardiac cdh     cleft   ni      swallow

Root node error: 757/9734 = 0.078

n= 9734

      CP nsplit rel error xerror  xstd
1 0.044      0      1.00   1.00 0.035
2 0.021      3      0.87   0.87 0.033
3 0.015      4      0.85   0.85 0.032
4 0.011      5      0.83   0.84 0.032
5 0.010      6      0.82   0.83 0.032
```

In the training subset sampled here, the root node error is 0.08. The relative cross-validation error decreases from 1 to 0.82 as the CP decreases from infinity to 0.01 (Figure 104).

Figure 104: Relative cross-validation error decreases and plateaus on a minimal error rate as complexity parameter decreases from infinity



The minimum error appears to correspond to a CP of 0.01. This corresponds to 6 splits and a tree size of 7 nodes. We can now prune the table to this CP value.

We find that there remain 6 inner nodes and 7 terminal nodes. In fact, the pruned tree is plotted and is identical to the unpruned tree dt.raw (Figure 102, p291).

The algorithm's default number of cross-validation runs is ten. If we ran the tree many more times, would get different results? We re-run the algorithm using 50 cross-validation trials.

The result is much the same with inner nodes and 7 terminal nodes. Furthermore, we can identify the number of splits for each cross-validation run.

Table 112: Number of trials per split when cross-validation runs are set to 50

Splits	Number of trials
5	19
6	31
Total	50

We find that 19 of 50 runs resulted in an optimal split of 5. There were 31 runs that recommended a split of 6. There is no great variation in the number of recommended splits. This suggests that the splitting variables of importance have been identified and are contained in the model. Presenting further variables (increasing complexity) does not reduce misclassification error.

We test the performance of the pruned model against the test data we find that there is no improvement in predictive performance.

Table 113: Confusion matrix and performance metrics of dt.pruned

		Actual fundoplication status	
		True	False
Predicted	True	90	233
	False	41	3804

Statistics	
Accuracy	0.93
Sensitivity	0.28
Specificity	0.99
PPV	0.69
Negative predictive value (NPV)	0.94

Pruning the model using the complexity parameter has made no improvements to sensitivity or accuracy. The dt.raw model performs well at classifying patients who do not need fundoplication. However, it performs poorly when assigning patients to fundoplication.

Optimising the classifier using random forest

Pruning using cross-validation by complexity parameter did not improve the model. If the model was susceptible to cross-validation errors, one strategy to overcome is application of ensemble technique.

DTs are high-variance classifiers. This means that perturbing the data by a changing the attributes of a few observations can result in a completely different tree being constructed. The DT algorithm optimises the tree to best fit a particular data set. Therefore, when exposed to a second dataset that varies from the training dataset, the algorithm is biased towards the training data, resulting in misclassification errors.

This is a bias-variance tension that can be mitigated by training multiple trees (i.e. an ensemble) on multiple data subsets of data. The sampling of each subset is random. The collective term for an ensemble of trees is a forest. Therefore, this ensemble technique is appropriately named a 'random forest'.

The random sampling technique used to create data subsets is 'bagging' i.e. bootstrapping and aggregation. This technique was discussed in Section IV, Introduction (3 p. 301). In summary, observations are randomly selected with replacement from the main dataset. A second element of randomisation is found in the random forest splitting procedure. Each tree is presented with a random selection of split criteria. For example, we have 29 variables that could be used for splitting at each node. We can 'instruct' the algorithm to utilise only 3 randomly selected split criteria at each node. Therefore, each tree in the forest will be randomly different to the others. The output of a random forest is an aggregate classification result i.e. the average class probability for an observation.

Like simple DTs, individual trees in random forests maximise information gain. The Gini index of impurity can also be used as a 'best split' metric. For each tree, the decrease in Gini index associated with each split criterion is estimated. For the ensemble, the Gini index for each split criterion can be averaged out. The split criteria can then be ranked according to importance based on the mean decrease in Gini index.

Growing a random forest

A saturated formula is used for the random forest model i.e. dt.rf

Equation 18: Formula for dt.rf random forest DT model

```
dt.rf<- randomForest(fundo~tracheal+cleft+sleep+trachy+aspiration+cld+cdh+asthma+cardiac+ni+oatof+achalasia+swallow+chrom+consang+prem+haem+endocrine+renal+skeletal+bone+dental+oncology+metabolic+aresp+immune+cardsurg+eb+ simplegord,
        data=train)
```

We automate this process using the R package "randomforests"(262). We used the default number of DTs i.e. 500. There were 5 variables tried at each split. The criterion for choosing a split variable is the Gini index of impurity.

Figure 105: Implementation of a random forest model

```
Call:
 randomForest
        (formula = fundo ~ tracheal + cleft + sleep + trachy + aspiration + cld + cdh + asthma + cardiac +
        ni + oatof +      achalasia + swallow + chrom + consang + prem + haem + endocrine + renal +
        skeletal + bone + dental + oncology + metabolic +    aresp + immune + cardsurg + eb + simplegord,
        data = train)

Type of random forest: classification

Number of trees: 500
No. of variables tried at each split: 5
OOB estimate of  error rate: 6.06%
```

The out-of-bag error rate (OOB) is the 'majority vote' collated from all trees. It is a useful reckoning parameter. For the dt.rf model, the OOB indicates that 6% of observations would be classified NF, where F was the correct classification.

Table 114: Confusion matrix and performance statistics for dt.rf model

		Actual fundoplication status	
		True	False
Predicted	True	211	546
	False	44	8933

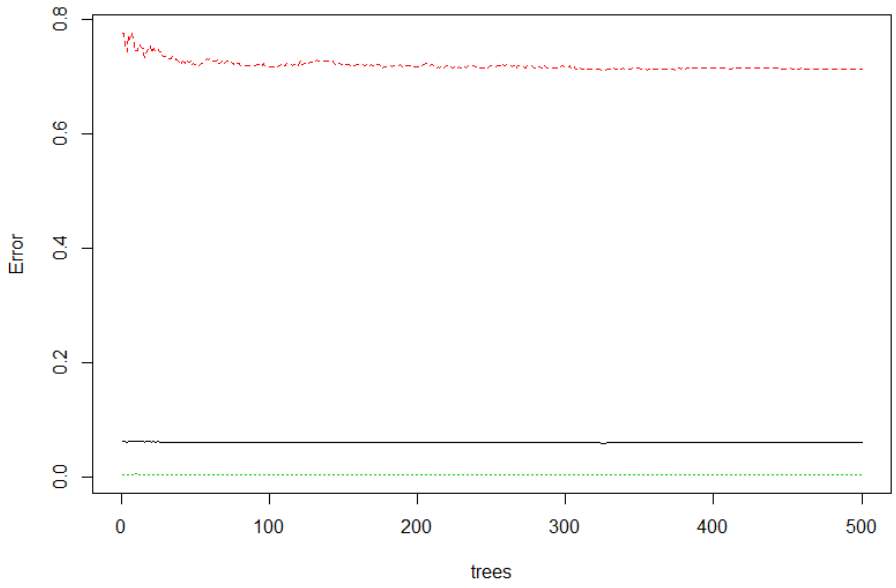
Statistics	
Accuracy	0.94
Sensitivity	0.28
Specificity	0.99
PPV	0.85
Negative predictive value (NPV)	0.94

The most prominent feature from the confusion matrix is the increase in the proportion of true positive classifications i.e. patients assigned correctly to fundoplication. We therefore expect an increase in both sensitivity and accuracy when performance is measured against the test data.

Using a random forest approach improved the sensitivity of the model only marginally from 27.8 to 28.5%. The model was marginally more accurate at 94% (was 93%). This corresponded to the event rate in the data. The model remained highly specific with a false positive rate of <1%.

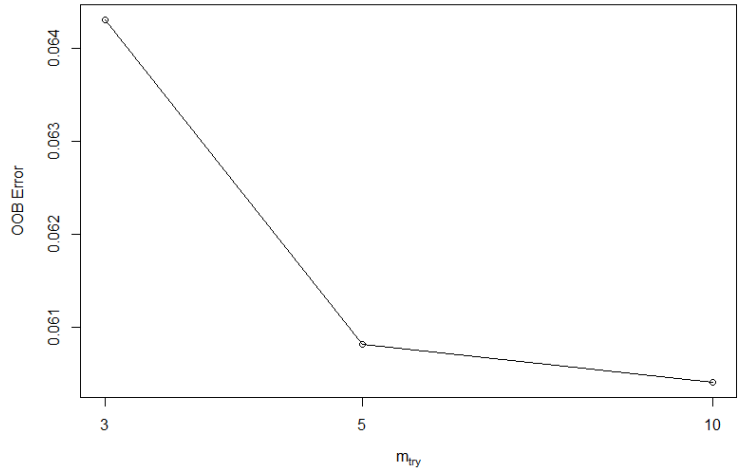
It was perhaps surprising that increasing the number of trees and randomising subsets of observations and splits did not improve model performance. However, reviewing the relationship between the error rate and the number of trees provided an explanation.

Figure 106: Error rate response as the number of trees in the ensemble is increased



We can see from the number of trees versus error rate plot that there is indeed a decrease in error rate between 1-50 trees. However, beyond 100 trees, the error rate is stable.We can also use the number of split criteria presented at each split as a forest tuning parameter. The out-of-box error rate varies as the number of split criteria is varied. We utilise this to identify the optimal number of split criteria that minimises the OOB. The algorithm searched for the OOB error for split criteria ranging from 3-50. We found that increasing the number of split criteria presented resulted in a marginal decrease in the OOB from 6.08% to 5.96%. This decrease in OOB error was achieved at a nadir of 10 split criteria.

Figure 107: Out-of-box (OOB) error rate falls as the number of split criteria at each node is increased



Therefore, we regenerated the random forest, this time presenting 10 randomly selected split criteria at each node. This model is called dt.rftuned.

The confusion matrix for the resulting random forest is presented below.

Table 115: Confusion matrix and performance statistics for dt.rftuned

		Actual fundoplication status	
		True	False
Predicted	True	92	231
	False	18	3827

Statistics	
Accuracy	0.94
Sensitivity	0.28
Specificity	0.99
PPV	0.84
NPV	0.94

Comparing the basic versus tuned random forest, the difference in sensitivity, specificity and accuracy was negligible. The classifier sensitivity was found to be low at 28.5%. This classifier would not be useful in identifying children who need a fundoplication. The specificity was very high suggesting that the classifier would be useful in identifying children who do not need a fundoplication i.e. most children.

We ranked the importance of each variable using mean decrease in Gini index. The top five variables are summarised in Table 116 below. Neurological impairment was by far the most important variable.

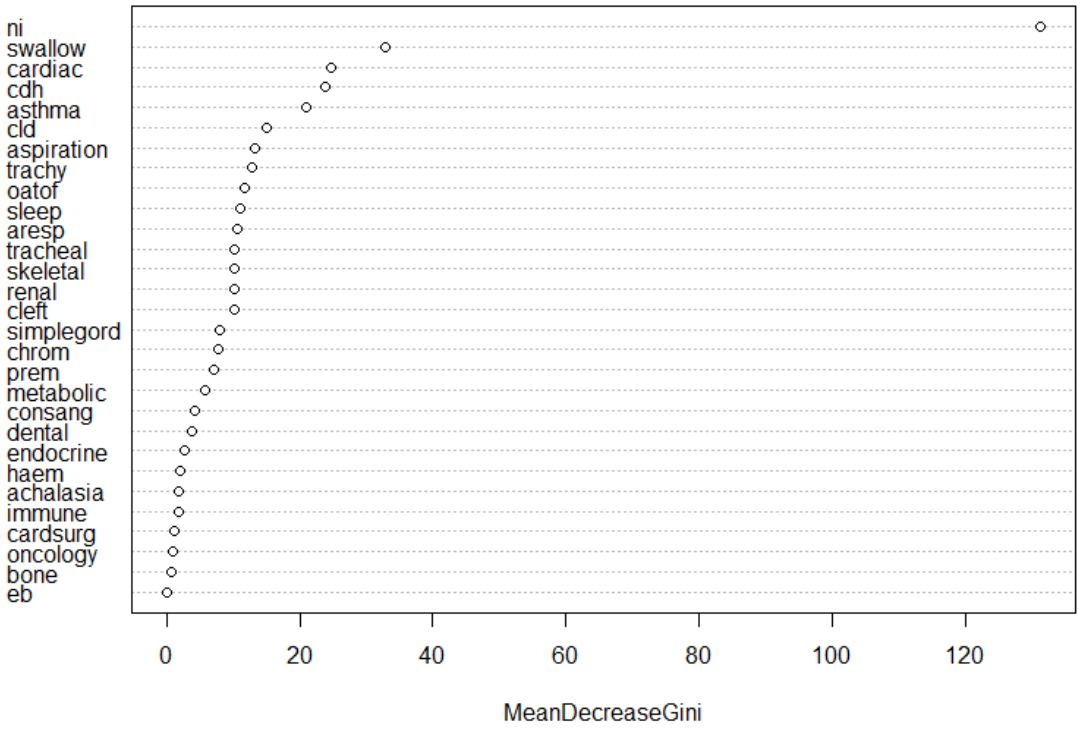
This was followed by swallowing disorders, cardiac disease, CDH, asthma and CLD. Notably, these variables were those included in inner nodes in prior DT plots.

Table 116: Variable importance ranks arising from random forest analysis

Rank	Variable	Mean Decrease Gini index
1	NI	131.4
2	swallow	32.9
3	cardiac	24.7
4	CDH	23.7
5	asthma	20.9

We observed a separation in mean Gini index between ni and swallow. This separation was more clearly demonstrated in the chart below (Figure 108). There was also separation to a lesser extent between asthma and cld.

Figure 108: Variable importance arising from the mean decrease in Gini index of the random forest classifier



Our analysis demonstrated no increased classifier performance with the ensemble random forest approach. There are several reasons for this. Firstly, the random forest approach is useful for large data sets where there is computational efficiency in splitting the data and running multiple parallel DT processes. The fundoplication training and test sets are not large enough to allow these efficiencies.

Secondly, the random forest did not perform better than a predictor that characterised all children as having the majority outcome. This indicates that the outcome is rare. This suggests that a sampling technique that accounts for the rarity of this outcome is indicated.

Optimising the classifier by introducing misclassification cost

Focusing on the confusion matrix, we noticed that the false negative classification rate was equivalent to the event rate in the data i.e. 6%. As our outcome is rare, falsely classifying all true positives as false negatives would still result in a highly accurate classifier (94%).

'Penalising' the algorithm for false negative classifications would reduce this misclassification and consequently improve sensitivity.

A loss matrix is a 2 by 2 matrix that corresponds to true positive, false positive, false negative and true negative respectively in the confusion matrix. The matrix assigns zero cost true correct observations i.e. true positive and true negative. The matrix assigns a positive cost to the cross-diagonals i.e. false positive and false negative misclassifications. To increase sensitivity of the classifier, we could either increase the true positive rate or decrease the false negative rate. We chose to penalise false negative misclassification by assigning a higher cost (20) to false negative classification, and a lower cost (1) to false positive classification. Hence, we established a loss matrix of [0, 1, 20, 0] (Figure 34).

Table 117: The loss matrix applies positive costs to false positive (blue) and false negative (red) costs

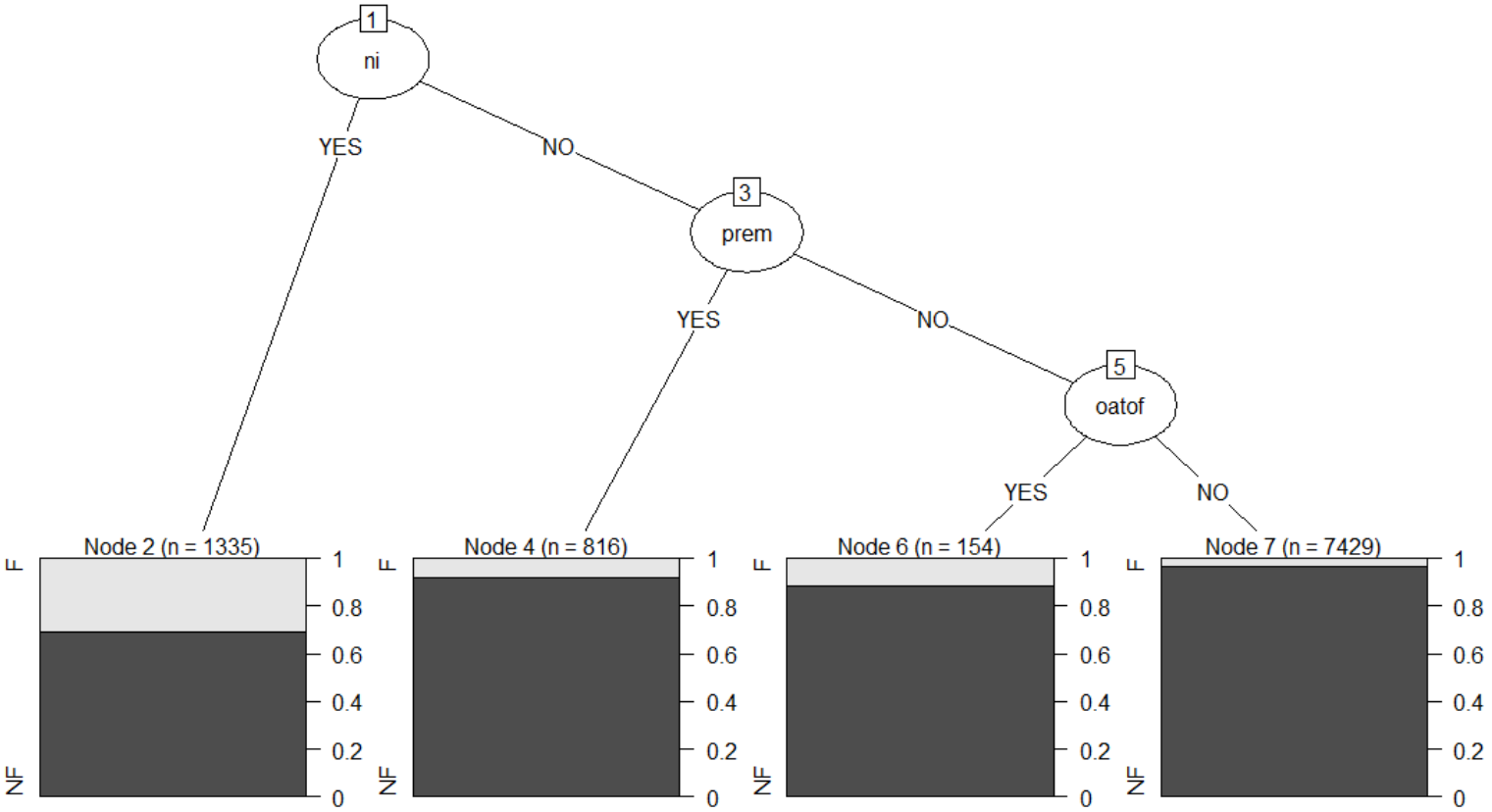
Confusion matrix			**Loss matrix**		
	Actual fundoplication status			**Cost**	
Prediction	F	NF	Cost	F	NF
F	90	41	F	0	1
NF	233	3804	NF	20	0

This loss matrix can also be rationalised on clinical grounds. Morbidity associated with untreated reflux in patients (false negative) may be greater than that experienced by patients who undergo unnecessary fundoplication (false positive), although one could also argue that unnecessary fundoplication exposes to greater risk of complications so should be penalised more. We applied the cost matrix [0,1,20,0] to the DT and observed the differences.

The resulting tree (dt.cost) included fewer branches and terminal nodes compared to the basic dt.raw, which had 6 inner nodes and 7 terminal nodes. The dt.cost had 3 inner nodes and 4 terminal nodes. algorithm resulted in 5 inner nodes and 6 terminal nodes. Furthermore, 1 of the 4 terminal nodes (No OATOF) was nearly pure.

We also observed that the split criteria were different in content and order compared to the dt.raw and dt.pruned and dt.rfpruned models. Therefore, the list of variable importance would vary too.

Figure 109: dt.cost with a false negative misclassification cost of 20 applied



Key

F= Fundoplication NF= No fundoplication YES = comorbidity is present NO= comorbidity is absent

The confusion matrix from the cost-adjusted model was very different from the cost-free DTs. The model was far more sensitive (78% versus 28%). The gross number of false negative classifications was reduced from 223 to 85.

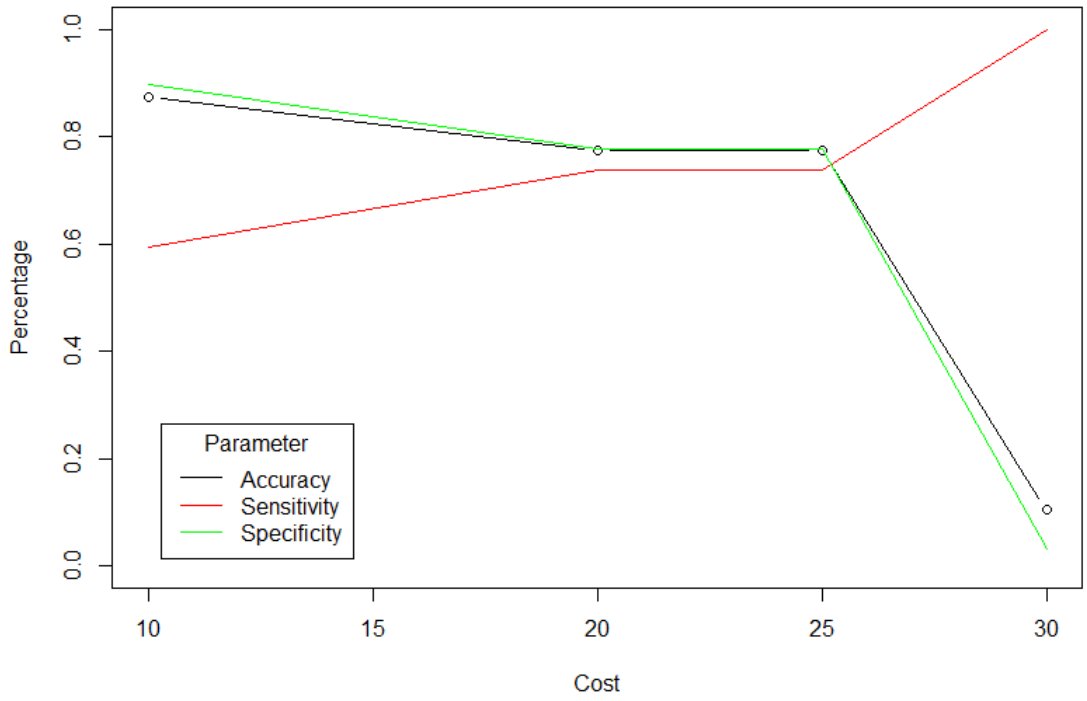
Table 118: Confusion matrix for a dt.cost, a DT with a false negative misclassification cost of 20 applied

		Actual fundoplication status	
		True	False
Predicted	True	238	859
	False	85	2986

Statistics	
Accuracy	0.78
Sensitivity	0.74
Specificity	078

The loss matrix was arbitrarily specified. To identify the cost matrix that would best optimise sensitivity, specificity and accuracy we repeated the procedure for a range of false negative costs. False positive cost is fixed at 1. To visualise the optimal value, we plotted cost against these parameters.

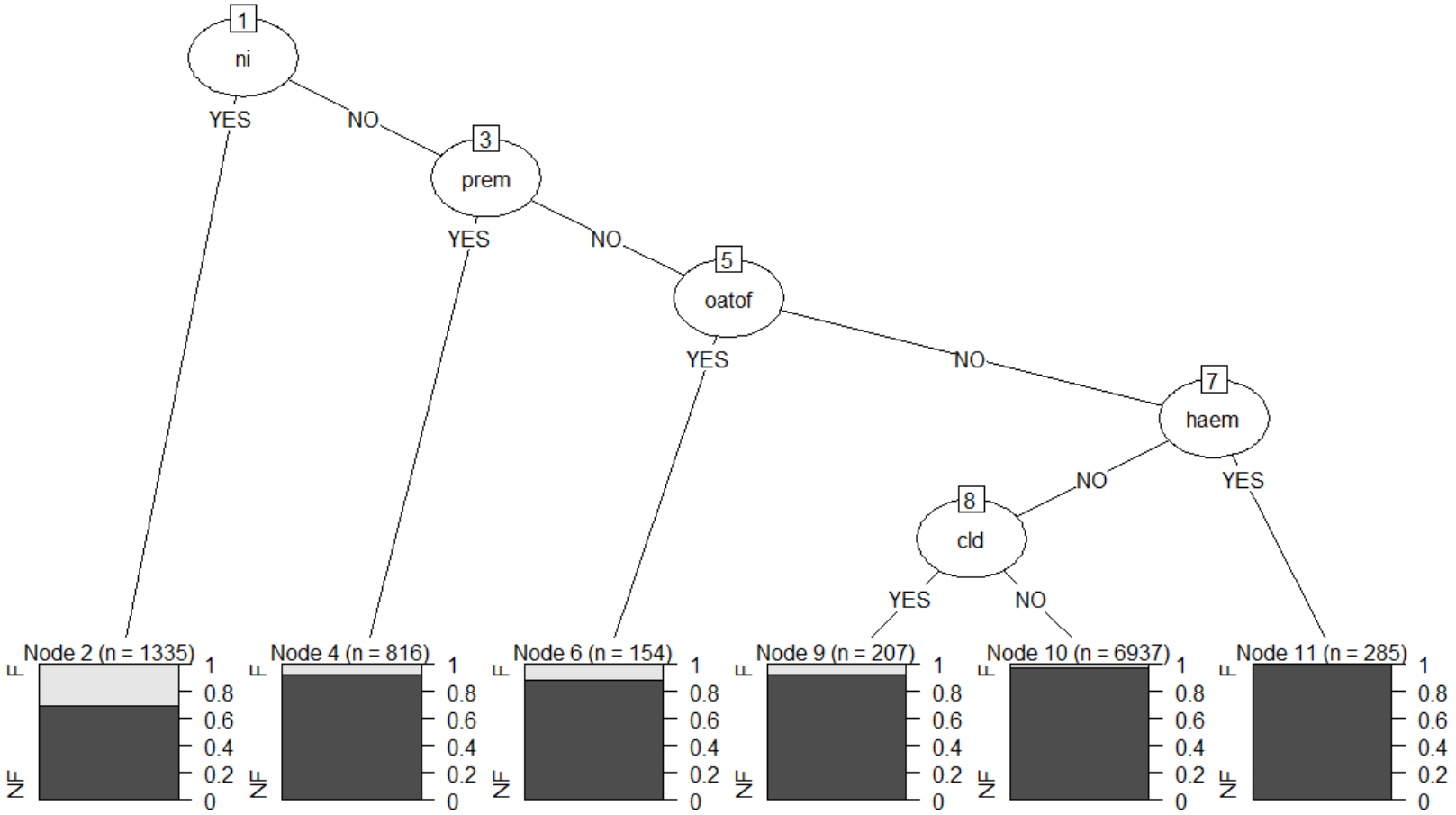
Figure 110: Change in sensitivity, specificity and accuracy as false negative cost is increased



As false negative cost increased, accuracy decreased. The optimal value that maximised both sensitivity and accuracy was 25. The plot appears optimal at a false negative cost of 25. As cost increases beyond this, accuracy falls and specificity fall. We regenerated the DT using the cost matrix [0,1,25,0] resulting in below.

We can summarise the loss-adjusted model as follows: as in all the DTs, neurological impairment is the key criterion for predicting fundoplication. When neurological impairment is present, 15% of patients will receive fundoplication.

Figure 111: dt.cost25 has a false negative misclassification cost of 25



Key

F= Fundoplication NF= No fundoplication YES = comorbidity is present NO= comorbidity is absent

In the absence of NI, prematurity, OATOF and CLD increase the risk of fundoplication. Haematological conditions reduce the risk of fundoplication. This is clinically intuitive as the operative risk for children with haematological disorders (e.g. thrombophilia, sickle cell disease) are significant e.g. bleeding risk, sickle cell crises, need for transfusion etc. In this subpopulation, in the absence of neurological disease, the operative morbidity of fundoplication may be perceived by the clinician to outweigh the benefit.

We ranked the variables by Gini index to understand their importance to predicting fundoplication risk (Table 119). On this cost-adjusted model, there was less of a gap between the NI and subsequent variables, suggesting that the cost-adjusted model was less biased by this single variable.

Table 119: Variables important for predicting fundoplication, ranked by decrease in Gini index

Rank	Variable	Mean Decrease Gini index
1	NI	774.8
2	CLD	197.2
3	prem	150.8
4	OATOF	97.1
5	swallow	96.4
6	haem	83.9
7	oncology	72.9
8	tracheal	58.5
9	achalasia	22.3

We assessed the performance matrix of the cost-adjusted [0,1,25,0] model. While the model was less specific, a false negative cost adjustment of 25 increased sensitivity and NPV.

Table 120: Confusion matrix for a DT with a false negative misclassification cost of 25 applied

		Actual fundoplication status	
		True	False
Predicted	True	238	859
	False	85	2986

Statistics	
Accuracy	0.78
95%CI	0.76-0.79
Sensitivity	0.74
Specificity	078
PPV	0.22

NPV	0.97

The AUC for this model is 0.76, making it a ‘fair’ performer (Figure 112). The optimal probability cut-off was 0.4 This is associated with a sensitivity of 74% and a specificity of 78%.

At 74% sensitivity, this costed model is more sensitive than prior models. The clinical implication is the improved ability to identify patients who require fundoplication. However, the model is less specific (78%). Therefore, the false positive rate is now 12%. This means that 12 of 100 patients will receive a fundoplication that is not indicated. The AUC is lower than the logistic models which have an AUC of around 0.81. The NPV is high at 97%. Therefore, a costed model will be likely to identify patients who do not require fundoplication.

Figure 112: Receiver operating characteristics for the loss-adjusted model

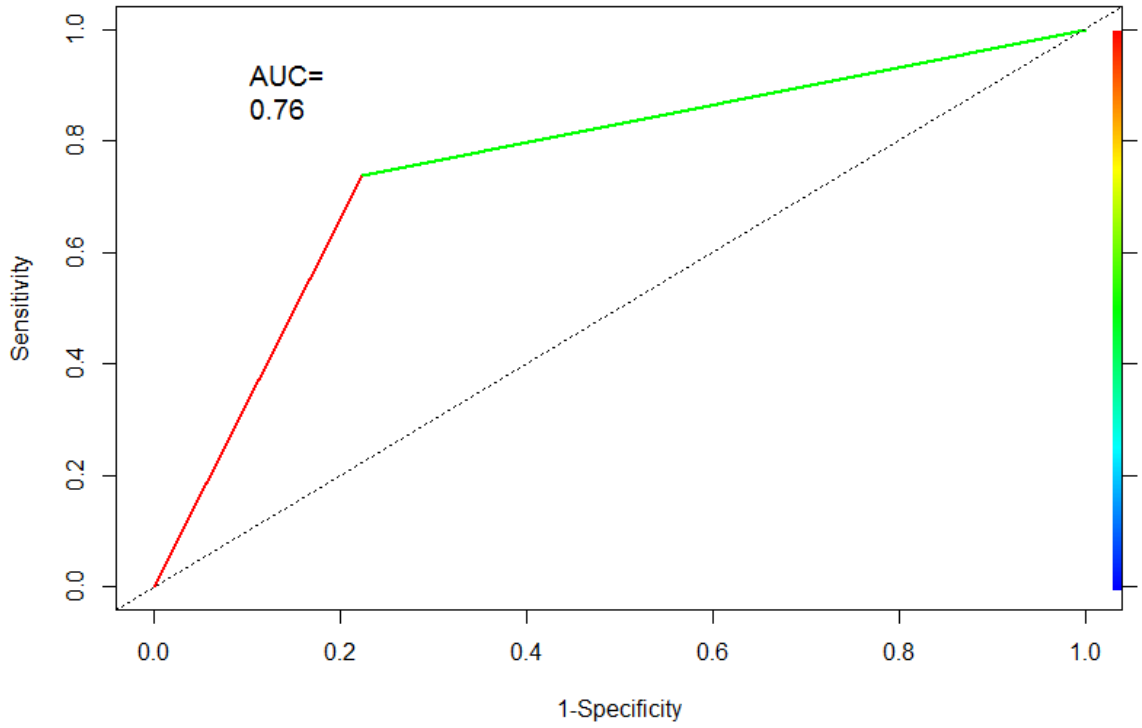


Table 121: Summary of DT models

Model	Description	No. parameters	Predictive comorbidities	Sensitivity	FPR
dt.raw	Basic DT	6	Ni, cardiac, swallow, CDH, asthma, cleft	28%	1%
dt.pruned	Pruned with complexity parameter	6	Ni, cardiac, swallow, CDH, asthma, cleft	28%	1%
dt.rftuned	Random forest classification 500 trees, with 10 random split criteria	6	Ni, cardiac, swallow, CDH, asthma, cleft	29%	0.4%
dt.cost25	Cost-adjusted	5	Ni, prematurity, OATOF, CLD are predictive for fundoplication. Haematological disease is protective	78%	12%

.

In summary, DT modelling yielded several models, of which the best performer included cost adjustment. The benefit of this modelling technique is the ability to weigh tree splitting based on clinical priorities or model heuristics. Furthermore, the output is algorithmic and hence intuitive to clinicians used to visualising clinical pathways.

Compared to logistic regression models explored prior, DT sensitivity is overall lower. Next, we consider a different modelling approach using association rules.

MODELLING WITH ASSOCIATION RULES

In the previous section, we used two supervised learning approaches i.e. logistic regression and DTs to model the fundoplication. In this section, we report on our findings using an unsupervised learning approach i.e. association rule learning(249).

Association rule learning, a.k.a. "market basket" approach is commonly experienced when shopping for items online. For example, when shopping on Amazon, a shopper is often prompted with buying suggestions e.g. "Customers who bought this item also bought" and "Frequently bought together" hyperlinks. Marketers use data to understand shopper characteristics and identify purchasing patterns. These patterns can then be used to lay out the store and set up promotions.

Studying a shopper's basket might result in the following rule:

$$\{nappies + baby\ formula\ milk\} \rightarrow \{baby\ wipes\}$$

I.e. if nappies and baby formula are purchased, then baby wipes are likely to be purchased too. This expression can then be used to offer baby wipe promotions to other purchasers of nappies and formula milk.

There is specific nomenclature applied to association rule (AR) learning(249,263). The equation above summarises a 'transaction'. In this example, nappies and baby formula milk form an 'item set'. The item set is based on the frequency of the co-occurrence of individual items in the data. The relationship between item sets and other items in the market basket e.g. baby wipes, can also be measured and described. The rule has a right-hand side (RHS) and a left hand side (LHS). The strength of associations between the RHS and LHS can be quantified using the metrics of support, confidence and lift.

Support is a measure of how frequently an item or item set occurs in the data. It is summarised as:

Equation 19: Support

$$Support(X) = \frac{Count(X)}{N}$$

Where:

- Count (X) is the number of transactions in which the item or item set X appears
- N is the total number of transactions

Confidence is a measure of a rule's accuracy. Take a rule (X→Y). Confidence is the proportion of transactions where item set X is associated in item set Y. It is calculated as:

Equation 20: Confidence

$$Confidence\ (X \rightarrow Y) = \frac{Support(X,Y)}{Support(X)}$$

The confidence equation is non-reversible. This means that the confidence associated with (X→Y) is not the same as (Y→X).

Lift is a measure of relative frequency(264) . It is a measure of frequency of observation of item set X given prior knowledge that item set Y has been selected. In a commercial context, lift tells us how much more likely the shopper is to select nappies, given they already have baby wipes in their basket. The greater the lift, the stronger the association. Lift has utility as a comparative parameter. It can be used to assess which combinations are most probable. Lift can be summarised as:

Equation 21: Lift

$$Lift\ (X \rightarrow Y) = \frac{Support\ (X \rightarrow Y)}{Support\ (X)}$$

Application

In our clinical context, the shopper is the surgeon surveying a waiting room full of patients with varying characteristics. She samples symptoms, comorbidities and investigations before deciding whether to add fundoplication to the basket. What are the association rules governing this choice?

This modelling approach will result in a descriptive, rather than a predictive model. Descriptive models, nonetheless, have value as a decision support tools. We can summarise the "black box" synthesis of knowledge, experience and intuition that goes on in the decision-maker's mind into a statement relating comorbidities to outcome. Furthermore, descriptive modelling minimises bias. It is a simple pattern recognition approach that enables knowledge discovery without the potential bias of heuristics introduced when models are trained to fit data.

The main disadvantage of descriptive modelling is the performance measurement limitation. There is no objective way to measure the performance of association rules. As such, there is a risk of reaching spurious conclusions. Another important caveat is the limited number of items available for sampling. Due to the nature of our retrospective database, we are only able to see the comorbidities being sampled. It may be that parent characteristics (e.g. level of educational attainment) may be a factor in whether a child gets a fundoplication. However, the same limitation is present for the other approaches used.

A second caveat to consider is the dynamic nature of comorbidities. We assume that all comorbidities are in play on the day of decision making. Therefore, a patient with asthma as comorbidity is assumed to have active asthma at the time the decision was taken to perform fundoplication. In reality, we cannot know from our data what the severity of asthma was at the time the decision was taken and how much of a factor it was in the decision-making process.

Data selection, pre-processing and transformation

For the first trial of AR learning, we use the complete dataset. As AR provide descriptive rather than predictive models, we cannot apply a model developed on training data and verified on test data.

We utilised the following packages to automate the association rules analysis.

1. arules(265)
2. Matrix(266)

The first task was to reformat the data into a sparse matrix. The sparse matrix "basket.dta" contains all the data available in our metaphorical shopper's basket.

To form a sparse matrix, the database is indexed by patient. Each row represents a patient, and each patient represents a transaction.

A random selection of 5 transactions in the sparse matrix are reviewed. The patient with TransactionID 1 has 7 comorbidities and had a fundoplication. The fifth patient had simple GORD and no fundoplication.

Table 122: A random selection of transactions

TransactionID	Items
1	1{cardiac, cld, fundo, NI, oatof, prematurity, renal, tracheal}
10	2{aresp, aspiration, cardiac, cld, fundo, NI, prematurity, tracheal}
100	3 {consanguinity, fundo, NI, prematurity, renal}
1000	4 {NI, prematurity}
10000	5 {SimpleGORD}

The most common item set size is 1 i.e. most patients have just 1 comorbidity- simple GORD. The largest transaction contained 8 items i.e. there were 10 patients who had 8 comorbidities.

Table 123: Itemsets and sizes

	Sizes							
Items	1	2	3	4	5	6	7	8
Itemset size	11514	1544	494	167	102	46	26	10

Most patients have simple GORD and therefore, simple GORD was the most common item in the basket (7). Interestingly, cardiac disease appeared more frequently than NI as a comorbidity.

Table 124: Sparse matrix data demonstrating that simple GORD was the most common item, followed by cardiac disease.

Transaction ID	
Item	**n**
Simple GORD	6797
Cardiac	2593
NI	1940
Prematurity	1417
Fundo	1080
CLD	517
(Other)	3455

As we were specifically interested in the relationship between comorbidities and fundoplication, we subset the data to exclude transactions where the only comorbidity was simple GORD.

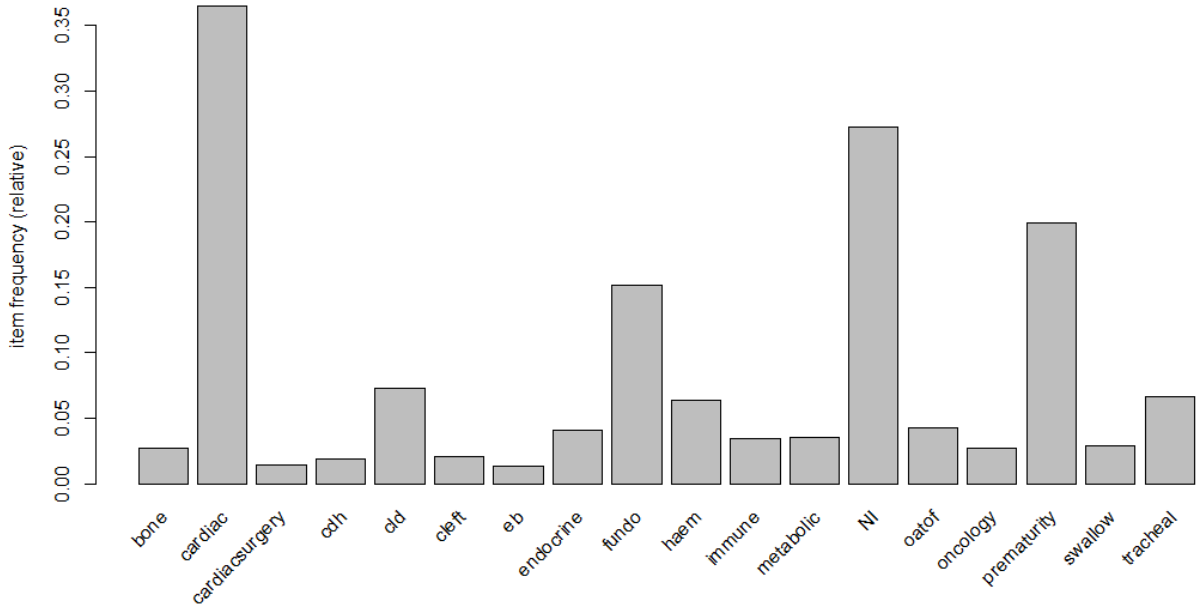
Excluding the transactions with simple GORD as the only item reduced the dataset (onlyco) to 7105 transactions. The most common item in the reduced dataset was now cardiac disease.

Figure 113: Output – item sets and sizes of comorbidities in patients with GORD

```
summary(onlyco)

transactions as itemMatrix in sparse format with
 7106 rows (elements/itemsets/transactions) and
 30 columns (items) and a density of 0.05161

most frequent items:
   cardiac         NI prematurity       fundo        cld     (Other)
      2593       1940       1417       1080        517       3456

element (item set/transaction) length distribution:
sizes
  1        2       3     4      5     6     7    8
4717    1544    494  167    102    46    26   10
```

Item frequency was visualised in a histogram. It was impractical to plot every single data point on the histogram. Therefore, we set a threshold of comorbidities that affect at least 10% of the database population. This parameter is known as common support. Common support was configured at 0.1.

Figure 114: Relative item frequency for transactions in RetrospectiveGOR database.



Modelling the data

The simplest procedure for association rule learning is a stepwise approach. Firstly, all possible item sets are identified. The frequency of each itemset is estimated. For each itemset, all other items associated with this item set are identified. This is done in a stepwise way, working from the most frequent to the least frequent itemset, till all associations are identified.

In our database, there are 29 and 1 outcome. Therefore, there are 2^{29} possible itemsets (536,870,912) for the learner to evaluate. Such a stepwise approach would be cumbersome, processor-hungry and time consuming. Instead, educated guesses – heuristics- are applied.

The most commonly applied heuristic for association leaning is the *a priori* algorithm(267). Stated briefly:

"All subsets of a frequent item set must also be frequent."

Expressed in the transactional terms of nappy/formula milk and baby wipe transaction above:

For [nappies and baby formula] to be classified as a frequent item set, both items must be individually frequent.

We applied the *a priori* restriction to identifying association rules. For the first trial, we have set a low threshold of support i.e. 0.001. This means that the rule should apply to at least 0.1% of transactions.

This first trial resulted in 213 rules. On reviewing the rules, we noted multiple rules that did not have fundoplication on either the left or right hand side. As we wanted to identify only rules related to fundoplication, we further restricted the algorithm to search for rules containing the variable "fundo".

Removing rules not related to fundoplication results in 165 rules. However, within this subset there is some redundancy e.g. where the right and left items in one rule match the left and right items in another rule respectively e.g.

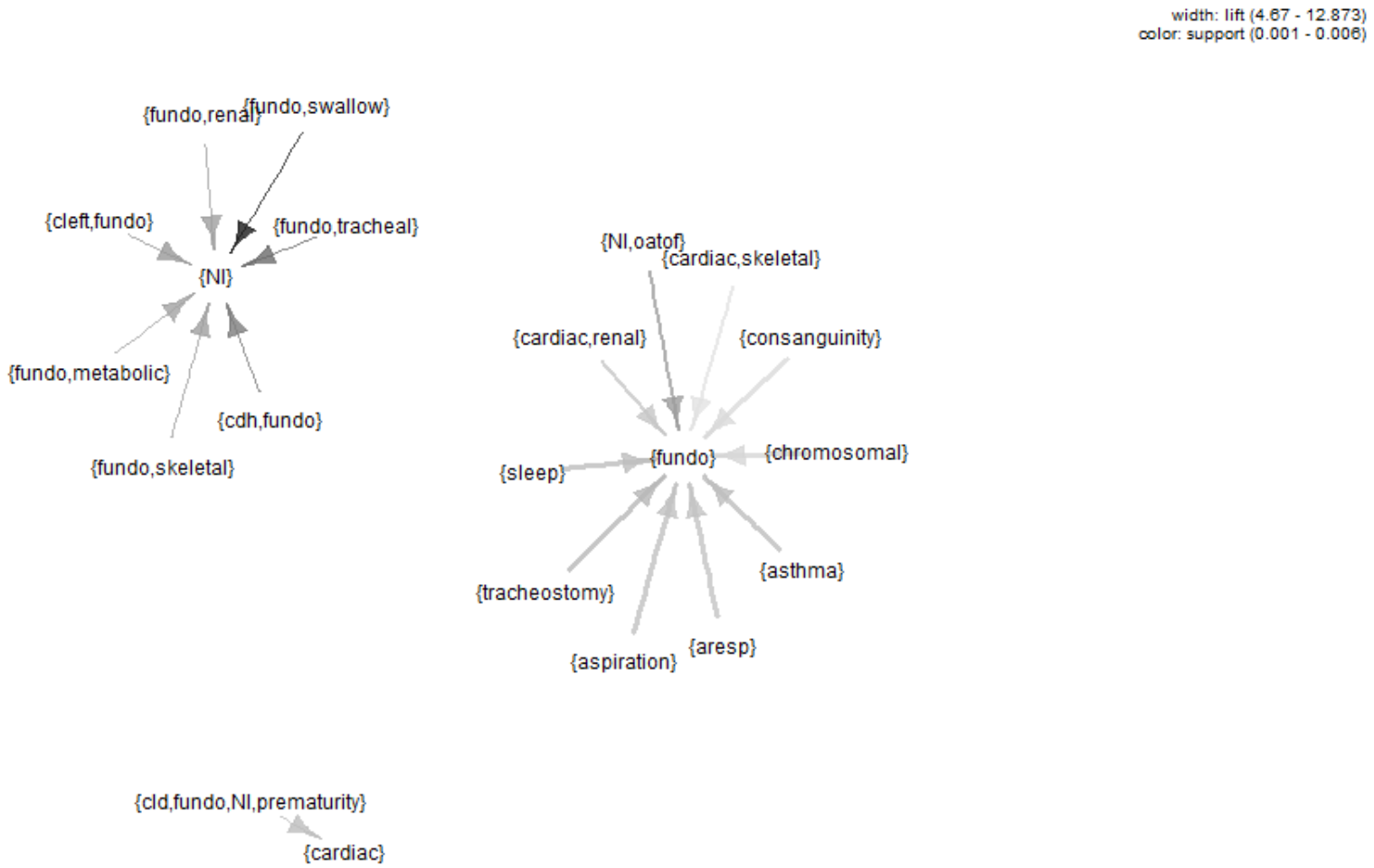
$$\{fundo\} \rightarrow \{NI + cardiac\}$$

$$\{NI + cardiac\} \rightarrow \{fundo\}$$

Therefore we identified and pruned the redundant rules.

Pruning redundant rules results in 18 rules. These rules were visualised in a relationship (Figure 115). The arrows and their shading transmit information about two measures of association. The distance away from the central node is a measure of the lift. The closer item set to the central node, the greater the lift. The greyscale of the arrow is a measure the level of support. The darker the greyscale, the greater the level of support.

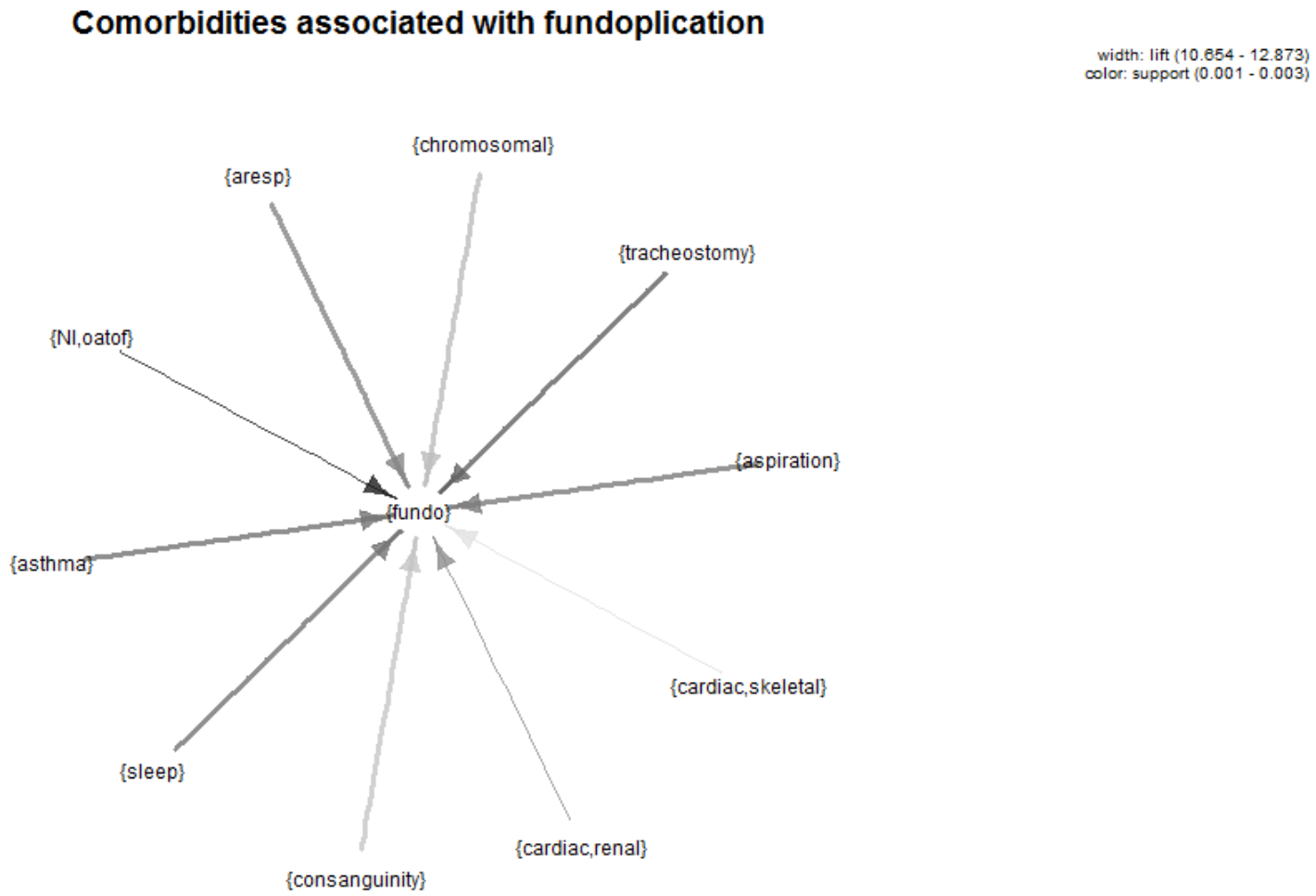
Figure 115: Association rules in a cohort of patients with comorbidities, some of whom undergo fundoplication



We observed an association between cardiac disease and chronic lung disease, neurological impairment and fundoplication. We also observed an association between neurological impairment and patients who have fundoplication, with prior comorbidities e.g. swallow disorders. However, this relationship map was difficult to interpret as there were three centres i.e. NI, cardiac disease and fundoplication. For the purposes of understanding fundoplication risk, we further limited our focus to only rules centred around fundoplication. Therefore, the association rules were simplified and re-stated to include only those with fundoplication as a right hand side term.

The resulting subset of rules centred on fundoplication was visualised the relationship map below (Figure 116).

Figure 116: Association rules centred on patients undergoing fundoplication



The darker the greyscale , the greater the support. The closer the item set is to the centre, the stronger the support. In total, there are 12 comorbidities that were associated with fundoplication. We observed that NI+OATOF in combination had the strongest association with fundoplication.

These rules were summarised in Table 125 below, sorted in order of support.

Table 125: Association rules for comorbidities related to fundoplication

Rules	Support (%)	Confidence (%)	Lift
{NI, oatof} => {fundo}	0.53	86.36	5.68
{tracheostomy} => {fundo}	0.41	100.00	6.58
{asthma} => {fundo}	0.38	100.00	6.58
{sleep} => {fundo}	0.38	100.00	6.58
{aspiration} => {fundo}	0.37	100.00	6.58
{aresp} => {fundo}	0.35	100.00	6.58
{cardiac, renal} => {fundo}	0.34	82.76	5.45
{chromosomal} => {fundo}	0.27	100.00	6.58
{consanguinity} => {fundo}	0.25	100.00	6.58
{cardiac, skeletal} => {fundo}	0.21	88.24	5.81
{cdh, cld} => {fundo}	0.17	80.00	5.26
{oatof, swallow} => {fundo}	0.14	100.00	6.58
{cld, renal} => {fundo}	0.13	90.00	5.92
{cld, prematurity, tracheal} => {fundo}	0.13	81.82	5.38
{dental, NI} => {fundo}	0.11	80.00	5.26
{prematurity, skeletal} => {fundo}	0.11	80.00	5.26

Interpretation

The first rule can be summarised as follows: in 0.5% cases, we can conclude with 86% confidence that that the combination of NI and OATOF comorbidities is associated with fundoplication. This rule held true for 35 of 7105 patients (support 0.5%). This rule has a lift of 5.68 (range 5.26-6.58) suggesting that it is neither the strongest nor the weakest of associations.

The immediate application for this information is comparing a dramatic visualisation of pathways to fundoplication. In Figure 116 we *see* the feeder services for fundoplication clinics. We can also see which comorbidities bring patents ‘closer’ to a fundoplication outcome.

Secondly, association rules also deepen understanding of operative populations. Research in this area is often limited by the fact that surgeons cannot compare outcomes as they cannot be sure that the operations were performed on comparable populations. A market basket approach enables surgeons to visualise an operative population. It also provides useful metrics for comparison. For example two

surgeons can compare the support and confidence metrics for their NI+OATOF populations as a way of understanding variations in morbidity and mortality figures.

We also observe that respiratory conditions e.g. asthma, sleep apnoea, aspiration, acute respiratory failure, chronic lung disease comprise many of the itemsets represented here. This information can be useful in planning services. For example e.g. pre-operative assessment for fundoplication could include review by a respiratory physician to ensure respiratory disease was controlled and optimised.

MODEL SELECTION: RISK OF FUNDOPLICATION

In this chapter we explored several methods of estimating fundoplication risk. In Table 126below, we summarise the best models from each approach.

Both the logistic regression model and the DT rules have similar accuracy and sensitivity. The logistic regression model with bias reduction had the best balance of accuracy and sensitivity. It would be the model of choice for analysing fundoplication risk. The association rules model, although visually arresting, provided no quantitative measures or predictive parameters for comparison with other models.

The other modelling approaches are contributory. The DT model returned the fewest predictive comorbidities and therefore, is perhaps the simplest to interpret. The association rule model was the most intuitive to understand, once visualised.

It is also noteworthy that the leading predictive comorbidities indicated are broadly similar for these three models. Despite the different modelling assumptions and heuristics applied, the usual suspects remained the same i.e. NI, prematurity, OATOF, CLD. This lends strength to our conclusions about the key comorbidities predicting risk of fundoplication. These comorbidities will be discussed at length at the end of this chapter and in the next.

In the next section, we focus on modelling the risk of failure of fundoplication.

Table 126: A summary of the modelling methods used to assess fundoplication risk

Model	Accuracy	Inputs	Predictive comorbidities	Probability cut-off	Sensitivity	FPR	AUC	AIC
Logistic regression with bias reduction (brfit)	80%	20	14*	0.05	74%	22%	0.81	2016
Decision tree with false negative cost adjustment (dt.cost25)	77%	29	5†	0.40	78%	12%	0.76	-
Association rules (fundorulesfinal)	N/A	29	12‡	N/A	N/A	N/A	N/A	N/A
* Predictive: NI+ cleft, NI + tracheal disease, NI + swallow disorders, CDH, renal disease, skeletal anomalies, NI +cardiac disease, OATOF, prematurity, CLD Protective: Haematological disease, cardiac and tracheal disease in isolation								
† Predictive: NI, prematurity, oatof, cld are predictive for fundoplication. Protective: Haematological disease								
‡ Predictive: NI+ oatof, tracheostomy, asthma, sleep, cardiac+renal, chromosomal and consanguineous conditions, acute respiratory failure, cdh, cld, tracheal anomalies								

CHAPTER 3: RISK OF FUNDOPLICATION FAILURE

This risk associated with failure of fundoplication is two-fold. Firstly, continuing symptoms perpetuate morbidity for the child. Secondly, re-do fundoplication will not be on virgin tissue. Altered anatomy and adhesions limit the surgeon's approach. This may increase the risk of intra-operative injury to bowel and abdominal viscera. There is also the inevitable morbidity arising from pain and anaesthetic exposure.

It is important to understand factors pre-disposing to fundoplication failure and hence redo. There are patient factors e.g. poor patient selection. If the operation was ineffective the first time, why should it work the second time? Perhaps it is the wrong operation for this particular presentation. Secondly, there are operator factors to consider. If an operation needs to be re-done, perhaps it was not done properly in the first place. Identifying risk factors for failure of fundoplication is a priority for both clinician and patient.

As discussed previously, our cohort is limited to the examination of patient factors. We had insufficient granularity of operator data to allow a viable inclusion of operating surgeon into the database. Furthermore, GOSH is a teaching hospital. The surgeon responsible for the patient's care as noted in the operating theatre database may not be the primary operating surgeon. In a retrospective review of data, it was not practical to obtain primary operator information of any veracity.

Therefore, this analysis was focused on the available patient factors in the retrospective GOR database that can predict failure of fundoplication leading to revision surgery.

LOGISTIC REGRESSION

Re-do fundoplication is a binary outcome. Therefore, logistic regression is an applicable statistical method.

Pre-processing and data sampling

We selected from the database observations of patients who had a fundoplication. There were 1080 patients who had fundoplication. Of these 75 (7%) had a RF.

We set a seed and partitioned the data, utilising a 70:30 training to testing ratio. Training data were sampled without replacement. The RF training set ("redotrain") contained 756 observations with 29 comorbidities described. The test data were the remnant subset of the following sampling training data. The testing data contained 324 observations of 29 comorbidities.

Modelling

Step 1: Summarising the redo training data

The training data are summarised below. There were no missing data. Data were available on all 29 comorbidities. There were 577 patients (76%) who had one fundoplication and 179 patients (24%) required a redo procedure.

Figure 117: Training dataset for RF

```
summary(redotrain)
```

```
 PatientID           hadfundo         fundonumber                    fundocategory

Length:756         Mode:logical      Min.   :1.00                   GREATER THAN ONE: 52
Class :character   TRUE:756          1st Qu.:1.00                   ONE           :704
Mode  :character   NA's:0            Median :1.00
                                     Mean   :1.08
                                     3rd Qu.:1.00
                                     Max.   :4.00

 tracheal       cleft       sleep        trachy        aspiration
 ABSENT :715   ABSENT :725   ABSENT :735   ABSENT :737   ABSENT :741
 PRESENT: 41   PRESENT: 31   PRESENT: 21   PRESENT: 19   PRESENT: 15


      cld          cdh       asthma         cardiac       ni
 ABSENT :675   ABSENT :716   ABSENT :736   ABSENT :558   ABSENT :329
 PRESENT: 81   PRESENT: 40   PRESENT: 20   PRESENT:198   PRESENT:427

     oatof       achalasia     swallow      chrom         consang
 ABSENT :707   ABSENT :749   ABSENT :681   ABSENT :742   ABSENT :741
 PRESENT: 49   PRESENT:  7   PRESENT: 75   PRESENT: 14   PRESENT: 15

      prem          haem       endocrine       renal        skeletal
 ABSENT :617   ABSENT :752   ABSENT :732   ABSENT :732   ABSENT :733
 PRESENT:139   PRESENT:  4   PRESENT: 24   PRESENT: 24   PRESENT: 23

      bone         dental       oncology     metabolic       aresp
 ABSENT :754   ABSENT :745   ABSENT :753   ABSENT :718   ABSENT :740
 PRESENT:  2   PRESENT: 11   PRESENT:  3   PRESENT: 38   PRESENT: 16

     immune       cardsurg         eb        simplegord     redo
 ABSENT :745   ABSENT :750   ABSENT :755   ABSENT :577   Mode :logical
 PRESENT: 11   PRESENT:  6   PRESENT:  1   PRESENT:179   FALSE:704
                                                         TRUE :52
                                                         NA's :0
```

Step 2: Apply logistic regression

The model is called logitr.

Equation 22: Formula for logitr regression model

```
fmredo<redo~tracheal+cleft+sleep+trachy+aspiration+cld+cdh+asthma+cardiac+ni+oatof+achalasia+swallo
w+chrom+consang+prem+haem+endocrine+renal+skeletal+bone+dental+oncology+metabolic+aresp+immu
ne+cardsurg+eb+simplegord
```

The regression formula for logitr (above) was applied with the following result. The variables predicting RF were sorted in order of estimate size. Variables not meeting the p< 0.05 level of significance was excluded from the summary result.

```
summary(logitr)


Call:
glm(formula = fmredo, family = binomial(logit), data = redotrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3393  -0.3512  -0.1785  -0.0001   2.6900

Coefficients:
```

	Estimate	Standard Error	z value	Pr(>|z|)	
ni	2.64	1.05	2.52	0.01165	*
skeletal	1.73	0.57	3.03	0.00242	**
aspiration	1.71	0.71	2.40	0.0162	*
cdh	1.59	0.45	3.49	0.00049	***
tracheal	1.54	0.52	2.96	0.00309	**
oatof	1.25	0.51	2.46	0.01386	*
cld	-1.37	0.66	2.07	0.03824	*

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

 Null deviance: 378.73  on 755  degrees of freedom
Residual deviance: 268.18  on 726  degrees of freedom
AIC: 328.2

Number of Fisher Scoring iterations: 18
```

Step 3: Statistical inference

NI has the most positive estimate, and therefore was associated with the greatest increase in risk of RF. CLD has a negative estimate and therefore reduces the risk of RF. The data were presented as odds ratios with 95% confidence intervals and ranked in order of effect size. These odds ratio data offer a more intuitive appreciation of effect size. NI increased risk of RF by a factor of 14.

Table 127: Odds ratios and confidence intervals for comorbidities predictive of risk of RF

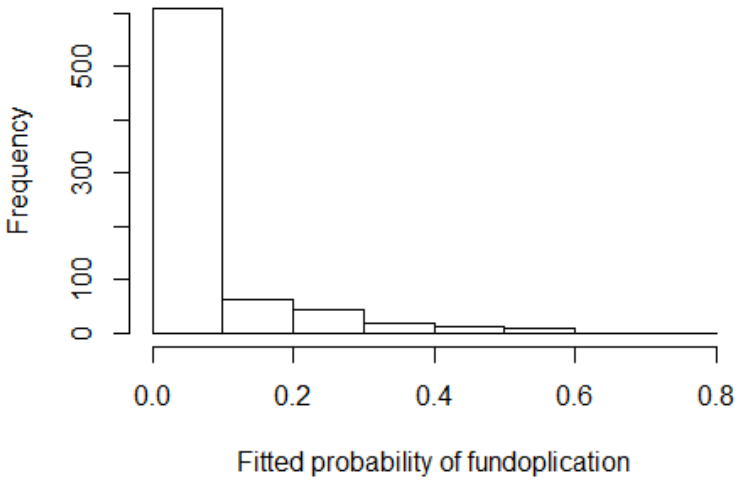
		95% Confidence interval	
	Odds ratio	2.5(%)	97.5(%)
ni	13.97	2.75	256
skeletal	5.67	1.74	16.89
aspiration	5.50	1.25	21.36
cdh	4.88	1.96	11.82
tracheal	4.64	1.62	12.61
oatof	3.50	1.24	9.31
cld	0.25	0.06	0.83

Note that presence of chronic lung disease is associated with a four-fold decrease in the risk of RF compared with no comorbidities.

Step 4: Goodness of fit

Fitted values for probability of RF were reviewed in a histogram. The probability of RF is <0.1 for >600 patients. This corresponds to the observed event rate of 75% and suggests that a probability of at least >0.1 is the threshold for RF.

Figure 118: Fitted probabilities for the logitr regression model of RF



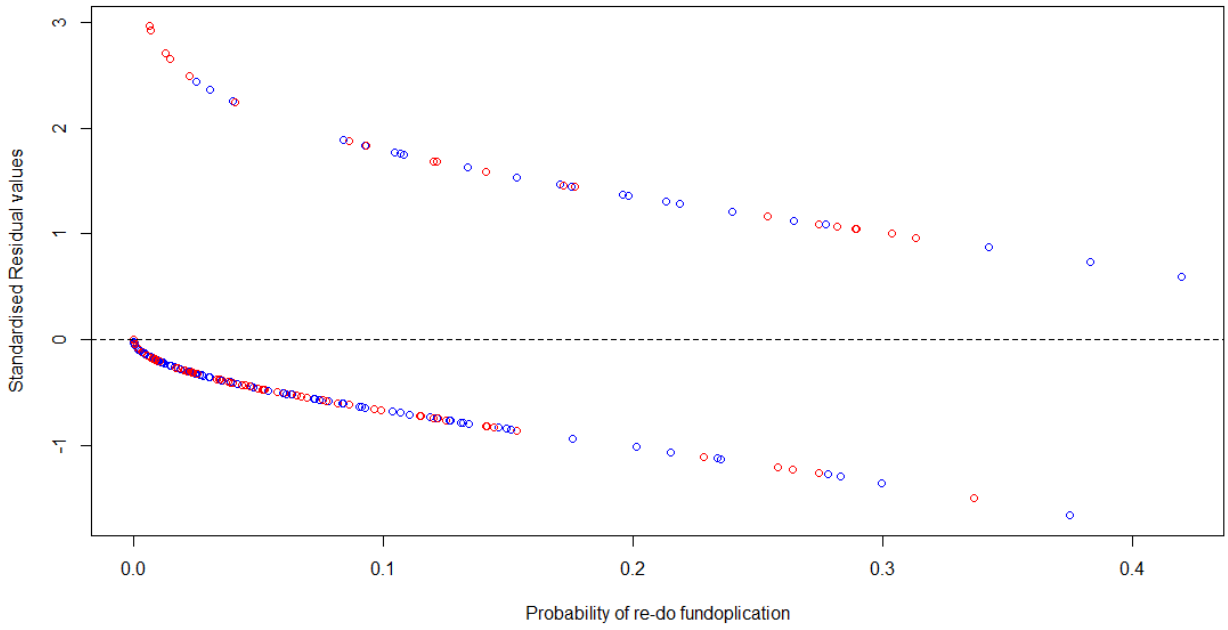
NULL VS. RESIDUAL DEVIANCE

From the summary of the logistic regression, we observe that the residual deviance (268.18) is less than the null deviance (378.73). The p-value for the chi squared analysis of difference in deviance is <0.001. Therefore, we can accept the logitr model as better than a null model.

RESIDUAL VS. FITTED DISTRIBUTION

The standardised residual vs. fitted plot for the logitr model is plotted below.

Figure 119: Standardised residual vs fitted probability plot for the logitr model



Standardised residuals clustered around the even horizontal, although the curved pattern suggested a non-linear relationship between variables and outcome. This was an early indication that the model may be mis-specified or influential variables or interactions were at play.

AIC AND BIC

The AIC for the logitr model was 328.18. The BIC was 467.02. These values are useful comparators for subsequent models.

Step 5: Predictive performance

Next, we assessed the predictive performance of the logitr model using the test data. The testing data ("redotest") contained 324 observations of 29 comorbidities. The event rate for RF was the test data was 22%.

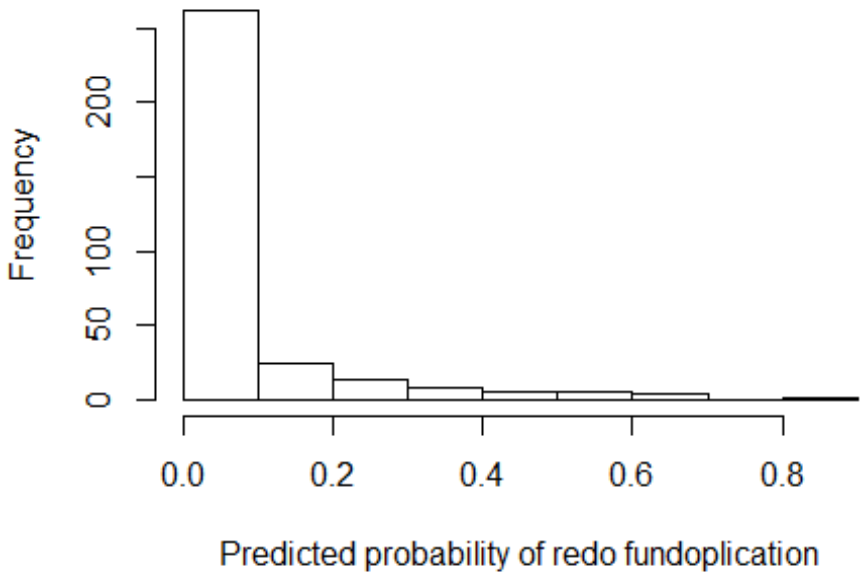
Figure 120: Test data set for RF

```
 PatientID          hadfundo         fundonumber                fundocategory
Length:324         Mode:logical     Min.   :1.00               GREATER THAN ONE: 23
Class :character   TRUE:324         1st Qu.:1.00               ONE           :301
Mode  :character   NA's:0           Median :1.00
                                    Mean   :1.08
                                    3rd Qu.:1.00
                                    Max.   :3.00

  tracheal        cleft           sleep           trachy          aspiration
ABSENT :304   ABSENT :313   ABSENT :318   ABSENT :314   ABSENT :313
PRESENT: 20   PRESENT: 11   PRESENT:  6   PRESENT: 10   PRESENT: 11

    cld             cdh            asthma          cardiac          ni
ABSENT :280   ABSENT :310   ABSENT :317   ABSENT :231   ABSENT :140
PRESENT: 44   PRESENT: 14   PRESENT:  7   PRESENT: 93   PRESENT:184

   oatof          achalasia       swallow         chrom           consang
ABSENT :298   ABSENT :320   ABSENT :305   ABSENT :319   ABSENT :321
PRESENT: 26   PRESENT:  4   PRESENT: 19   PRESENT:  5   PRESENT:  3

prem             haem           endocrine       renal           skeletal
ABSENT :256   ABSENT :323   ABSENT :314   ABSENT :310   ABSENT :310
PRESENT: 68   PRESENT:  1   PRESENT: 10   PRESENT: 14   PRESENT: 14

   bone           dental         oncology       metabolic        aresp
ABSENT:324    ABSENT:324    ABSENT:324    ABSENT :317   ABSENT :315
                                           PRESENT:  7   PRESENT:  9

   immune         cardsurg        eb              simplegord       redo
ABSENT :313   ABSENT :322   ABSENT :323   ABSENT :252   Mode :logical
PRESENT: 11   PRESENT:  2   PRESENT:  1   PRESENT: 72   FALSE:301
                                                       TRUE :23
                                                       NA's :0
```

.

We apply the logitr model to the test data and generate predictions of the probability of fundoplication. The logitr model predicts that most observations in the test data will have a fundoplication probability of <10%. Reviewing the fitted values for the test data confirms this expectation.

Figure 121: Histogram of predicted values for logitr model



Therefore, we assign observations with probabilities of >0.1 to a class prediction of redo-fundoplication. We create a new vector containing the predicted classification. Based on the probability cut-off of 0.1 and generate a confusion matrix.

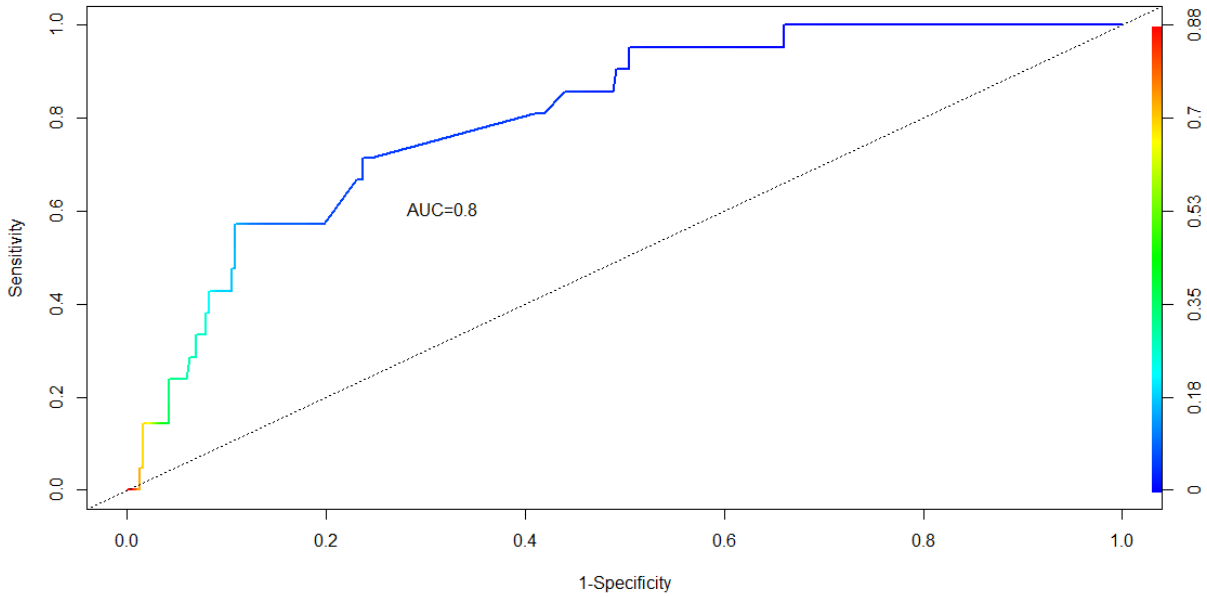
Table 128: Confusion matrix and performance metrics for logitr model

		Actual redo fundoplication status	
		True	False
Predicted	True	16	46
	False	7	255

Statistics	
Accuracy	0.84
95% CI	0.79-0.88
Sensitivity	0.70
Specificity	0.85
PPV	0.26
NPV	0.97

The optimal sensitivity/specificity cut-off was at a probability of 0.05. At this cut-off, sensitivity was 87% and specificity was 71%. Notably, as there are fewer data, the logitr ROC curve is more jagged than curves previously plotted for primary fundoplication.

Figure 122: Receiver operating characteristics for logitr model



The clinical implication can be summarised as follows: Where the probability of RF was >5%, the model accurately assigned 87% of patients to fundoplication.

The area under the ROC curve (AUC) was a 0.8 i.e. better than random chance and a 'good' predictor. The false positive at this cut-off was 29%. The implication was redo surgery would be incorrectly predicated for approximately 1 in 3 patients. The residual vs. fitted plot was also suggestive of non-linear interactions or influential variables. Considering these implications, it was clear the logitr model should be further optimized to reduce the false positive rate.

Strategies to improve the fit of the logitr model included:

1. Identifying the effects of interactions

2. Reducing noise arising from unnecessary inclusion of uninformative parameters

3. Addressing the bias associated with the rarity of the fundoplication event.

Improving the model: the search for interactions

We searched for interacting parameters influential in predicting RF risk. Due to the large number of comorbidities reported here, automated analysis for n-way interactions was limited by processing memory. Therefore, we limited the search to 2-way interactions. The search was performed in a manual, serial fashion illustrated below.

From prior analysis using the logitr model, we observed that NI was the key comorbidity associated with increased risk of fundoplication. To begin, we interrogated the data for an interaction between NI and sleep apnoea.

There was no interaction identified between NI and sleep apnoea. This process was repeated for all possible combinations. No interaction significantly affecting the outcome was identified for this or any other variable combinations.

Improving the model: Stepwise regression

Next, we tried to optimize this model by removing non-contributory parameters. This was achieved using stepwise regression.

Stepwise regression entails introducing model variables in a serial fashion. If the introduced variable improves the fit of the model, it is retained in the formula. If it does not, it is discarded. To assess improvement of fit, several criteria can be used e.g. Akaike information criterion (AIC), Bayesian information criterion (BIC).

Stepwise regression using AIC

The logitr formula (Equation 23) was used as the base formula for stepwise regression which was called stepredo_aic.

Figure 123: Summary output from calling the stepredo_aic model

```
redostep<- glm(fmredo,
        family=binomial(logit),
        data=redotrain)
stepredo_aic<-step(redostep,
            direction="both",
            k=2)

summary(stepredo_aic)
```

```
Call:
glm(formula = redo ~ tracheal + trachy + aspiration + cld + cdh +
    cardiac + ni + oatof + swallow + renal + skeletal, family = binomial(logit),
    data = redotrain)


Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.345  -0.325  -0.162  -0.068   2.946
```

The regression coefficients are reported in below. Eleven comorbidities are were retained in the stepwise. Of these, 6 were associated with a significant increase (NI, OATOF, CDH, tracheal and skeletal anomalies) or decrease (CLD) in the risk of RF.

Table 129: Regression coefficients for stepwise regression using AIC as the reduction parameter

Coefficient	Estimate	Std. Error	z value	Pr(>|z|)	p
ni	3.16	1.03	3.06	0.00218	<0.01
skeletal	1.56	0.55	2.84	0.00452	<0.01
cdh	1.54	0.43	3.53	0.00041	<0.001
tracheal	1.31	0.498	2.63	0.00857	<0.01
aspiration	1.22	0.65	1.86	0.0629	0.1
oatof	1.09	0.49	2.21	0.02744	0.05
renal	0.97	0.605	1.58	0.11378	n.s.
swallow	0.72	0.378	1.89	0.05844	0.1
cardiac	0.66	0.34	1.93	0.05329	0.1
cld	-1.31	0.61	-2.15	0.03134	0.05
trachy	-1.56	0.89	-1.74	0.08117	<0.1

To assess goodness of fit we compared residual deviances. As stepredo_aic is a nested model of logitr, ANOVA applies. There was no significant difference in residual deviance observed. Therefore, stepredo_aic, compared to logitr, does not improve fit.

Figure 124: ANOVA comparing residual deviance for logitr and stepredo_aic models

```
anova(stepredo_aic, logitr, test="Chisq")

 Analysis of Deviance Table

 Model 1: redo ~ tracheal + trachy + aspiration + cld + cdh + cardiac +
    ni + oatof + swallow + renal + skeletal
 Model 2: redo ~ tracheal + cleft + sleep + trachy + aspiration + cld +
    cdh + asthma + cardiac + ni + oatof + achalasia + swallow +
    chrom + consang + prem + haem + endocrine + renal + skeletal +
    bone + dental + oncology + metabolic + aresp + immune + cardsurg +
    eb + simplegord


  Resid. Df    Resid. Dev   Df    Deviance    Pr(>Chi)
1      744        282
2      726        268          18     14.1         0.72
```

To assess predictive performance, we assessed the reduced models against the test data.

The confusion matrix demonstrated the inferior accuracy of the stepredo_aic model compared to logitr. In the test data, there were 23 patients who had RF. The stepredo_aic model correctly predicted fundoplication status for 18 of these. There were also 301 patients who did not receive RF. There were 67 patients were wrongly predicted to receive redo-fundoplication by this classifier, giving a false positive rate of 12% and a low PPV of 21%.

ROC analysis demonstrates that stepredo_aic is a 'good' predictor with an AUC of 0.86. The optimal probability cut-off was 0.05. At this probability, the classifier had a sensitivity of 78% and a specificity of 77%. The sensitivity was comparable to logitr, but the stepredo_aic model was less specific.

Table 130: Confusion matrix and performance metrics for stepredo_aic model

		Actual redo fundoplication status	
		True	False
Predicted	True	18	67
	False	5	234
Statistics			
Accuracy	0.78		
95% CI	0.73-0.82		
Sensitivity	0.78		
Specificity	0.78		
PPV	0.21		
NPV	0.98		

Next we examined whether the model fit could be improved by using BIC rather than AIC as the reduction criterion.

Stepwise regression using BIC

As discussed previously, the Bayesian information criterion (BIC) differs from AIC in that there is a greater penalty for number of parameters included. The BIC reduced model is called stepredo_bic.

The logitr base formula (Equation 23) is used with the resulting coefficients (Table 131).

Table 131: Coefficients predicting risk of RF using BIC as a reduction criterion

	Estimate	Error	z value	Pr(>|z|)	p
(Intercept)	-5.81	1.00	-5.80	6.70E-09	<0.001
ni	3.64	1.02	3.59	0.00033	<0.01
cdh	1.27	0.39	3.21	0.00133	<0.01
skeletal	1.7	0.46	3.76	0.00015	<0.001
tracheal	1.29	0.42	3.06	0.0022	<0.01

Using BIC pared the model down to 2 parameters i.e. NI and CDH. To select the best model, we compared the predictive performance of the AIC and BIC reduced models.

To assess goodness of fit, we compared the BIC-reduced model with the full (logitr) model.

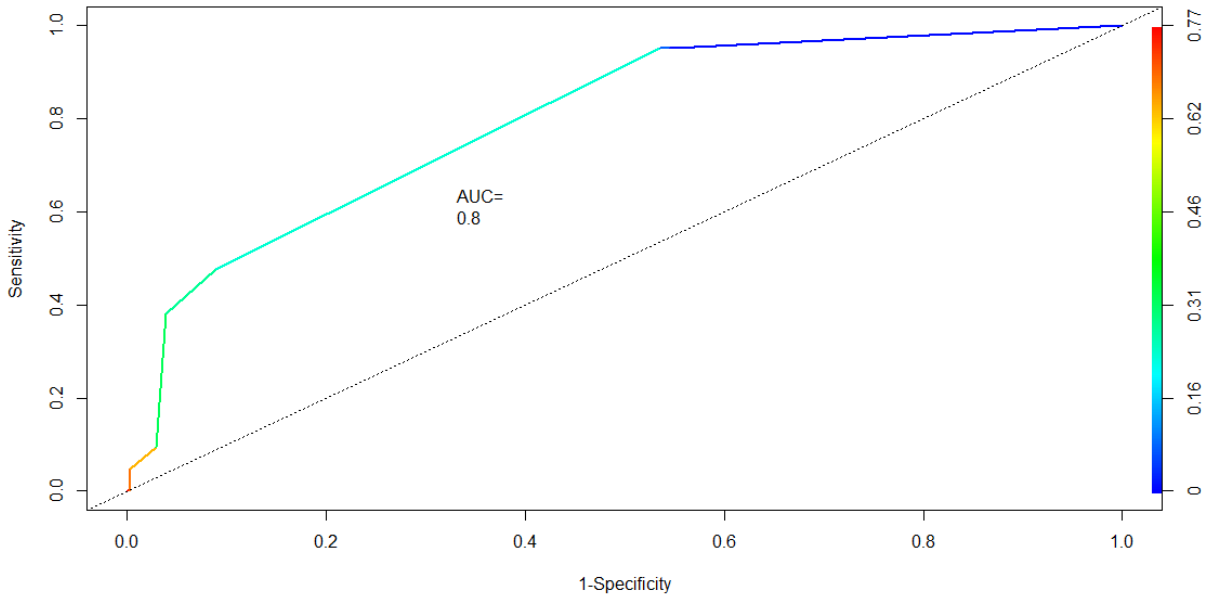
Figure 125: ANOVA comparing residual deviances of logitr versus stepredo_bic models

```
Analysis of Deviance Table

Model 1: redo ~ cdh + ni
Model 2: redo ~ tracheal + cleft + sleep + trachy + aspiration + cld +
    cdh + asthma + cardiac + ni + oatof + achalasia + swallow +
    chrom + consang + prem + haem + endocrine + renal + skeletal +
    bone + dental + oncology + metabolic + aresp + immune + cardsurg +
    eb + simplegord
  Resid. Df  Resid. Dev  Df  Deviance  Pr(>Chi)
1       753        317
2       726        268  27      48.9    0.0061 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The BIC reduced model has the greater residual deviance (317 vs. 268), therefore we expect it would fit the data less well compared to logitr.

The BIC-reduced model has an AUC of 0.78, making this a 'fair' classifier. The optimal cut-off is identified as 0.96. At this cut-off, the sensitivity is 95% and false positive rate is 46%.

Figure 126: Receiver operating characteristics for stepredo_bic model



The clinical implications were clear from the confusion matrix data. Using this classifier and a probability cut-off of 0.96, 22 of 23 children from the test data were correctly predicted for RF. However, the classifier incorrectly predicted 162 of 301 children to a receive a RF.

stepwise regression : AIC vs BIC

Analysis of variance (ANOVA) was used to examine whether the AIC and BIC reduced models differed significantly from each other. As the BIC-reduced model contains all parameters of the AIC-reduced model we can consider it to be a nested model.

```
anova(stepredo_bic, stepredo_aic, test="Chisq")
```

```
Analysis of Deviance Table

Model 1: redo ~ cdh + ni
Model 2: redo ~ tracheal + trachy + aspiration + cld + cdh + cardiac +
    ni + oatof + swallow + renal + skeletal


  Resid. Df  Resid. Dev  Df  Deviance  Pr(>Chi)
1       753       317
2       744       282   9      34.8  6.6e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The BIC-reduced model had the greater residual deviance, therefore we expect it to fit the data less well.

The goodness of fit parameters and predictive power for the stepredo_aic and stepredo_bic models are summarised below.

Table 132: Comparing the model goodness of fit and prediction parameters for AIC and BIC stepwise regression

	Logitr	**stepredo_aic**	**stepredo_bic**
Parameters	29	11	3
Null deviance	379	379	379
Residual deviance (ANOVA p-value)	268 (<0.01)		317
Residual deviance (ANOVA p value)		282 (<0.001)	317.00
Residual deviance (ANOVA p value)	268 (>0.5)	282	
AIC	328	306	323
BIC	467	362	337
AUC	0.84	0.86	0.78
Sensitivity	70%	78%	95%
False positive rate	15%	12%	54%

The AIC-reduced model had the highest AUC (0.86) and the lowest false positive rate (12%). Therefore, we accepted the stepredo_aic model as the optimal specification of the logistic regression model for RF. However as sensitivity remained low (78%) we investigated whether the reduced model could be improved further using bias reduction techniques.

Improving the model: bias reduction

RF is a rare event in both the training (52/736) and test data sets (23/324). We investigated whether bias reduction techniques can result in improved sensitivity and specificity.

We will use the reduced formula generated from AIC stepwise reduction as the base formula for bias reduction. As before, this process is automated using Kosmides'(256) brglm package for R. The formula used is based on the reduced model with 11 variables arising from stepwise (AIC) regression.

Equation 23: Formula for bias reduction regression model -brfitr

```
fm.r2<-redo~tracheal+trachy+aspiration+cld+cdh+cardiac+ni+oatof+swallow + renal + skeletal
```

Figure 127: Bias reduction modelling of redo fundoplication risk -brfitr

```
brfitr<-brglm(fm.r2, family=binomial(logit), data=redotrain, method="brglm.fit")

summary(brfitr)

Call:
brglm(formula = fm.r2, family = binomial(logit), data = redotrain,
    method = "brglm.fit")
```

The resulting coefficients are presented in the table below:

Table 133: brfit model coefficients

	Estimate	Error	z value	Pr(>|z|)	p
ni	2.81	0.83713	3.353	0.0008	<0.001
skeletal	1.93	0.48477	3.975	7.04E-05	<0.001
cdh	1.67	0.42661	3.907	9.35E-05	<0.001
aspiration	1.59	0.64552	2.458	0.01398	<0.05
renal	1.51	0.53455	2.821	0.00479	<0.01
oatof	1.31	0.46455	2.821	0.00479	<0.01
tracheal	1.20	0.49141	2.442	0.01461	<0.05
cardiac	0.31	0.34386	0.899	0.36892	
swallow	0.07	0.43021	0.159	0.87389	
tracheostomy	-0.47	0.66977	-0.703	0.4821	
cld	-1.39	0.61171	-2.273	0.02306	<0.05

Coefficients are transformed to odds ratios with 95% confidence intervals. Only predictive comorbidities are summarised below.

Table 134: Odds ratio and confidence intervals for the brfitr bias reduction model

		95% Confidence interval	
	Odds ratio	**2.5(%)**	**97.5(%)**
ni	15.9	4	423

skeletal	4.7	1.5	13.4
cdh	4.5	1.9	10.7
tracheal	3.6	1.3	9.6
oatof	2.9	1.1	7.6
cld	0.3	0.07	0.9

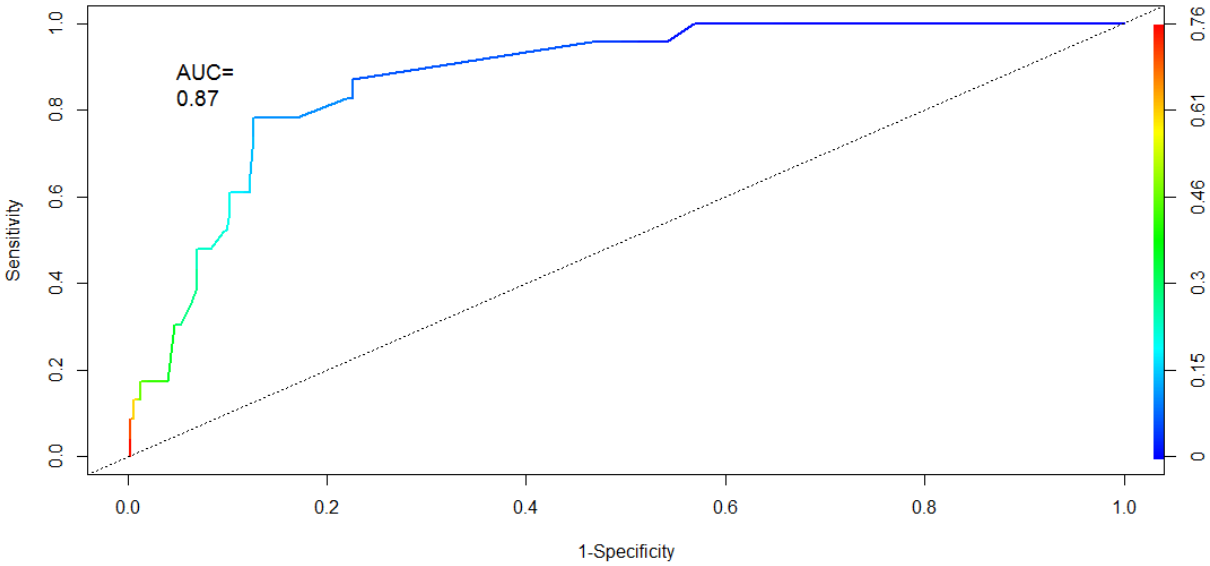
To assess predictive performance, we fit the test data using the brfitr model.

To assess goodness of fit, we compared information criteria for the various models. The BIC of the brfitr model was 362. This is 105 units less than that of the logitr model. This provides 'very strong' evidence that the data are better fitted by this model when compared to logitr model.

The brfitr model has a similar BIC to the stepredo_aic model (also 362) suggesting that there were no improvements in fit from brfitr compared to the stepredo_aic model.

On reviewing receiver operating characteristics, we found that brfitr has an AUC of 0.87. This indicates that its predictive performance is better than a coin toss, and marginally better than the stepredo_aic model (AUC=0.86)

Figure 128: ROC curve for the bias reduction model for RF (brfitr)



The probability cut-off that optimised sensitivity and specificity was 0.05. The model was 87% sensitive and 71% specific.

Summarising logistic regression models

The models are summarised and compared in Table 135 below:

Table 135: Logistic regression models for predicting risk of RF

Model	Description	Parameters	Predictive comorbidities	BIC	Sensitivity (probability cut-off)	FPR	AUC

Logitr	Basic logistic regression	29	7*	467	87% (0.05)	29%	0.84
Stepredo_aic	Stepwise reduction using AIC	11	6 **	362	83% (0.07)	12%	0.86
Brfitr	Bias reduction	11	6**	362	87% (0.05)	29%	0.87
* NI, tracheal, CDH, OATOF, skeletal, aspiration, CLD							
**NI, tracheal, CDH, OATOF, skeletal, CLD							

Compared to the logitr model, stepwise reduction pared down the input parameters. Interestingly, the predictive comorbidities and hence clinical inferences were very similar. One comorbidity i.e. aspiration was a significant predictor in the logitr model but not after stepwise reduction. The BIC for the logitr model was found to be the highest. The stepredo_aic model had the lowest lower false positive rate. Differences in AUC between the logitr and stepredo_aic models were marginal.

Comparing the stepredo_aic model and the brfit model, we found similar BIC and AUC. However, the stepredo_aic model has a lower false positive rate. Using the brfit model would result in prediction of twice the number of unnecessary RFs.

Therefore, the stepredo_aic model was selected as the best logistic regression model for predicting RF. The stepredo_aic formula summarises the 11 comorbidities of interest when predicting the risk of fundoplication.

Equation 24: Logistic regression formula containing parameters predictive for redo fundoplication

```
redo~tracheal+trachy+aspiration+cld+cdh+cardiac+ni+oatof+swallow+renal+skeletal.
```

The formula above has clinically important implications. The stepwise regression formula can be used as a tool to focus history-taking. Instead of working through a long list of comorbidities, the clinician can focus the interview on these key comorbidities.

The stepredo_aic model also identified comorbidities influencing risk of RF. These are summarised in the Table 136 below.

Table 136: Comorbidities that increase or decrease the odds of RF as identified by the stepredo_aic model

		95% Confidence interval	
	OR	2.5%	97.5%

ni	23.46	4.82	423.18
skeletal	4.74	1.52	13.42
cdh	4.64	1.93	10.75
tracheal	3.70	1.34	9.62
oatof	2.96	1.08	7.57
cld	0.27	0.07	0.80

Neurological impairment is the key risk factor predisposing to RF. However, note the wide confidence interval for the odds ratio associated with NI. This indicates that we cannot be precise about the effect size of NI on RF. This wide confidence interval hints at unmeasured / unobserved variables at work and probably reflects the wide spectrum of presentation of children with NI. Notably, the search for interactions predicting RF risk did not identify any interactions. This may mean that there are no 2-way interactions. It may also mean that there are more complex i.e. >2-way interactions.

Notably, NI was also a key factor in predicting the first fundoplication. As fundoplication is a prior requirement for RF, NI could be a confounding factor as it influences prior treatment.

Chronic lung disease is consistently found to reduce the risk of RF. In the discussion chapter, we will explore the literature and suggest explanations for these observations.

More immediately, we explore classification and regression methods in predicting risk of RF.

DECISION TREE MODELLING

Selecting and pre-processing the data

We partitioned the subset of patients of 1080 patients who had fundoplication. We observed a 70:30 training to testing ratio. This generated the training dataset “redotrain” with 756 observations of 29 variables and the test dataset “redotest” with 324 observations of 29 variables.

We used the same packages listed in the section on decision tress for primary fundoplication (page 291).

Modelling

A DT was fitted to “redotrain”.

Equation 25: Formula for decision tree model of RF

```
dt.redoraw<-rpart(redo~tracheal+cleft+sleep+trachy+aspiration+cld+cdh+asthma+cardiac+ni+oatof+achalasia+swallow+chrom+consang+prem+haem+endocrine+renal+skeletal+bone+dental+oncology+metabolic+aresp+immune+cardsurg+eb+ simplegord,
        data=redotrain)
```

The summary output of the model (dt.redoraw) is as follows of interest. All training data observations have been classified into the root node with no splits.

```
dt.redoraw
```

```
n= 756

node), split, n, loss, yval, (yprob)
      * denotes terminal node

1) root 756 52 FALSE (0.93122 0.06878) *
```

The error rate is 7%. This finding arises because there is no particular feature strongly predictive of outcome that can be used to divide the data.

This is a convergence problem i.e. the classification heuristic resolves the data to one terminal node. There are two explanations for this observed convergence. First, data are unbalanced with a relatively few observations of the outcome of interest (RF rate is only 7%). Secondly, there may be highly predictive variables causing bias towards a singular terminal node.

Where convergence is created by a low event rate (e.g. RF rate = 7%), one mitigating strategy is improving the mis-classification cost.

Improving the model: introducing misclassification cost

In the confusion matrix, we observed that the false negative rate was high. Therefore, one solution to the conversion problem is introducing a misclassification cost for false negative predictions.

Table 137: Confusion matrix and performance statistics for dt.redoraw

		Actual redo fundoplication status	
		True	False
Predicted	True	0	0
	False	23	301
Statistics			
Accuracy	0.93		
Sensitivity	1		
Specificity	0		
PPV	0.29		

Therefore, using a cost matrix, we introduced a cost to false negative classification. False negative classifications were weighted arbitrarily at 10, false positive classifications were weighted arbitrarily at 1. Clinically, this model behaviour is analogous to penalising a surgeon who fails to perform a redo for a patient who needs this intervention.

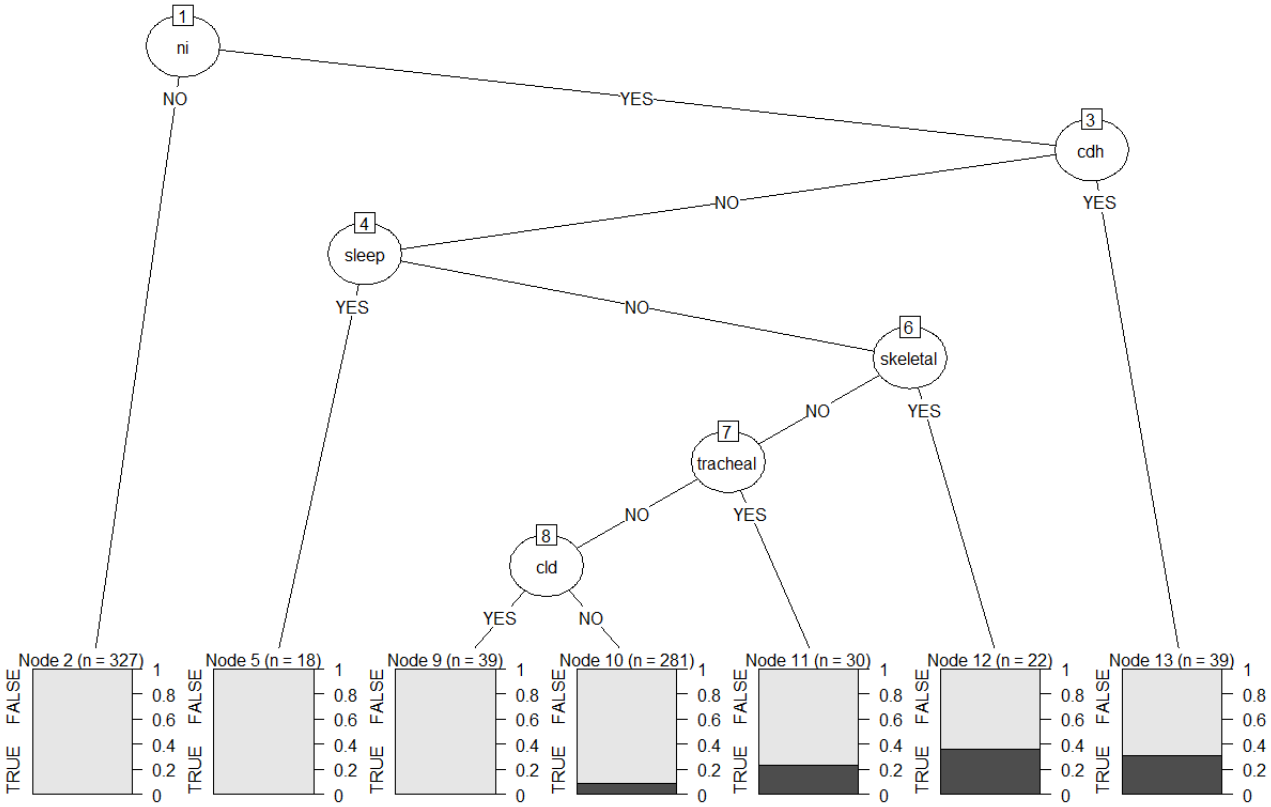
We introduced a false negative misclassification cost of 20 and re-modelled the training data (dt.redocost).

Equation 26: dt.redocost- a DT model of RF with a false negative misclassification cost of 20

```
dt.redocost<-
rpart(redo~tracheal+cleft+sleep+trachy+aspiration+cld+cdh+asthma+cardiac+ni+oatof+achalasi
a+swallow+chrom+consang+prem+haem+endocrine+renal+skeletal+bone+dental+oncology+m
etabolic+aresp+immune+cardsurg+eb+ simplegord,
        data=redotrain,
        method="class",
        parms = list(loss = matrix(c(0, 20, 1, 0), nrow = 2)))
```

This cost resulted in branching of the DT. The root node was split to 4 inner nodes and 5 terminal nodes.

Figure 129: DT algorithm with a misclassification cost applied to false negative classifications



.

Key

YES= had primary fundoplication NO= no primary fundoplication TRUE= had RF FALSE = No RF

This costed model is notable for its first split. The absence of neurological impairment leads to a pure terminal node. No neurologically normal patients had a RF.

The confusion matrix from the costed model demonstrates that a model with false negative cost improves classification. The model is less accurate but more sensitive (54%, was 0%). The number of false negative classifications has reduced. However, we note an increase in false positive classification.

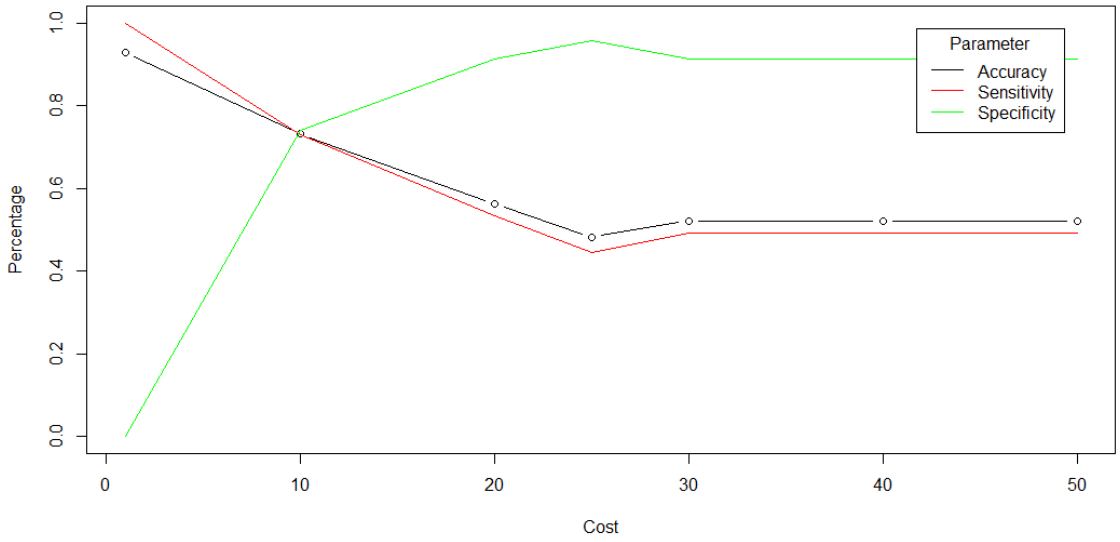
Table 138: Confusion matrix and performance metrics for dt.redocost which adds a false negative misclassification cost of 20

		Actual fundoplication status	
		True	False
Predicted	True	21	140
	False	2	161
Statistics			
Accuracy	0.56		
Sensitivity	0.53		
Specificity	0.93		

A misclassification cost was arbitrarily chosen to prove the concept that classification cost could resolve the convergence problem. Furthermore, only one of the terminal nodes is pure, suggesting that the model can be further optimised.

To identify the optimal value for false negative cost, we assessed performance metrics for a range of classification costs.

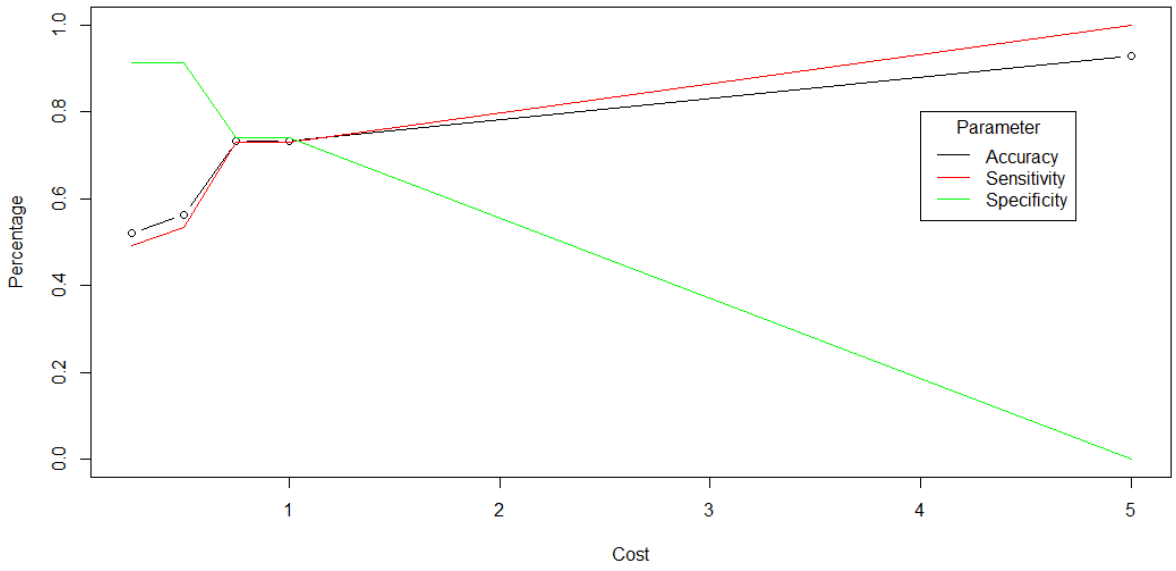
Figure 130: False negative cost versus [sensitivity, specificity and accuracy]



As false negative cost increased, specificity increased and both accuracy and sensitivity decreased. The optimal false negative cost is where the both sensitivity and specificity are maximised i.e. a weight of 10.

We also investigated the impact of a false positive cost. Maintaining the negative cost parameter at 10, we vary the false positive cost matrix parameter from 0-50.

Figure 131: False positive cost vs [accuracy, sensitivity and specificity]



Both the sensitivity and accuracy of prediction increases with false positive cost, peaking at 5. Sensitivity and specificity converged at a false positive cost of 0.75. As the value is less than one, this means that the algorithm 'rewards' rather than penalises misclassification of observations as false positive. The clinical correlation of this observation is that to optimise both sensitivity and specificity some unnecessary RFs will occur. This implication is both realistic and pragmatic.

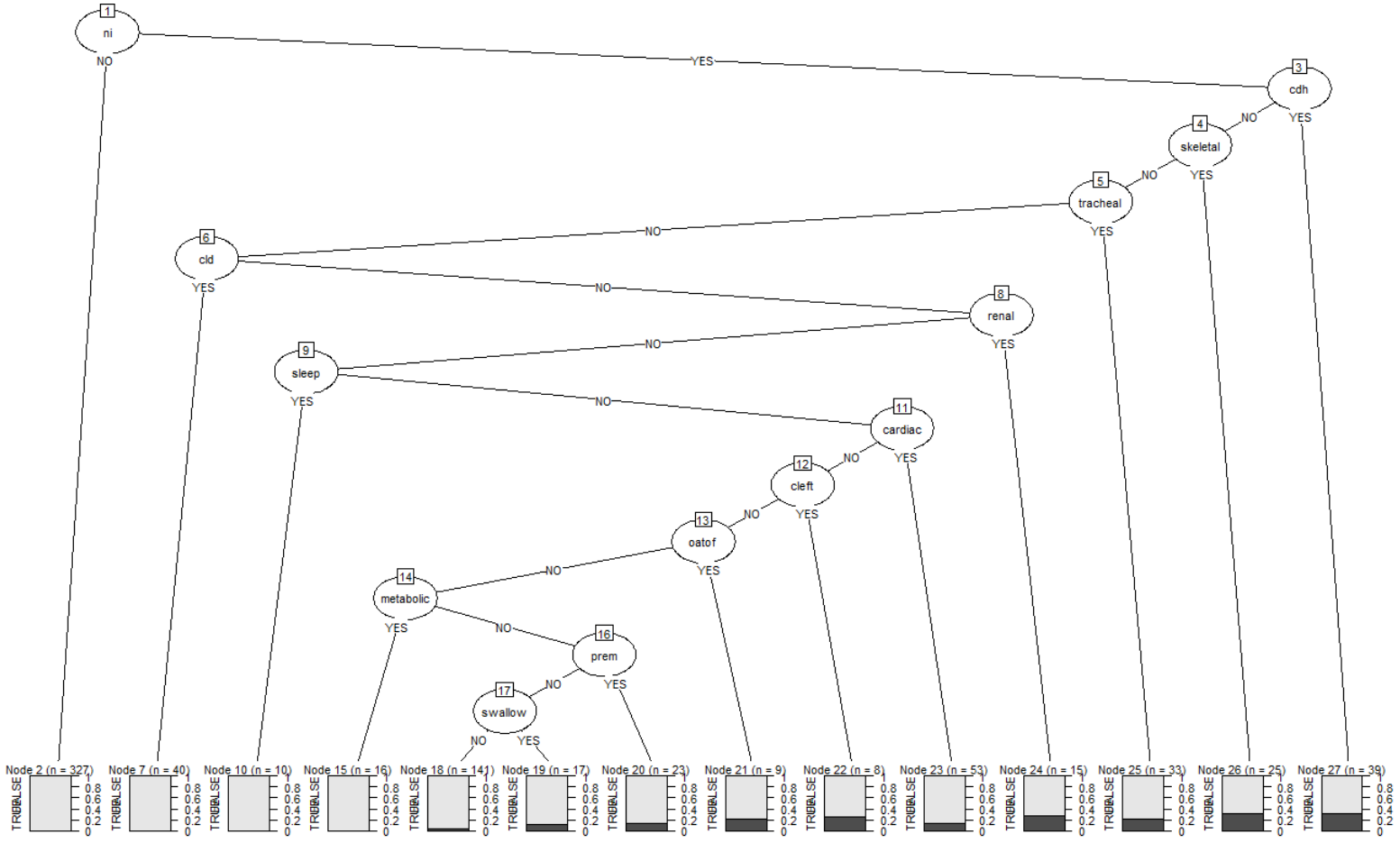
The model was restated (dt.redocost1).

Equation 27: dt.redocost1 has a false negative cost of 20 and a false positive cost of 0.75

```
dt.redocost1<-
rpart(redo~tracheal+cleft+sleep+trachy+aspiration+cld+cdh+asthma+cardiac+ni+oatof+achalasi
a+swallow+chrom+consang+prem+haem+endocrine+renal+skeletal+bone+dental+oncology+me
tabolic+aresp+immune+cardsurg+eb+ simplegord,
        data=redotrain,
        method="class",
        parms = list(loss = matrix(c(0, 10, 0.75, 0), nrow = 2)))
```

The model was re-drawn resulting in the algorithm below.

Figure 132: dt.redocost1 has a false negative cost of 20 and a false positive cost of 0.75

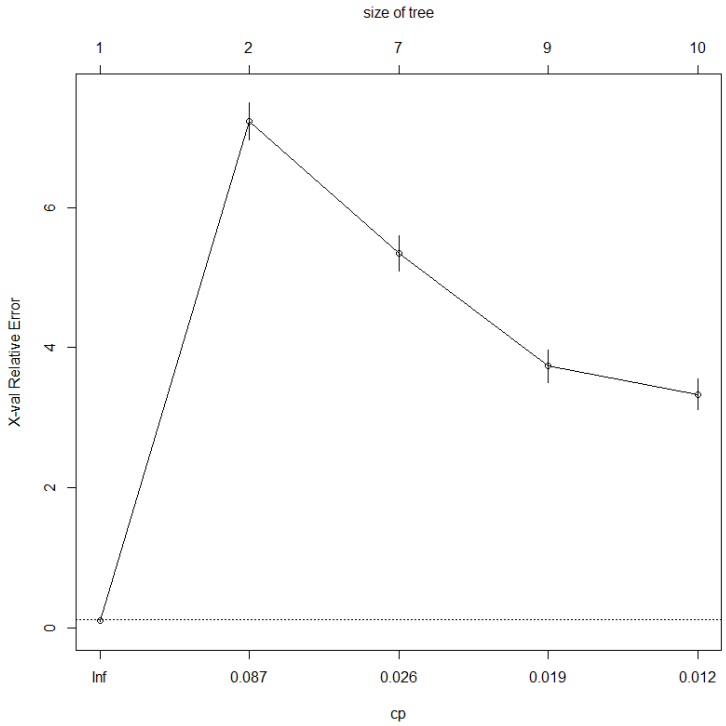


Key

YES= had primary fundoplication NO= no primary fundoplication TRUE= had RF FALSE = No RF

The resulting tree has 14 terminal nodes with multiple inner node branching. There are too many branches to garner useful information. This classifier could be further improved by pruning. Post-pruning using the complexity parameter was applied.

Improving the model: pruning



Cross-validation error was lowest at a complexity parameter of 0.01. This was associated with 9 splits.

Table 139: Complexity parameter (CP) metrics for pruning the dtredocost1 model

	CP	N splits	Relative error	Cross-validated error	Std. deviation
1	0.43	0	1	20	0.20
2	0.013	1	0.56	10.7	0.39
3	0.010	6	0.48	8.7	0.38
4	0.015	8	0.53	3.2	0.216
5	0.01	9	0.52	3.2	0.218

The relative error value did not fall below the error level. Therefore, we did not expect to see improvement in the model as the complexity was increased.

Figure 133: Using complexity parameter to identify the optimal size of the cost-adjusted RF DT

The tree was pruned to a complexity parameter of 0.01 and re-plotted (dt.pruneredo). There remained 9 inner nodes and 10 terminal nodes. Pruning by complexity parameter did not change or improve the tree nodes, nor did it improve the error rate for each terminal node. The resulting tree was identical to **Figure 133**.

Improving the model: cross-validation

We wanted to find out if increasing the number of trials could improve the tree. We applied manual cross-validation using 50 trials (dt.redo50). This resulted in the same tree achieved in Figure 132 above. Therefore, pruning the model further by cross-validation did not improve either sensitivity or accuracy.

Predictive performance

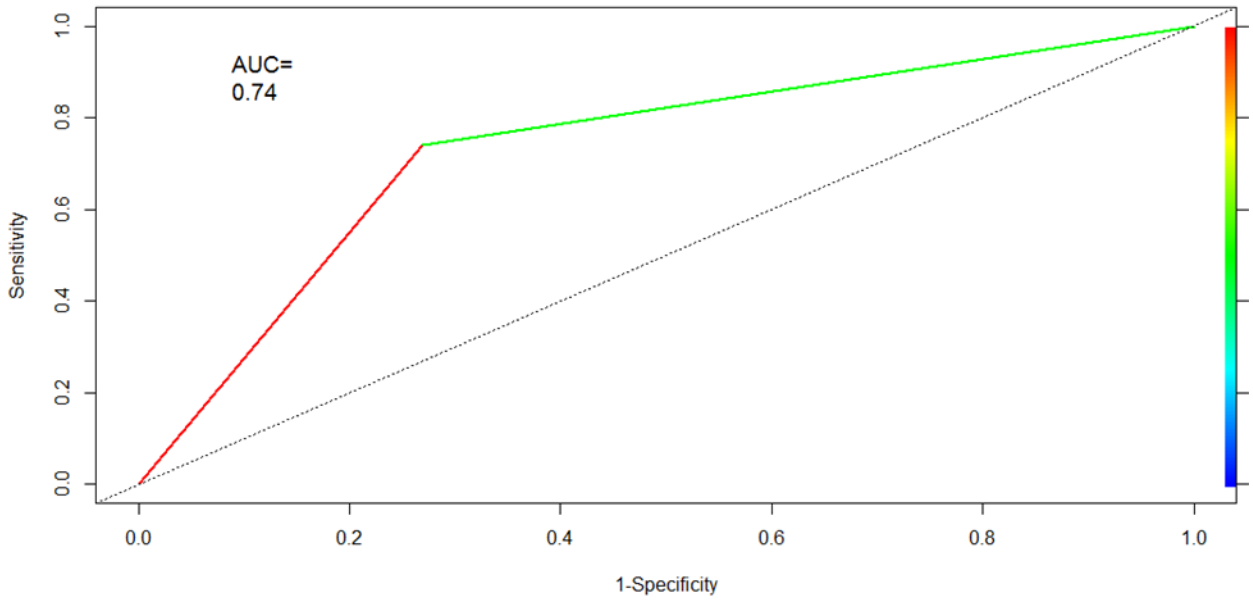
While the PPV of dt.redo50 was observed to be high (97%), the negative predictive value was low at 17%. The clinical implication is that patients requiring redo procedures would not be identified in 4 of 5 cases if this algorithm was applied.

Table 140: Confusion matrix and performance metrics for dt.redo50, pruned by complexity parameter

		Actual redo fundoplication status	
		True	False
Predicted	True	17	6
	False	81	220
Statistics			
Accuracy	0.73		
Sensitivity	0.73		
Specificity	0.74		
PPV	0.97		
NPV	0.17		

An ROC curve was used to assess how this cost-adjusted classifier compared to other classifiers of RF. With an AUC of 0.74, the cost-adjusted DT was a fair predictor of RF risk. The probability cut-off was 0.6. This corresponded to a 73% sensitivity and 73% specificity.

Figure 134: ROC profile for dt.redo50, a cost-adjusted, pruned and cross-validated model of RF



Summarising DT modelling

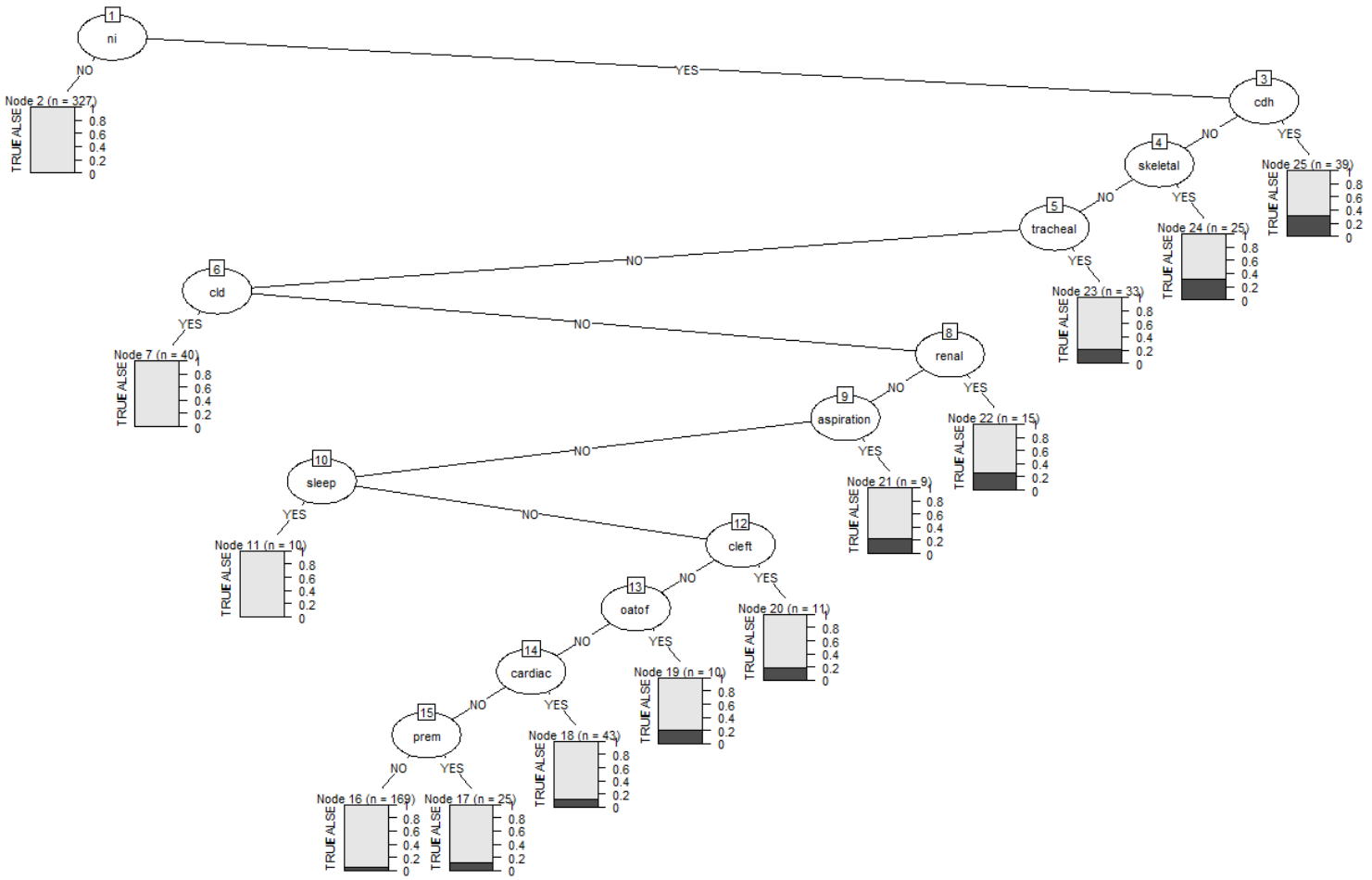
Dt.redo50 , a cost-adjusted, pruned and cross-validated tree, is accepted as the best specification of the RF using DT modelling. Variables predictive of RF were summarised in order of importance as splitting criteria. The variable importance is calculated by comparing mean decrease in Gini impurity index for each comorbidity.

Table 141: Variable importance as splitting criterion for dt.redo50 model

Comorbidity	**Overall**
ni	92
cdh	40
skeletal	27
cld	11
cardiac	10
asthma	9.6
aspiration	6.8
chrom	5.7
cleft	3.4
metabolic	3.0

In common with the logistic regression model for RF, NI was the most important factor predicting RF. The dt.redo50 tree is presented in below.

Figure 135: dt.redo50 model, a cost-adjusted, cross-validated DT for RF



Key

YES= had primary fundoplication NO= no primary fundoplication TRUE= had RF FALSE = No RF

The clinical implications of the loss-adjusted model were summarised as follows:

1. Neurological impairment was the key criterion associated with redo-fundoplication. In the absence of neurological impairment, 97% of patients did not receive a RF.
2. When NI is associated with CDH, 30% of patients received RF.
3. Notably, chronic lung disease appeared protective. When chronic lung disease was present, 98% of patients did not receive a RF.

This resulting algorithm was a fair predictor of RF. Notably, after the first node, all further splits only described 7% of the data. Inevitably, this convergence problem arose due to the low event rate of RF i.e. 7%.

MODEL SELECTION: RISK OF RF

We modelled the risk of RF using both logistic regression and DT modelling. To summarise, we compared the predictive performance of each model. The stepredo_aic was the best implementation of logistic regression modelling. The dt.redo50 was the best implementation of DT modelling.

Model	**Description**	**Number of parameters**	**Predictive comorbidities**	**Sensitivity (probability cut-off)**	**False positive rate**	**AUC**
Stepredo_aic	Logistic regression following stepwise reduction using AIC	11	6*	83% (0.07)	12%	0.86
dt.redo50	DT modelling	29	10~	74%(252)(1)	17%	0.74
* NI, skeletal, CDH, OATOF, CLD, tracheal						
~ NI, skeletal, CDH, skeletal anomalies, OATOF , cardiac, tracheal, CLD, cleft, swallow, prematurity						

Of the two modelling approaches, logistic regression yielded the simpler model. The DT models had poorer predictive performance than the logistic regression models overall. The DT model, although algorithmic, had too many terminal nodes to be useful in clinical setting.

The DT model did, however, generate variable importance data. This is confirmatory of the yield from logistic regression modelling. The comorbid variables predicting RF common to both models are NI, skeletal anomalies, OATOF, CLD and tracheal anomalies.

SUMMARY: KNOWLEDGE DISCOVERY FROM DATA MODELLING

We return to where we began: A surgeon is in her clinic and is counselling a parent of a child with GORD. Can we answer the questions set at the beginning of this modelling exercise? The odds yield from the logistic regression models has clinical utility in answering this question and will be useful to inform consent.

A parent might ask: “Given my child’s complex needs, are they likely to require fundoplication?”. We can now answer that NI increases risk of fundoplication at least 5-fold. We have odds estimated for other comorbidities too.

A parent may ask : ” What is the risk of failure of fundoplication?”. The surgeon might answer: “As your child has an NI, they are 20 times more likely to have a failed fundoplication compared to a child without NI.

In a circumstance where a fundoplication has already taken place, the question might be: “What is the risk of this fundoplication failing?” . The surgeon may answer: “As your child has NI, they are 60 times more likely to have a redo compared to a child without this condition. “.

Data modelling has also presented a clearer picture of the relative influence of comorbidities. These key comorbidities are:

- Neurological impairment
- Prematurity
- OATOF
- CDH
- CLD
- Skeletal anomalies

Interestingly, we have evidence of interaction of comorbidities. NI interacts with cleft and tracheal anomalies, cardiac disease and swallowing disorders to increase fundoplication risk.

Data mining has also identified risk factors for RF. These are:

- NI
- Skeletal anomalies
- CDH
- OATOF
- CLD
- Tracheal anomalies

NI is a key comorbidity predicting risk of first fundoplication. It is interesting that it persists as *the* key comorbidity predicting RF.

Another finding of interest here is oesophageal disease. There is a high preponderance of patients with comorbidities associated with oesophageal dysmotility that require RF. These comorbidities are CDH, OATOF and swallowing disorders. The role of oesophageal dysmotility will be explored further in the discussion chapter.

Prematurity is implicated in the first fundoplication. However, it is not a risk factor for RF. This observation may be linked to the phenomenon of catch-up growth. As premature infants move from infancy into childhood, those that survive experience a balancing of morbidity. Their health status approaches that of term peers with passing months. Therefore, prematurity may no longer be a risk factor by the time the child is old enough to be considered for a RF.

There are also respiratory conditions implicated in predicting RF. These are CLD, CDH, OATOF and tracheal anomalies. Explanations may include the cross-over between respiratory and reflux symptoms. Also, there may be competition between feeding and breathing functions in children with chronic respiratory disease, leading to failure to thrive.

In summary, this data mining exercise has succeeded in elucidating and enumerating risk factors for fundoplication. It has also been useful in understanding how the disease construct can be better modelled. In Section VI: Discussion (p388), we apply the knowledge gleaned from this data mining exercise to present a risk stratification algorithm for patients with GORD. Having explored which children need an operation, we turn our focus to which operation should be done and when. The next section addresses the timing of gastrostomy and fundoplication in children with NI.

SECTION V: GASTROSTOMY WITH OR WITHOUT FUNDOPLICATION: REMOS TRIAL

Children with neurological impairment frequently have GORD and often require tube feeding for inadequate or unsafe oral intake. Initially, nasogastric tube (NGT) feeding is used. If prolonged NGT feeding is established, the procedure to insert a gastrostomy tube (GT) is undertaken.

Historically, gastrostomy placement was done through an open Stamm gastrostomy tube (OGT) procedure. Developments in endoscopy led to the development of paediatric endoscopy and gastrostomy tubes. First described in 1980(268), percutaneous endoscopic gastrostomy tube (PEG) placement has become the standard. More recently, laparoscopic gastrostomy tube (LGT) has emerged as an option, as has the radiologically inserted gastrostomy (RIG).

Previous studies have demonstrated the feasibility and safety of laparoscopic gastrostomy(269). Laparoscopy offers the advantage of direct visualisation of the junction between the stomach and abdominal wall. Visualisation avoids inadvertent interposition of colon between the two, therefore avoiding the known complication of gastro-colo-cutaneous fistula. Another advantage of laparoscopic gastrostomy is placement of a balloon-retained low profile tube in the first instance. This avoids a second anaesthetic to change from a flange-retained tube (e.g. PEG) to a balloon retained tube. For children with NI, anaesthetic-sparing procedures are particularly attractive(269).

LGT, GT and PEG procedures require a general anaesthetic. The question then arises: in a child with GORD, should concomitant fundoplication be performed at the time of gastrostomy?

When open gastrostomy was the standard of care i.e. before the advent of PEG and laparoscopy, it was feasible to extend the incision to perform a concomitant fundoplication. Indeed, the fundoplication was considered an adjunct to OGT(270). In a 1996 review of GORD in childhood, Professors Fonkalsrud and Ament(271) wrote:

"… for children with neurologic impairment who require a feeding gastrostomy and who frequently benefit from a concomitant fundoplication with or without modified pyloroplasty. In view of the low cost and low risk of complications after fundoplication, most paediatric surgeons favour performing operation early in children with symptoms before more severe complications of GER develop."

With the development of the PEG, the question of whether to perform a fundoplication concomitantly with gastrostomy became more troublesome. The minimally invasive endoscopic approach used to insert a PEG did not easily extend to open or laparoscopic fundoplication. Furthermore, PEG placement could be done by both gastroenterologists and surgeons. However, a surgeon would be required if concomitant fundoplication were indicated.

With the advent of LGT, the ease of concomitant laparoscopic fundoplication (LF) at the time of LGT has re-emerged. Although data from meta-analysis(272) favour LGT where outcomes are intra-and post-operative complications, PEG remains a quicker and more popular procedure(273).

Variations in surgical practice demonstrate the lack of consensus. In a survey, paediatric surgeons working in Denver, Colorado reported performing concomitant fundoplication in 52% of patients during their reporting period(205). Indications for fundoplication included GORD, failure to thrive, risk of aspiration, neurological impairment and chronic respiratory disease. Failure of medical management was reported to be a factor for 72% of patients, although evidence for failure of medical therapy was

found in only 15% of patient notes reviewed. Furthermore, aspiration was stated as an influence in 81% of cases. However, there was actual clinical history of aspiration in only 6% of cases. These findings suggest that perceived rather than observed risk was a deciding factor. In a subsequent paper, the same authors compared concomitant fundoplication rates in a neighbouring state and found that the Colorado rate (60%) was higher than the north Carolina (43%) rate for similar patient groups(274).

In summary, review from uncontrolled studies supports both points of view. The overlap of symptoms of feeding difficulties and GORD in children with NI has led to conflicting data(52). Generalizability of data is further limited by its fragmentation by the type of gastrostomy device utilised (52). Notably, a Cochrane review by Vernon-Roberts and Sullivan(52) failed to identify any RCT addressing this question. Authors were unable to come to a definitive answer on concomitant fundoplications with gastrostomy placement and called for an RCT targeted at this question.

EFFECT OF GASTROSTOMY FEEDING ON GORD

One argument for concomitant fundoplication is that gastrostomy feeding may exacerbate GORD symptoms. Hament et al(58) compared post-operative reflux, assessed by pH study, in children with PEG +/- fundoplication. A positive pH study was found in 2.5% of the PEG only group compared with 8% of the PEG with fundoplication group. Post-operative vomiting was higher in the PEG plus fundoplication group (41%) compared to the PEG only group (18%). Although both groups had improved reflux rates, there were two patients who did not vomit pre-op who had vomiting post PEG-only placement. In a similar study, Launay et al(211) studied 20 children having PEG placement. Reflux index improved or normalised in 6 children with a prior pathological pH study. In thirteen children with a prior normal study, the post procedure study remained normal. Only one child, who had a normal pre-PEG pH study, had an abnormal study after PEG. In a retrospective case series, Peters et al(275) found that a PEG in combination with medical therapy controlled reflux in the majority of cases.

Wilson et al(276) identified 28 patients with pathological pH study who also required PEG for inadequate oral nutrition intake. Of these, 25 had symptoms of GORD which completely resolved or decreased in frequency following PEG placement. They concluded that concomitant fundoplication is not necessary. However, assessment of symptoms was not systematic or blinded and no control (PEG+fundoplication) group was available for comparison.

Using pH-MII, Toporowskja-Kowalska et al(212) measured pre- and post-PEG reflux episodes in 15 patients with neurological impairment. The mean time pH-MII was performed after PEG formation was 7.35 (±0.71) months. They found no significant difference in pre- and post- gastrostomy MII parameters.

NEED FOR FUNDOPLICATION AFTER GASTROSTOMY

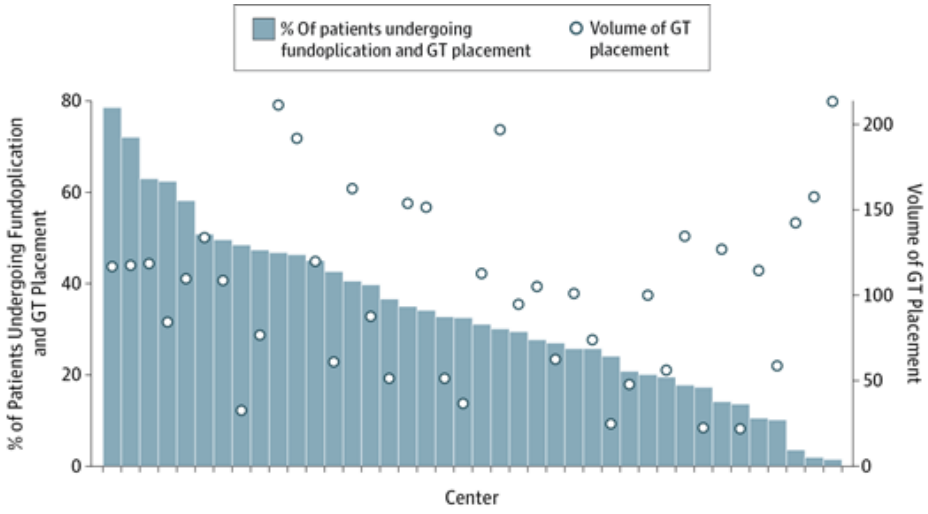
A second argument for concomitant fundoplication is that, in children with GORD, a significant proportion of them will go on to require fundoplication later. Therefore, concomitant fundoplication limits the episodes under general anaesthetic.

Vishwanath et al.(277) found that only 17% of children who had a gastrostomy went on to require fundoplication. Wadie et al(278) compared patients having gastrostomy alone with those having gastrostomy plus fundoplication, and found that although GORD symptoms were more frequent in the gastrostomy only group (16% versus 10%), overall post-operative complications were higher in the

gastrostomy with fundoplication complication group (49% vs. 33%). This group recommended against prophylactic fundoplication together with gastrostomy placement. In a retrospective series of 70 children who had PEG, Van der Merwe(279) found that 17% required fundoplication post-PEG placement. Langer et al(217) studied 50 patients and found that 17% required subsequent fundoplication. Wheatley et al(280) had prospectively studied 148 children referred for OGT. Of the 43 with a normal pre-operative pH-study, only 6 developed GORD symptoms and 5 (11%) required fundoplication.

Contrary evidence emerges from a large retrospective review. Barnhart et al(281) searched administrative databases of 42 children's hospitals in the US. They identified 4163 infants who underwent either gastrostomy with or without fundoplication. Using propensity score matching, they compared outcomes between groups. The key finding was an increased rate of pneumonia, aspiration and oesophagitis admissions in the first year following gastrostomy in infants who had concomitant fundoplication.

Figure 136: Percentage of Infants Who Underwent Concomitant Fundoplication by Center and Overall Volume of Gastrostomy Tube (GT) placement in 42 children's hospitals between 2005-2010



Reproduced with permission from: DC, Hall et al(282).

The study by Barnhart et al(281) reveals large variations in practice. Concomitant fundoplication rate varied from 2-80 % depending on institution polled. These variations are borne out in literature, where fundoplication rates after gastrostomy placement range from 14-91%(283). The figure above is important because it suggests a significant role for surgeon preference. Despite the absence of controlled studies or strong prospective study evidence, some surgeons will offer concomitant fundoplication.

One reason for this practice tradition is an expectation that medical therapy will fail in NI children. Children with NI are more likely to require fundoplication than non-NI children(49) and therefore are over-represented amongst children undergoing fundoplication, with a cohort prevalence of 20-40%(35).

This has created a tautology: NI children are offered fundoplication because it is believed that they are more likely to require one later. However, evidence from controlled studies to support this surgical belief is lacking. Data from uncontrolled study may have severity bias, where patients with more severe reflux were allocated to gastrostomy with fundoplication. This likely bias makes the data difficult to interpret. Furthermore, contrary to the opinion offered by Fonkalsrud and Ament(271), fundoplication is not a 'low risk' procedure. Children who have fundoplication with gastrostomy experience more frequent complication(64). Fundoplication is associated with a higher recurrence rate, morbidity and mortality in this group of children (2,44). Indeed, fundoplication is a risk factor for gastrostomy revision(284) thus threatening the primary indication for surgery in the first place.

In a meta-analysis, Noble et al(285) commented on the quality of research papers attempting to answer this question. Most published evidence arises from small case series. The data are limited by unsystematic pre and post gastrostomy diagnostics. NI is poorly characterised in most articles reviewed. Changes to the feeding regime pre- and post- gastrostomy are not documented. For these reasons, a study like the REMOS trial was urgently needed.

CHAPTER 2: THE REMOS TRIAL

The REMOS trial (Reflux: Medical or Surgical) was established in 2008. The aim was to investigate whether fundoplication or medications are more effective in controlling reflux in NI children post gastrostomy.

The hypothesis stated:

"Fundoplication is superior to medical treatment of gastro-oesophageal reflux in children with neurological impairment requiring a gastrostomy."

The REMOS trial was commenced by my predecessor at the ICH (Mr R. Peters 2008-2010). When I took over as trial coordinator (April 2010), key elements were already in place i.e. trial protocol and research and ethics committee (REC) review. My predecessor had begun to recruit (a single patient) when he decided to leave the project. Subsequently, my tasks became patient recruitment, data collection, analysis and report writing. Therefore, majority of the work presented in this opus is my own.

As trial design was established prior to my involvement, I can take neither credit nor responsibility. However, this distance from trial design allows a dispassionate assessment of strengths and limitations. The trial protocol is available for review in Section V Appendix (p517).

CONSORT 2010(286) guidelines for reporting randomised controlled trials have been observed in the methods and results presented below.

DESIGN

The trial was designed as a non-blinded randomised controlled trial. Patients were randomly allocated to either:

- have gastrostomy with fundoplication and no medical therapy, or
- gastrostomy without fundoplication but with medical therapy.

Laparoscopic gastrostomy (LGT) was selected rather than PEG or GT to allow easy adjunct of the fundoplication procedure.

The trial site was as UCL Institute of Child Health/ Great Ormond Street Hospital, London. The principal investigator was were Professor Agostino Pierro. The trial was registered with the institutional Research and Development department and approved by the London Bloomsbury Research Ethics committee (08/H0713/99). The trial was funded by SPARKS charity.

PARTICIPANTS

The target population was children with neurological impairment and feeding difficulties referred to Great Ormond Street Hospital paediatric surgery and gastroenterology departments for gastrostomy.

Inclusion criteria

Inclusion criteria were stated as fulfilment of all three criteria below:

1. neurological impairment with a referral for gastrostomy and/or fundoplication
2. presence of GOR symptoms including one or more of:

a. vomiting

b. recurrent aspiration pneumonia (>3 episodes in a year requiring antibiotic administration)

c. anemia (defined according to WHO criteria(287))

d. failure to thrive (weight < 2nd percentile for age for at least 6 months);

3. documented GOR on 24h oesophageal pH monitoring (reflux index ≥10%).

Exclusion criteria

Children with the following conditions were excluded

1. structural anomalies

a. hiatus hernia (> one vertebral body)

b. intestinal malrotation

c. gastric outlet obstruction or severely delayed gastric emptying on upper GI contrast study;

2. previous gastrostomy, fundoplication or other abdominal surgery;

3. acute life-threatening events and/or apnoea associated with GOR and requiring urgent operative treatment.

Recruitment and informed consent

Patients were identified at general surgery and gastroenterology clinics. Patient notes were reviewed to identify patients newly referred primarily for a gastrostomy. With the approval of the clinician and the patient/carer, the researcher would sit in during the consultation. If inclusion criteria were met and no exclusion criteria identified, the researcher would then request time to consult with the parent.

The researcher would describe the study, its risks, benefits and alternatives. Parents were given the following materials:

- PARENT INFORMATION LEAFLET Version 2.0 7/11/2008 (Section V Appendix, p524)
- PARENTAL CONSENT FORM Version 2.0 22/02/2013 (Section V Appendix, p524)

Parents were given time to consider. At the next clinic visit, parents willing to participate would sign the informed consent form.

Investigations

Following informed consent, an assessment of the following parameters was made.

1. 24-hour pH- MII study

2. upper GI contrast study to exclude gastric outlet obstruction, intestinal malrotation and hiatus hernia

3. gastric emptying time measured by octanoate breath test

4. gastrointestinal symptoms and quality of life questionnaires

In patients meeting inclusion criteria, baseline assessment of nutritional indices was performed. These were:

1. assessment of dietary intake (carbohydrate, fat and protein) using a 3-day diet diary

2. weight centile , weight velocity, height, BMI

3. body composition measured by 3 methods to increase precision

a. skinfold thickness to measure subcutaneous fat depot

b. bioelectric impedance analysis which measures impedance of the body to a small electric current, thus estimating fat free mass

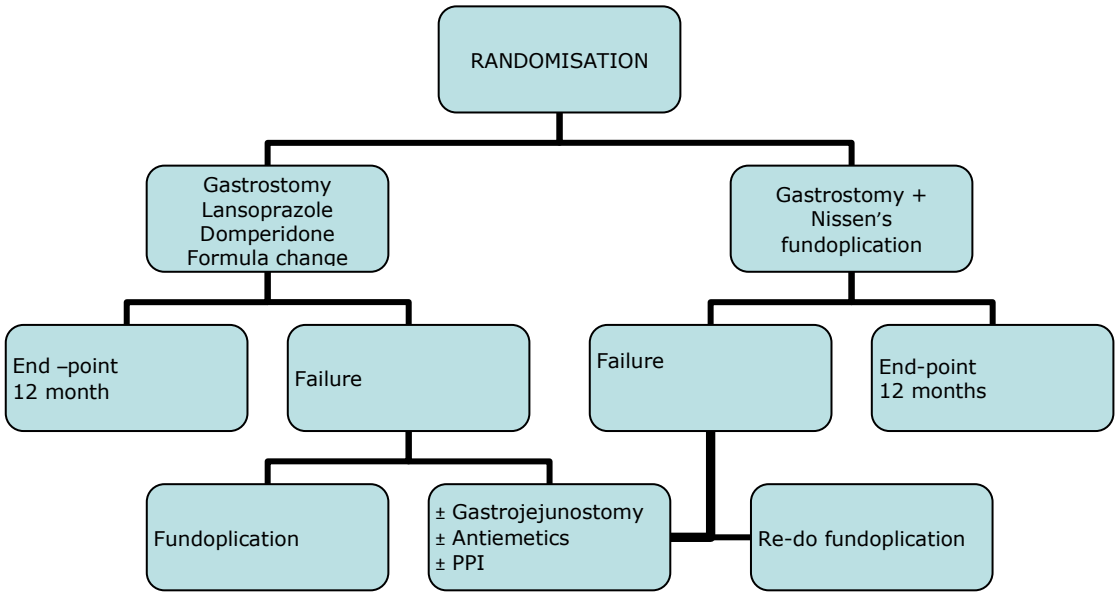
c. deuterium dilution test to estimate total body water and fat free mass.

RANDOMIZATION

Patients were randomised to either:

1. laparoscopic gastrostomy with maximal medical treatment for GOR (Medical / Group A)
2. laparoscopic gastrostomy and Nissen fundoplication (Surgical / group B)

Figure 137: Schema of patient pathway in the REMOS trial following randomisation



Acknowledging risk factors for GORD, a matching procedure was utilised to ensure that participants in groups A and B were as similar as possible. The matching procedure used was randomisation with minimisation. At enrolment, patients were be ranked by minimisation criteria. A weight would be assigned for each criterion met. Patients would then be allocated progressively to ensure minimisation scores in either group matched closely.

The minimisation criteria were stated as:

1. age [<1 year; 1-4 year; >4 year]
2. severity of reflux index at pH study [5-10%, 10-20%; >20%]
3. weight Z-score [<-2.2, -2.2 to -1.6 , >-1.6 to -0.5, >-0.5])
4. triceps skinfold thickness Z-score [<-1.6, -1.6 to 0.5, >0.5]
5. degree of neurological impairment based on GMFCS score [levels I-III, level IV, level V]
6. presence of associated anomalies (yes/no);
7. medical treatment for GOR [none, <6 months before entry, >6 months before entry].

Sequence generation

The minimisation applet (SIMIN) was developed at ICH by in collaboration with Prof. A. Wade (ICH Population, Policy & Practice Programme)(288). This applet was the implementation of a matching

method that enabled prospective randomisation with minimisation as patients were recruited. When a new patient was recruited, the algorithm would 'consider' the imbalance in selected prognostic factors in the prior cohort. The new recruit would be weighted according to the prognostic factors. Based on the weighting, they would be allocated. The heuristic is to make treatment groups as similar as possible with respect to prognostic factors i.e. minimise the difference. An element of pure randomization is also included in the minimization program to ensure that patient allocation is not completely predictable, which would introduce a bias. This is analogous to flipping a biased coin.

The inputs were the patient identifier and the minimisation characteristics. The output was the randomisation arm.

Allotment and concealment mechanism

Once a patient was allocated to a randomisation arm, this would be documented in trial documentation and patent notes. The patient consent form included in the notes was marked with A or B depending on the randomisation arm. Further, the surgeon would be informed directly of the randomisation outcome. As the study was neither single or double blinded, no concealment procedures were necessary. Blinding was thought to be neither ethical nor practicable.

Intervention

Under a general anaesthetic, all children had a laparoscopic gastrostomy. Depending on randomisation arm, some children would have a concomitant fundoplication (Group B/ Surgical). Others would have no fundoplication (Group A/ Medical). The medical group would continue anti-reflux medication (ARM) post-procedure.

For the purposes of the trial, the ARM protocol was:

- omeprazole up to 3mg/kg/d or a maximum of 80mg/day)

1 plus

- domperidone (up to 2.4mg/kg/d up to 80mg/day)

The surgical group would stop ARM post procedure. In children randomised to fundoplication a laparoscopic Nissen fundoplication was performed. The protocol stipulated that the operation would be conducted by 1 of 5 consultant paediatric surgeons practising at GOSH. In these children, pre-operative acid suppression medication would be discontinued post-operatively. The protocol stated:

"The patients randomised to fundoplication will not receive medical treatment postoperatively unless gastro-oesophageal reflux is re-occurring".

Criteria for diagnosing recurrent GORD (rGORD) was symptom based i.e. were recurrent chest infections, vomiting, retching or retrosternal pain. In addition to rGORD, anticipated complication for the surgical group was gas bloating and intestinal obstruction.

OUTCOME MEASURES

The primary outcome measures were defined as:

1. Quality of life: this was measured using a questionnaire devised by O'Neill et al(289) to measure parental opinion of quality of life in children post gastrostomy. QOL would be measured at 6 and 12 months post-operatively.

2. Reflux index. This would be measured by pH-MII at 12 months following surgery.

The secondary outcome measures were defined as:

1. pH-MII parameters:

a. number of reflux events

b. number of reflux events in the first two hours after the feed;

c. average reflux height

d. average minimum proximal and distal pH;

e. total acid clearance time per hour, total reflux duration per hour

f. DeMeester score.

2. A record of gastrointestinal symptoms

3. Daily use of anti-reflux medication

4. Gastric emptying measured by octanoate breath test

5. Peri-operative factors

a. Duration of the operation

b. disposable instruments used (number and type)

c. medications given for GOR and antibiotics during admission into hospital (number and dose)

d. duration of hospital stay (days);

6. Operative complications: conversion to open procedure, bleeding, intestinal perforation, pneumothorax

7. Post-operative complications: wound infection, bronchopneumonia, dysphagia, gas bloating, retching, vomiting, dumping syndrome).

8. Nutritional indices

a. assessment of dietary intake (carbohydrate, fat and protein) using a 3-day diet diary

b. weight centile, weight velocity, height, BMI

c. body composition (as above)

9. Recurrence of GORD (rGORD)

a. Medications for GOR (dose and number during the first postoperative year);

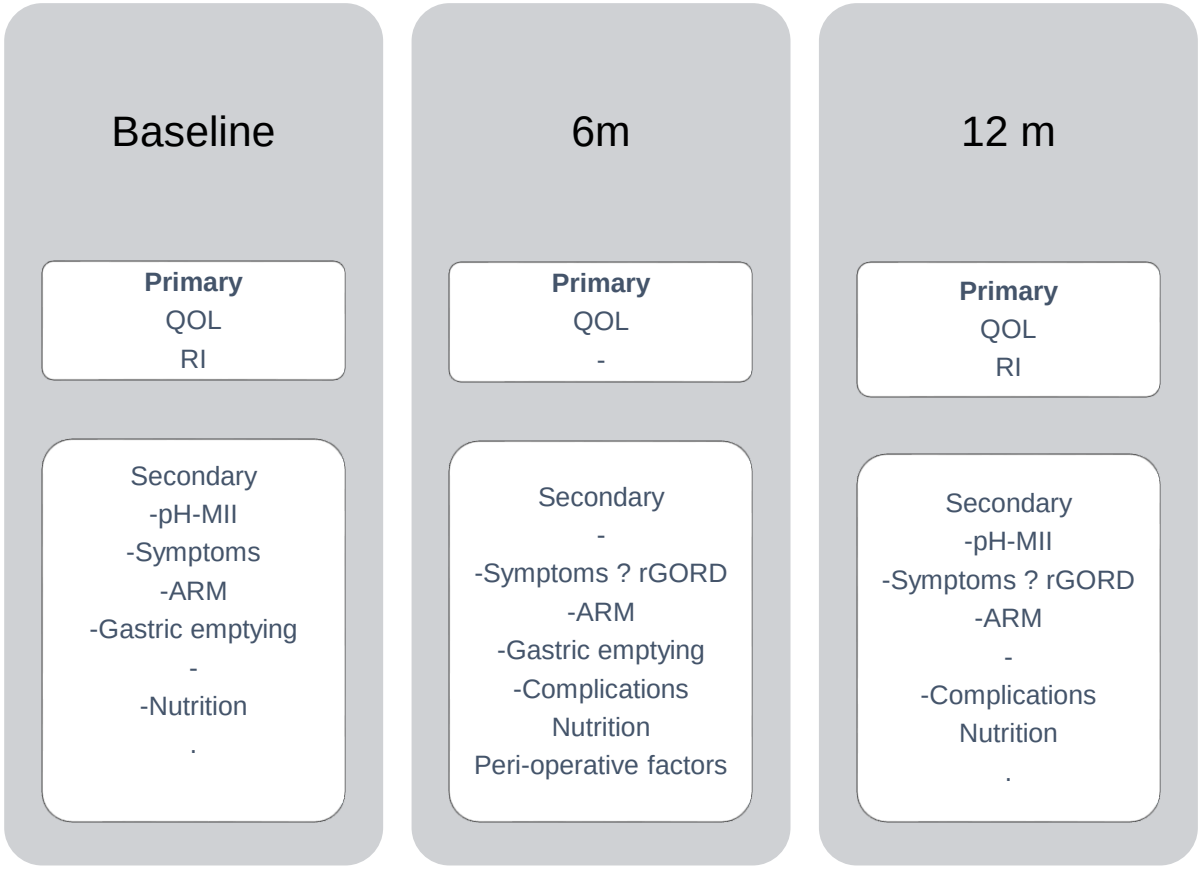
b. Episodes of chest infection (number requiring antibiotic treatment during the first postoperative year),

c. Admissions to hospital after randomisation (number during the first postoperative year);

10. Cost of care during the first year after randomisation (based on variables 8, 9, 10 and 11).

Patients were assessed at the baseline, 6 months and 12 months. The schematic below summarises the patient pathway through assessments.

Figure 138: REMOS schedule of investigations



SAMPLE SIZE

Sample size estimation as based on a previous study of quality of life measures in children who had laparoscopic of open Nissen fundoplication(290). In this prior study, there was a significant improvement of 32±19 points on the QOL scale. Assuming that children treated with gastrostomy plus medical therapy have only a 16-point improvement in QOL score, 30 patients in each arm, as originally proposed using pH study as the primary end point, would give us 90% power at the 0.05 level to detect this difference (Appendix B). In Great Ormond Street Hospital, a large number of laparoscopic Nissen fundoplications and gastrostomies are performed in NI children; approximately 90 every year. 70 of these children were expected to be eligible for the trial each year and are it was therefore expected that 60 patients would be recruited in two years.

DISCONTINUATION/ WITHDRAWAL

Failure of therapy was defined as

- either suspicion of ongoing severe GOR
  - at least 3 hospital admissions, or
  - one admission to the intensive care unit for aspiration pneumonia requiring antibiotic administration
- falling weight centile at the 6-month evaluation (as defined in minimisation criteria)
- incapacitating vomiting after at least 6-month evaluation

Failure would trigger a 24-hour pH study. Those failing in the medical gastrostomy group would be considered for laparoscopic fundoplication. Conversely, those failing in the fundoplication group would be considered for either medical therapy or RF. This decision would be made by the clinician responsible for the patient's care, rather than study protocol.

A Trial Steering Committee was convened to supervise the conduct of the trial to 'rigorous scientific, clinical and ethical standards'(291). This committee was to comprise an independent Chairman, two independent members (Paediatric Surgeon and Paediatrician), a nurse representative, a parents' representative, trial co-ordinators and a representative of the Data Monitoring and Ethics Committee (DMEC).

The DMEC would comprise persons independent of the trial organisers and those providing therapy to patients in the trial. DMEC remit was defined as:

a. review assumptions underlying sample size consideration

b. modify or close the trial.

Overwhelming evidence of superiority of one arm over another would trigger an early end to the trial.

During ethics review prior to trial commencement, the protocol was amended to address discontinuation criteria. The initial protocol suggested an assessment of treatment outcomes when 30 patients were recruited. This was estimated to be at 2 years into the trial. However, the ethics committee highlighted

that, in a patient failing to thrive following gastrostomy, assessment for surgery after 2 years could be considered to be withholding of treatment. To address this, the protocol was amended to read:

"Patients who do not respond to gastrostomy plus medical treatment will be offered fundoplication when their treatment has been identified as failing".

Furthermore, the protocol was amended to enable DMEC review after 1 year of recruitment. Should evidence of overwhelming superiority of treatment emerge, the trial would be stopped.

Overwhelming superiority was defined as a significant difference ($p<0.01$) for any of these outcomes (i.e. quality of life, 24h pH study and failure of intervention),

STATISTICAL METHODS

Groups A and B will be compared for statistical differences. For primary outcomes, questionnaire data will undergo median comparison. RI data will be compared for means. For secondary outcomes, nutritional indices are normally distributed (height, weight and BMI z-score). Means will be compared for these parameters. However, given the small sample size anticipated (60 patients) qualitative review of parameters may be sufficient.

AMENDMENTS TO PROTOCOL

In October 2010, after a few months of recruitment, the researcher alerted the principal investigator of some inconsistencies in the trial protocol. These amendments were made through the medium of a FILE NOTE -OCTOBER 2O10 (Section V appendix, p528). Briefly, the amendments are summarized below.

1. Using pH-MII: the researcher encountered patients who had a pathological pH-MII study but had a RI between 5-10%. These children would not meet the inclusion criteria of the trial as no impedance parameters were considered and the RI threshold was >10%. Therefore, the protocol was modified to from: "documented GOR on 24h oesophageal pH monitoring (reflux index ≥10%)" to "documented GOR on 24h pH or oesophageal impedance monitoring".
2. The minimisation criteria included "severity of reflux index at pH study [5-10%, 10-20%; >20%]". However, as pH-MII study was to be considered, the minimisation criteria were amended to read "[severity of reflux index at pH study <10%, 10-20%, > 20%]". This amendment had the effect of lowering the RI threshold for inclusion to <10%, as long as impedance GOR criteria were met.
3. It was realized belatedly that the minimisation program SIMIN had not included triceps skinfold thickness as a criterion. The minimisation criteria included weight z-score which also served as a surrogate for nutritional status and was less prone to operator variability. Rather than modify the application after the trial had begun, a decision was taken to minimize without utilising this criterion. Therefore, the file note stated: "Triceps skin fold thickness z-score will not be utilized as a minimization criterion. Triceps thickness will remain a secondary outcome measure."
4. Collaboration with the gastroenterology department presented an opportunity to perform oesophageal manometry on trial patients. Therefore, the amendment read: "In addition to oesophageal impedance, we shall also perform oesophageal manometry. This will give us a pre- and post-operative assessment of the oesophageal peristalsis wave and pressure at the transition zone and lower

oesophageal sphincter." It is important to state that oesophageal manometry would have required a general anaesthetic in our population of children.

5. The protocol was also amended to allow patients in the Surgical group to have ARM in the period immediately following surgery. The chief investigator considered that some ARM cover was required due to the risk of stress ulcers related to surgery. ARM was to continue for 2 weeks post operatively then to be discontinued after discharge. Therefore, assessing which patients re-started ARM or failed to discontinue ARM in the Surgical group became an important outcome.

In December 2012, secondary to the concerns discussed above, it became necessary to undertake a substantial amendment to the protocol. The changes are documented in NOTICE OF SUBSTANTIAL AMENDMENT (non-CTIMP), Section V appendix p531). Per trial protocol, the hospital Research Ethics Committee (REC) must be notified of any Substantial amendments. REC review and approval was carried out in January 2013.

1. To meet safety standards, investigators could no longer utilise non-medical grade octanoic acid and deuterated water. Obtaining medical grade octanoic acid and deuterated water was cost prohibitive. Furthermore, the operational environment had changed with the introduction of CTIMP (Clinical trial of an investigational medicinal product) guidelines. The Medicines for Human Use (Clinical Trials) Regulations 2004 required REC review of ingested substances where there was study of rates of metabolism or excretion. To administer any medicinal products in the process of a clinical trial, full trial re-authorization was required with categorisation of these products as either a food or a medicine. There was little safety data about octanoate in particular to guide this process. The investigators, balancing the costs and benefits, chose to omit these investigations from the study. These investigations were secondary outcomes and, though of interest, were deemed not to be sufficiently crucial to impair the scientific value of the study.

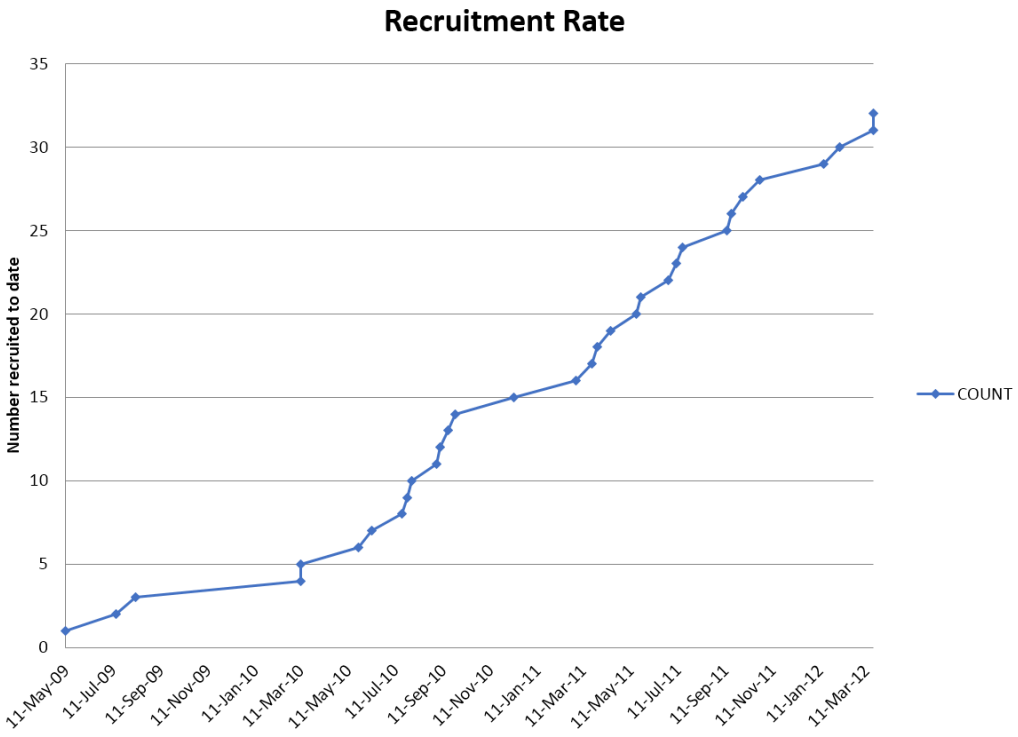
2. The changes to inclusion criteria i.e. including pH-MII findings, lowering the RI threshold and omitting triceps measurement, were communicated to the hospital REC.

CHAPTER 3: RESULTS

RECRUITMENT

Recruitment was commenced in April 2009 by the first trail co-ordinator. Unfortunately, the co-ordinator left the employ of the unit by October 2009. Recruitment was recommenced in April 2010. By December 2011, we had recruited 50% of patients required for the study and an interim analysis was performed.

Figure 139: Timeline of recruitment.



The initial trial co-ordinator conducted recruitment from May 2009-Jan 2010. I conducted recruitment May 2010 -March 2012. Recruitment in the interval was conducted by a further Clinical Research Fellow.

Table 142: Key trial stages and dates

Stage	Date
Initiation	August 2008
Ethics consideration	October 2008
Ethics Approval	December 2008
Initial Trial Recruitment	May 2009
Trial Re-start Date	February 2010
Proposed End Date	December 2012

The interim analysis results are available for this trial. We achieved 50% recruitment by December 2011 and presented the results of the interim analysis to the Data Monitoring and Ethics Committee (DMEC). At interim review, the results of 32 patients were available for analysis.

Limitations to recruitment

1. The initial recruitment began in 2009. Unfortunately, the initial trial coordinator left the employ of the Institute by October 2009. There was a hiatus between his leaving and the hire of a new coordinator (this researcher). When I re-started recruitment in May 2010, it was clear that some patients had been lost to follow-up as timely re-assessment was not possible. Only 1 patient recruited by the initial coordinator could be tracked and retained within the study.
2. The initial trial entry criteria required a reflux index of 10% on pH study. These levels of acidity are uncommon in children with neurological impairment as they are often exclusively milk fed. The recruitment criteria made no account for alkaline reflux episodes. Initial recruitment was slower than expected. An amendment to the inclusion criteria was made to address this issue. pH-MII parameters were utilized rather than RI alone (see amendments to protocol below). This intervention resulted in an improvement in recruitment rates.
3. It was necessary to perform pH impedance on patients prior to recruitment. Unfortunately, this investigation is invasive and poorly tolerated. Furthermore, this test required a 24-hour admission period in hospital. Obtaining admission for this investigation was often subject to availability of beds, which was a factor beyond the control of the researchers. In mitigation, patients were admitted to GOSH accommodation rather than hospital beds. This intervention improved recruitment rates.

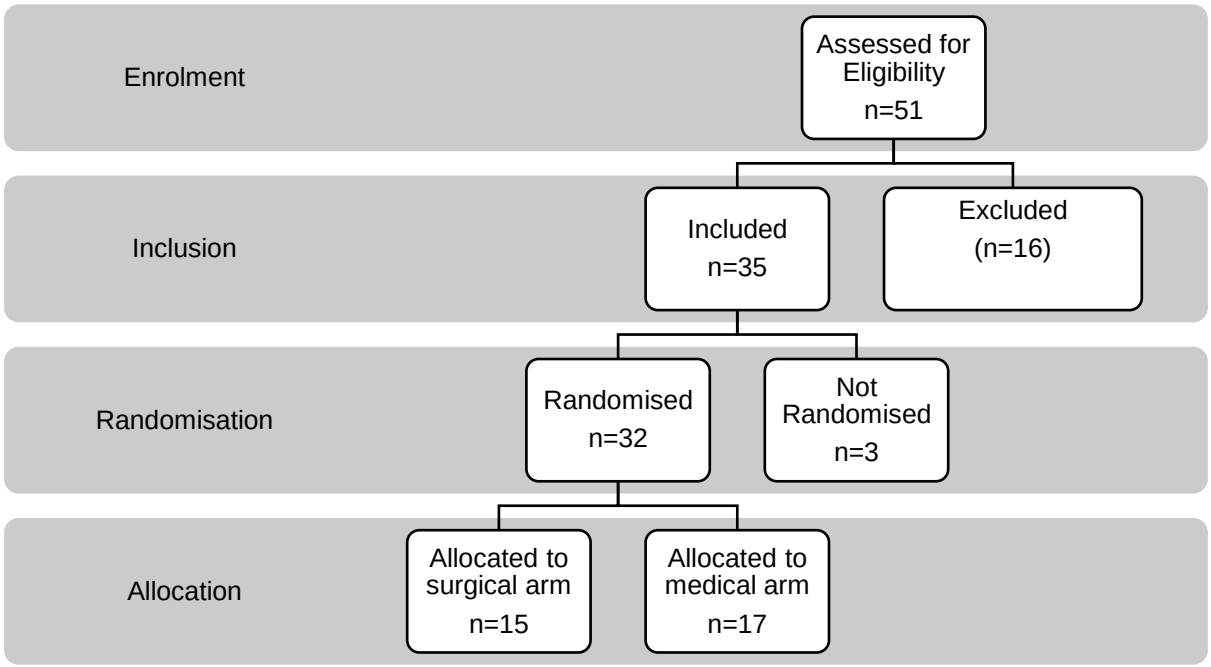
Summary of participants and baseline data

During the study period, 51 patients were assessed for eligibility and 35 were recruited.

There were 8 patients who were excluded due to lack of equipoise i.e. clear indication for either gastrostomy alone or fundoplication. A further 8 patients were excluded due to language barrier (n=2), RI <5% on initial pH study (n=4), inevitable loss to follow-up / international patients (n=2).

There were 3 patients who met inclusion criteria who were recruited but not randomised. One patient died prior to randomisation. Two patients were designated as palliative prior to randomisation. Thus, 32 patients were randomised.

Figure 140: Flow chart summarising allocation of participants



The median age at randomization was 18 (3.4- 93.6) months. The median age at surgery was 18 months (3.8- 96). The median follow-up (May 2010 – Dec 2015) was 40.1 (1.8-71.1) months. There were 13 girls and 19 boys.

Patients were randomised to either Medical (gastrostomy +ARM) or Surgical (gastrostomy + fundoplication) groups. The causes of the neurological impairment are as summarised below:

Table 143: Description of neurological impairment

	Medical (n=17)	Surgical (n=15)
Cerebral palsy	5	6
Pervasive Developmental Disorder	7	4
Birth Asphyxia	1	1
Central core myopathy	1	0
Mitochondrial myopathy	2	0
Myopathy, unspecified	1	0
Larsen's syndrome	0	1
Microcephaly	0	1
Downs' syndrome	0	1
Hydrocephalus	0	1
Total	17	14

Randomisation with minimisation appeared to be successful as the groups were well matched for the minimisation criteria i.e. age, RI, weight, severity of neurological impairment (GMFCS) and duration of prior medical treatment.

Table 144: : Distribution of minimisation criteria in recruited children following randomisation

Minimisation criterion		**Medical (n=17)**	**Surgical (n=15)**	**Total**
GMFCS	I	0	0	0
	II	1	2	3
	III	7	5	12
	IV	5	6	11
	V	4	2	6
Age	<1 year	6	6	12
	1-4 years	7	7	14
	>4 years	4	2	6
Reflux index	5-10%	11	10	21
	10-20%	6	4	10
	>20%	0	1	1
Weight z score	<-2.2	4	2	6
	-2.2 to -1.6	7	7	14
	>-1.6 to -0.5	6	6	12
Duration of ARM	None	1	0	1
	<6 months	3	1	4
	>6 months	13	14	27

Randomisation with minimisation to weight z-score strata resulted in an even distribution of patients across the weight brackets. The degree of neurological impairment in each patient was categorized using the GMFCS score. We found that most patients recruited in the trial had more severe neurological impairment e.g. GMFCS scores III-V. This skewing reflects the serious nature of comorbidities seen at Great Ormond Street Hospital.

All patients had an UGI prior to randomisation. No structural anomalies were identified in any of the patients investigated.

Upper Gastro-intestinal study	**Medical (%)**	**Surgical (%)**	**P***

Reflux identified on study	13/17 (76.4%)	9/15 (60%)	>0.05
Hiatus hernia	0	0	-
Malrotation	0	0	-
Delayed Gastric emptying	0	0	-

PRIMARY OUTCOME

Median at randomization was 18 months (3.4- 93.6). Median age at randomisation was surgery was 18 months (3.8- 96) 21.5 . Duration of follow-up at interim analysis was 40.1 months (1.8-71.1).

Quality of Life measures

In comparing primary outcomes in our two groups, we found that quality of life scores were similar in both groups at baseline and at 6 month review. Both groups demonstrated improvements in symptom scores.

Table 145: Questionnaire response scores at baseline assessment

(Mean ± standard deviation)	**Medical**	**Surgical**	**P**
Gastro-intestinal symptom score (maximum 25)	8.8 (± 3.6	9.7 (± 4.2)	0.5
Child and Parent Quality of Life score (maximum 50)	27.0 (±4.6)	28.3 (±3.7)	0.4
Parenting stress index –depression scale (max 55)	28.2 (±4.6)	29.2 (± 3.6)	0.5

*Independent sample t-test

Notably, there was no difference in growth achieved at 6 months following surgery.

Table 146: Questionnaire responses at 6 months

(Mean ± standard deviation)	**Medical**	**Surgical**	**P**
Gastro-intestinal symptom score (maximum 35)	5.7 (± 3.0)	5.5 (± 4.3)	>0.05
Child and Parent Quality of Life score (maximum 50)	30.2 (±4.1)	32.5 (±5.0)	>0.05
Parenting stress index –depression scale (maximum 55)	29.0 (±5.1)	27.3 (± 5.9)	>0.05

*Independent sample t-test

There were some differences between groups when comparing secondary outcomes. Vomiting was less frequent in patients who had undergone fundoplication.

Table 147: Comparison of vomiting at baseline and at 6 months

	Medical			**Surgical**		
	Baseline	**6 months**	***P value**	**Baseline**	**6 months**	**P**
% Vomiting	14 (82.3%)	8 (53.3%)	0.07	14 (93.3%)	4 (26.7%)	0.0005
Fundoplication for vomiting	-	1	-	-	1	

In patients who received medical treatment, there was a subjective improvement in gastro-intestinal symptoms and in quality of life over time. In patients who had an anti-reflux procedure, there was a subjective improvement in symptoms over time. However, there was no improvement in parenting stress scores in either group.

Table 148: Comparison of questionnaire responses at baseline and at 6 months

Mean (maximum score)	**Medical**			**Surgical**		
	Baseline	**6m**	***p**	**Baseline**	**6m**	***P**
Gastro-intestinal symptom score (35)	8.8	5.7	0.013	9.7	5.5	0.015
Child and parent Quality of Life score (40)	27.0	30.2	0.023	28.3	32.5	>0.05
Parenting stress index –depression scale (55)	28.2	29.0	>0.05	29.3	27.3	>0.05

*Paired-sample t-test

Reflux Index and other pH-MII parameters

At entry to the trial, the RI between group were as follows:

	Medical	**Surgical**	**P**
Reflux index (%)	10.5 (±4.5)	8.9 (±4.4)	>0.05
Reflux episodes	89 ((±24.4)	78(±26.8)	>0.05
Reflux episodes/hr.	4((±1.31)	4((±1.47)	>0.05

SECONDARY OUTCOMES

Nutritional indices at baseline and again at 6 months were as follows:

Table 149: Comparison of weight, height and BMI pre / post intervention

	Medical			**Surgical**		
mean z-score	**Baseline**	**6 months**	***P value**	**Baseline**	**6 months**	***P value**
weight	-0.98 (±0.88)	0.1 (±0.92)	0.00027	-0.97 (±1.27)	0.87 (±1.98)	0.0058
height	-1.27 (±1.04)	-0.26 (±0.79)	>0.05	-0.89 (±0.92)	0.01 (± 0.8)	>0.05
BMI	-0.19 (±1.36)	0.33 (± 1.3)	>0.05	-0.61 (±1.67)	0.98 (± 1.96)	>0.05
*Paired-sample t-test						

Both groups demonstrated improvements in symptom scores. Both groups demonstrated growth following surgery. Both groups had a weight z-score of approximately -1 at baseline. The Surgical group appeared to achieve the greatest difference in growth with weight and BMI z-score approaching 1. This may suggest that in children with a gastrostomy, fundoplication is more effective in promoting catch-up growth compared to ARM.

Both groups demonstrated catch-up growth following surgery. Notably, there was no difference in symptoms scores, or growth achieved at 6 months following surgery.

Table 150: Comparison of weight, height and BMI pre / post intervention, standardised by z score

(Mean ± standard deviation)	**Medical (n=17)**	**Surgical (n=15)**	**P ***
z-score weight	0.1 (±0.92)	0.87 (±1.98)	>0.05
z-score height	-0.26 (±0.79)	0.01 (± 0.8)	>0.05
z-score BMI	0.33 (± 1.3)	0.98 (± 1.96)	>0.05

*Independent sample t-test

Focusing on nutritional indices, we demonstrated a significant improvement in weight in both groups. Body mass index (BMI) remained static over time.

ANCILLARY ANALYSIS

At 6 months, vomiting improved in both groups. In the medical group, vomiting rates fell from 82% to 53% and the difference was not significant. In the Surgical group, vomiting rates fell significantly from 93% to 27%. The proportion for patients with persistent vomiting was higher in the medical group at 6 months.

Table 151: Frequency of vomiting post intervention

	Medical			**Surgical**		
	Baseline	**6 months**	***P value**	**Baseline**	**6 months**	***P value**

Vomiting	14 (82.3%)	8 (53.3%)	0.07	14 (93.3%)	4 (26.7%)	0.0005

We identified 4 patients in the Surgical group who continued to require ARM at 6 months following surgery.

In 2016 (after the end of the trial) a colleague reviewed the recruited patients to identify patients requiring further operative intervention. None of the Medical patients required a fundoplication. None of the Surgical patients required RF. However, gastrojejunostomies were placed in recruited patients. One child in the Medical and 2 children in the Surgical group received gastrojejunostomy for severe GORD. Three deaths in each group were secondary to the primary medical condition and not associated with to GORD.

Table 152: Long-term follow-up of symptoms post REMOS trial

	Medical (%)	**Surgical (%)**
Vomiting	8 (53.3%)	4 (26.7%)
Need anti-reflux medications after surgery	N/A	4 (27%)
Gastrojejunostomy	1 (6%)	2 (13%)
Mortality (unrelated to GORD)	3 (17%)	3 (20%)

HARMS

An 11-year-old girl who met the inclusion criteria on initial assessment, the trial co-ordinator was unable to pass the pH probe following the usual protocol. The procedure was abandoned, and we asked a senior consultant to pass the tube under X-ray guidance. This task proved impossible even under X-ray guidance. The patient was subsequently referred to the Ear Nose and Throat Surgeons to investigate nasal anatomy with possible choanal stenosis. This has been documented as a serious and adverse event because there was significant distress caused by repeated attempts at intubation. It was an adverse event because we have been unable to quantify the degree of reflux in this patient using the pH study. This has limited and delayed further treatment.

A 4-year-old girl developed a blocked NG tube during a pH-MII. Unfortunately, this was not reported till the end of the pH-MII study 24 hours later. This study was repeated.

There were no peri-operative complications in either group. There were some minor and anticipated complications. These are documented in the table below.

Table 153: Minor complications

	Medical (%)	**Surgical (%)**
UGI before 12 months	2 (11.7%)	4 (26.7%)
Infection at site	2 (11.7%)	1 (6.6%)
Tube displacement	1 (5.8%)	0

Patients received an UGIC prior to the 12-month follow-up due to clinical suspicion of wrap failure or herniation. None of the patients who had UGIC had these complications.

CHAPTER 4: DISCUSSION OF REMOS TRIAL

FINDINGS

The differences between Medical and Surgical groups can be summarised as :

- Significant reduction in the incidence of vomiting in the surgical group
- Significant improvement in QOL score in medical group
- Cross-over in the fundoplication group with 27% requiring ARM following fundoplication.

The similarities between the groups can be summarised as:

- Significant reduction in GI symptom score in both groups
- No difference in rates of re-operation so far

The data collected suggests that treatment of children with NI and GORD results in similar outcomes for quality of life. Objective measures suggest that there is an improvement in vomiting. Following interim review, we continued recruitment to archive 39/60 patients. However, the safety and procurement issues highlighted resulted in a hiatus that coincided with the end of the trial funding period. The trial was therefore unfortunately terminated prematurely with no clear prospect of re-starting recruitment and completing follow-up.

The preliminary findings detailed suggest a change in the approach to children with NI and GORD.

- In children for whom vomiting is a key feature of the GOR pathology, anti-reflux procedures may prove more efficacious than medical management.
- In children with other features of GORD e.g. failure to thrive and feeding intolerance, gastrostomy and medical treatment may be the advocated treatment.

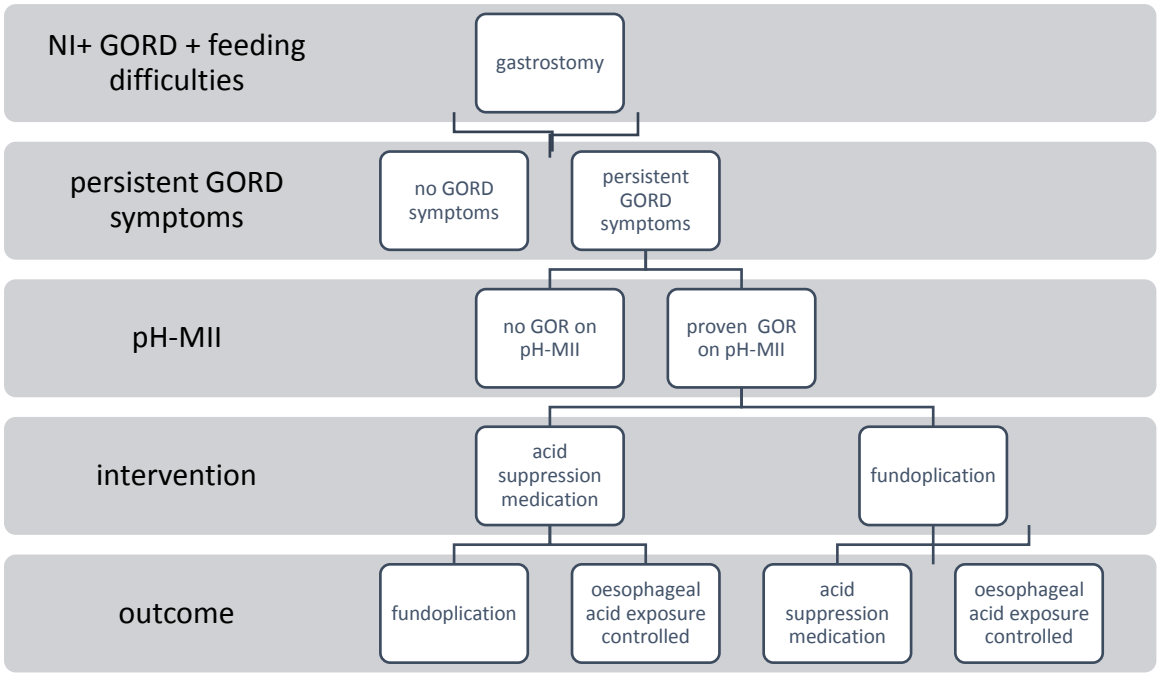
However, completion of recruitment and analysis of data from the whole cohort, and publication of data is required before this approach can be endorsed fully.

As part of understanding the problem of gastro-oesophageal reflux in children and providing background data for this trial, we have investigated the role of radiological investigations and surgical interventions.

LIMITATIONS

Design

How might such a study be designed to answer the question sufficiently? Firstly, as we cannot establish equipoise, a simple randomised trial is not indicated. Instead, we suggest a cross-over design, with double blinding.



Trial entry criteria should be children with demonstrable reflux, meeting a defined threshold of reflux episodes per hour. The trail design should be naïve to acid exposure. It then follows that one both arms will have an intervention to reduce oesophageal acid exposure. pH-MII enables the researcher to enumerate reflux episodes and acid exposure separately. Therefore, the confounding effect of addressing both reflux and acid exposure in a single intervention is removed. Once group will have acid suppression medication, the other will have fundoplication. Should participants in either group fail, they can then cross-over to the alternate intervention.

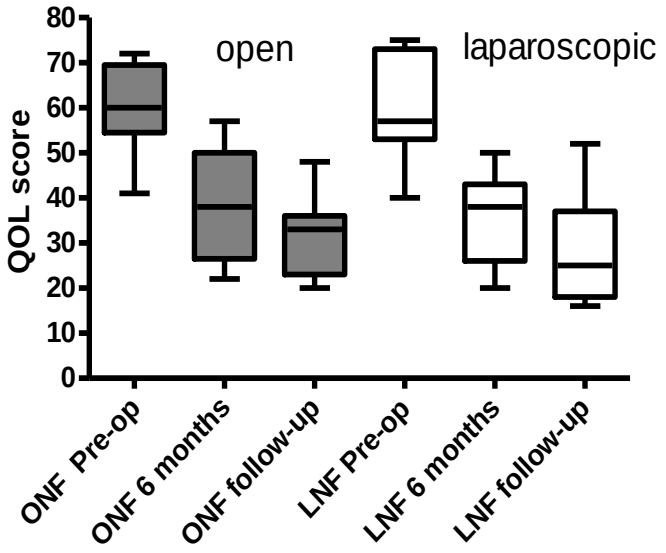
Proportions achieving oesophageal acid exposure control after each intervention would be compared. Control could be defined at a reflux index threshold. The more effective intervention would be that which achieved control with the fewest patients crossing over to the alternative trial arm.

Sample size calculation

The sample size calculation was based on a QOL questionnaire score difference. This difference was identified in a comparison of laparoscopic versus open fundoplication. The data are unpublished.

A mean score or 32+/- 19 was the observed difference before and after laparoscopic fundoplication. It was assumed that the observed difference before and after gastrostomy without fundoplication could be estimated by halving the effect observed with fundoplication. However, in the original, on which this calculation is based, some patients (it not clearly enumerated) had a gastrostomy as part of their procedure. Therefore, the effect of a gastrostomy is ‘baked into’ the reported effect size(290).

Figure 141: "Quality of life score in our randomised controlled trial comparing open and laparoscopic fundoplication". Reproduced with permission from the REMOS trial protocol.



Parental / caregiver concerns and QOL measures do matter. However, if the question is which operative or pharmacological strategy best controls GORD symptoms, then its other suitable effects should be sought. For example, rGORD rates before and after gastrostomy, rGORD rates before and after fundoplication.

Amendments

Trial protocol amendments are a common and unavoidable problem when running a clinical trial. There are few studies on protocol amendments. However, a review of pharmacological trials based at Tufts Center for the Study of Drug Development and found that trials with complex protocols had, on average, 3.2 amendments(292). In a follow-up study, the same authors pooled data from 17 US-based large pharmaceutical and biotechnology trial operators. This study yielded data from 3410 protocols conducted between January 2006 and December 2008. Competed trials had an average of 2.3 amendments each. Amendments were classified as: Completely Avoidable," "Somewhat Avoidable," "Somewhat Unavoidable," or "Completely Unavoidable." Based on this, 34% of amendments were categorised as somewhat or completely avoidable(292).

We have summarised 5 amendments to the trial protocol, of which 2 were substantial enough to be result in file notes and steering committee notification. Amendments to protocol based on pH-MII thresholds and data were unavoidable. At the time of trial design, the full implications of pH-MII were not understood as the methodology had not yet been implemented. Furthermore, no normative data on pH MII were available. However, it was clear that use of pH-metry alone would be ethically unsound as pH-MII was, increasingly, the clinical standard for GORD diagnosis. The study would be mis-specified in remaining agnostic to non-acid reflux in both design and analysis.

The re-institution of ARM in the first 2 post-operative months could be considered a somewhat avoidable amendment. The rationale for providing post-operative cover against acid -related stress ulcers is sound. However, this amendment concerned a parameter central to the study design. Such an

amendment, in my opinion, was substantial and should have triggered re-visiting the trial design with the hospital research and ethics committee. However, I do not believe this amendment had any safety implications. However, it made the analysis of data difficult to interpret as it implemented a cross-over methodology on a two-arm trial design.

An alternative design approach would have been use of placebo ARM. Following the operation, the Medical group could continue ARM and the Surgical group could have a placebo ARM. The ethical approval process for such a trial, if undertaken today, would be as a CTIMP because the efficacy of a medicinal intervention was being tested.

Other amendments regarding use of octanoate for gastric emptying measurement were completely unavoidable. The change in the regulatory environment retroactively introduced elements to the REMOS trial that might render it a CTIMP trial by MHRA standards. Removing administration of contrast / isotope measurements from the trial was the most prudent approach to ensure trial completion.

The amendment of RI from 10% to 5% was somewhat avoidable. The RI of 10% was simply too high to enable brisk recruitment. The 10% threshold was based on prior studies(134) but was just as arbitrary as 5%. This consideration was not factored into projections of recruitment rate. Indeed, if the primary outcome and recruitment criteria were based on pH-MII criterion of reflux episodes alone, there may have been no issues with recruitment. However, this was not possible because there were (and remain) no robust normative data on reflux episodes on pH-MII in children.

In general, the secondary endpoints were too numerous and often vague. For example, the study was simply not funded to enable the researcher to perform any serious assessment of the cost of a participants' care during the first year after randomisation. GOSH is a quaternary referral hospital. Therefore, to robustly follow-up patients and their cost of care, the researcher would have needed to be in regular contact with all participants and their local hospitals to estimate care costs.

Adequate control

The Medical group continued to receive domperidone following the procedure while the Surgical did not. In a review by Lorenzo and Orenstein, authors discussed post-fundoplication syndrome and the role of medications to control these symptoms. In particular, motility agents such as domperidone, and erythromycin are recommended for hypomobility and delayed gastric emptying post-surgery. Although fundoplication may address LOS tone, delayed gastric emptying may persist (or worsen due to intra-operative vagotomy), necessitating use of motility agents. Therefore, the Medical group received a treatment for a condition which the intervention group (B) was likely to develop. This introduced a confounding factor.

The mechanisms for preventing oesophageal acid exposure between the Medical group and the Surgical were different. The Medical group had medication to reduce degree of acidity and volume of acid. The Surgical group did not have any change in degree or acidity, or volume of acid produced. Instead, they had a physical barrier to prevent acid exposure of the oesophagus. Equipoise was based on the assumption that the child's symptoms were due to oesophageal acid exposure. However, we now have evidence that non-acid reflux is associated with symptoms of GORD. Therefore, The Medical

group remained untreated for reflux. Secondly, symptoms of feeding difficulty could arise from gastric acid exposure. The Surgical group remained untreated for gastric acid exposure.

Blinding

Double blinding the REMOS trial would have been desirable. Primary outcome measurement included a questionnaire for parents. Prior beliefs on the efficacy of fundoplication might alter vigilance levels and questionnaire response of unblinded parents. For example, parents who know their child has not had fundoplication may be vigilant for signs that fundoplication is required. This may explain the higher rates of reported vomiting in The Medical group at 6 months.

In a previous study comparing laparoscopic versus open fundoplication, parents and healthcare providers were blinded to procedure performed. This was achieved using a large occlusive dressing(293) after the operation was complete. In the REMOS trial, laparoscopic gastrostomy was conducted with or without laparoscopic fundoplication. The external differences between procedures would have been minimal, although a laparoscopic fundoplication concomitantly with gastrostomy would require one or two extra 1cm ports. Therefore, a standardised, large dressing could be sufficient to blind parents, caregivers and the researcher. Although this was technically feasible, it would have introduced further complexity as the researcher would not have been able to randomize patients, and it would have been difficult to maintain blinding post-operatively. It is also questionable whether blinding the parents and caregivers would have been ethically acceptable. For example, a child presenting to a secondary hospital with post-operative complications would require different investigations and treatments if the operation was gastrostomy alone or fundoplication.

Symptom criteria for inclusion

To be included patients had to have at least one of 4 symptoms and signs of GORD. This stipulation imposes a false equivalence on these criteria. For example, a patient with vomiting may have a primary problem with GORD. A patient with failure to thrive may have a primary problem with bulbar incoordination, renal impairment or metabolic disease. Recruitment was inappropriately agnostic to primary symptoms.

Vomiting, the primary symptom of GORD, was problematic as a criterion for inclusion. There was no enumeration of frequency or duration required to meet the inclusion threshold.

Recurrent pneumonia (3 episodes requiring antibiotic treatment) was an inclusion criterion, but ALTE episodes led to exclusion. Therefore, there is a contradiction lies in the fact that aspiration pneumonia may be a criterion for both inclusion and exclusion.

Anaemia as defined by the WHO(287), was a criterion for inclusion. However, in children with feeding difficulties and other complex needs, anaemia may not be attributable to GORD alone. Furthermore, the WHO definition classifies mild, moderate and severe anaemia. The protocol failed to specify at which threshold a patient can be included.

In mitigation, a minimal threshold for RI on pH study was set. This measure ensured that all a patient recruited had a demonstrable element of acid reflux.

Primary outcome measurement instrument

The QOL questionnaire was devised by O'Neill et al.(294) and published in 1996. It had two elements – the parental attitudes towards daily care and parental stress index – depressions scale. The latter appears to have been adapted from the parental stress index (PSI) by Richard Abedin(295). This QOL instrument was included in outcome assessment of parental/caregiver concerns before and after fundoplication. The first element (parental attitudes towards daily care) was not validated before its application to the REMOS study. Respondents are expected to respond to questions like "ease of feeding", "pneumonia or other pulmonary condition" by choosing from a Likert scale ranging from "Excellent/1 "to "Terrible/5". Setting aside internal validity, no attempts at external validation are made.

The inclusion of the second element (PSI) of the questionnaire was even more problematic. As the researcher responsible for delivering this questionnaire, I remain concerned about its use as a tool for measuring operative outcomes. In my opinion, some questions were intrusive, inappropriate and irrelevant.

e.g.

"When I think about the kind of parent I am, I often feel guilty or mad at myself. "

"I am unhappy with the last purchase of clothing I made for myself. "

Inclusion of this tool assumes that parental stress is only indexed to the child's status. Inclusion also assumes that the researcher is warranted access to this information. Parents attend hoping to get a feeding intervention or fundoplication from the encounter with the researcher. They may feel coerced to answer intrusive questions about their mental health. As a doctor, my duty of care is to the child patient. Inevitably, I encountered parental depression and stress. Having elicited these questionnaire responses about parental depression, did that create a duty of care towards the parent?

Thus, this questionnaire, although previously used in a similar population, was not validated, seems overly intrusive by today's standards, and if a further study were to be carried out, should be replaced by a more appropriate, validated, QOL instrument.

Summary

On this modest sample size, we can hesitantly conclude that fundoplication may control vomiting better in children with gastrostomy. We can also conclude that, as both groups demonstrated improvement in nutritional outcomes, gastrostomy alone is a suitable intervention for overcoming feeding difficulties and failure to thrive.

Based on these findings, should children with GORD should have gastrostomy with fundoplication? This conclusion cannot be drawn from the study performed. Equally, the premature halting of the study means that data are insufficient to make any strong recommendations.

We have identified study design and implementation issues that limited the yield and generalisability of this trial. These are discussed below. Further limitations in study design and interpretation are discussed extensively in Section VI, “Discussion: Chapter 2: Lessons from the REMOS trial”, p426. Recommendations for improvements to trial design are made.

In many ways, the conduct of this trial illustrates the central issue in paediatric GORD. Of the multiple parameters – symptoms, signs, investigations, comorbidities- which factors can safely be used to stratify risk and, when breached, commit a patient to fundoplication?

In the next and final chapter, we address this question using the entirety of the experiments and investigations that comprise this opus.

SECTION VI: DISCUSSION

CHAPTER 1 : A SYNTHESIS

In the opening chapter, we set a practical challenge i.e. solving the 'the surgeon's dilemma'. We envision a surgeon surveying a roomful of children with GORD. The children have an ungoverned mixture of symptoms, comorbidities, diagnostic tests and expectations. How can the surgeon decide which child will experience 'greatest-benefit-for-lowest-risk' from fundoplication?

In the period during which I have been working on this thesis, I have progressed in my surgical training to become that very surgeon surveying the waiting room. Hence this 'surgeon's dilemma' is now a personal, professional challenge that carries responsibilities and implications.

The aim of this PhD project was to identify a risk stratification method for children with GORD where surgical treatment is considered. In hindsight, this was an ambitious aim given the limitations of available research data and size of cohorts available for study. Some headway, however, has been made. In this chapter, we present key findings and discuss their relationship to existing research knowledge. These key findings are:

1. Demography: The younger the patient, the higher the risk of RF.
2. Diagnostics: Shifting paradigms have resulted in lack of standardized reflux symptom diagnostics and reflux-symptom association metrics. This work contributes through use of digital research tools to develop a reproducible symptom tracking method.
3. Comorbidities: As comorbidities interact, specialists must interact too.
4. Surgical approach: Criteria for surgery must be standardized and met. In the face of contradictory evidence, a staged approach i.e. gastrostomy with or without concomitant fundoplication, appears most prudent.
5. Lessons from the REMOS trial: pragmatic surgical RCT design can yield answers to well-set questions.

I will also review trends in this field that have emerged since beginning this work. These include:

- A move away from acid suppression in paediatric patients
- A move away from complete to partial fundoplication
- A move away from fundoplication
- The gastrojejunal feeding tube as an alternative to fundoplication
- Feeding children blended food per gastrostomy
- Increased use of database studies and data mining approaches
- Increased use smartphones in healthcare

To consolidate, I will propose a new management algorithm for GORD that stratifies patients accounting for the major risk factors identified in this research project. The algorithm proposes is keenly focused on a paediatric paradigm of GORD. This algorithm also incorporates the digital symptom collection methods developed in this project. We believe this pathway will ensure that only children who will experience 'greatest-benefit-for-lowest-risk' are offered fundoplication.

DEMOGRAPHY: THE YOUNGER THE PATIENTS THE HIGHER THE RISK OF RF.

In our retrospective cohort study, we could estimate the age at presentation for 8897 (64%) of patients. The mean age at presentation was 2.6 (±2.01) years. The median age at fundoplication was 4.6 (0.28-14) years. The youngest patient to have fundoplication was 3 ½ months old. There were 37 children under the age of 2 years who had fundoplication. At this age, we would still expect physiological reflux to be present.

What reasons can be given to justify operating at such a young age? Firstly, sicker children will need surgery sooner. In our cohort, the outliers had experienced acute life-threatening events (ALTE). Indeed, we have published data from our centre where surgery was prompted by ventilator dependence(95). We reported on a series of 26 ventilator-depended infants (median age 5.8 months (0.8-19.4) months who had fundoplication. Triggers for surgery included ALTE (19%) and respiratory collapse (31%). All children were weaned off ventilation following fundoplication.

A second reason is a historic trend to perform early fundoplication in children with OATOF and CDH in anticipation of severe GORD. In fact, in children with CDH, some surgeons would perform fundoplication at the time of primary diaphragm repair. Evidence arises from an earlier paper from our centre. Kubiak et al(296) presented 66 patients less than 4 months old who had fundoplication (1986-1997). Majority (56/66) patients had associated anomalies including OATOF (n = 19), congenital diaphragmatic hernia (n = 4) and neurological disorders (n=29). This paper also belies the trend towards young-age surgery in patients with NI. This trend is observed in both our cohort and in the wider literature.

Age at fundoplication should be immaterial if the outcomes are similar across the age spectrum. A key finding the relationship between age and fundoplication success rate. Univariate analysis of our cohort found that decreasing age was an independent risk factor for fundoplication. In our case series of 26 ventilator dependent infants, 5 children subsequently had revision of fundoplication within a year of first surgery(95). One child had oesophago-gastric dissociation. Other authors have demonstrated similar poor outcomes(160,297). In a large retrospective review (832 children undergoing fundoplication) Baerg et al(188) found that the mean age at first fundoplication was significantly lower in children who went on to have RF.

Several explanations for this phenomenon have been suggested. Surgery on younger children may be more technically difficult. This association is difficult to demonstrate from literature but there are hints. For example, Shah et al(160) reviewed laparoscopic fundoplication in children under 5 kilograms at Children's Hospital of Pittsburgh. The age at surgery ranged from 2 weeks to 3 years (It is startling that there would be a 3-year-old weighing less than 5 kilograms!). Notably, 50 (41%) children in this series had a history of prematurity. The diagnostic modality used was the UGIC in 44% of patients. There was a high rate of conversion to open fundoplication (7%). The reason for conversion was given as "dense intra-abdominal adhésions". There were serious intra-operative complications i.e. bleeding from the accessory hepatic artery due to retraction (n=1) and gastric perforation (n=1). Operative times ranged from 45 to 320 minutes. The authors stated that indication for surgery was failure of medical management. However, reflux episodes are often physiological in infants. It is unknown with time for maturation would have resolved reflux in some of the patients included. Furthermore, children had

surgery between 2003 and 2007 when pH-MII was not established. This perhaps justifies the use of UGIC as an objective diagnostic in the preamble to performing fundoplication in a child weighing less than 5 kg.

In a large retrospective database review of the patient health information system (PHIS), McAteer et al(298) identified 141190 children (age <18 years discharged between 2002-2010) with primary diagnosis of GORD. Surgery was more likely for patients under 2 months compared to older children. However, younger age at fundoplication was associated with an increased length of stay.

Poorer outcomes in younger children may be explained by truncal growth in the first year. It is conceivable that the natural elongation of the trunk distracts against the fixed point of the wrapped fundus and draws it into the chest. The comments of Hogan and Shaker(299) (albeit made regarding adults) are salient here: “An oesophagus disabled by an inappropriate or dysfunctional fundoplication wrap is a terrible price to pay for control of acid reflux.”

Not all workers in this field are supportive of fundoplication in infancy. Kubiak et al(296) state: “Fundoplication in early infancy is unsuccessful in a high proportion of patients.” McAteer et al(298) go even further: “anti-reflux procedures are most commonly performed in children during a period of life when regurgitation is normal and physiologic and objective measures of GERD are difficult to interpret”.

Given increasing concerns on the effect of general anaesthetic in infancy on neurocognitive outcomes, any surgery, let alone 5-hour surgery in this age group is not undertaken trivially(300). Furthermore, at the age of a year, less than 5% of infants have symptoms associated with reflux e.g. regurgitation(39–42).

This understanding of the risk of younger age in fundoplication informs our recommendation for our new risk stratification algorithm. To discourage fundoplication in infancy, we propose that a diagnosis of GORD should only be made in infants > 1 year. Specifically, corrected gestational age of >1 year should be achieved before a diagnosis of GORD is reached. This accounts for the overlap between prematurity and GORD.

Setting an age bar for diagnosis has some precedent. The British Thoracic Society(301) recommends a diagnosis of asthma after the age of 2 years. Although this protocol concerns a different disease condition, similarities in the natural history of presentation of the two diseases make comparison illustrative. Firstly, wheezing is loosely used by parents to describe all manner of noisy breathing in children. Secondly, infants and children have narrow airways compared to adults. Therefore, viral illnesses will result in wheezing as the airways fill with inflammatory fluid. Wheezing, although present, is not pathognomic of reactive airway wheeze characteristic of asthma in this age group. Thirdly, cohort studies have demonstrated that children who present with wheeze before the age of two are usually asymptomatic by end childhood. Lastly, lung function testing is unreliable for children under the age of 6 years.

Similarly, refraining from diagnosis of GORD in the under 1-year-old will lead to restraint in investigation and treatment of normal physiology. With any innovation, an inadvertent arms race to push the envelope can develop. Laparoscopic fundoplication in ever-smaller and ever-younger children may

reflect this tendency(160). A policy of restraint from surgery until after child's first birthday reverts focus to patient-centred indications.

There will be children who are at risk of aspiration and acute life-threatening events that cannot be expected to run the gamut of mortal risk until their first birthday. The approach to these difficult symptoms will be discussed later in this chapter.

DIAGNOSTICS: THE SYMPTOM NEBULA IS TAMED BY A REPRODUCIBLE SYMPTOM TRACKING METHOD.

To select the right patient for the right operation, we require a clear understanding of the symptom profile of GORD. Therefore, an unimpeachable diagnosis of GORD is central to this endeavour. Diagnostics standards can comprise symptoms, signs and results of investigatory tests.

In the Chapter 2, on symptoms of GORD, we explored the various diagnostic criteria. Typical reflux syndrome i.e. 'frequent of severe heartburn' as described by the K21 disease code (WHO ICD 10(302)) applies to competent and forthcoming patients. This definition has little utility in a preverbal child, or one with NI.

For the purposes of this work, we accepted the Montreal consensus definition of GORD in children i.e.

"...a condition that develops when the reflux of stomach contents causes troublesome symptoms and/or complications.(22)"

Developing on this tenet, we see that there are two key elements in this definition. Firstly, there must be reflux episodes. Secondly, there must be symptoms and/or complications. Implicit in this definition is an association between the reflux and the symptoms.

Therefore, it follows that a diagnostic standard for GORD in children is founded on three pillars:

1. defining symptoms
2. demonstrating reflux
3. demonstrating association between symptoms and reflux

Defining symptoms

As a preamble to this work, we first demonstrated the sheer breadth and frequency of symptoms associated with GORD. This was done in two ways.

Firstly, we reviewed the literature to identify common symptoms described by researchers and clinicians in this field. The most common symptoms were regurgitation, vomiting, food refusal, crying, irritability

Secondly, we used digital research tools to develop a patient-centred symptom tracking methodology- the TARDIS:REFLUX app to collect reflux symptoms(303).

The TARDIS:REFLUX app is a novel, efficient and systematic way for collecting data on "troublesome symptoms". It is a key contribution to the methodology of GORD diagnostics. Smartphones--Powerful, handheld computers- creates opportunities to re-think traditional methods of data collection. We have updated the established method (i.e. paper symptom questionnaires) in several key ways.

- We changed the temporal pattern of recording data. TARDIS:REFLUX enables real-time and prospective tracking of reflux symptoms. This eliminates recall bias.
- By changing the medium, it has been possible to expand quantity data points that can be collected. This improves both data density and breadth, compared to existing questionnaires for paediatric GORD.
- The app makes transmission and storage of symptom data more efficient and secure.
- Only comparisons of frequency and pattern of symptoms can be made. Deliberately, no gradations of severity are possible using the app questionnaire. This eliminates the severity bias, where only severe symptoms are recorded, or severe symptoms trigger treatment and intervention.

- The questionnaire is reproducible.
- As the diagnostic instrument remains with the patient/ parent, then there is no observer variability or bias. The clinician is at risk of collecting and documenting symptoms as seen through the lens of their disease paradigm.

Self- or carer-monitoring of symptoms presents an opportunity for therapeutic behavioural intervention. For example, a self-monitoring teenager may notice symptoms of reflux after eating certain foods. A mother may notice that her baby vomits with certain feeds but not others. In both these circumstances, an opportunity to avoid the perceived trigger and monitor the symptom is presented.

Significantly, this proof-of-concept work also demonstrates how NHS data protection regulations can be navigated to enable full and safe engagement of patients with health professionals over cyberspace. This will be discussed in greater details in a subsequent section.

When the app was tested by parents of children with GORD in the pilot group, it was found to be more user friendly and acceptable than paper-based questionnaires. This is a promising finding. One might expect that a change of medium of questionnaire delivery would be unwelcome, and that busy parents would prefer a maintenance of status quo. Instead, we found that the app generated more responses and a better data density, despite the hurdles of the initial set up. We attribute this to the pre-existing popularity and acceptability of apps in general as a way of transacting in everyday life. Frequent users of apps would find use of paper forms anachronistic.

The TARDIS:REFLUX app was subjected to focus group testing(304). This yielded some promising results. Qualitative assessment of both utility and usability was positive. Enthusiasm was expressed for the general application of apps as research tools. A particular insight was the desirability of apps designed by experts and tightly focused to address a specific condition. Focus group participants envisioned the application of the apps like ours for episodic rather than constant use.

Focus-group respondents also liked the ability to store data for themselves. They also liked the portability feature. Data could be easily shown to the GP, the paediatrician, the surgeon. This portability feature is a reaction against current Data Protection regulations and “safe haven” rules which can make sharing of patient data between healthcare workers (HCW) difficult to achieve.

In addition to pilot study, focus group testing and expert review, the app was made available on general release on the iTunes store. The app was downloaded 482 times(303). From this group, 188 users proceeded to register and 161 symptom episodes were submitted. From our cohort of iTunes app users, we have generated a second list of common symptoms of GORD. The top five symptoms that raise concern in the population are:

1. Vomiting
2. Regurgitation
3. Food refusal
4. Crying
5. Coughing

Although the sample size is modest, the data above is of crucial importance. These self-reported data are true population-based data. Despite minimal advertising, the app as downloaded in all 5 continents.

This surprising finding demonstrates the potential reach of such a tool. The app therefore provides a robust method for researchers to sample symptom data directly from the population. This method can be customised with researchers defining cohorts of interest to poll.

The contrast between the top population symptoms and the top symptoms concerning HCW (demonstrated by the symptoms nebula) is interesting. We find it reassuring that the top 4 symptoms concerning HCW and parents/caregivers (P/CG) in the population are the same. This suggests commonalities in the concept of GORD symptoms between the two groups. However, coughing features higher in the population-based list of symptoms compared to the HCW.

One explanation could be the non-specific nature of this symptom. P/CG and HCW alike know that children cough for many reasons, and sometimes for no reason at all. Vomiting and regurgitation, on the face of it, appear more specific to the reflux phenomenon.

Another explanation could be the difference between population samples and research/reported samples. HCW see a cohort in which children with NI are heavily represented.

In summary, the TARDIS:REFLUX app provides a novel, robust and acceptable way of collecting symptom data for children with GORD. However, collecting symptom information is just one part of the diagnostic triad. The next challenge is objectively demonstrating reflux episodes.

Demonstrating reflux

In the introduction, we described the history of GORD from the first descriptions of 'peptic ulcer of the oesophagus' to pH-impedance measurement of bolus traffic across the GOJ. We noted how each new technical development in diagnostic modality shifted the disease paradigm.

We discussed how Cannon's in 1911 work established barium contrast studies as a powerful visual demonstration of the reflux phenomenon(9). The cardinal symptoms of hiatus hernia (HH) were typical reflux syndrome i.e. heartburn, particularly after ingestion of food. This association was strengthened with the advent of barium contrast studies which powerfully demonstrated reflux episodes in HH patients(10). Reflux and HH became almost synonymous terms. Indeed, it is an oft forgotten fact that Nissen first performed fundoplication for a patient with an incarcerated paraesophageal HH. For 20 years hence, the diagnosis and treatment of GORD was focused on narrowing the diaphragmatic hiatus, restoring the cardia-oesophageal angle and reinforcing the GOJ sphincter with gastropexy.

The HH-based paradigm applies poorly to children as HH- congenital or acquired- is a rare condition in children. Despite this, the paradigm continued to be applied, although the use of barium was substituted for water-soluble contrast of the upper GI contrast study (UGIC).

Today, the MII is well established as the gold standard for GORD diagnosis in children (136). However, when this research study began, MII was a new and competing technology. Indeed, the use of the UGIC as the diagnostic study for reflux was still prevalent.

A key contribution of this work has been addressing the continued of the UGIC in diagnosis reflux in children. In published work, we demonstrated that reflux is still matter of diagnostic comment in children undergoing UGIC at our institution(156). Radiologists reported on the presence of reflux across the GOJ and the height achieved by the refluxate. Indeed, some UGIC operators would go so far as to perform provocation manoeuvres e.g. gastric distension and Valsalva manoeuvre, in order to demonstrate reflux. Enumerating the problem, we found that, compared to the MII, the UGIC had a sensitivity of 42.8% and a negative predictive value of 24%. Furthermore, the prevalence of hiatus was low (1%) in our high-risk cohort of children with symptoms of GOR, reflecting the low prevalence of HH in the general population of children.

This work confirmed the continuing influence of the HH paradigm as late as 2010, in a leading institution. This phenomenon is not unique to our institution. As recently as 2013, Barnhart et al(281) published a database review of more than 4000 infants with neurological impairment undergoing gastrostomy tube insertion with (34%) or without (66%) fundoplication. Data were amalgamated from 42 children's hospitals in the US health region. In this study, the upper GI contrast study was the modality of choice. Only 9% of infants undergoing fundoplication had ph.-metry prior to surgery.

These examples also illustrate a theme: disease paradigms affect clinicians approaches to management of conditions.

Notably, the HH diagnostic paradigm has been abandoned in adult patients, A functional paradigm now accepted i.e. inappropriate transient relaxation of the lower oesophageal sphincter (LOS). This paradigm shift can be traced to two key insights made in the late 1950's. Firstly, oesophageal manometry finally led to the demonstration of the LOS in human patients. A LOS dysfunction model of

GORD began to take root. Secondly, Tuttle and Grossman's used a pH probe to measure acid oesophageal acid exposure(10). Johnson and DeMeester formalized prolonged pH measurements from the oesophagus into a GOR diagnostic(15). There was a shift of focus away from HH and towards oesophageal acid exposure due to LOS dysfunction. During the next few years, the gold standard for demonstration of reflux was measurement of reflux index i.e. the percentage of time of oesophageal acid exposure during 24-hour pH-Metry. Advances in endoscopy led to a coupling of endoscopic and histological findings. Demonstration of the relationship between Helicobacter Pylori infection and gastric and duodenal ulceration in 1982 led to a Nobel Prize for Marshall and Warren(305). But it also mandated application of endoscopy and biopsy in the catalogue of GORD diagnostics. The phenomenon of reflux oesophagitis- where there was endoscopic, then histological evidence of oesophagitis – superseded the concept of the 'peptic ulcer of the oesophagus'.

Modern oesophageal manometry has also led to a deeper understanding of the functional model of GORD. Echoing the early studies by Hirschowitz, we find that transient relaxation of LOS occurs throughout the day(16). Sometimes it is part of the swallow response or to allow belching. The concept of appropriate and inappropriate LOS relaxation is now established. This has challenged the 'mechanical dysfunction of the LOS due to HH' paradigm of GORD. This in turn leads on to the concept of reflux as a physiological phenomenon. In the face of 'physiological GOR' demonstrating a link between reflux episodes and troublesome symptoms becomes all the more crucial.

These modalities, robust as they are, have limited application in paediatric patients. Prolonged pH metry is invasive and unpleasant and requires a cooperative child. Endoscopy requires general anaesthetic in most paediatric patients.

The gold standard for demonstrating reflux episodes today is pH-MII. Measurement of reflux episodes and bolus pH independently has disrupted the fundamental 'peptic ulcer of the oesophagus' paradigm. Although acid exposure of the oesophagus does occur, the non-acid refluxate phenomenon has been demonstrated irrefutably(36,149). Non-acid reflux is particularly prevalent in children. Indeed, the demographic most likely to have GORD, i.e. milk-fed children with NI, are also most likely to have non-acid reflux. This has challenged the paradigm that symptoms of GORD are explained oesophageal acid exposure. It also challenges the confabulation of acid suppressant medications as "anti-reflux" therapies.

pH-MII: Solutions and controversies

The demonstration of reflux using pH-MII is not without its problems. Introduction of pH-MII in the late 1990's was marred by variability in test conditions. Experts in this field responded to this variability in nascent practice with efforts to establish consensus e.g. the Porto Consensus.

In 2002, 11 experts in paediatric GORD held a workshop in Porto to discuss the conceptual problems of paediatric GORD in the light of pH-MII(26). The emerging consensus document contained useful recommendations for the measurement of reflux episodes. These have been explored in depth in the introductory chapter. Key principles established include defining rate of pH sampling frequency, recommending that dataloggers digitising at <1Hz. The consensus also recommended exclusion of post-prandial recordings as the fall in pH could be due to either refluxate or food bolus. On the question

of the impedance drop sufficient for refluxate, the experts disagreed with convention requiring a 50% drop in impedance as significant of liquid reflux. Crucially, the parameters for gas refluxate in adults were accepted for paediatric use: children and adults were considered sufficiently similar.

The Porto consensus did not address normative values for pH-impedance in children. This remained a perennial problem. In 2012, a consensus statement released by the European paediatric impedance working group(142) (EUROPIG) made an attempt to address this deficit. The aim was to come up with standardised indications, methodology and interpretation of MII. By the authors own admission, the report was based on expert opinion and consensus due the paucity of evidence in this area.

The Euro-PIG consensus(145) also reiterates the arbitrary normal values for reflux episodes proposed by the German paediatric impedance group i.e. >73 episodes if age >1year, >100 episodes if age <1 year. In this statement, they appear to move away from emphasis on symptom-reflux association:

Quoting a previous paper, they stated:

"MII-pH monitoring detects acid, weakly acid, and non-acid reflux episodes, it is superior to pH monitoring alone for evaluation of the temporal relation between symptoms and GER" (136); however, this statement does not endorse that the correlation between symptoms and GER is good."

Furthermore:

"In general demonstration of temporal association is relevant for symptoms of short duration e.g. cough, apnoea, desaturation), whereas for symptoms of long duration (e.g. laryngitis, hoarseness, bronchitis) a global interpretation of MII is more relevant. "

Here, the author must register concern that 'global interpretation' introduces a subjective aspect. If the interpreter and the treating clinician are the same person, the interpretation may lack objectivity e.g. where an opinion has been formed on risk of aspiration and suitability for fundoplication. Confirming the relationship between reflux episodes and reflux symptoms appears to be the key diagnostic tenet.

Rather than loosening the diagnostic standard(145) as suggested by the EUROPIG consensus, this author recommends that an objective and high threshold is maintained. Neither the Porto nor the EUROPIG consensus opine on cessation of acid suppression before pH-MII testing. In our risk stratification algorithm, we recommend maximal medical therapy as a pre-requisite for ph.-MII testing. The pH-MII allows us to separate the often-conflated issue of acid suppression and reflux episode management. The question we seek to answer is does the patient need a fundoplication to control their reflux symptoms. If the test is done off medications, the patient loses this acid suppression therapy benefit. This may lower the threshold for surgery. Furthermore, we recommend testing children with aspiration history to demonstrate whether episodes are related to impaired gag reflex, depressed consciousness or GI disorder e.g. oesophageal dysmotility and GORD.

Emerging limitations in the REMOS trial diagnostic protocol

The problems of an undetermined diagnostic standard were illustrated in the conduct of REMOS trial. It is important to say that the trial protocol as written in 2008. The trial was initiated by another researcher in 2008. After his departure, the trial was restarted (with my participation) in 2010. The trial was

recruiting patients till April 2013. Therefore, the trial existed in a changing landscape as more data available on pH-MII challenged previous understanding.

Once the trial began, it became necessary to revise the inclusion criteria for the REMOS trial. The initial trial protocol, written in 2008, had a reflux index threshold of 10% as an inclusion criterion. However, we found – in reviewing pH impedance studies performed prior to trial start and the literature- that reflux indices this high would be outliers. Early measurements in trial participants yielded reflux indices <5%. Therefore, a protocol amendment was made and authorized and the reflux index threshold was lowered to 5%.

However, the 5% threshold was arbitrary. In the absence of validation studies, we can only guess at the degree of oesophageal acid exposure associated with pathology. Indeed, studies of endoscopic 'erosive esophagitis' tend to find evidence of erosive oesophagitis and Barrett's oesophagus in adults in the 6th Decade of life(306). This suggests that it is cumulative oesophageal acid exposure rather than the reflux index on the test day that is linked to oesophageal pathology. Contrarily, once metaplasia is identified, it does not appear to progress in treated subjects. This raises the question of whether field change is a unique and non-progressive insult in the oesophageal mucosa(306).

This inclusion criterion only factored in acid reflux as a possible source of pathology. At the time of trial formulation, it was not clear whether non-acid reflux was related to symptoms. Therefore, setting up the trial as agnostic to non-acid reflux and controlling for non-acid reflux results in analysis was a deliberate and sound strategy. However, we now have evidence that non-acid reflux can be associated with symptoms of GORD. We now know that pH-MII captures ~50% more reflux episodes compared to pH-metry alone(96). Therefore, non-acid reflux **and** frequency of reflux episodes should be part of the diagnostic criteria.

At the time the REMOS trial received ethical approval as a clinical trial, there were no normative data available for the pH-MII. The role of non-acid reflux and its association with symptoms was simply not understood. The interpretation of the tests was based on expert assessment. In our institution, the expert interpretation set the reflux episode threshold at 73 reflux events in a 24-hour study period (EURPIG consensus). This threshold was based on study of 60 healthy adult volunteers(36,96). Understanding the differences in the pathophysiology of GORD in adults and children, we see in hindsight that this threshold is, at best, arbitrary. Diagnostic separation of reflux index and reflux episodes created interesting circumstances e.g. where pH-MII would demonstrate a 'normal' reflux index e.g. 3%, with an 'abnormal' number of predominantly non-acid reflux episodes e.g. 124. Should this be interpreted as a normal or abnormal study? Should this patient be included in the trial?

Since then, limited 'normative' data have since been published(36,96). These data have come too late to be applicable to the REMOS trial. However, some of the data are based on study of symptomatic children and selection of those with no acid reflux as control data(36). Other 'normative' studies are based on outlier populations i.e. preterm neonates(96).

In summary, using pH-MII as a key diagnostic for REMOS at a time when normal values were not known was a fundamental error in research trial protocol. Prospective long-term observational studies of MII

in both symptomatic and asymptomatic patients are required to define and validate diagnostic MII criteria for GORD.

Indeed, establishing and validating normative data will only be a first step. The next and perhaps most crucial step will be establishing the link between symptoms and reflux episodes. This was foreshadowed by the NASPHGAN consensus (2009)(22):

“This test (MII) detects acid, weakly acid and non-acid reflux episodes. It is superior to pH monitoring alone for evaluation of the temporal relation between symptoms and GER.”

Therefore, the key to GORD diagnosis is the demonstration symptom-reflux association.

Demonstrating reflux-symptom association

One of the calculable parameters from pH-MII recoding is the symptom association probability. We propose strict adherence to pH-MII symptom association probability (SAP) as the gold standard for demonstrating association between symptoms and reflux. An SAP >95% will be considered positive.

As detailed by Loots et al(307), expert observers demonstrate variability in characterisation of reflux events when analysing pH-MII traces(307). Literature also reveals variations in automated analysis settings. Some authors define reflux-symptom temporal association as occurring within a 30-second window(26). This leads to inter observer variability in pH-MII analysis. On this issue of analysis, we are supportive of the EUROPIG consensus(145) i.e. a 2 minute window for reflux-symptom association(96,137,145).

The verity of the study depends on faithful recording of symptoms by patients, parents and caregivers. Data from both adult and paediatric studies have shown that symptom recording inaccuracies(100). To mitigate this, researchers continue to seek objective ways to record symptoms during MII. For example, video recording for seizure activity and pulmonography for apnoea have been reported. Acoustic cough recording during MII is a non-invasive and objective way for recoding cough episodes during MII(100). In the absence of these enhanced test conditions, clinicians should endeavour to optimise test conditions and train parents carefully in symptom recording.

Objective and validated methods of symptom recording e.g. our TARDIS:REFLUX app can be used in conjunction with the pH-MII data-loggers to create an accurate temporal log of symptom episodes. Compared the Sandhill datalogger, the TARDIS:REFLUX app allows recording of 21 symptoms rather than being limited to 4 symptoms.

A new diagnostic criterion

Based on this work, we propose a diagnostic standard for GORD in children. This new diagnostic standard is a development on Montreal consensus(22) that better delineates diagnostic criteria for GOR and GORD.

The Montreal consensus (4)- defines GORD as:

“...a condition that develops when the reflux of stomach contents causes troublesome symptoms and/or complications.”

We recognize that GOR can be suspected without the actual demonstration of reflux. We acknowledge that a nebula of symptoms has been associated with GOR. To separate the suspicion from the

diagnosis, a second construct is required when there is a demonstrable association between GOR symptoms and reflux episodes.

Therefore, we redefine:

GOR Syndrome: 2 of more symptoms suspected to be associated with reflux

GOR Disease: when association of symptoms and reflux episodes is demonstrated empirically on pH-MII (SAP >95%).

We recommended in a prior section that diagnosis of GORD be made only after corrected gestational age of > 1 year. Concurrently, pH-MII testing is recommended after this threshold. This is to mitigate measurement of physiological reflux in the first year of life as leading to operative intervention.

The use of UGIC is not recommended as a diagnostic. As discussed in the introduction, this test is a hangover from hiatus hernia diagnosis in adults. It is poorly sensitive for GORD and low yield for structural anomalies. Notably, NICE guidelines(308) published in 2015 agree with this position. The statement is summarised as:

"Do not offer an upper gastrointestinal (GI) contrast study to diagnose or assess the severity of GORD in infants, children and young people."

Limitations: ALTE as an absolute indication for GORD

Apparent life-threatening events (ALTEs) are episodes of apnoea, colour change, hypotonia, choking and gagging(309). Although rare (<1% of emergency visits to hospital in infants under 12 months), failed resuscitation can result in death. Clinicians may be hesitant to postpone diagnosis of GORD till the first birthday in an infant with ALTE.

The association between ALTEs and GORD is thought to be aspiration. As discussed in chapter 1, aspiration can be due to impaired gag reflex, depressed consciousness status or gastrointestinal causes e.g. GOR or oesophageal dysmotility. Where aspiration is observed in the presence of uncoordinated swallowing or tracheal anomalies, fundoplication would not improve these factors. In fact, it may make symptoms worse(123).

Fear of aspiration may be more salient than demonstrable proof of aspiration. In 2012, a survey of 6 surgeons treating 169 patients referred for gastrostomy at the University of Colorado was published(1). Surgeons stated aspiration as a reason for concomitant fundoplication in 81% of cases. However, in review of case notes, only 6% had a clinical history of aspiration. Furthermore, there is no standardized, sensitive test for aspiration events. In summary, the relationship between GORD and aspiration events is rarely demonstrated objectively prior to intervention.

This is reflected in the NICE guidelines that state:

"GOR only rarely causes episodes of apnoea or ALTEs, but consider referral for specialist investigations if it suspected as a possible risk factor…"

What are the implications for the risk stratification pathway? It is reasonable that children at risk of ALTE children are investigated promptly. Waiting for the 1st birthday is not recommended. In fact, these children will benefit from early assessment by an MDT. In addition to early pH-MII, bronchoscopy and pepsin assay testing is recommended. New methods in microbiomics are expected to supplant these

methods Where feasible, fluoroscopic and manometric assessment of oesophageal motility may be contributory. Where aspiration is linked to neurological status or oesophageal dysmotility, fundoplication is not indicated. Where aspiration can be clearly shown to arise from gastroesophageal reflux, children may be considered for early fundoplication. However, temporising with distal feeding into the jejunum until the age of physiological reflux has passed would be a recommended strategy. This is discussed in detail in the next section.

COMORBIDITIES: INTERACTING COMORBIDITIES REQUIRE INTERACTION OF SPECIALISTS

To understand GORD as holistically as possible, we sought to estimate the effect of comorbidities as risk factors for fundoplication. NICE guidelines(308) published in 2015 take a similar panoramic approach. The guidelines state:

"When deciding whether to investigate or treat, take into account that the following are associated with an increased prevalence of GORD.

- *Premature birth*
- *Parental history of heartburn or acid regurgitation*
- *Obesity*
- *Hiatus hernia*
- *History of congenital diaphragmatic hernia (repaired)*
- *History of congenital oesophageal atresia (repaired)*
- *A neuro-disability…"*

To investigate comorbidities and their relationship with GOR, we took a data mining approach. Since work began on this project, several authors in this field have published database studies illustrating this approach. In 2013, Barnhard et al(281) published applied a data mining analysis to the Paediatric Health Information System. This is an administrative database with demographic, diagnostic and procedure data. They reviewed patient admission data retrospectively (2005-2010). The aim of the study was to identify patients undergoing gastrostomy tube and fundoplication procedures. Specifically, they searched for patients undergoing repeat hospitalisation in the same institution. Although physiological data are not available, there are demographic, diagnostic and procedure data within this database. The key outcome measure was repeat hospitalisation within the same institution. Their aim was to address the dilemma clinical of gastrostomy prior to or concurrent with fundoplication. Like our database, the PHIS was limited by absence of clinical symptom data. Authors also acknowledge that although diagnostic codes for neurological impairment could be identified, the severity of neurological impairment could not be discerned from simple from ICD codes.

Goldin et al(310) also reviewed the Washington State Comprehensive Hospital Abstract Reporting System database (CHARS) to identify children undergoing anti-reflux procedures. Data was captured for the period 1987 to 2001. The data contained demographic, hospital episode and procedural variables. ICD-9 coding was also available. They used the database to compare pre- and post- anti-reflux procedure hospitalization rates.

In a more recent database study, McAteer et al (298)reviewed the PHIS. Data was captured for the years 2002 to 2010. The aim of this study was to describe characteristics of paediatric patients (age <18years) undergoing ARPs. The database characteristics are remarkably similar to our institutional database, describe below. Data on demographics, comorbidities, investigations, procedures and hospital admission episodes were available.

A key contribution of the retrospective database review and subsequent data modelling is identification of comorbidities that increase a patient's likelihood to receive fundoplication.

These are:

- Neurological impairment
- CDH
- OATOF
- CLD
- Prematurity
- Skeletal anomalies
- Renal anomalies

We also identified comorbidities, which, in interaction, increase risk of fundoplication. These are:

- NI and cleft anomalies
- NI and tracheal anomalies
- NI and cardiac disease
- NI and swallowing disorders

Our methodology findings are echoed by those of McAteer et al(298). These authors used a similar methodology i.e. data mining a retrospective a large administrative database. Their review of the PHIS between2002-2010 identified >140000 children a primary diagnosis of GORD. They reviewed demographic and comorbid factors in patients undergoing ARS and identified factors increasing risk of ARS. Comorbidities are ordered from greater to lesser risk, with only significant hazards included.

Table 154: Comparison of factors increasing risk of anti-reflux surgery identified in retrospective database review studies

RetrospectiveGOR cohort	**McAteer et al**
Neurological impairment	Neurodevelopmental delay, Cerebral palsy
CDH	CDH
CLD	Cardiopulmonary disorder, Aspiration pneumonia
OATOF, Prematurity, **Skeletal anomalies,** **Renal anomalies**	-
-	Hiatal hernia, Failure to thrive, Intestinal fixation anomaly Chromosomal anomaly

In contrast to our findings McAteer et al(298) did not identify OATOF, skeletal or renal anomalies as increasing risk of ARS. Furthermore, data on gestation age were not available for analysis in the McAteer study. Lastly, the McAteer study is based on amalgamation of data from 41 children's hospitals over a large region. This reflects large variations in practice compared to our single center database review. Therefore, any further comparison e.g. of hazard ratios, would not stand scrutiny.

Below, we discuss the implicated comorbidities in turn, contrasting our findings with evidence in literature. Finally, we discuss implications or this understanding on the risk stratification algorithm for children with GORD.

Neurological impairment (NI)

Our retrospective database findings concur with those in literature on the interaction of NI and GORD. Children with NI have a higher incidence of reflux(52)(53,54,63). They are more likely to experience failure of medical therapy(311). They are also more likely to undergo surgery(52) and more likely to have surgical failure and require RF(312)(2,312). NI also increases mortality risk following fundoplication (200).

Although much has been written about NI and GORD, a key limitation is the variation in definition of NI. In the numerous studies documenting the effect of NI on GORD, few authors offered a single standard definition of NI(60)(281), . In the REMOS trial, we used the GCMFS scale to qualify neurological impairment. Knattens et al(60) followed the definition established by Barnhart et al(281) i.e. NI is:

"a static or progressive, central or peripheral neurological condition" associated with intellectual disability and or functional impairment".

Barnhart also published ICD 9 codes corresponding to NI e.g. cerebral palsy, hydrocephalus, and epilepsy. They further defined qualifying criteria for severe NI i.e. presence of cerebral palsy, mostly or exclusively tube fed, daily medications and non-ambulance based on the GCMFS scale. In contrast, Bohmer et al(313) utilized a standard of based on intelligence quotient (IQ). Patients with IQ<50 were included in their study. Indeed, most authors simply neglect to define NI. Therefore, any conclusions about NI and GORD carry a caveat i.e. NI is a broad term that encompasses multiple and disparate conditions.

Does the lack of standardized definition matter? Clinical experience informs us that the symptomatology of flaccid paralysis (e.g. myotonic dystrophy) varies from that of spastic paralysis (e.g. cerebral palsy). Should peripheral myopathies be included with brain anomalies under the umbrella of NI? To answer this question systematically, a study investigating patterns of symptoms and pH-impedance findings in patients with a variety of NI diagnoses is suggested.

Children with neurological impairment are more likely to experience failure of medical therapy. Children with NI are more likely to undergo surgery. Our retrospective database study identified an increased risk of ARS in children with NI (OR 5.3, 95%CI 4.3-6.4). Symptomatology is a suggested explanation for this phenomenon. Subjective symptoms e.g. abdominal pains, heartburn, are less likely to expressed in a child with NI. Therefore, when objective symptoms are observed e.g. vomiting, hematemesis, failure to thrive, they may represent a more advanced stage of GORD(312). Patients with complex symptomatology may be more likely to be selected for surgery.

Children with NI are more likely to have surgical failure and require RF. Patients with NI have more early post-operative complications, particularly respiratory infection(312). Authors suggest that this may be due to poor handling of oropharyngeal secretions, which remain unmitigated by fundoplication. A key data-mining finding was the interaction of NI and swallowing disorders to increase risk of

fundoplication. Often swallowing disorder and GORD are treated in separate clinical silos. As will be discussed later, interacting comorbidities demand the interaction of specialists.

In patients with multiple and complex comorbidities, failure of fundoplication may be related to operative intervention for inappropriate symptoms. The study by Kawahara et al(314) is illustrative. Authors (314) presented 58 children with NI selected for laparoscopic surgery. Majority (51) had pH-metry (although results are not reported). Although surgery-controlled vomiting in most patients, respiratory symptoms persisted in 52% of patients ultimately resulting in death in 22% of patients. This study is historic, but illustrates how fundoplication can be applied in a hopeful attempt to palliate symptoms unrelated to GORD.

Mechanical failure of the fundoplication is more common in children with NI. Intra-thoracic migration of the wrap occurs more frequently in children with NI(312,315,316). Explanations offered include spasticity, seizures and scoliosis(2,312). Peri- and post-operative strategies to mitigate these risk factors should be considered. For example, we propose a neurologist on the pre-operative multi-disciplinary team (MDT) who can advise on peri- and post-operative seizure control to protect the wrap.

Children who retch and vomit due to activation of the emetic reflex pre-operatively will continue to do so post operatively(125). In a small but important study, Richards and Miller(125) found that patients with retching symptoms on a pre-operative questionnaire continued to do so post fundoplication. In a much larger retrospective cohort analysis of 832 children who had fundoplication, Baerg et al(188) identified retching as an independent risk factor (OR: 3.59 (95% CI: 1.56–8.25)) predicting failure of fundoplication.

This post-operative retching is thought to contribute to mechanical failure and wrap herniation. Fundoplication has no known mechanism of action against the central emetic reflex. Therefore, this finding should be unsurprising and reverts importance problem of symptom definition. Where vomiting is an indication for fundoplication, an attempt to differentiate the effortless vomiting of reflux from the effortful, pallid vomiting of the emetic reflux must be made. A trial of centrally acting anti-emetics e.g. Ondansetron is prudent and has been recommended by some workers in this field[7]. The TARDIS:REFLUX app has a role to play in enabling parents to characterize symptoms. A desirable augmentation of the app could be the incorporating of video guides of demonstrating objective signs. Furthermore, the app could be modified for video-capture of patient symptoms for later discussion with health professional.

Re-operation rates and mortality are higher post fundoplication in NI children compared to NN children(2). Indeed, the retrospective database review identified a first revision rate of 20%. The second revision rate was 1.7%. One patient (0.2%) had four fundoplication in total. A valid question raised is whether re-operation is appropriate. If the mechanisms causing wrap disruption persists, repeating the operation would only lead to repeated poor results. Therefore, a comprehensive re-assessment of symptoms and repeating of objective test (UGIC, pH-impedance) is necessary before RF. Secondly, other procedures e.g. gastro-jejunal tube placement and oesophago-gastric dissociation should be considered as second-line therapy.

To summarize, we can apply our findings to the patient investigation and management algorithm mitigate risk in children with NI and GORD. Firstly, a consensus statement on definition of NI in relation to GORD is required. We adopt the Barnhart et al ICD 9 criterion. Although this criterion encompasses a range of diagnoses, a broad inclusive criterion bears more utility. Having met the inclusion criteria, a patient's symptom definition is key. A symptom tracker can be applied to clearly define a child's pattern of symptoms. Next, association between symptom and reflux should be demonstrated. Lastly, given the high coincidence of GORD and NI, we recommend that a neurologist should be a fixture of the MDT panel assessing patients for fundoplication.

Prematurity

Our database study demonstrated an increased risk of fundoplication associated with prematurity (OR 2.1, 95%CI 1.6-2.6). To our knowledge, ours is the first study identifying prematurity as an independent risk factor for progression to fundoplication.

GORD is often diagnosed in infants with prematurity due to delay in attainment of enteral independence. Ventilation may be a hidden risk factor. Pepsin, a marker for gastric aspiration, has been identified in tracheal aspirates of mechanically ventilated children(317,318). Furthermore, fundoplication is a proven strategy for reducing ventilator dependence in premature infants(95).

One might assume that in infants with a history of prematurity physiological reflux resolves commensurate to corrected gestational age. Given the lack of comparative studies, we cannot make any conclusions on this observation. However, it may be more useful to consider GORD as part of a prematurity sequence. Cohort studies have demonstrated the sequelae of prematurity i.e. intra-ventricular haemorrhage, sub-glottic stenosis(319), cerebral palsy, neuro-cognitive and motor delay and associated with feeding and swallowing difficulties(320). Fundoplication is useful in weaning premature infants off ventilation(95) thus locating GORD within the prematurity sequence.

The mechanism by which prematurity pre-disposes to severe GORD and fundoplication is unclear. For the purpose of risk stratification history of prematurity should be noted. However, at this nascent stage in our understanding, we recommend that this risk factor is not used to influence decision-making for/against fundoplication. A future research direction is assessment of fundoplication risk whilst adjusting for gestation as a baseline covariable.

Whilst it is not possible to mitigate the effects of congenital anomalies e.g. OATOF, one can certainly consider the effect of prematurity. Therefore, the we have recommended a corrected gestational age threshold of >1year for investigation of reflux symptoms.

Children with prematurity or NI often achieve motor and cognitive milestones with delay. Therefore, rather than age specific criteria, a movement towards milestone criteria is recommended. More research investigating the relationship between measurable milestones and the resolution of physiological reflux is urgently required. Furthermore, the surgeon must have an understanding of the expected progress of a child with prematurity or NI through their milestones. Here, a multi-disciplinary approach offers a ready remedy.

Congenital diaphragmatic hernia (CDH)

In our modelling exercise, we identified CDH as a comorbidity that increases both risk of fundoplication (OR 6.3, 95%CI 3.7-10.5) and risk of RF (OR 4.9, 95%CI 1.9-11.8). This increased risk is borne out in literature. The reported incidence of GOR symptoms in after CDH repair ranges from 20-88%(83) (321)%. In studies of adults with a history of CDH, the incidence of oesophagitis and Barrett's oesophagus is high(83).

It is difficult to ascertain the direction of causality. Is increased risk of fundoplication due to severity of GORD in CDH, or evidence of surgeon's bias? As with other comorbidities, the surgeon may be more likely to select a patient for fundoplication knowing the attendant risk associated with CDH. This is particularly likely when there is no standardization of reflux testing as observed in the Retrospective GOR cohort.

Several authors have tried to identify CDH factors that pre-dispose to fundoplication. Univariate analysis has identified several factors including:

- Pre-natal diagnosis of CDH(90,322)
- Gestational age >37 weeks (90)
- Size of defect(90,323), Patch repair of CDH(89,194)
- Severity of pre-operative respiratory compromise
  - Longer duration of ventilation(322)
  - High frequency oscillatory ventilation (89)
  - Severity of pulmonary hypertension, extra-corporeal membrane oxygenation (ECMO) (89), use of inhaled nitric oxide (90)
  - Apgar score at 5 minutes <6 (90)
- Pre-operative thoracic position of the stomach prior to CDH repair(322).
- Pre-operative thoracic position of liver(194)
- Tube feeding at discharge(90)
- Longer duration of hospitalisation(322)
- Fetoscopic endoscopic tracheal occlusion (89)

All these factors have been identified from observational studies applying univariate analysis to small cohorts. Literature in this field is characterized by selected sub-group analysis e.g. long-term outcomes of CDH repair in children who have had ECMO at a single institution(321). Therefore, we cannot make any conclusions about predictive CDH-factors. However, these data highlight that the nature, presentation and treatment of the CDH defect should be factored in when assessing patient risk for fundoplication.

Several mechanisms have been suggested to explain the association between GORD and CDH. It is known, however, that the angle of His, the length of intra-abdominal oesophagus and the diaphragmatic crura contribute to the integrity of the lower oesophageal sphincter mechanism. These three elements are disrupted in CDH. Sigalet et al(323) suggested that the absence or insufficiency of the left diaphragmatic crus are contributory to post-operative reflux. Kieffer et al(322) noted that diaphragmatic repair does not restore the intra-abdominal portion of the oesophagus. Fasching et al(195) examined

diaphragm motility with ultrasound following CDH repair. They observed a reduced motility on the side of repair. They suggested that an increase in the abdominal-to-thoracic pressure gradient post repair predisposes to both reflux and hiatus hernia. This mechanism might explain the higher rates of both primary and re-do fundoplication rate in children with CDH.

CDH could be a sentinel sign for abnormal development of the oesophagus. Pederiva et al(324) demonstrated scarcity of oesophageal neural fibres in the Sprague rat nitrofen-induced CDH model. In earlier post mortem studies of human infants, Pederiva et al(325) further demonstrated a decrease in neural fibres in the inter-muscular plexus in the lower third of the oesophagus, compared to controls. Arena et al(197) reported a reduction in type I and II peristaltic waves in CDH patients thought the significance in the absence of normative data is unclear.

Some surgeons will perform 'prophylactic fundoplication' at the time of CDH repair in anticipation of GORD and growth failure. Dariel et al (91) assessed growth outcomes of infants who had received prophylactic fundoplication. They found that survival with growth failure was more likely in the group that did not have prophylactic fundoplication and conclude that fundoplication has a role in preventing growth failure. However, the definition of growth failure in this study is problematic. Authors define growth failure as "weight-for-height and height-for-age ratios <−1.5 standard deviations (SD), because a −2 SD threshold would have eliminated too many cases, and −1 SD would not have been discriminatory". Another limitation of this study is timing of weight measures. Surgery occurred between the day 2 to day 10 of life. 'Growth failure' was diagnosed at follow up at 6 months and 1 year. Within the cohort, there will be a spread of birth weights. Some babies will, constitutionally, be born at the 2nd centile. However, no comparison with birth centiles or z-scores is reported.

It is not clear if prophylactic fundoplication makes a difference to GOR symptoms in CDH patients. Operative factors may play a role. GORD appears to be more common in patients who have a patch repair of the CDH(195). Without fundoplication, many CDH patients demonstrate resolution of reflux symptoms like other infants in the general population. In a prospective cohort study of 77 infant with a range of CDH severity, Verbelen et al(89) found that 21% of patients required anti-reflux surgery after CDH repair. However, 45% of patients were off 'anti-reflux medications' after a year of treatment.

Studies that better define the relationship between CDH and GORD is required e.g. an RCT in which infants get randomized to CDH repair with or without fundoplication. Such a study was performed by Maier et al(198). They followed up 79 neonates with CDH some of whom were randomly assigned to have 'prophylactic fundoplication'. GOR was defined as daily vomiting and/or recurrent chest infections with failure to thrive. Patients meeting these criteria had UGIC and ph.-metry. At 6 months, 35% of patients who had fundoplication and 60% that did not report GOR symptoms (p=0.055). By 24 months, 20% of patients in each group had symptoms. Overall, this study does not support prophylactic fundoplication. Like the study by Verbelen et al(89), this study also demonstrates that GOR symptoms improve with time in most patients with CDH.

What are the implications for risk stratification of the CDH patient? Based on the evidence, we acknowledge the increased risk of GORD and progression to fundoplication. However, we note that symptoms resolution is possible with increasing patient age. We also note the increased risk of

fundoplication failure. On balance, we would not recommend prophylactic fundoplication. Patients with CDH should have their primary repair and close follow-up with standardized objective investigations. This approach is described in the management algorithm proposed at the end of this chapter. A standardised approach ensures that although CDH is taken into consideration, criteria for fundoplication will be similar to criteria patients with other comorbidities. This process eliminates surgeons bias.

Oesophageal atresia with / without tracheoesophageal fistula (OATOF)

Our study identified increased risk of fundoplication in patients with a history of OATOF (OR 4.2, 95%CI 2.8-6.2). Patients with OATOF continue to experience respiratory morbidity(82,326) as well as GORD long after primary repair. Possible causes of ongoing pulmonary morbidity include pulmonary hypoplasia, pleural scarring from neonatal thoracotomy and scoliosis associated with thoracotomy(83). There is increasing evidence that underlying oesophageal dysmotility may be the mechanism for ongoing symptoms of GORD(78,79,81). Historically, however, an association between the respiratory morbidity and GORD has been assumed based on this co-incidence(82).

This co-incidence has not been substantiated by objective testing. A historic limitation is the methods used to demonstrate the relationship between GORD and pulmonary disease. Pulmonary function tests(327), are not validated for children under 8 years. Diagnosis of reflux is based on pH-metry and UGIC(82,326) does not directly demonstrate aspiration. Reflux aspiration demonstrated on UGIC may be masked by, or secondary to narrowing at the oesophageal anastomosis(82,326).

We cannot determine from retrospective review if increased fundoplication risk in OATOF patients is due to GORD severity or surgeon bias. However, unlike CDH, there is no tradition of 'prophylactic fundoplication' with primary OATOF repair. This is because division of TOF requires a thoracic approach and fundoplication requires an abdominal approach. Hence no RCTs have been identified in which patients are randomised to OATOF repair with or without fundoplication.

The association between OATOF, GORD, pulmonary disease and oesophageal dysmotility needs urgent re-examination with application of up to date techniques e.g. pH-MII and app-based symptom trackers. Symptom tracking is essential to identify of the OATOF symptom profile differs from that of other GORD patients. Oesophageal manometry is recommended in all OATOF patients prior to fundoplication. Objective tests for aspiration are fundamental in demonstrating GORD-pulmonary disease association. Historic biomarkers for aspiration e.g. LLM and pepsin will be supplanted by biomic methods e.g. 16S sequencing of gastric, oropharyngeal and pulmonary aspirates to identify phylogeny of microbial flora(328).

What are the implications for risk stratification of patients with a history of OATOF? It would be unfortunate to perform fundoplication in vain hope of improving symptoms caused by either oesophageal dysmotility or underlying pulmonary disease. Indeed, a background of oesophageal dysmotility may render symptoms worse and predispose the patient to fundoplication failure. Therefore, our management algorithm recommends – in addition to standardised objective tests- oesophageal manometry for all OATOF patients and MDT panel prior to fundoplication.

Chronic lung disease (CLD)

In our retrospective cohort study, we found that CLD increases the risk of fundoplication (OR 2.7, 95%CI 1.9-3.8). This finding is unsurprising. Ventilator dependence is a feature of chronic lung disease particularly in premature infants. The practice of fundoplication to assist in weaning off mechanical ventilation is established. In 2003, Barnes et al(103) also reported on a series of 10 (corrected age 11 .5 weeks (term -22 weeks) who had fundoplication for ALTE episodes requiring ventilation. All 10 had a reflux index of at least 10% demonstrated on double-probe ph.-metry.

More recently, this author(95) published a study of 26 ventilator-dependent children (median age 5.8 (0.8 - 19.4 months) who were weaned off ventilation following fundoplication(95). In this study, 13 children had emergency fundoplication following ALTE or respiratory deterioration attributed to GORD. None of the children had pre-operative MII or ph.-Metry. The decision to perform fundoplication was made on clinical assessment of symptoms. Following fundoplication, the median time to extubation was 4 (2 - 18 days).

In a smaller study in a smaller infant, Varal et al(329) report fundoplication in an ex-30-week gestation, 9-week infant. GOR symptoms persisted despite trans-pyloric feeding, alginate and pro-kinetic therapy. Although ph-metry demonstrated GOR, the indication for surgery was aspiration. A chest X-ray demonstrated pneumonic infiltration and atelectasis. However, these radiographic findings can be seen in CLD and are not specific to aspiration. There was no objective test for aspiration. This infant was weaned off ventilation by the 3rd post-operative day.

In these three series, a common feature is the lack of objective evidence for the association of GORD and aspiration events. As discussed in the introduction, there is no clear evidence to support the use of either pepsin(330), or lipid-laden macrophages as biomarkers for aspiration. These tests remain supportive but inconclusive.

New methods are emerging that offer some hope. A tripartite of symptom-sign-reflux association will provide a stronger basis for diagnosis of GORD. Djeddi et al(331) used polysomnongraphy combined with pH-MII to assess the relationship between autonomic nervous system activity and reflux events in 19 neonates Rosen et al(328) have recently reported on the use of 16S ribosomal RNA sequencing to match bacterial colonies in the gastric, broncho-alveolar lavage, and oropharyngeal fluids obtained through broncho-alveolar lavage. Notably, Blondeau et al(104) demonstrated how cough and GOR were monitored simultaneously through the use of a manometric MII system. Temporal association was noted when cough on manometry and reflux on MII occurred within a 2-minute time window. An association was considered positive if, over a tracing period, the ([Cough+Reflux] or [Reflux+Cough]) symptom association probability was >95%. [Cough+Reflux] was demonstrated in 19 patients, [Reflux+Cough] was demonstrated in 18 patients. Although the study is too small to make any conclusions, it demonstrates a viable method for investigating individual patients with chronic cough.

In our database study, we were surprised to find no association between asthma and GORD. This is interesting negative as there are multiple adult studies reporting the relationship of asthma, chronic cough and GORD. The putative mechanism is the triggering of bronchospasm by micro-aspirations acting on reactive airways. The co-incidence of chronic cough (persisting for more than 12 weeks) and GORD is estimated at 20-40% in adult practice(332). A meta-analysis of 28 observational and

interventional studies addressed this question(332). Studies in which adults had both asthma and GORD symptoms, ph.-metry or endoscopic findings consistent with GORD were included. Overall, the pooled odds for asthma risk in adult patients with GORD was 2.26 (1.8-2.8 95%CI). A diagnosis of asthma preceded GORD in some studies, and vice versa in others. Therefore, causality and its direction were unproven.

Paediatric data is scant. A systematic review and meta-analysis of GORD RCTs by Chang et al(105) was limited by an insufficiency of paediatric data. Interventional studies in small cohorts were too dissimilar to compare. Many studies considered multiple symptoms associated with GORD and determining the specific effect of GORD treatments on cough was difficult. The review mentions a single study(179) in which the effect of lansoprazole and placebo on cough outcomes in children was compared. The placebo group was favoured. Yet despite paucity of supportive data, acid-suppression treatment is suggested in children with chronic cough(333).

Chang et al(208) make the point that both cough and GORD are common diseases. Cough is the most common symptom presented in general practice(208). Their co-incidence is statistically unsurprising and does not necessarily imply association(208).

In our retrospective cohort, CLD was not associated with increased risk of RF. This finding is reassuring. We could generously impute that there were few fundoplication failures in patients with CLD. More likely, however, is the effect of time. Like GORD, most infants with CLD simply 'grow out of it'.

Skeletal anomalies

Data mining identified an increased risk of ARS with skeletal anomalies (OR 4.2, 95%CI 2.3-11.7). This association has not been enumerated before. This subset of patients reflects idiosyncrasies of a quaternary referral centre. As detailed in the introduction, the orthopaedic team at GOSH treat children with scoliosis, developmental dysplasia of the hip, osteogenesis imperfecta and limb anomalies requiring reconstruction e.g. clubfoot. Of these, children with scoliosis are most likely to require fundoplication. Few centres in the UK undertake paediatric spinal surgery. Therefore, we find there is no literature available for comparison.

The challenges associated with fundoplication in the scoliotic child are recognized. As reported by Drucker et al(334) , the gastroesophageal junction lies high, hidden beneath the costal margin. The costal margin approaches the iliac crest, preventing adequate exposure. To mitigate these issues, authors proposed an open trans-thoracic approach i.e. a lateral incision between the 7th and 8th right rib space. Both Van der Zee et al(335) and Tatekawa et al(336) propose a laparoscopic approach to improve visibility and access to the diaphragmatic hiatus. Given these operative challenges, it is perhaps unsurprising that skeletal anomalies were also a risk factor for RF.

Renal disease

We identified an increased risk of fundoplication in patients with renal disease (OR5.3, 95%CI 2.3-11.8). This is a surprising finding. Unlike OATOF and CDH, there is little evidence in literature for an association for between GORD and renal disease.

GOSH is one of a few centres in the UK offering renal replacement therapy (hemofiltration, dialysis) for children. Therefore, this finding reflects a GOSH idiosyncrasy. Patients with renal disease often have

feeding difficulties and failure to thrive. This is due to appetite suppressant effects of underlying chronic disease and medications. Failure to thrive results in referral for gastrostomy tube (GT) placement. A surgeon receiving a referral for GT appropriately considers whether concomitant fundoplication (GTCF) is indicated, particularly if vomiting, poor feeding and abdominal pain are confounding symptoms.

GOSH has led the field in the management of growth and nutrition in children with severe chronic kidney disease. In 2002, Lederman et al(337) compared rates of peritonitis in patients with GT placement before or after onset of peritoneal dialysis. In 2010, Mekhali et al(338) reported on a cohort 101 children with chronic kidney disease treated at GOSH since 2000. They found that 66% of patients had a feeding for feeding for a period (median 1.7, range 0.1-6.9 years). The GT rate was 37% and the GTCF rate was 13%. The indications for fundoplication were not stated. Similarly, Humphrey and Najmaldin(339) reported 28 laparoscopic GT placements in children referred for feeding difficulties, failure to thrive and drug administration. Specific indications for laparoscopic GTCF in 16 of 28 patients- “mainly with neurological impairment”(339)- were not explicitly stated.

Neither our study nor the literature can explain why renal disease increases risk of fundoplication. However, GT /GTCF surgery is preferable before rather than after PD commencement(340). This is due to the risk of peritonitis associated with abdominal surgery in PD patients. Post-peritonitis adhesions inside the peritoneal cavity may render PD impossible. Therefore, we speculate that clinicians may set the threshold for opportunistic GTCF in renal patients who need GT lower.

Interaction of comorbidities requires interaction of specialists

Our data mining exercise included a search for interacting comorbidities that increased fundoplication risk.

We found increased risk where NI interacts with:

- cleft anomalies
- tracheal anomalies
- cardiac disease
- swallowing disorders.

Previous studies of fundoplication risk assume comorbidities are independent variables(52,65,72). In contrast, we took a factorial approach and investigated the data for interaction effects. Such an approach better reflects reality and better captures the complexity of the data. This approach also improves the signal-to-noise ratio. Single variable analysis suggested that 21 comorbidities had a significant effect on fundoplication risk. Searching for interactions pared the list of individual risk factors from 21 to 10. Furthermore, the confidence interval of the odds ratio was narrower for some individual risk factors (e.g. skeletal anomalies, renal disease) indicating increased precision.

Modelling to include interactions revealed some surprising risk complexes e.g. the co-incidence of NI and cleft anomalies (OR 108, 95%CI 9.5-4242). Although cleft palate is known association of GORD, the interaction of cleft anomalies and NI is not established. This finding may reflect the domicile of the study i.e. at a quaternary referral centre. Children with rare cleft syndromes (e.g. Cornelia de Lange(341) , CHARGE(342)) are over-represented at our institution and may skew the data. Notably,

the confidence interval for the increase in odds is extremely large, suggesting poor precision of this estimate.

Similarly, the observation of NI, cardiac disease and GORD in this cohort reflects the prevalence of congenital cardiac conditions at a quaternary centre. GORD is a feature of some cardiac syndromes e.g. Velo-cardio-facial syndrome(343), single ventricle physiology(344). Indeed, fundoplication has been used to correct growth failure and improve nutritional status as part of single ventricle palliation(344,345). The mechanism associating NI, cardiac disease and GORD is not known. Some have posited underlying autonomic dysfunction, with a decrease in parasympathetic tone observed prior to reflux episodes measured using impedance(331).

The co-incidence of NI and swallowing dysfunction also increases fundoplication risk (OR 12.8, 95%CI 3-91.4). Foregut dysmotility and immaturity are suggested mechanisms underlying this association(346). The direction of effects in the tri-complex of GORD, NI and swallowing dysfunction is not clear. Impedance manometry has demonstrated that children with GORD symptoms and pathological pH-MII parameters have prolonged oesophageal bolus transit time and slower transit velocity(148). Conversely, and perhaps surprisingly, improved swallowing function has been observed after treatment for GORD in infants and children(116).

Tracheal anomalies in childhood are rare and can be congenital or acquired (e.g. sub-glottic stenosis secondary to neonatal intubation)(119). Tracheal anomalies are managed at specialist centres to achieve a concentration of experience. The observation of significant populations of children with NI, tracheal anomalies and GORD is probably idiosyncratic to specialist centres with tracheal reconstruction services(117,347). Indeed, a similarly specialist centre (Children's Hospital in Philadelphia) identified GORD in fifth of patients with tracheal anomalies(117). Furthermore, the management pathway for tracheal anomalies often involves actively seeking and treating GORD as a prophylactic against aspiration(118) and recurrence of stenosis(347).

These data reiterate the importance of NI as a risk factor for GORD. These findings have implications for our risk stratification model. We have recommended an MDT criterion in the risk stratification algorithm. Given the importance of NI in GORD, both as an independent variable and a co-factor, the author recommends the permanent inclusion of a paediatric neurologist in this MDT panel.

It is important to acknowledge a significant methodological limitation. We performed a search for 2-way interactions i.e. comorbid couplets. However, as complex interplay between comorbidities is likely, a search for n-way interactions would be superior. This would identify clusters or diseaseomes (348) rather than couplets. Two-way interactions of 29 comorbidities generated 812 different permutations, and hence 406 combinations for analysis. The search for 2-way interactions was completed within 6 hours. Searching for n-way interactions was untenable due to memory and processing limitations of the authors computer (Intel core 2.7gHx CPU, 6G RAM, Windows OS) and software package (R, version 3.1.0, 64-bit). However, the strategy adopted -albeit limited- is both illustrative and pragmatic.

Limitations of comorbidity analysis

As described in Chapter 2, GOSH is a quaternary referral hospital that does not have an Accident and Emergency department. Patients must be referred from another hospital to GOSH for a specialist

service. The patterns of disease seen at GOSH does not, inherently, reflect the distribution of disease in the population. Instead, it reflects the services available at GOSH which, in turn, reflect disease rarity, severity and NHS specialist provision policies.

However, GORD is a chronic disease. A general paediatric clinic sample of patients would probably be the best representation of the comorbidities associated with GORD. Unfortunately, GOSH did not offer a general paediatric specialty team during the retrospective cohort capture period. Therefore, cohort patients have been filtered through other specialists and may represent patients on the complex end of the spectrum. Fundoplication rates are probably unrepresentative.

Despite these sampling limitations, the sample of population is probably representative of populations seen by other surgical centres. Paediatric surgery units are based at tertiary centres. Therefore, most paediatric surgeon will see a filtered sample of the population. Children with cardiac and renal disease are subpopulations that may be idiosyncratic to GOSH and other centres treating complex congenital heart disease and providing renal replacement therapy respectively. Other special groups represented in our cohort are patients with epidermolysis bullosa, tracheal clefts and osteogenesis imperfecta. However, the common comorbidities implicated in fundoplication risk (NI, CDH, TOFOA and CLD) would be commonly found in most paediatric surgery centres.

SURGICAL APPROACH: A STAGED SURGICAL APPROACH APPEARS SAFEST

In the introductory chapter, we summarized indications for fundoplication as:

- reflux symptom control
- opportunistic
- prophylaxis

Ideally, the surgeon wishes to operate for symptom control. However, as patients often present with both feeding difficulties and reflux, fundoplication cannot be considered without first addressing gastrostomy formation. The surgeon's decision point is often: gastrostomy with / without fundoplication?

Does GT increase risk of GORD?

Prior to 2010, as noted in a systematic review by Noble et al(285), the evidence available to answer this question was conflicting and often of poor quality. The most consistent finding was that GT placement did not lead to an increase in adversely affect ph.-metry(57,211,349) or reflux symptoms(276,350).

This paucity of evidence before 2010 is due to retrospective study methodology, small sample sizes and inconsistency in pre-GT testing. After 2010, when the REMOS trial was well underway, more authors addressed this question. The retrospective review or practice published by Toporowskja-Kowalska et al(351) in 2013 illustrates a continuing problem with methodology. They identified 348 polish children who had GT placement (2000-2010). They found that only 51% of patients had pre-operative testing. Of these, a fifth had ph.-metry and 5% had MII. Another fourth had scintigraphy and a further third had UGIC. As discussed, these latter tests are not appropriate for diagnosis of GORD.

More recently, as pH-MII availability has increased, pre- and post- GT MII parameters have been reported. In a later study, Toporowskja-Kowalska et al(212) found no significant difference in MII results in 15 patients with NI who had repeat MII 6-8 months post gastrostomy.

Findings from the REMOS study are contributory. In subset analysis of our patients with NI, we assessed reflux symptoms, MII and nutritional outcomes pre-and post-GT placement. In the GT group, there was no change in reflux symptoms at 6- and 12-months' post GT placement. The trial identified no difference in early (6-month, 12 month) patient-centred outcomes. There was no difference in scores for the gastro-intestinal symptom questionnaire, the child and parent quality of life score and the parenting stress index-depression scale. Objective tests for reflux post-procedure yield incomplete data. MII testing post-GT placement was available for only 6/16 patients. This is due to withheld consent due to the invasive nature of the test. However, in the few patients for whom data was available, we found no adverse change in MII parameters. Lastly, all patients demonstrated improved nutritional outcomes following GT placement.

We conclude that GT does not increase risk of GORD. Therefore, anticipation of worsening reflux after GT placement cannot be used to justify GTCF.

How many patients require fundoplication post gastrostomy?

The REMOS trial methodology was established to systematically address this question. All patients had objective reflux testing (pH-MII) prior to intervention with pH-MII. Unfortunately, however, long-term data on redo or progression of fundoplication are not available (as the study halted after 16 months due to specific concerns regarding a gastric emptying test substrate). At 12 months' follow-up, however, none of the patients in the GT group were being considered for fundoplication.

On the face of it, this is a simple question to answer. For example, Catto-Smith et al(352) reviewed 173 children with NI who had PEG placement at the Royal Children's hospital Melbourne between 1990-1997. They found that only 4(2.3%) of these went on to require subsequent fundoplication. However, like many authors, the primary indication for GT is feeding difficulty. There was no indication to routinely investigate patients and quantify degree of reflux pre-GT placement. Absence of pre- and post-operative objective measures of reflux is a widespread problem. Conflation of feeding interventions (GT) and reflux control (GTCF) also confounds interpretation of data.

Barnhart et al(281) review data from 42 centres in the US, they identified 4163 infants with NI who had GT or GTCF placement whilst on a neonatal intensive care. The majority, 66% of infants, had GT only. Of these, only 4.7% had subsequent fundoplication within the follow-up year. This is comparable to Catto-Smith et al(56) (2.3%).

Given this low rate of subsequent fundoplication in the GT only cohorts, we recommend a staged approach. GT should be offered for patients with feeding difficulties. If reflux control fails after GT placement, despite ASM and a trial of jejunal feeding, fundoplication can then be considered. However, demonstrating association between symptoms and reflux remains a necessary pre-requisite for most patients. This staged approach is detailed below in the risk stratification algorithm.

Does fundoplication improve reflux symptoms or test parameters?

This is central question which few studies have answered systematically. Literature has focused on surgeon specific outcomes e.g. hiatus hernia, dysphagia with little published on symptom control. As discussed, this is because systematic methods of capturing pre- and post-operative symptoms have been lacking. Here again, the REMOS data is contributory. In the GTCF group, we found improved scores post on GORD symptom questionnaires post fundoplication. For those GTCF patients agreeing to MII post-surgery, there was demonstrable improvement in MII parameters.

At 12-month follow-up, none of the GTCF group had revision of fundoplication. This finding is attributable to a latency in outcome and longer-term follow-up is required. In our retrospective cohort study, the mean time to first revision was 22 months. Similarly, other studies have found a mean time to first revision of 16(214), 18(353), 27(188) , 30(189) months

The GT and GTCF groups had similar mean GORD and quality of life scores before and after surgery. Notably, both groups received a feeding intervention that has known benefit to the quality of life of the P/CG(285). Therefore, we speculate that improved symptoms scores in both groups are confounded by this 'feeding benefit'. Nutritional parameters improved in both groups with no significant difference. Again, this can be attributed to 'feeding benefit' to both groups.

Findings from literature are contradictory. Fundoplication does not appear to reduce subsequent reflux related hospitalization(310) or episodes of aspiration pneumonia(218),(354). A closer examination of the large retrospective cohort study by Barnhart et al(281) is warranted. In this study, data arise from an administrative database (Paediatric Health Information System) between 2005 and 2010. Although physiological data are not available, there are demographic, diagnostic and procedure data within this database. The primary outcome was subsequent reflux-related hospitalizations within the same institution in the first year of follow-up.

Propensity score matching was utilized to pseudo-randomise patients in the GT vs. GTCF group. Before matching, the cohort groups were significantly different. For example, on pneumonia and aspiration-key indication for fundoplication - the GTCF group had significantly higher rates. The GT group had significantly higher rates of gastrointestinal comorbidities. This could be interpreted as follows: the GT group had feeding difficulties and the GTCF group had airway difficulties. To mitigate this, authors use propensity matching on nearest neighbour demographics (e.g. sex, race, birthweight) as well as comorbidities (e.g. aspiration pneumonia, tracheostomy, ICD9CM codes for GORD and GORD medications). However, in the absence of a GORD diagnostic, propensity matching tells us that the patient groups had comparable morbidity. It does not confirm that patient groups had were matched for GORD severity status.

Lack of GORD matching is secondary to absence of objective measures of GORD. Indeed, Barnhart et al(281) report "extreme variation in surgical practice". Only 9.4% of the GTCF cohort had pH monitoring. Most patients had UGIC and 4.3% had an endoscopy. Authors report inconsistencies in diagnosis of pathological GORD in neonates. Therefore, GORD was not an inclusion criterion in this study.

Addressing this data limitation, propensity scoring yielded 1027 infants in each group who could be matched on demographics and non-GORD comorbidities. On the key outcome measure, authors found

no difference between reflux-related hospital admission in the first year between the GT and GTCF group.

Lee et al(354) also found that there was no difference in pre and post-ARS hospitalisation rates. Specifically, patients were readmitted at similar rates for aspiration and other pneumonia, respiratory distress, apnoea and failure to thrive both pre- and post-fundoplication.

Given the low rate of subsequent fundoplication in the GT only cohorts, and similar rates of reflux-related admission in the GTCF group, we recommend a staged approach i.e. 'GT, wait-and-see '. Surprisingly, Barnhart et al(281) draw the opposite conclusion. They conclude:

"In essence, these operations can be performed prophylactically in infants who may or may not have experienced complications of gastroesophageal reflux."

This conclusion is not supported by the evidence presented in their paper. No analysis is made of fundoplication related complications. These authors acknowledge that they were unable to capture complications of fundoplication that do now require re-admission e.g. "recurrent gastroesophageal reflux, gas bloat, and dumping syndrome". Furthermore, as reflux is not objectively demonstrated either before or after the fundoplication, can we be certain that reflux symptoms have been addressed by prophylactic fundoplication, rather than increasing age? Lastly, the authors were unable to prove treatment effect of GTCF on future reflux related hospital admissions. In this circumstance, surely opting for the less invasive procedure is most prudent? This is particularly crucial when considered the mortality risk associated with GTCF.

Post fundoplication mortality

Between 1994 and 2010, the mortality rate in the RetrospectiveGOR cohort is 6.93%. The mean age at death was 5.17 years (s.d.= 4.96, n=964). One startling finding was the increased rate of mortality in patients who underwent fundoplication. As expected, patients at GOSH had a higher mortality rate compared to the general population, particularly in infancy.

Compared to the GOSH total mortality data, the our Retrospective GOR cohort had a lower mortality rate.

Table 155: Comparing mortality rates for our cohort against overall GOSH total mortality

Type of mortality	**RetrospectiveGOR.db**	**GOSH ONS data**
Neonatal	0.14 (0.15)	3.5 (0.4)
Infant	1.6 (1.3)	5.2 (0.7)
Childhood death	5.4 (3.4)	1.4 (0.3)
*Data are mean (s.d.)		

We would expect fundoplication to stave off mortality. Indeed, reducing risk of death by aspiration is a rationale offered for the procedure. However, risk of mortality was increased in patients with fundoplication compared to those without fundoplication (odds ratio = 2.6, 2.1 -3.1 95% CI). The mean age at death was also significantly lower in patients who had undergone fundoplication. Our data revealed a 5-year survival rate of 99% for all children with GOR. The 10 and 20-year survival rates were

92% and 87% respectively, demonstrating a drop-in survival in adolescence. We speculated whether excess mortality in patients with fundoplication was a marker of severity of underlying respiratory disease. Risk of mortality was greatest in patients with acute respiratory failure, aspiration, tracheostomy and other tracheal and laryngeal anomalies. Subset analysis demonstrated that 5-year survival of children with NI was 80%. The 5-year survival was lowest for children with a history of aspiration pneumonia (40%). A history of cardiac disease or cardiac surgery was also associated with increased mortality risk.

Others have also reported the finding of increased mortality associated with fundoplication. Hebsen et al(355) , in a retrospective analysis, 334 neonates and infants undergoing single ventricle palliation. There was a significant increase in inter-stage (between 1st and 2nd stage operation) mortality in patients who underwent a Nissen fundoplication and gastrostomy tube placement, compared to those who persisted with nasogastric tube feeding. Authors proposed that this increased risk reflected the increased severity of illness, which predisposed patients to the fundoplication pathway a priori. However, in both this study and our own, causation is neither proven nor disproven. These findings may belie a contributory relationship between fundoplication and mortality risk.

In Hebsen's patients, mortality was attributable to respiratory illness in most patients who underwent ARS. This leads to the suspicion that fundoplication is not effective in preventing reflux-related mortality. Another possible explanation is that fundoplication is performed too late, after the onset of significant and life-threatening lung disease.

Mortality has also been attributed to neuro-disability. Hutton et al(356) reviewed survival of children with cerebral palsy in the north east of England. They identified adverse mortality with 74% of children with severe NI surviving to the age of 20 years. Similarly, Wockenforth et al(200) examined mortality post ARS and identified cerebral palsy as a risk factor. In a prospective cohort of 255 children who had surgery by one surgeon, they identified a 20% mortality rate. The 5-year survival post fundoplication was 59%. Regression analysis demonstrated that cerebral palsy in combination with pre-existing gastrostomy increased relative mortality risk 11-fold (95%CI 3.16-38.63). Published mortality rates post fundoplication are variable. Mortality appears to be increased in children with NI(218) who have had fundoplication (Table 157).

Table 156: Mortality rates following fundoplication

Author, Date	**Procedure**	**30-day post-operative survival**	**Overall mortality rate, follow-up time, 5-year survival (5YS)**
Wockenforth et al, 2011(200)	Open and laparoscopic (2) fundoplication, pyloroplasty in 4%.	11%	20%, -, 59% 5YS
Smith et al(216), 1992	Open fundoplication	26%	51%, 30 months, -
Fonkalsrud et al**(63)**, 1998.	Combined hospital study. Laparoscopic and open Nissen, Thal and Toupet fundoplications. Pyloroplasty in 11% of patients.	0.07% NN, 0.8% NI	-
Tovar et al(193), 2007.	Laparoscopic and open fundoplication between 1992-2006. Selective pyloroplasty.	1%	6.7%, 51 months, -
Rothenberg**(189)**, 2005.	Laparoscopic	1 in 1050	-
Pearl et al, 1990**(2)**.	Open fundoplication		28% NI, 6% NN 53% 5YS
* Due to malfunction of ventriculo-peritoneal shunt			

It would be inaccurate to suggest that there is no mortality risk associated with GT placement only. Straus et al(357) investigated survival of Californian children with cerebral palsy. Of these, those with a gastrostomy had a median survival of 7 years. In an update to this work(358) capturing the years 1983-2010, the group found that mortality rates in children with CP improved with overall childhood mortality rates in the US. However, these improvements disappeared in adolescence and adulthood, with patients with cerebral palsy at risk of early and excess mortality compared to the general population.

Barnhart et al(281) report mortality rates in a similar population. They reviewed a database to find 4163 with neurological impairment undergoing either GT or GTCF procedures. Overall mortality was 6.2%. Mortality was higher, but not significantly so, in the GT only group (6.4%) compared to the GTCF group (5.8%).

Catto-Smith et al(352) reported a 5YS of 61% following GT placement in children with neurological impairment. Sleigh et al(359) published a systematic review on risks of gastrostomy of jejunotomy feeding compared to oral feeding in children with cerebral palsy. These authors also identified an increased mortality risk following GT placement in children with cerebral palsy. However, the excess mortality is dramatically reduced when data is adjusted for degree of neuro-disability (hazard ratio from 38 to 3). This suggests that GT may be a surrogate of severity of neuro-disability, rather than an independent risk factor for mortality.

It is not possible to conclude that fundoplication confers no survival advantage based on these data. A case-control study comparing patients carefully matched for comorbidities and disease severity would be an ideal study design to answer this question.

Failure of surgery

Perhaps the most subjective reason against fundoplication is the RF rate. As a surgeon, I would hesitate to obtain consent for a procedure with a high failure rate. In the retrospective cohort study, we identified a first revision rate of 6.5%. Literature review identified a first revision rate of 1.6-18%).

To place this in context, the revision rate following division of tracheooesophageal fistula is 3-10%. The revision rate following congenital diaphragmatic hernia repair is 5-12%. When faced with high revision rates, the surgeon is dutybound to examine factors leading to failure of surgery.

These factors can be summarised as patient factors and operative factors. The Retrospective GOR cohort identified similar findings to those in literature (discussed in the introduction). We identified several patient comorbidities that are consistently implicated in predicting first revision fundoplication, irrespective of modelling method utilised. These are neurological impairment, skeletal anomalies, CDH and cardiac disease.

The finding of cardiac disease is noteworthy. In our retrospective cohort study, 3 of 4 models identified cardiac disease as a predictor of failure of fundoplication. The DT model was accepted as the optimal specification for the RF data. Using this loss-adjusted, cross-validated model identified cardiac disease as the second most important factor predicting RF. The logistic regression model identified a 4-fold increase in redo risk associated with cardiac disease.

This association is probably an idiosyncrasy of our cohort. The care of children with complex congenital heart disease takes place in tertiary and quaternary centres. This is a deliberate strategy to concentrate experience and expertise. Therefore, our cohort will have an over-representation of children with cardiac disease compared to a study based on community or district general hospital sampling. Single ventricular physiology tetralogy of Fallot, coarctation of the aorta and transposition of great vessels are amongst conditions requiring cardiac surgery in infancy and childhood. GOR has been reported in 30% of patients with single ventricular physiology(95) .

How can we explain the finding that cardiac disease was a predictor of fundoplication failure? One explanation could be open surgery. There exists a reluctance to perform laparoscopic surgery in patients with cardiac disease due to concerns about reduced venous return and cardiac embarrassment associated with pneumoperitoneum. Knott et (360) al observed that laparoscopic pneumoperitoneum, although tolerated, results in a fractional shortening in patients with single ventricular physiology i.e. the heart is less well filled and appears shorter at the end of diastole.

Interestingly, our modelling did not find cardiac disease to independently increase risk of first fundoplication. This was surprising. Changes in cardiac procedures may explain this phenomenon. Modifications to cardiac procedures to spare the recurrent laryngeal nerve have led to reduced rates of vocal fold palsy and laryngo-pharyngeal dysfunction(361). This in turn enables safer oral feeding and reduces aspiration events.

Secondly, failure to thrive in patients undergoing staged cardiac surgery is a long-observed and established phenomenon. Clinicians may be more accepting of poor growth velocity in this cohort and defer management till cardiac palliation is complete.

Can these patient factors increasing risk of failure be modified? As patients with these comorbidities make up a large proportion of patients referred for surgery, a strategy of exclusion is invalid. However, clinicians can aim to delineate symptoms related to comorbidities from symptoms related to GORD. Therefore, the indications for ARS need to be standardised, stated and met.

In our retrospective review, we were unable to identify indications for surgery. This is a pitfall of an administrative database study. In the REMOS trial, indications for surgery were clearly stated albeit later revised! Authors in literature fare no better. Literature review has demonstrated that there is no standardisation of surgical indications(339,362). Indeed, analysis of surgical outcomes is greatly limited by the varied indications for surgery.

The retrospective GOR cohort had a second revision rate of 0.7%. Again, this was comparable to a range of 0.3-2.2% in literature. The third revision rate was 0.2% (0-0.6% in literature). Although these numbers are low, revision surgery demands an interrogation of the construct. What factors can be modified on revision that were not addressed in the first operation?

Pacilli et al(284) have reviewed factors leading to failure of redo surgery. The only significant patient factor identified was NI. Operative factors were open fundoplication and absence of gastrostomy placement at first fundoplication. Patient NI may be a modifiable risk factor. NI, as discussed prior, is a poorly defined and much applied term in this field. This finding highlights the need for a deeper understanding of NI. It may be that specific types of NI are associated with recalcitrant failure of fundoplication. If identified, then careful patient selection can mitigate this issue.

The operative factors associated with failed redo (open surgery and GT placement at first fundoplication) can certainly be modified.

Laparoscopic versus open surgery: confounded by patient selection

In the introduction, we discussed the role of laparoscopic versus open surgery extensively. Laparoscopic access offers benefits of minimal access approaches i.e. better post-operative pain management, lower risk of adhesions. However, evidence is confounded by variability in patient selection and surgeon practice. The evidence cannot be adequately assessed do to heterogeneity of indications for surgery.

In our retrospective review, we chose not to focus on laparoscopic/open surgery as the study spans a time of transition from open to laparoscopic surgery. The effect of learning curves is undeniable and confounding. At the time of the REMOS trial, evidence for benefits of laparoscopic surgery was widely accepted. In fact, laparoscopic fundoplication was the only modality offered to trial participants.

It is interesting to conclude with an observation by Tovar et a(193). These authors suggest that the perception of lower risk with minimal invasive laparoscopic surgery has probably loosened indications for fundoplication. The temptation to perform opportunistic, prophylactic fundoplication for a vomiting infant whose requires GT placement is acknowledged(193).

Implications for algorithm

Understanding the differences in morbidity and outcome with GT and GTCF has informed our risk stratification algorithm. Crucially, our risk stratification algorithm separates chronologically feeding

interventions (GT, gastrojejunostomy tubes) with interventions for reflux (ASM, ARS). We believe that the evidence available best supports this approach. Staged application of these interventions avoids conflation of indications and obviates the dilemma of opportunistic concomitant fundoplication.

Our algorithm also sets an age threshold (corrected gestational age >1 year) both investigation and intervention. This is supported by evidence that younger age at fundoplication is a risk factor for symptom recurrence.

Exceptions to this recommendation are envisioned. There will be children whose anaesthetic risk is great and who are at high risk of GORD complications e.g. children with cardiac disease and CDH, children with a history of aspiration/ALTE are also excepted. In these children, MDT discussion and appropriate investigations are recommended where GTCF is being considered.

CHAPTER 2: LESSONS FROM THE REMOS TRIAL

The REMOS trial addressed the pragmatic presentation often seen in clinic. A patient has both feeding difficulties and reflux. Will a gastrostomy tube improve both problems? Should fundoplication be performed? What is the severity of reflux symptoms that justifies a concomitant fundoplication?

WAS 'EQUIPOISE' JUSTIFIED?

When performing an opportunistic operation, a dilemma arises. A patient who requires a feeding intervention is perceived to be in at-risk category for GORD e.g. NI. Opportunistic fundoplication is considered with gastrostomy placement (GTCF). This latter dilemma was the basis of the REMOS trial. In writing the trial protocol, a position of equipoise was taken based on literature review.

At the time of inception of the REMOS study, the GT versus GTCF was one of the central dilemmas in this area of research. The REMOS trial was established from a position of both research and clinical equipoise i.e. viewing GT versus GTCF as similar interventions based on morbidity and outcome. Data from the REMOS interim review and other studies certainly support this assertion that outcomes i.e. reflux, nutrition and growth, are similar. It is true that there was 'genuine uncertainty within the expert medical community...about the preferred treatment", hence 'clinical equipoise'(363). However, the as the interventions are significantly dissimilar, this equipoise must be questioned.

The difference between theoretical and clinical equipoise is a much-exercised topic. In our trial interventions, equipoise is imbalanced by a calibration issue(364). For theoretical equipoise to exist, we must believe that- if the patient know which treatment was best, they would easily choose it over the other treatment. If GT was the better supported treatment, parents would certainly have no qualms accepting this over GTCF. It is a short endoscopic procedure with few minor complications. However, the converse is less supported. If GTCF were a better treatment, parents would be less eager to choose this treatment given the burden of morbidity and mortality.

As detailed in the introductory chapter, fundoplication changes the mode of operation (GT can be endoscopic, GTCF is laparoscopic with risk of laparotomy). Fundoplication lengthens duration of operation. Fundoplication is associated with more severe morbidity and mortality, and has a higher re-operation rate. Morphologically, fundoplication surgery is non-trivial and destructive. As detailed in the introduction, fundoplication involves dissection of the oesophageal hiatus, and may involve inadvertent injury to phrenic and vagus nerves(365). Reconstruction of the oesophageal hiatus only poorly mimics the natural state and does not address nerve injury. Not surprisingly, wrap hiatus herniation is the most common mechanical complication following surgery. Fundoplication also involves mobilization of the gastric cardia with / without division of short gastric vessels. Consequentially, the morbidity and mortality associated GTCF is not comparable to that associated with GT alone.

RANDOMISATION AND UNCONTROLLED VARIABLES

Randomizing patients to dissimilar interventions, although problematic, is ethical if patients understand differences in interventions and consent to treatment. For example, patients with aortic aneurysms have been randomised to open surgery vs endovascular stent insertion. In the former, a large incision is made into the abdomen and the vessel is visualized and handled directly. In the latter, a stent is introduced through a small incision in a peripheral artery and guided into place using serial X-rays.

Therefore, there is established precedent for dissimilar interventions where due ethical review has been conducted.

However, we must draw attention salient contrast. In a stent vs, open surgery trial, although the technical aspects and potential morbidity are dissimilar, both intervention arms have the aneurysm addressed. This trial methodology, when contrasted with the REMOS trial, illustrates a central methodological issue.

The primary outcome for the REMOS trial was reflux symptom control. GT, however, is a feeding intervention. Therefore, the GT group received no treatment for reflux. Effectively, randomization to GT vs GTCF was analogous to a 'no reflux intervention' vs. 'reflux intervention' study. To have comparative treatment arms both addressing reflux would require strict adherence to a protocol where one arm received GT+ acid suppression medications (ASM) and the other arm received GTCF and no medication. This was initially envisioned in the trial protocol:

"Gastrostomy with medical treatment versus gastrostomy with fundoplication in children with neurological impairment."

However, patients recruited for fundoplication were already on ASM before surgery. An early amendment was made to trial protocol to continue ASM for the GTCF arm in the post-operative period, and discontinue therapy at first clinic review. This modification essentially led to cross-over and confounded trial results.

Another methodological pitfall in surgical RCTs is controlling intervention. Even where an intervention is highly protocolized, each surgeon performs a different procedure by virtue of individuality and approach. Indeed, a single surgeon performing the same procedure will probably demonstrate some variations each time, adapting for patient anatomy, conditions on the day etc.

In the REMOS trial, performance of GT or GTCF was limited to consultant surgeons at our institution. This reflects the pragmatic approach taken. The protocol controlled for technique-related variations by involving only the top tier of surgeons available(364). Another safeguard against operator variability would be adding performing surgeon to the minimisation criteria. Pragmatic approaches have precedence in surgical trials. As stated by Vernon-Roberts et al(52):

"Pragmatic clinical trials (or well-constructed observational studies) characterized by sufficient statistical power; clinically appropriate, well-defined patient variables; and comprehensive, meaningful patient outcomes have the greatest potential to identify ideal surgical candidates."

CHAPTER 3: TRENDS SINCE BEGINNING THIS PROJECT

Work on this Thesis began in May 2010 and was completed in December 2016. In the interim, there have been some major changes in the paradigm, perception and management of GORD. These changes include concerns regarding use of acid suppression in neonates, decreasing use of fundoplication and increased use of database studies.

Move away from acid suppression in paediatric patients

Acid suppression therapies have long been synonymous with 'anti-reflux medication'. However, this established tenet is under threat from three emergent facts. Firstly, non-acid reflux is demonstrably a significant component of reflux episodes in children. Secondly, the efficacy of acid suppression in controlling reflux symptoms in children is in question. Lastly, there is increasing evidence of risk of infection associated with acid suppression.

Non-acid reflux comprises a significant component of reflux episodes in children. As early as 2001, Wenzl et al had demonstrated a temporal association between 5-10 second apnoea episodes and non-acid GOR in infants. Rosen found that non-acid reflux was correlated to respiratory symptoms in older children. In infants with apnoea, both Mousa et al(127) found that non-acid reflux episodes constitute 48% of all reflux events. However, only 15% of apnoea episodes in this study were temporarily associated with reflux. Furthermore, there was no difference in acid versus non-acid reflux association with apnoea. In a similar study in preterm infants with asthma, Magista et al(130) identified 76% of reflux episodes as non-acid in 6 preterm infants. Rosen et al(147) found that respiratory symptoms were more strongly correlated with non-acid reflux in 21 children. At the time of the REMOS study, although symptom association had not been conclusively demonstrated, the contribution of non-acid reflux to acid reflux episodes, particularly in infants, was clearly significant. More recently, in 2012, Shin et al (366) published a small study of 23 infants (median gestation 32 (28-38 weeks) with MII. In this group 59% of reflux episodes were non-acid. They identified a positive apnoea-reflux symptom association in 4 of 21 patients using MII.

We still cannot conclusively draw association between non-acid reflux and symptoms based on these studies. The study numbers are small and there is heterogeneity in definitions of symptoms. What we can conclude is that non-acid reflux forms a significant proportion of reflux episodes, particularly in milk fed infants(97).

Given the significance of non-acid reflux in paediatric GORD, the role and efficacy of acid suppression in controlling reflux symptoms in children is increasingly being questioned. In a recent Cochrane meta-analysis, Tighe et al(124) investigated the efficacy of ASM for control of reflux symptoms in children. There were only 12 RCT studies assessing use of PPIs in children (Omeprazole -3, Lansoprazole-2, Pantoprazole-4). Most studies compared PPIs against other acid suppression, or compared different dosages of the same PPI. For H2 antagonists, there were only 4 relevant RCTs in children(124).

Overall, this meta-analysis concludes that there is only moderate evidence to support the use of PPIs for GORD. The evidence for H2 antagonists was even less compelling. Evidence is poor due to lack of independent, well placebo-controlled trials. The review authors found that manuscript preparation by pharmaceutical companies was commonplace. Crucially, there were no RCTs on pharmacological

management of treatment with GORD and neurodisability identified. The findings of this Cochrane review are destabilising in two key areas: Firstly, a key indication for surgical management of GORD is failed medical management. It may be that medical management fails due to inefficacious medications. Secondly, children with neurodisability are over-represented in patients selected for surgical management for GORD. Absence of evidence for ASM in this subgroup leaves a large gap in our understanding.

We now understand the importance of non-acid reflux in children and have a suitable diagnostic i.e. MII. Therefore, a pharmacological efficacy RCT comparing pre- and post-ASM pH-MII as in outcome measure urgently required. The reason for urgency is two-fold: Firstly, H2 antagonists are commonly prescribed in this population group. Malcolm et al (367) reported a prescription rate of 25% for extreme low birth weight and premature infants. Such prescription practices are not without risk. Secondly, there is increasing evidence of risk of infection associated with gastric acid suppression.

Increased rates of infection particularly necrotising enterocolitis (NEC)(177,368,369) and late-onset septicaemia (370,371) have been identified in preterm neonates. The putative mechanism is the absence of the bactericidal action of gastric acid leading to overgrowth and imbalance of gut flora. Advances in genomics and understanding of the microbiome are improving our understanding of this field. Gupta et al(372) found that ranitidine exposure lowered faecal microbiota variety in 76 premature infants studied. The balance was shifted towards proteobacteria overgrowth. These authors suggest that this imbalance may predispose towards NEC. The slow motility of the preterm infant GI tract may also be contributory(373).

In this oeuvre, we have studiously avoided the term ‘anti-reflux medication’ and have used acid suppression instead. This has been done to place emphasis on delineation of acid control versus reflux control. Both acid suppression and reflux control reduce oesophageal acid exposure may improve symptoms of reflux. However, conflation of the two concepts leads to acceptance of ‘reflux symptoms’ as evidence of reflux events. As we now have objective tests that can demonstrate reflux-symptom association, effects of oesophageal acid exposure versus effects of bolus reflux can be better understood. We predicted that, as the role of non-acid reflux and the risk of acid suppression in children is better understood, the term ‘anti-reflux medication’ will fall out of use.

Move away from complete to partial fundoplication

NI may have played a role in the longevity of this operation. In adult literature, there is a trend away from complete to partial fundoplication due to dysphagia and gas bloat complications. Adults and (in my experience) neurologically normal children are able to verbalise these symptoms. Bitter complaint will certainly influence the surgeon's choice of procedure. However, paediatric surgeons performing this procedure on NI children lack the insight provided by such direct feedback. In my personal experience, observing the misery of dysphagia and gas bloat in a neurologically normal 15-year-old was an education in how debilitating post-fundoplication syndrome can be.

Move away from fundoplication

Our database review identified trend i.e. yearly reductions in fundoplication procedures performed. Others have also observed this trend. Goldin et al(374) queried the PHIS and collected data from 36

children's hospitals captured between 2001-2006. They found that the number of fundoplication has performed decreased year on year. These data have been updated more recently. Brun et al(375), 2016, queried the PHIS between 2004-2013. There was a general decrease in frequency in index cases e.g. fundoplication, nephrectomy, etc. Fundoplication fell from 1848 cases in 2004 to 842 cases in 2013. Authors suggest that data availability on failure rates has resulted in a preference for medical management and fewer referrals for surgical treatment.

This impression has been mirrored by Gundeti(376), who states that:

« … the pendulum has swung the other way, and surgeons are performing fewer fundoplications in recent years than previously. «

In s systematic review, Fortunato and Cuffari(377) wrote:

"… there seems to be agreement that concomitant anti-reflux surgery should not automatically be performed … and that routine investigation for GER in asymptomatic children should be avoided."

Hament et al(58) go further and recommend investigation of GORD only after gastrostomy should symptoms emerge.

Based on anecdotal clinical experience, this swing away from fundoplication has been driven by poor outcomes. Whilst some published series report complications of less than 1%(190), this is not the experience of surgeons I have worked with nor patients I have seen. Furthermore, there is a lack of long-term data in the literature(308). Long after authors publish short term results, the long-term outcomes are apparent in clinic, where patients return with recurrent or intractable symptoms. A notable exception is the study comparing laparoscopic Nissen versus laparoscopic Thal by Kubiak et al(378). Early outcomes (6 weeks post procedure) suggested a higher operative complication and dysphagia rate with laparoscopic Nissen fundoplication. However, 10-year follow-up of the same cohort demonstrated higher recurrence and redo rates with laparoscopic Thal fundoplication(182).

Another reason for the trend away from fundoplication is the demonstrated efficacy of gastrojejunal feeding in NI patients with reflux.

Gastrojejunal tube placement as an alternative to fundoplication

Increasingly, patients with GORD are being managed with gastrojejunal tube (GJT) placement. The rationale is simple: distal feeding reduces bolus volume available for gastroesophageal reflux. Compared to fundoplication, the GJT is a far less invasive intervention. GJT placement is achieved by advancing a jejunal tube through a new or pre-existing GT site. Advancement can be done either under endoscopic or fluoroscopic image guidance.

Several studies have compared GJT as an alternative to fundoplication. In a retrospective review by Wales et al(217) , no differences were identified in recurrent reflux symptoms or failure to thrive when GJT versus fundoplication patients were compared. Children who had GJT were more likely to continue using ASM. Crucially, 15% of patients with GJT patients had symptom improvement and had their GJT removed. This aspect of reversibility is certainly encouraging, especially in children in whom GORD symptom abatement with age is considered. GJT are not without complications. Authors identified minor complications of tube blockage and displacement. Patients re-attended 1-2 times per year for

replacement of the GJ under general anaesthetic. Major complications of obstruction and intussusception were also observed with GJT.

Srivastava et al(218) found no differences in aspiration pneumonia or mortality in children who had either first GJT or first fundoplication. A systematic review in 2015(379) identified a trend towards more frequent but minor complications with GJT compared to fewer but major complications with fundoplication. We agree with the authors that a large scale, prospective and comparative studies are required. Given the current equipoise, this question is ideal for a randomised control trial. Inclusion should be children with NI, pre-existing gastrostomy, on-going symptoms of reflux and positive SAP demonstrated on MII.

Blended feed per gastrostomy

This is a mode of feeding where enteral nutrition (EN) formulas are substituted or replaced with pureed whole foods. This approach has been driven by patients and P/CG with gastrostomy feeds and is beginning to gain traction with clinicians and nutritionists. Putative benefits of pureed whole foods include high fibre content and plant-based nutrients (phytonutrients) that are unavailable in dairy based enteral nutrition formulations (380) .

As this approach is consumer led, a lot of evidence is anecdotal. However, there are some formal studies. Pentiuk et al(381) at Cincinnati Children's hospital selected 33 children for a trial of pureed by gastrostomy diet. All children had symptoms of retching and gagging post fundoplication. A customised diet to meet nutrition goals was designed for each child. These researchers found that P/CG reported a >50% reduction in retching and gagging for 73% of children, all the while meeting weight goals. Interestingly, 57% of this small cohort had increased oral intake at the end of the trial. This hints at peripheral benefits in addressing oral aversion and improving appetite. The pleasure that comes from sharing a meal with family cannot be quantified. Adult of pureed diets report a slight taste sensation on burping(382).

Nutritionists strike a word of caution, however. It is much less convenient to calculate enteral intake and balance this against requirements on a pureed diet. However, most children healthy children diets governed by guidelines rather than strict measurements of protein and calories. Unlike health children, gastrostomy fed children will have other disabilities that make it more difficult to maintain homeostasis e.g. inability to express satiety, inability to exercise. Therefore, a switch to pureed diet should be undertaken with clinical supervision and support.

Furthermore, pureed foods are not sterilized. It is recommended that they should be avoided in immune-compromised patients and administered within 2 hours of preparation(59). Clinicians worry about tube clogging. In children, flange retained gastrostomy tubes are changed under a general anaesthetic. Therefore, a blocked tube introduces a degree of morbidity that may be deemed unacceptable.

We are hopeful about this approach for long term gastrostomy feeding. Early evidence is promising specifically in children with feeding difficulties retching. We have discussed earlier how retching has been found to predict failure of primary(125) and revision fundoplication(284). The ‘soft’ benefits i.e. eating with the rest of the family, taste, variety, are immeasurable and yet attractive. However, for this approach to bear fruit, parents and caregivers need to be supported by clinicians and nutritionists.

Database studies and data mining

At the start of this project, the status of the database study in research was nascent.

In this field, we have seen an increasing interest in published data arising from database reviews. Notably, Barnhart(383) , Brun(375) and McAteer et al(298) published findings based on review of data amalgamated from the Patient Health Information System (PHIS). Certainly, multicentre amalgamations of data suffer loss of granularity from variations in centre, regional and surgeon practice. Administrative databases contain some clinical data e.g. demographics, diagnostic comorbidity codes, billing for investigations. However, administrative databases are often lacking in key clinical information e.g. results of investigations, indications for surgery. What is lost in detail may be gained in breadth. For example, the national trend towards fewer fundoplications identified by Brun(375) and Gundeti et al(376) may be missed in data from a single centre maintaining status quo.

We predict that future applications for research approval and clinical trial registration will require a database review of existing clinical data to establish a snapshot of current practice. Availability of data mining tools will also make collaborative, multicentre and population-based studies more feasible and attractive to researchers.

This thesis resulted in the TARDIS:REFLUX app, a proof-of-concept project that demonstrates how NHS data protection regulations can be navigated. At the time, there were few apps available for clinical use. Since then there has been an explosion in number and variety of clinical apps available.

Despite this explosion in app-availability, there remain a lack of systematic study into clinical effectiveness of app use. Components of mobile health have been reviewed. For example, there are several Cochrane reviews addressing use of mobile messaging(384,385). In a study on communication of medical results, authors could only identify one RCT. In this study there was indeed a reduction in anxiety scores where women receive early test result scores(384). In a similar review into phone-based interventions on adherence to contraceptive use, authors identified 5 RCTs but no clear benefit in contraceptive adherence for mobile messaging over standard communication methods e.g. face-to-face, mail(385). Perhaps unsurprisingly, the smartphone as a medium is simply as competitive as other prior mediums. This may reflect the fact that it is the message rather than the medium that has any health intervention benefit. For example, Semper et al(386) found that apps were as efficacious as other self-monitoring methods for weight management in obese adults.

There are good examples of efficacious practice. For example, Phillips et al(387) used a smartphone platform to deliver HIV risk reduction videos. The format proved feasible and acceptable. Delivery of sensitive patient education content to personal devices appeared to enhance acceptability.

Some smartphones models have inbuilt hardware that has found applications in research. Manufacturers have made hardware available for app developers to incorporate into apps. For example, Jenny(388) used the smartphone accelerometer to measure joint angles. Precision and accuracy were ‘good’ compared to a gold standard navigation system used for peri-operative patients undergoing total knee arthroplasty.

Some smartphones contain accelerometers that measure distance travelled over time. These have been entrained into apps to measure activity i.e. actigraphy. Examples include pedometers for ambulation(389,390) and sleep actigraphy (391) that detect eye movement. Some smartphones contain gyroscopes which detect movement on both horizontal and vertical axes can be detected. A pedometer utilising a gyroscope is able to track walking upstairs(392). Activity profiles can be generated based on pattern recognition algorithms e.g. walking, versus running versus jumping.

Although tools are available, validation of methodology is often lacking. In a review of sleep apnoea related apps, Behar et al(391) noted a proliferation in sleep screening apps. However, few apps utilised validated measurement tools. Variability in results depending on the phone used, the type of patient, the use of an in-phone or remote sensor and the user’s environment was noted.

Smartphone applications can be designed to be context aware. Apps with Global Positioning System (GPS) enabled allow localisation of the user and the users activity. This reflects a trend in ‘ecological’ data i.e. where a subject is understood in the context of their environment. This could lead to exciting developments in the study of public health. Information about disease incidence or prevalence can be enriched by understanding subject location. Behavioural interventions could be targeted based on

ecological data. For example, a user could receive suggestions of restaurants with healthy eating options, or suggestions for exercise locations based on their GPS positioning.

Several researchers have promoted the body area network as an infrastructural technology central to the development of telehealth(393). A body area network is generated by wearable sensors that capture health-related information e.g. g blood pressure and blood glucose. The smartphone can be used as the data storage and processing hub for these sensors. Smartphones provide both the portability and computing power and mobility required to support a body area network. This is predicted to be a growing field in mobile health.

The effect of sharing health-related data on social networks has been studied. In theory, peer networks can be used to provide motivation and reinforce positive health messages. For example, researchers found that patients with depression participating heavily in an online social network had greater improvement on depression scales compared to occasional users(394). In practice, results are difficult to analyse due to quality of studies and lack of standardised methodology. Eysenbach et al(395) reported a systematic review on the effect of virtual networks on health behaviour. The found the effect of social networks difficult to evaluate because of poor study design, comparative studies without a randomised element and the relegation of the social network intervention to secondary outcome status.

Uptake amongst medical professionals may be a limiting factor. In a survey of 182 colorectal surgeons working in the UK, Ireland and Europe, Smart(396) identified a high proportion of smartphone ownership (83.5%). The age-distribution of smartphone ownership followed the general population trend. Respondents in the 21-50-year age group were significantly more likely to own a smartphone than respondents in the 51-70-year age group (88.8%vs 72.7% respectively P = 0.02).

A primary concern for health professionals is data protection rules. A key contribution of this work has been demonstrating feasibility. This proof-of-concept work also demonstrated how NHS data protection regulations can be navigated to enable full and safe engagement of patients, data and health professionals in cyberspace. When we began this project, the status quo for data transmission was unwieldy and not suited to our purposes.

All NHS organisations are required to observe ‘safe haven’ procedures in storage and sharing of data. The two key pieces of legislation and guidance are the Data Protection Act 1998(397) and the NHS Code of Practice: Confidentiality Annex A1-Protect Patient Information(398). A safe haven is a location within an organisation where arrangements have been made to ensure that person-identifiable data can be held, received and communicated. For example, if person-identifiable information is collected by a HCW on a paper form, the form can only be shared to another safe haven. The data can only be shared with other HCW involved in the patient. The named Data Protection Officer (DPO) at the institution is responsible for overseeing the security and limiting arrangements for sharing the data. The transmission of data beyond the secure arrangements within the institution is prohibited. Where exceptions are made, permission must be sought from the DPO and consent must be sought from the patient/ parent. Therefore, should a parent complete a symptom questionnaire for the gastroenterologists, the parent must rely on the gastroenterologists to pass on this questionnaire to the surgeon. Prior permission is required to view the data via email, fax or post. The process and

permissions can be cumbersome. Therefore, data remains in the silo of those who have collected it in the first place.

We overcame this limitation by enabling the P/CG to maintain ownership. Symptom data collected on the P/CG device belongs to them and they can share it with whomever they please. There remains an onus on the researcher to store data securely once received. However, responsibility and choice for sharing data remained with the P/CG. This change of paradigm simplifies the process of sharing data amongst HCW and bypasses cumbersome safe haven regulations. It is an illustration of a patient-centred innovation that has made it simpler for P/CG to report symptoms to HCW.

Currently, HCW can view data on the P/CG device or on the TARDIS:REFLUX database server. In future iterations of the TARDIS:REFLUX app, we would like to build functions to email symptom reports directly to HCW. Given financial means, online access to the researcher's database online for both HCW and P/CG would be feasible goal. These measures will obviate the need for the HCW to be in the same physical space as the patient in order to view their symptom data. For P/CG who experiences a high burden of visits to hospital appointments, this will be a welcome development.

Data protection in healthcare has been regulated by the EU Data Protection Directive 95/46/EC since 1995. An amendment to the directive, the General Data Protection Regulation was adopted in April 2016. This directive was to supersede the Data Protection Directive from 25 May 2018. However, in July 2016, a referendum in the UK signalled UK exit from the European Union. The future of this regulatory landscape is now uncertain. It will be interesting to note what direction protection of health data takes. Will individual health authorities e.g. NHS England, NHS Wales take individual responsibility and create different regulatory frameworks? Will pre-existing EU Law already adopted and adapted for parliament remain ratified as national law?

We expect that in the future, smartphone apps will be an integral part of the healthcare landscape, delivering healthcare and research tools. As researchers are increasingly trained in digital skills, the default methods for data collection and analysis will evolve as a second-hand consequence. Equally, patient expectations are taken from the real world, rather than the healthcare sector. We predict that, to ensure patient participation, health researchers gravitate towards digital tools e.g. online surveys, research trial apps etc.

CHAPTER 5: CONCLUSION

IMPACT: A NEW ALGORITHM FOR SURGICAL RISK STRATIFICATION

When I began this task, I envisioned the research output at the end would be an elegant formula enumerating risk according to comorbidities and investigation findings. The Nottingham prognostic index (NPI) for breast cancer was an inspiration.

My envisioned formula would enumerate the probability of successful fundoplication based on scoring demographic, comorbid and operative factors. However, since beginning the work, I have realized that much preliminary work on the disease paradigm is required before we a prognostic model is possible. The surgeon's dilemma is a problem whose nidus is symptomatology.

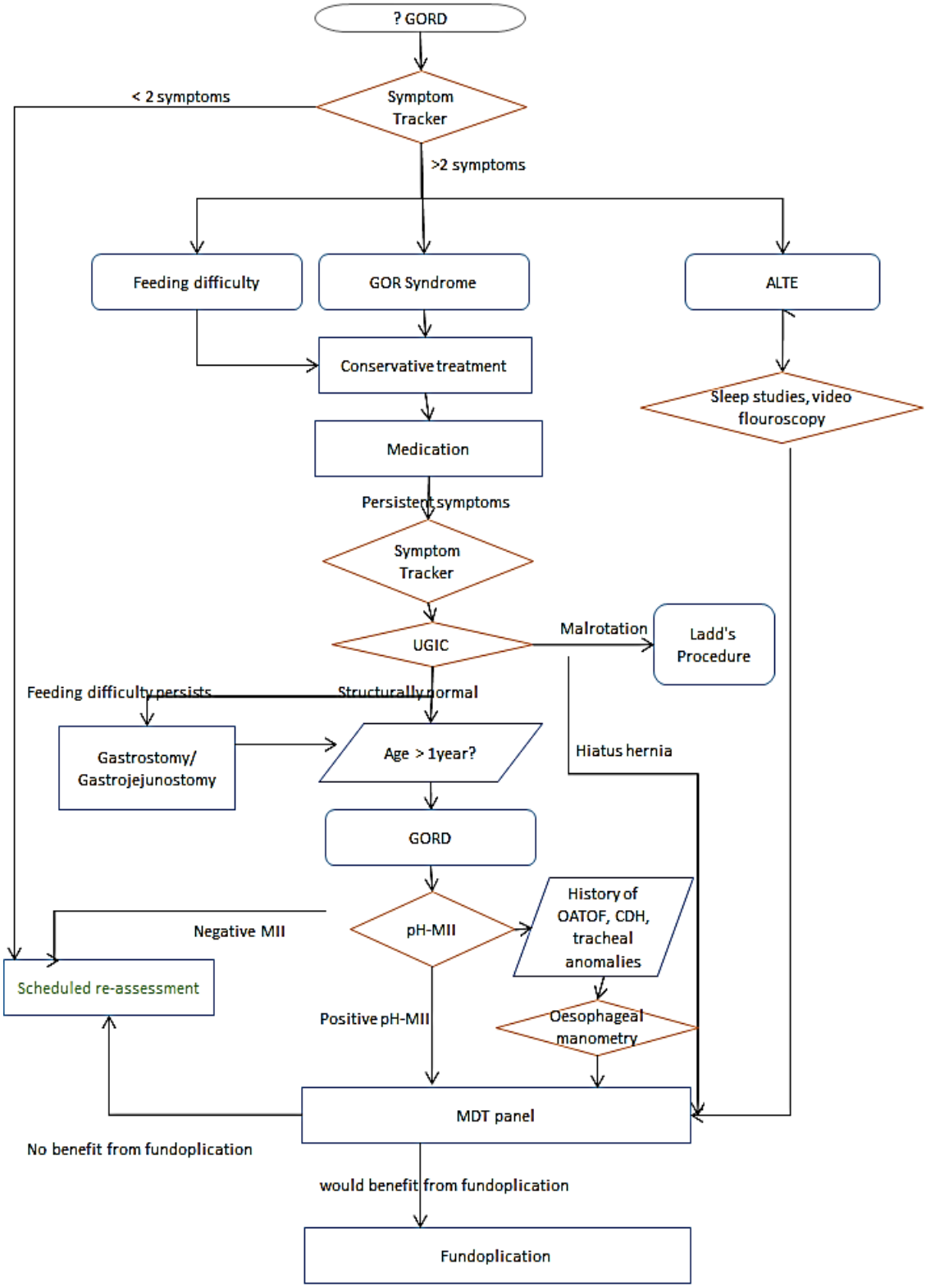
Therefore, in consultation with my supervisors, I adjusted the scope and ambition of the project. This work has focused on the basic building blocks of diagnosis i.e. correct identification of symptoms, understanding the impact of demography and comorbidities and appropriate choice of investigations.

The Paediatric GORD Surgical Risk Stratification Algorithm (PG-SRSA) proposed below is the result of synthesis of the multiple projects in this work. The aim of the algorithm is dynamic process of risk stratification. As patient's progress along a pathway, they are filtered to investigations and treatments suited to their cluster of comorbidities, symptoms and signs. The algorithm repeatedly utilizes the symptom tracking method developed in this work. The understanding of diagnostics developed in running the REMOS trial informs the judicious use of clinical investigations. Lastly, what we have learnt about comorbidities informs both the investigation process and the MDT panel recommendation.

I believe that this algorithm will be of greater value for the working surgeon compared to any prognostic model we could have devised. It presents a pragmatic way of mitigating patient risks. For the surgeon surveying her waiting room full of children with GORD and a mixture of comorbidities, the task is simplified. The surgeon simply locates the patient's position in the algorithm. As the patient progresses along the algorithm, evidence-based and defensible decision-making will emerge. If adopted, this standardised approach will also lead to a body of clinical evidence where indications for surgery are similar and patient cohorts are comparable.

In contrast to the ESPGHAN(136)guideline (See Section I Appendix items, p477) that was extant when this work was done, the PG-SRSA addresses age at intervention and investigation. Emphasis is also placed on prospective symptom tracking. Unlike the ESPGHAN guideline, routine endoscopy and biopsy is not included in this pathway. This is because histological oesophagitis rates are low(37) . Correlation between endoscopic and histological pathology is poor(163) . Crucially, in children surgery is indicated for symptom control as the cumulative effects of acid exposure i.e. Barrett's oesophagus are rarely observed(136). Furthermore, patients are often already filtered to surgical referral through gastroenterology clinics. Therefore, this investigation, where indicated may already be done.

Figure 142: Paediatric GORD Surgical Risk Stratification Algorithm



The ESPGHAN guideline includes endoscopy for the purposes of diagnosing eosinophilic oesophagitis(163). This is a controversial diagnosis amongst paediatric gastroenterologists. Whilst true rates of eosinophilic oesophagitis are not known, my inclination would be to retain this study for select

patients. Furthermore, in patients who progress to having a gastrostomy placed, the opportunity to take biopsies presents.

In contrast to the ESPHGAN guideline, the UGIC is included in this algorithm. As a reflux investigation, the UGIC is not indicated. However, from a surgical perspective, ruling out hiatus hernia and malrotation in patients prior to consideration for gastrostomy or fundoplication is important. Although both diagnoses are rare (~1%)(156) finding these conditions on during the operation, or worse still, not recognising these conditions would be detrimental to the patient. For example, a fundoplication in a patient with intermittent volvulus from malrotation could lead to closed-loop obstruction.

The ESPGHAN guideline has recently been updated(136). Recommendations are broadly similar to those made in 2009. However, anachronistic features of the new recommendations are inclusion of 'Barium' contrast and pH-metry as potential investigations.

FUTURE DIRECTIONS

REMOS Trial

There remains a need for an RCT comparing ARS with acid suppression medication (ASM) for control of reflux symptoms in children. Much preliminary work is required before this can be achieved. Literature review revealed that a myriad of symptoms is associated with GORD. The Montreal / NASPGHAN list of symptoms is expansive. Using pH-MII and symptom trackers like the TARDIS:REFLUX app, symptom-reflux association can be established for large cohorts of children with symptoms. This evidence-based list can then be standerdised and adhered to as a list of indications for intervention. Furthermore, the role of acid suppression in paediatric GORD requires interrogation. As noted in the Cochrane meta-analysis by Tighe et al(124), evidence for pharmacotherapy is poor due to heterogeneity in both symptom description and outcome measurement.

Were we to redesign the REMOS trial, key recommendations would be:

1. True equipoise i.e. ASM vs ARS in patients with pre-existing GT in situ. Cross –over design would be used to address calibration bias i.e. patients who have persistent symptoms on ASM cross over to ARS, and vice-versa.
2. Standardized pre- and post- intervention diagnostics for GOR (pH-MII), aspiration or respiratory compromise. Robust symptoms tracking pre-and post-operatively e.g. TARDIS:REFLUX symptom tracker app
3. Randomization with minimization for demographics as well as key comorbidities
4. Standardised, objective and validated indications for surgery
5. Strict adherence to trial protocols with intention-to-treat analysis

TARDIS:REFLUX

There has been discussion whether future versions of the app should allow free-text entry of symptoms. However, the author feels that this will introduce subjective bias. One parent's description of pain will be another parent's discomfort. A better solution is to broaden the list to include more peripheral symptoms identified in the symptom's nebula.

Other improvements can be made. Introducing an immediate feedback feature for parents would be useful. This could be in the form of notification is sent to the app administrators. Augmenting the TARDIS:REFLUX app to utilise smartphone hardware would be desirable. GPS could be enabled to understand ecology of symptoms. Patterns may emerge e.g. retching symptoms at school rather than at home. Image and video capture could be incorporated to enrich the documentation of a child's symptoms. This would be particularly useful for neurological symptoms e.g. dystonia versus seizure.

Another planned improvement is enabling users to send compare data with other users. We plan to host a website where patients can compare their data anonymously. We would also like to enable users to securely send data to HCW beyond our institution involved in their care.

We believe that engaging a wide range of HCW will be key to dissemination and adoption of the PG-RSA.

SUMMARY STATEMENT

I began my work at the UCL Institute of Child Health as a Clinical Research Associate for the REMOS trial. I applied to improve my prospects of getting into surgical training. Once my training contract was secure, I chose to continue with my PhD as my interest was piqued. This body of work presented here has been a powerful vehicle for my development as an academic and surgeon.

In my role as trial coordinator, I learnt the mechanics of clinical trial research. Growing in confidence, I instigated the TARDIS:REFLUX pilot study and the REFLUX UX workshops. Both these endeavours gave me practical experience of ethics and governance processes. The supportive environment at UCL enabled me to a novel research and clinical tool using up to date digital research methods.

Perhaps the most useful experience has been the data mining study. Learning the computational and statistical methods necessary has been the steepest learning curve of my adult life. However, this journey been rewarded in robust, transferrable skills in statistical methods and data analysis that continue to prove useful.

In parallel, I have been developing as a surgeon. This has provided important insights into the children discussed herein. It has also deepened my sense of urgency- that we surgeons must continually interrogate the evidence that underpins our operations.

Most importantly, the work presented has changed my aspect. The final chapter of the PhD journey coincides with my penultimate year of training as a paediatric surgeon. My academic and surgical pursuits have interacted, with one perspective enriching the other. The end of this journey finds me deeply motivated to begin another - as an academic surgeon.

1. Fox D1, Barnard J, Campagna EJ, Dickinson LM, Bruny J KA. Fundoplication and the pediatric surgeon: implications for shared decision-making and the medical home. - PubMed - NCBI. Acad Pediatr [Internet]. 2012;12(6):558–66. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22981670

2. Pearl RH1, Robie DK, Ein SH, Shandling B, Wesson DE, Superina R, Mctaggart K, Garcia VF, O'Connor JA FR. Complications of gastroesophageal antireflux surgery in neurologically impaired versus neurologically normal children. - PubMed - NCBI. J Pediatr Surg [Internet]. 1990;25(11):1169–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/?term=pearl+antireflux+surgery+neurologically

3. Gilger MA1, Yeh C, Chiang J, Dietrich C, Brandt ML E-SH. Outcomes of surgical fundoplication in children. - PubMed - NCBI. Clin Gastroenterol Hepatol [Internet]. 2004;2(11):978–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15551250

4. Modlin IM, Eng F, Kidd M, Lye KD. Historical perspectives on the treatment of gastroesophageal ref lux disease. Gastrointest Endosc Clin N Am [Internet]. 2003 [cited 2018 Mar 22];13:19–55. Available from: http://www.giendo.theclinics.com/article/S1052-5157(02)00104-6/pdf

5. Albers. Einnerung an Aertze uber die Behandlung des Akuten Wasserkopfs. J d Chir Augenhelk. 1833;(27):537–40.

6. Fatović-Ferenčić S, Banić M. No Acid, No Ulcer: Dragutin (Carl) Schwarz (1868–1917), the Man Ahead of His Time. Dig Dis [Internet]. 2011 [cited 2018 Mar 22];29(5):507–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22095019

7. Friedenwald J, Feldman M ZW. Peptic Ulcer of the Esophagus. Am J Med Sci. 1929;177(1):1–14.

8. Tilleston. Peptic ulcer of the oesophagus. Am J Med Sci. 1906;132:240.

9. WB C. Conditions affecting the activities of the cardia. In: The Mechanical Factors of Digestion. New York: Longmans, Green, and Co.,; 1911. p. 32–44.

10. DENT J. Review article: from 1906 to 2006 ? a century of major evolution of understanding of gastro-oesophageal reflux disease. Aliment Pharmacol Ther [Internet]. 2006 Nov [cited 2016 Jun 27];24(9):1269–81. Available from: http://doi.wiley.com/10.1111/j.1365-2036.2006.03122.x

11. ALLISON PR. Reflux esophagitis, sliding hiatal hernia, and the anatomy of repair. Surg Gynecol Obstet [Internet]. 1951 Apr [cited 2018 Mar 22];92(4):419–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14835197

12. Palmer ED. The hiatus hernia-esophagitis-esophageal stricture complex: Twenty year prospective study. Am J Med [Internet]. 1968 Apr 1 [cited 2018 Mar 22];44(4):566–79. Available from: https://www.sciencedirect.com/science/article/pii/0002934368900570

13. WINKELSTEIN A, Winkelstein A, Chevalier J., Friedenwald J, Winkelstein A. PEPTIC ESOPHAGITIS. JAMA J Am Med Assoc [Internet]. 1935 Mar 16 [cited 2016 Jun 28];104(11):906. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.1935.02760110034008

14. BERNSTEIN LM, BAKER LA. A clinical test for esophagitis. Gastroenterology [Internet]. 1958 May [cited 2018 May 10];34(5):760–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/13538145

15. Johnson LF, Demeester TR. Twenty-four-hour pH monitoring of the distal esophagus. A quantitative measure of gastroesophageal reflux. Am J Gastroenterol [Internet]. 1974 Oct [cited 2016 Jun 28];62(4):325–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/4432845

16. Hirschowitz BI. Endoscopic examination of the stomach and the duodenal cap with the fiberscope. Lancet. 1961;282(7304):1074–8.

17. Fox D, Morrato E, Campagna EJ, Rees DI, Dickinson LM, Partrick DA, et al. Outcomes of laparoscopic versus open fundoplication in children's hospitals: 2005-2008. Pediatrics [Internet]. 2011 May [cited 2018 Mar 13];127(5):872–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21502226

18. Caro JJ, Salas M, Ward A. Healing and relapse rates in gastroesophageal reflux disease treated with the newer proton-pump inhibitors lansoprazole, rabeprazole, and pantoprazole compared with omeprazole, ranitidine, and placebo: evidence from randomized clinical trials. Clin Ther [Internet]. 2001 Jul [cited 2018 Apr 25];23(7):998–1017. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11519776

19. Van Oijen AH, Verbeek AL, Jansen JB, De Boer WA. Review article: treatment of Helicobacter pylori infection with

ranitidine bismuth citrate- or proton pump inhibitor-based triple therapies. Aliment Pharmacol Ther [Internet]. 2000 Aug [cited 2018 Apr 25];14(8):991–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10930892

20. INGELFINGER FJ, KRAMER P, SANCHEZ GC. The gastroesophageal vestibule, its normal function and its role in cardiospasm and gastroesophageal reflux. Am J Med Sci [Internet]. 1954 Oct [cited 2018 Mar 22];228(4):417–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/13197391

21. Fyke J. FE, Code CF, Schlegel JF. The Gastroesophageal Sphincter in Healthy Human Beings. Digestion. 1956;86(3):135–50.

22. Vakil N, van Zanten S V, Kahrilas P, Dent J, Jones R. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. Am J Gastroenterol [Internet]. 2006 Aug [cited 2015 Feb 5];101(8):1900–20; quiz 1943. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16928254

23. Dent J, Holloway RH, Toouli J, Dodds WJ. Mechanisms of lower oesophageal sphincter incompetence in patients with symptomatic gastrooesophageal reflux. Gut [Internet]. 1988 Aug [cited 2018 Mar 22];29(8):1020–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3410327

24. Mittal RK, Balaban DH. The Esophagogastric Junction. Epstein FH, editor. N Engl J Med [Internet]. 1997 Mar 27 [cited 2018 Apr 9];336(13):924–32. Available from: http://www.nejm.org/doi/10.1056/NEJM199703273361306

25. Silny J. Intraluminal Multiple Electric Impedance Procedure for Measurement of Gastrointestinal Motility. Neurogastroenterol Motil [Internet]. 2008 Jun 28 [cited 2018 Apr 9];3(3):151–62. Available from: http://doi.wiley.com/10.1111/j.1365-2982.1991.tb00061.x

26. Shay S, Tutuian R, Sifrim D, Vela M, Wise J, Balaji N, et al. Twenty-four hour ambulatory simultaneous impedance and pH monitoring: a multicenter report of normal values from 60 healthy volunteers. Am J Gastroenterol [Internet]. 2004 Jun [cited 2016 Jul 30];99(6):1037–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15180722

27. Attwood SEA, DeMeester TR, Bremner CG, Barlow AP, Hinder RA. Alkaline gastroesophageal reflux: implications in the development of complications in Barrett's columnar-lined lower oesophagus. Surgery 1989;106:764-770. Surgery. 1989;106:764–70.

28. Beck IT, Champion MC, Lemire S, Thomson AB, Anvari M, Armstrong D, et al. The Second Canadian Consensus Conference on the Management of Patients with Gastroesophageal Reflux Disease. Can J Gastroenterol [Internet]. 1997 Sep [cited 2018 Apr 9];11 Suppl B:7B–20B. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9347173

29. Noh YW, Jung H-K, Kim S-E, Jung S-A. Overlap of Erosive and Non-erosive Reflux Diseases With Functional Gastrointestinal Disorders According to Rome III Criteria. J Neurogastroenterol Motil [Internet]. 2010 Apr 30 [cited 2018 Apr 9];16(2):148–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20535345

30. Johanson JF. Epidemiology of esophageal and supraesophageal reflux injuries. Am J Med [Internet]. 2000 Mar 6 [cited 2018 Apr 9];108 Suppl 4a:99S–103S. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10718460

31. Diamant NE. Pathophysiology of gastroesophageal reflux disease. GI Motil online, Publ online 16 May 2006; | doi101038/gimo21 [Internet]. 2006 May 16 [cited 2018 Mar 22]; Available from: https://www.nature.com/gimo/contents/pt1/full/gimo21.html

32. Mittal RK, Goyal RK. Sphincter mechanisms at the lower end of the esophagus. GI Motil online, Publ online 16 May 2006; | doi101038/gimo14 [Internet]. 2006 May 16 [cited 2018 Apr 9]; Available from: https://www.nature.com/gimo/contents/pt1/full/gimo14.html

33. Kessing BF, Conchillo JM, Bredenoord AJ, Smout AJPM, Masclee AAM. Review article: the clinical relevance of transient lower oesophageal sphincter relaxations in gastro-oesophageal reflux disease. Aliment Pharmacol Ther [Internet]. 2011 Mar [cited 2018 Apr 23];33(6):650–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21219371

34. Lidums I, Checklin H, Mittal RK, Holloway RH. Effect of atropine on gastro-oesophageal reflux and transient lower oesophageal sphincter relaxations in patients with gastro-oesophageal reflux disease. Gut [Internet]. 1998 Jul [cited 2016 Aug 5];43(1):12–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9771399

35. Sherman PM, Hassall E, Fagundes-Neto U, Gold BD, Kato S, Koletzko S, et al. A Global, Evidence-Based Consensus on the Definition of Gastroesophageal Reflux Disease in the Pediatric Population. Am J Gastroenterol [Internet]. 2009 May 7 [cited 2016 Aug 2];104(5):1278–95. Available from: http://www.nature.com/doifinder/10.1038/ajg.2009.129

36. Mousa H, Machado R, Orsi M, Chao CS, Alhajj T, Alhajj M, et al. Combined multichannel intraluminal impedance-pH (MII-pH): multicenter report of normal values from 117 children. Curr Gastroenterol Rep [Internet]. 2014 Aug [cited 2016 Jul 29];16(8):400. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25064319

37. Gilger MA, El-Serag HB, Gold BD, Dietrich CL, Tsou V, McDuffie A, et al. Prevalence of endoscopic findings of erosive esophagitis in children: a population-based study. J Pediatr Gastroenterol Nutr [Internet]. 2008 Aug [cited 2018 Apr 9];47(2):141–6. Available from: https://insights.ovid.com/crossref?an=00005176-200808000-00005

38. El-Serag HB, Bailey NR, Gilger M, Rabeneck L. Endoscopic manifestations of gastroesophageal reflux disease in patients between 18 months and 25 years without neurological deficits. Am J Gastroenterol [Internet]. 2002 Jul [cited 2018 Apr 9];97(7):1635–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12135011

39. Nelson SP, Chen EH, Syniar GM, Christoffel KK. Prevalence of symptoms of gastroesophageal reflux during infancy. A pediatric practice-based survey. Pediatric Practice Research Group. Arch Pediatr Adolesc Med [Internet]. 1997 Jun [cited 2015 Sep 24];151(6):569–72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9193240

40. Hegar B, Dewanti NR, Kadim M, Alatas S, Firmansyah A, Vandenplas Y. Natural evolution of regurgitation in healthy infants. Acta Paediatr [Internet]. 2009 Jul [cited 2016 Aug 1];98(7):1189–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19397533

41. Van Howe RS, Storms MR. Gastroesophageal reflux symptoms in infants in a rural population: longitudinal data over the first six months. BMC Pediatr [Internet]. 2010 [cited 2016 Aug 9];10:7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20149255

42. Kleinman L1, Rothman M, Strauss R, Orenstein SR, Nelson S, Vandenplas Y, Cucchiara S RD. The infant gastroesophageal reflux questionnaire revised: development and validation as an evaluative instrument. - PubMed - NCBI. Clin Gastroenterol Hepatol [Internet]. 2006;4(5):588–96. Available from: http://www.ncbi.nlm.nih.gov/pubmed/?term=kleinman+infant+gastroesophageal+reflux+questionnaire

43. Nelson SP, Chen EH, Syniar GM, Christoffel KK. One-year follow-up of symptoms of gastroesophageal reflux during infancy. Pediatric Practice Research Group. Pediatrics [Internet]. 1998 Dec [cited 2018 Apr 23];102(6):E67. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9832595

44. Dent J, El-Serag HB, Wallander M-A, Johansson S. Epidemiology of gastro-oesophageal reflux disease: a systematic review. Gut [Internet]. 2005 May 1 [cited 2018 Mar 23];54(5):710–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15831922

45. Shabib SM, Cutz E, Sherman PM. Passive smoking is a risk factor for esophagitis in children. J Pediatr [Internet]. 1995 Sep [cited 2018 Apr 23];127(3):435–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7658277

46. Koebnick C, Smith N, Coleman KJ, Getahun D, Reynolds K, Quinn VP, et al. Prevalence of Extreme Obesity in a Multiethnic Cohort of Children and Adolescents. J Pediatr [Internet]. 2010 Jul [cited 2018 Apr 23];157(1):26–31.e2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20303506

47. Koebnick C, Getahun D, Smith N, Porter AH, Der-Sarkissian JK, Jacobsen SJ. Extreme childhood obesity is associated with increased risk for gastroesophageal reflux disease in a large population-based study. Int J Pediatr Obes [Internet]. 2011 Jun [cited 2018 Apr 23];6(2–2):e257–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20615162

48. Sullivan PB. Gastrointestinal disorders in children with neurodevelopmental disabilities. Dev Disabil Res Rev [Internet]. 2008 [cited 2018 Apr 23];14(2):128–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18646021

49. Sullivan PB. Gastrostomy and the disabled child. Dev Med Child Neurol [Internet]. 1992 Jun [cited 2018 Apr 23];34(6):552–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1535323

50. Sullivan PB. Gastrointestinal disorders in children with neurodevelopmental disabilities. Dev Disabil Res Rev [Internet]. 2008 [cited 2018 May 2];14(2):128–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18646021

51. Tin W, Wariyar U, Hey E. Changing prognosis for babies of less than 28 weeks' gestation in the north of England between 1983 and 1994. Northern Neonatal Network. BMJ [Internet]. 1997 Jan 11 [cited 2018 Apr 23];314(7074):107–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9006468

52. Vernon-Roberts A, Sullivan PB. Fundoplication versus postoperative medication for gastro-oesophageal reflux in children with neurological impairment undergoing gastrostomy. Cochrane database Syst Rev [Internet]. 2013 Jan [cited 2015

Sep 5];8:CD006151. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23986351

53. Gangil A, Patwari AK, Aneja S, Ahuja B, Anand VK. Feeding problems in children with cerebral palsy. Indian Pediatr [Internet]. 2001 Aug [cited 2016 Aug 11];38(8):839–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11520994

54. Reyes AL, Cash AJ, Green SH, Booth IW. Gastrooesophageal reflux in children with cerebral palsy. Child Care Health Dev [Internet]. [cited 2016 Aug 11];19(2):109–18. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8462129

55. Sullivan PB. Gastrointestinal problems in the neurologically impaired child. Baillieres Clin Gastroenterol [Internet]. 1997 Sep [cited 2018 May 2];11(3):529–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9448914

56. Heine RG, Reddihough DS C-SA. Gastro-oesophageal reflux and feeding problems after gastrostomy in children with severe neurological impairment. Dev Med Child Neurol. 1995;34:34:320–329.

57. Razeghi S, Lang T, Behrens R. Influence of Percutaneous Endoscopic Gastrostomy on Gastroesophageal Reflux: A Prospective Study in 68 Children. J Pediatr Gastroenterol Nutr [Internet]. 2002 Jul [cited 2016 Aug 2];35(1):27–30. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00005176-200207000-00007

58. Hament JM, Bax NMA, van der Zee DC, De Schryver JEAR, Nesselaar C. Complications of percutaneous endoscopic gastrostomy with or without concomitant antireflux surgery in 96 children. J Pediatr Surg [Internet]. 2001 Sep [cited 2016 Aug 2];36(9):1412–5. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022346801392916

59. Edwards S, Davis AM, Bruce A, Mousa H, Lyman B, Cocjin J, et al. Caring for Tube-Fed Children: A Review of Management, Tube Weaning, and Emotional Considerations. JPEN J Parenter Enteral Nutr [Internet]. 2016 Jul [cited 2016 Aug 11];40(5):616–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25791833

60. Knatten CK, Kvello M, Fyhn TJ, Edwin B, Schistad O, Aabakken L, et al. Nissen fundoplication in children with and without neurological impairment: A prospective cohort study. J Pediatr Surg [Internet]. 2015 Dec [cited 2016 Feb 1]; Available from: http://www.sciencedirect.com/science/article/pii/S0022346815008271

61. Sulaeman E, Udall JN, Brown RF, Mannick EE, Loe WA, Hill CB, et al. Gastroesophageal Reflux and Nissen Fundoplication Following Percutaneous Endoscopic Gastrostomy in Children. J Pediatr Gastroenterol &amp Nutr [Internet]. 1998 Mar [cited 2016 Aug 2];26(3):269–73. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00005176-199803000-00006

62. Gold BD, Freston JW. Gastroesophageal reflux in children: pathogenesis, prevalence, diagnosis, and role of proton pump inhibitors in treatment. Paediatr Drugs [Internet]. 2002 [cited 2018 May 24];4(10):673–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12269842

63. Fonkalsrud EW, Ashcraft KW, Coran AG, Ellis DG, Grosfeld JL, Tunell WP, et al. Surgical treatment of gastroesophageal reflux in children: a combined hospital study of 7467 patients. Pediatrics [Internet]. 1998 Mar [cited 2016 Aug 2];101(3 Pt 1):419–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9481007

64. Martinez DA, Ginn-Pease ME, Caniano DA. Sequelae of antireflux surgery in profoundly disabled children. J Pediatr Surg [Internet]. 1992 Feb 1 [cited 2018 May 2];27(2):267-71; discussion 271-3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1532982

65. Lopez-Fernandez S, Hernandez F, Hernandez-Martin S, Dominguez E, Ortiz R, De La Torre C, et al. Failed Nissen fundoplication in children: causes and management. Eur J Pediatr Surg [Internet]. 2014 Feb [cited 2015 Sep 5];24(1):79–82. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23982817

66. Rice H, Seashore JH, Touloukian RJ. Evaluation of Nissen fundoplication in neurologically impaired children. J Pediatr Surg [Internet]. 1991 Jun [cited 2018 May 2];26(6):697–701. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1941460

67. Pimpalwar A, Najmaldin A. Results of laparoscopic antireflux procedures in neurologically impaired children. Semin Laparosc Surg [Internet]. 2002 Sep [cited 2015 Aug 27];9(3):190–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12407529

68. Capito C, Leclair M-D, Piloquet H, Plattner V, Heloury Y, Podevin G. Long-term outcome of laparoscopic Nissen-Rossetti fundoplication for neurologically impaired and normal children. Surg Endosc [Internet]. 2008 Apr 26 [cited 2018 May 2];22(4):875–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17963001

69. Engelmann C, Gritsa S, Ure BM. Impact of laparoscopic anterior 270° fundoplication on the quality of life and symptoms profile of neurodevelopmentally delayed versus neurologically unimpaired children and their parents. Surg Endosc [Internet].

2010 Jun 24 [cited 2018 May 2];24(6):1287–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20033727

70. Rossi V, Mazzola C, Leonelli L, Gandullia P, Arrigo S, Pedemonte M, et al. Long-term outcome and need of re-operation in gastro-esophageal reflux surgery in children. Pediatr Surg Int [Internet]. 2016 Mar 28 [cited 2018 May 2];32(3):277–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26711122

71. Martin K, Deshaies C, Emil S. Outcomes of pediatric laparoscopic fundoplication: a critical review of the literature. Can J Gastroenterol Hepatol [Internet]. 2014 Feb [cited 2018 Mar 13];28(2):97–102. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24288692

72. Legrand C, Michaud L, Salleron J, Neut D, Sfeir R, Thumerelle C, et al. Long-term outcome of children with oesophageal atresia type III. Arch Dis Child [Internet]. 2012 Sep [cited 2015 Sep 5];97(9):808–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22753768

73. Holschneider P, Dübbers M, Engelskirchen R, Trompelt J, Holschneider AM. Results of the operative treatment of gastroesophageal reflux in childhood with particular focus on patients with esophageal atresia. Eur J Pediatr Surg [Internet]. 2007 Jun [cited 2015 Sep 5];17(3):163–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17638154

74. Hassall E. Childhood Barrett's esophagus under the microscope. Am J Surg [Internet]. 1994 Mar [cited 2018 Apr 26];167(3):287–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8160898

75. Burjonrappa SC, Youssef S, St-Vil D. What is the Incidence of Barrett's and Gastric Metaplasia in Esophageal Atresia/Tracheoesophageal Fistula (EA/TEF) Patients? Eur J Pediatr Surg [Internet]. 2011 Jan 15 [cited 2018 May 18];21(01):25–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20954108

76. Hsieh H, Frenette A, Michaud L, Krishnan U, Dal-Soglio DB, Gottrand F, et al. Intestinal Metaplasia of the Esophagus in Children With Esophageal Atresia. J Pediatr Gastroenterol Nutr [Internet]. 2017 Jul [cited 2018 May 18];65(1):e1–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28248794

77. Koivusalo AI, Pakarinen MP. Outcome of Surgery for Pediatric Gastroesophageal Reflux: Clinical and Endoscopic Follow-up after 300 Fundoplications in 279 Consecutive Patients. Scand J Surg [Internet]. 2018 Mar 7 [cited 2018 May 15];107(1):68–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28385101

78. van Wijk M1, Knüppe F, Omari T, de Jong J BM. Evaluation of gastroesophageal function and mechanisms underlying gastroesophageal reflux in infants and adults born with esophageal atresia. - PubMed - NCBI. J Pediatr Surg [Internet]. 2013;48(12):2496–505. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24314193

79. Lemoine C, Aspirot A, Le Henaff G, Piloquet H, Lévesque D, Faure C. Characterization of esophageal motility following esophageal atresia repair using high-resolution esophageal manometry. J Pediatr Gastroenterol Nutr [Internet]. 2013 Jun [cited 2016 Aug 10];56(6):609–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23343933

80. Rintala RJ. Fundoplication in Patients with Esophageal Atresia: Patient Selection, Indications, and Outcomes. Front Pediatr [Internet]. 2017 [cited 2018 May 18];5:109. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28555181

81. Romeo C1, Bonanno N, Baldari S, Centorrino A, Scalfari G, Antonuccio P, Centonze A GC. Gastric motility disorders in patients operated on for esophageal atresia and tracheoesophageal fistula: long-term evaluation. - PubMed - NCBI. J Pediatr Surg [Internet]. 2000;35(5):740–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10813339

82. Chetcuti P, Phelan PD. Respiratory morbidity after repair of oesophageal atresia and tracheo-oesophageal fistula. Arch Dis Child [Internet]. 1993 Feb [cited 2016 Aug 10];68(2):167–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8481036

83. Peetsold MG, Kneepkens CMF, Heij HA, IJsselstijn H, Tibboel D, Gemke RJBJ. Congenital diaphragmatic hernia: long-term risk of gastroesophageal reflux disease. J Pediatr Gastroenterol Nutr [Internet]. 2010 Oct [cited 2015 Aug 27];51(4):448–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20512059

84. Fraga JC, Jennings RW, Kim PCW. Pediatric tracheomalacia. Semin Pediatr Surg [Internet]. 2016 Jun [cited 2018 May 18];25(3):156–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27301602

85. Valusek PA1, St Peter SD, Tsao K, Spilde TL, Ostlie DJ HG 3rd. The use of fundoplication for prevention of apparent life-threatening events. - PubMed - NCBI. J Pediatr Surg [Internet]. 2007 [cited 2018 May 18];42(6):1022–4. Available from: https://www.ncbi.nlm.nih.gov/m/pubmed/17560213/

86. Koivusalo AI, Pakarinen MP, Rintala RJ. Modern outcomes of oesophageal atresia: Single centre experience over the last twenty years. J Pediatr Surg [Internet]. 2013 Feb [cited 2018 May 18];48(2):297–303. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/23414855

87. Peetsold MG, Heij HA, Kneepkens CMF, Nagelkerke AF, Huisman J, Gemke RJBJ. The long-term follow-up of patients with a congenital diaphragmatic hernia: a broad spectrum of morbidity. Pediatr Surg Int [Internet]. 2009 Jan [cited 2016 Mar 14];25(1):1–17. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18841373

88. Koivusalo AI, Pakarinen MP, Lindahl HG, Rintala RJ. The cumulative incidence of significant gastroesophageal reflux in patients with congenital diaphragmatic hernia-a systematic clinical, pH-metric, and endoscopic follow-up study. J Pediatr Surg [Internet]. 2008 Feb [cited 2015 Jul 16];43(2):279–82. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18280274

89. Verbelen T, Lerut T, Coosemans W, De Leyn P, Nafteux P, Van Raemdonck D, et al. Antireflux surgery after congenital diaphragmatic hernia repair: a plea for a tailored approach. Eur J Cardiothorac Surg [Internet]. 2013 Aug 1 [cited 2016 Mar 14];44(2):263–7; discussion 268. Available from: http://ejcts.oxfordjournals.org.libproxy.ucl.ac.uk/content/44/2/263

90. Terui K, Taguchi T, Goishi K, Hayakawa M, Tazuke Y, Yokoi A, et al. Prognostic factors of gastroesophageal reflux disease in congenital diaphragmatic hernia: a multicenter study. Pediatr Surg Int [Internet]. 2014 Nov [cited 2015 Aug 27];30(11):1129–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25217138

91. Dariel A, Rozé J-C, Piloquet H, Podevin G. Impact of prophylactic fundoplication on survival without growth disorder in left congenital diaphragmatic hernia requiring a patch repair. J Pediatr [Internet]. 2010 Oct [cited 2015 Aug 27];157(4):688–90, 690.e1. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20633898

92. Diaz DM, Gibbons TE, Heiss K, Wulkan ML, Ricketts RR, Gold BD. Antireflux Surgery Outcomes in Pediatric Gastroesophageal Reflux Disease. Am J Gastroenterol [Internet]. 2005 Aug 1 [cited 2018 May 15];100(8):1844–52. Available from: http://www.nature.com/doifinder/10.1111/j.1572-0241.2005.41763.x

93. Hillemeier AC, Grill BB, McCallum R, Gryboski J. Esophageal and gastric motor abnormalities in gastroesophageal reflux during infancy. Gastroenterology [Internet]. 1983 Apr 1 [cited 2018 May 2];84(4):741–6. Available from: https://www.gastrojournal.org/article/0016-5085(83)90140-3/fulltext

94. Omari TI, Barnett CP, Benninga MA, Lontis R, Goodchild L, Haslam RR, et al. Mechanisms of gastro-oesophageal reflux in preterm and term infants with reflux disease. Gut [Internet]. 2002 Oct [cited 2018 Mar 29];51(4):475–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12235066

95. Macharia EW, Eaton S, de Coppi P, Curry J, Drake D, Cross K, et al. Fundoplication in ventilator-dependent infants with gastro-oesophageal reflux. Eur J Pediatr Surg [Internet]. 2012 Feb [cited 2015 Sep 5];22(1):91–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22434231

96. López-Alonso M, Moya MJ, Cabo JA, Ribas J, del Carmen Macías M, Silny J, et al. Twenty-four-hour esophageal impedance-pH monitoring in healthy preterm neonates: rate and characteristics of acid, weakly acidic, and weakly alkaline gastroesophageal reflux. Pediatrics [Internet]. 2006 Aug [cited 2016 Jul 30];118(2):e299-308. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16831894

97. Wenzl TG, Schenke S, Peschgens T, Silny J, Heimann G, Skopnik H. Association of apnea and nonacid gastroesophageal reflux in infants: Investigations with the intraluminal impedance technique. Pediatr Pulmonol [Internet]. 2001 Feb [cited 2016 Jul 31];31(2):144–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11180691

98. Mercado-Deane M-GG, Burton EM, Harlow SA, Glover AS, Deane DA, Guill MF, et al. Swallowing dysfunction in infants less than 1 year of age. Pediatr Radiol [Internet]. 2001 Jun 28 [cited 2015 Sep 22];31(6):423–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11436889

99. Rosen R, Amirault J, Johnston N, Haver K, Khatwa U, Rubinstein E, et al. The utility of endoscopy and multichannel intraluminal impedance testing in children with cough and wheezing. Pediatr Pulmonol [Internet]. 2014 Nov [cited 2018 May 2];49(11):1090–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24178927

100. Rosen R, Amirault J, Heinz N, Litman H, Khatwa U. The sensitivity of acoustic cough recording relative to intraesophageal pressure recording and patient report during reflux testing. Neurogastroenterol Motil [Internet]. 2014 Nov [cited 2016 Aug 10];26(11):1635–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25264213

101. Sifrim D, Barnes N. GERD-related Chronic Cough. J Clin Gastroenterol [Internet]. 2010 Apr [cited 2018 May 2];44(4):234–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20104183

102. M. Ngerncham, D.C. Barnhart RNH. Risk factors for recurrent gastroesophageal reflux disease after fundoplication in

pediatric patients: a case–control study. J Pediatr Surg. 2007;42:1478–85.

103. Barnes N, Robertson N, Lakhoo K. Anti-reflux surgery for the neonatal intensive care-dependent infant. Early Hum Dev [Internet]. 2003 Dec [cited 2016 Mar 17];75(1–2):71–8. Available from: http://www.sciencedirect.com/science/article/pii/S0378378203001488

104. Blondeau K, Mertens V, Dupont L, Pauwels A, Farré R, Malfroot A, et al. The relationship between gastroesophageal reflux and cough in children with chronic unexplained cough using combined impedance-pH-manometry recordings. Pediatr Pulmonol [Internet]. 2011 Mar [cited 2016 Mar 17];46(3):286–94. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20967945

105. Chang AB, Lasserson TJ, Kiljander TO, Connor FL, Gaffney JT, Garske LA. Systematic review and meta-analysis of randomised controlled trials of gastro-oesophageal reflux interventions for chronic cough associated with gastro-oesophageal reflux. BMJ [Internet]. 2006 Jan 7 [cited 2016 Aug 10];332(7532):11–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16330475

106. Thakkar K, Boatright RO, Gilger MA, El-Serag HB, Katz RM, Fonkalsrud EW, et al. Gastroesophageal reflux and asthma in children: a systematic review. Pediatrics [Internet]. 2010 Apr 1 [cited 2018 May 2];125(4):e925-30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7243506

107. Tolia V, Vandenplas Y. Systematic review: the extra-oesophageal symptoms of gastro-oesophageal reflux disease in children. Aliment Pharmacol Ther [Internet]. 2009 Feb 1 [cited 2018 May 2];29(3):258–72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19143046

108. Condino AA, Sondheimer J, Pan Z, Gralla J, Perry D, O'Connor JA. Evaluation of infantile acid and nonacid gastroesophageal reflux using combined pH monitoring and impedance measurement. J Pediatr Gastroenterol Nutr [Internet]. 2006 Jan [cited 2016 Jul 29];42(1):16–21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16385248

109. Palm K, Sawicki G, Rosen R. The impact of reflux burden on Pseudomonas positivity in children with Cystic Fibrosis. Pediatr Pulmonol [Internet]. 2012 Jun [cited 2018 Apr 27];47(6):582–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22162484

110. Robinson NB, DiMango E. Prevalence of Gastroesophageal Reflux in Cystic Fibrosis and Implications for Lung Disease. Ann Am Thorac Soc [Internet]. 2014 Jul 17 [cited 2018 Apr 27];11(6):964–8. Available from: http://www.atsjournals.org/doi/abs/10.1513/AnnalsATS.201401-044FR

111. Benden C, Aurora P, Curry J, Whitmore P, Priestley L, Elliott MJ. High prevalence of gastroesophageal reflux in children after lung transplantation. Pediatr Pulmonol [Internet]. 2005 Jul [cited 2018 Apr 30];40(1):68–71. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15880421

112. Sheikh S, Quach J, McCoy K. Nissen Fundoplication in Patients With Cystic Fibrosis. Impact on Nutrition. Chest [Internet]. 2011 Oct 23 [cited 2018 Apr 30];140(4):385A. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0012369216541385

113. T. Marquez, Thao & A. Saltzman, Daniel & J. Hess, Donavon & E. Regelmann, Warren & J. Schwarzenberg, Sarah & M. Brearley, Ann & D. Acton R (2009). Nissen Fundoplication in Children with Cystic Fibrosis and Gastroesophageal Reflux Disease: Effects On Pulmonary Function Reconsidered. In: American Academy of Pediatrics National Conference and Exhibition. 2009.

114. Boesch RP, Acton JD. Outcomes of fundoplication in children with cystic fibrosis. J Pediatr Surg [Internet]. 2007 Aug 1 [cited 2018 Apr 30];42(8):1341–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17706493

115. Srivastava R, Jackson WD, Barnhart DC. Dysphagia and Gastroesophageal Reflux Disease: Dilemmas in Diagnosis and Management in Children with Neurological Impairment. Pediatr Ann [Internet]. 2010 Apr 1 [cited 2018 Apr 30];39(4):225–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20411900

116. Suskind DL, Thompson DM, Gulati M, Huddleston P, Liu DC, Baroody FM. Improved infant swallowing after gastroesophageal reflux disease treatment: a function of improved laryngeal sensation? Laryngoscope [Internet]. 2006 Aug [cited 2015 Sep 22];116(8):1397–403. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16885743

117. Altman KW, Wetmore RF, Marsh RR. Congenital airway abnormalities requiring tracheotomy: a profile of 56 patients and their diagnoses over a 9 year period. Int J Pediatr Otorhinolaryngol [Internet]. 1997 Aug 20 [cited 2015 Sep 22];41(2):199–206. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9306176

118. Hadi U, Hamdan A-L. Diagnosis and management of tracheal stenosis. J Med Liban [Internet]. Jan [cited 2015 Sep 22];52(3):131–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16432969

119. Takamizawa S, Tsugawa C, Nishijima E, Muraji T, Satoh S. Laryngotracheal separation for intractable aspiration pneumonia in neurologically impaired children: experience with 11 cases. J Pediatr Surg [Internet]. 2003 Jun [cited 2015 Sep 22];38(6):975–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12778406

120. PACE F, PALLOTTA S, TONINI M, VAKIL N, BIANCHI PORRO G. Systematic review: gastro-oesophageal reflux disease and dental lesions. Aliment Pharmacol Ther [Internet]. 2008 Mar 27 [cited 2018 May 2];27(12):1179–86. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18373634

121. Wilder-Smith CH, Wilder-Smith P, Kawakami-Wong H, Voronets J, Osann K, Lussi A. Quantification of Dental Erosions in Patients With GERD Using Optical Coherence Tomography Before and After Double-Blind, Randomized Treatment With Esomeprazole or Placebo. Am J Gastroenterol [Internet]. 2009 Nov 4 [cited 2018 May 2];104(11):2788–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19654570

122. Ganesh M, Hertzberg A, Nurko S, Needleman H, Rosen R. Acid Rather Than Nonacid Reflux Burden Is a Predictor of Tooth Erosion. J Pediatr Gastroenterol Nutr [Internet]. 2016 Feb [cited 2018 May 2];62(2):309–13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26230904

123. Hassall E. Talk is cheap, often effective: symptoms in infants often respond to non-pharmacologic measures. - PubMed - NCBI. J Pediatr [Internet]. 2008;152(3):301–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/?term=hassall+talk+is+cheap

124. Tighe M, Afzal NA, Bevan A, Hayen A, Munro A, Beattie RM. Pharmacological treatment of children with gastro-oesophageal reflux. Cochrane database Syst Rev [Internet]. 2014 [cited 2016 Jul 31];(11):CD008550. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25419906

125. Richards CA, Milla PJ, Andrews PL, Spitz L. Retching and vomiting in neurologically impaired children after fundoplication: predictive preoperative factors. J Pediatr Surg [Internet]. 2001 Sep [cited 2016 Feb 1];36(9):1401–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11528615

126. Raghavendran K, Nemzek J, Napolitano LM, Knight PR. Aspiration-induced lung injury. Crit Care Med [Internet]. 2011 Apr [cited 2018 May 3];39(4):818–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21263315

127. Mousa H, Woodley FW, Metheney M, Hayes J. Testing the association between gastroesophageal reflux and apnea in infants. J Pediatr Gastroenterol Nutr [Internet]. 2005 Aug [cited 2016 Jul 31];41(2):169–77. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16056095

128. Monasterio FO, Molina F, Berlanga F, López ME, Ahumada H, Takenaga RH, et al. Swallowing disorders in Pierre Robin sequence: its correction by distraction. J Craniofac Surg [Internet]. 2004 Nov [cited 2018 Apr 30];15(6):934–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15547378

129. Committee on Fetus and Newborn. American Academy of Pediatrics. Apnea, sudden infant death syndrome, and home monitoring. Pediatrics [Internet]. 2003 Apr [cited 2016 Jul 31];111(4 Pt 1):914–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12671135

130. Magistà AM1, Indrio F, Baldassarre M, Bucci N, Menolascina A, Mautone A FR. Multichannel intraluminal impedance to detect relationship between gastroesophageal reflux and apnoea of prematurity. Dig Liver Dis. . 2007;39(3):216–21.

131. Sondheimer JM. Continuous monitoring of distal esophageal pH: a diagnostic test for gastroesophageal reflux in infants. J Pediatr [Internet]. 1980 May [cited 2016 Jul 29];96(5):804–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7365578

132. Reflux., Boix-Ochoa J, Lafuenta JM G-VJ. Twenty-four hour exophageal pH monitoring in gastroesophageal. J Pediatr Surg. 1980;15(1):74–8.

133. Vandenplas Y, Lepoudre R, Helven R. Dependability of esophageal pH-monitoring data in infants on cutoff limits: the oscillatory index. J Pediatr Gastroenterol Nutr [Internet]. 1990 Oct [cited 2016 Jul 30];11(3):304–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2246710

134. Vandenplas Y, Goyvaerts H, Helven R, Sacre L. Gastroesophageal reflux, as measured by 24-hour pH monitoring, in 509 healthy infants screened for risk of sudden infant death syndrome. Pediatrics [Internet]. 1991 Oct [cited 2016 Jul

29];88(4):834–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1896295

135. Cucchiara S, Staiano A, Gobio Casali L, Boccieri A, Paone FM. Value of the 24 hour intraoesophageal pH monitoring in children. Gut [Internet]. 1990 Feb [cited 2018 May 4];31(2):129–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2311969

136. Yvan Vandenplas Ã, Rudolph Committee Members CD, Di Lorenzo C, Hassall E, Gregory Liptak jj, Lynnette Mazur ô, et al. Pediatric Gastroesophageal Reflux Clinical Practice Guidelines.

137. Vandenplas Y, Sacré-Smits L. Continuous 24-hour esophageal pH monitoring in 285 asymptomatic infants 0-15 months old. J Pediatr Gastroenterol Nutr [Internet]. [cited 2016 Jul 29];6(2):220–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3694346

138. Jalal A, Payne HRJ, Jeyasingham K. The influence of age on gastro-oesophageal re¯ux: a re-appraisal of the DeMeester scoring system. [cited 2018 May 10]; Available from: https://watermark.silverchair.com/18-4-411.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAawwggGoBgkqhkiG9w0BBwagggGZMIIBlQIBADCCAY4GCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMXIEMTuY0CY81K_ySAgEQgIIBXyx2gCciEKjVpwp_Gu3eUGrGmqRvSyXR4tV27Yqt23h881

139. Cucchiara S, Santamaria F, Minella R, Alfieri E, Scoppa A, Calabrese F, et al. Simultaneous prolonged recordings of proximal and distal intraesophageal pH in children with gastroesophageal reflux disease and respiratory symptoms. Am J Gastroenterol [Internet]. 1995 Oct [cited 2018 May 4];90(10):1791–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7572896

140. MANCHESTER TC. EFFECT OF ORANGE AND LEMON JUICES ON ACTIVITY OF PROTEOLYTIC ENZYMES. J Food Sci [Internet]. 1942 Sep 1 [cited 2018 May 10];7(5):394–402. Available from: http://doi.wiley.com/10.1111/j.1365-2621.1942.tb17665.x

141. Sifrim D, Castell D, Dent J, Kahrilas PJ. Gastro-oesophageal reflux monitoring: review and consensus report on detection and definitions of acid, non-acid, and gas reflux. Gut [Internet]. 2004 Jul [cited 2018 May 11];53(7):1024–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15194656

142. Pilic D, Fröhlich T, Nöh F, Pappas A, Schmidt-Choudhury A, Köhler H, et al. Detection of gastroesophageal reflux in children using combined multichannel intraluminal impedance and pH measurement: Data from the German Pediatric Impedance Group. J Pediatr. 2011;158(4).

143. Wiener GJ, Richter JE, Copper JB, Wu WC, Castell DO. The symptom index: a clinically important parameter of ambulatory 24-hour esophageal pH monitoring. Am J Gastroenterol [Internet]. 1988 Apr [cited 2016 Jul 31];83(4):358–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3348191

144. Breumelhof R, Smout AJ. The symptom sensitivity index: a valuable additional parameter in 24-hour esophageal pH recording. Am J Gastroenterol [Internet]. 1991 Feb [cited 2016 Jul 31];86(2):160–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1992627

145. Wenzl TG, Benninga MA, Loots CM, Salvatore S. Indications, Methodology, and Interpretation of Combined Esophageal Impedance-pH Monitoring in Children: ESPGHAN EURO-PIG Standard Protocol.

146. Weusten BL, Roelofs JM, Akkermans LM, Van Berge-Henegouwen GP, Smout a J. The symptom-association probability: an improved method for symptom analysis of 24-hour esophageal pH data. Gastroenterology. 1994;107(6):1741–5.

147. Rosen R, Nurko S. The importance of multichannel intraluminal impedance in the evaluation of children with persistent respiratory symptoms. Am J Gastroenterol [Internet]. 2004 Dec [cited 2016 Jul 31];99(12):2452–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15571595

148. Di Pace MR, Caruso AM, Catalano P, Casuccio A, Cimador M, De Grazia E. Evaluation of esophageal motility using multichannel intraluminal impedance in healthy children and children with gastroesophageal reflux. J Pediatr Gastroenterol Nutr [Internet]. 2011 Mar [cited 2018 May 11];52(1):26–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21376190

149. Borrelli O1, Battaglia M, Galos F, Aloi M, De Angelis D, Moretti C, Mancini V, Cucchiara S MF. Non-acid gastro-oesophageal reflux in children with suspected pulmonary aspiration. Dig Liver Dis. 2010;42(3):115–21.

150. Greifer M, Ng K, Levine J. Impedance and Extraesophageal Manifestations of Reflux in Pediatrics. 2012 [cited 2018 May 11];122:1397–400. Available from: http://www.sbccp.org.br/arquivos/LG-2012-06-Impedance-and-Extraesophageal.pdf

151. Edoardo Savarino1, Albert J. Bredenoord2, Mark Fox3 JEP, Group SR and CPG on behalf of the IW, Function for D of GM and. GERD consists of troublesome symptom s or mucosal damage resulting from retrograde movement of gastric content through an incompetent oesophagogastric junc- tion (EGJ). [cited 2018 May 11]; Available from: https://www.nature.com/articles/nrgastro.2017.130.pdf?origin=ppub

152. Skopnik H, Silny J, Heiber O, Schulz J, Rau G, Heimann G. Gastroesophageal Reflux in Infants: Evaluation of a New Intraluminal Impedance Technique. J Pediatr Gastroenterol &amp Nutr [Internet]. 1996 Dec [cited 2016 Aug 2];23(5):591–8. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00005176-199612000-00014

153. van Wijk MP, Benninga MA, Omari TI. Role of the multichannel intraluminal impedance technique in infants and children. J Pediatr Gastroenterol Nutr [Internet]. 2009 Jan [cited 2016 Jul 29];48(1):2–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19172117

154. Dalby K, Nielsen RG, Markoew S, Kruse-Andersen S, Husby S. Reproducibility of 24-hour combined multiple intraluminal impedance (MII) and pH measurements in infants and children. Evaluation of a diagnostic procedure for gastroesophageal reflux disease. Dig Dis Sci [Internet]. 2007 Sep [cited 2016 Jul 30];52(9):2159–65. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17436090

155. Salvatore S, Hauser B, Vandemaele K, Novario R, Vandenplas Y. Gastroesophageal reflux disease in infants: how much is predictable with questionnaires, pH-metry, endoscopy and histology? J Pediatr Gastroenterol Nutr [Internet]. 2005 Feb [cited 2018 Apr 26];40(2):210–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15699699

156. Macharia EW. Comparison of upper gastrointestinal contrast studies and pH/impedance tests for the diagnosis of childhood gastro-oesophageal reflux. Pediatr Radiol [Internet]. 2012 Aug 4 [cited 2016 Jun 28];42(8):946–51. Available from: http://link.springer.com/10.1007/s00247-012-2405-3

157. Valusek P a., Peter SDS, Keckler SJ, Laituri C a., Snyder CL, Ostlie DJ, et al. Does an upper gastrointestinal study change operative management for gastroesophageal reflux? J Pediatr Surg [Internet]. 2010;45(6):1169–72. Available from: http://dx.doi.org/10.1016/j.jpedsurg.2010.02.083

158. Alvares BR, Della Torre H, Mezzacappa MA. Alvares BR et al. GI series sensitivity for gastroesophageal reflux in preterm newborn Sensitivity of upper gastrointestinal series in the diagnosis of gastroesophageal reflux in premature infants *. Jul/Ago [Internet]. 2011 [cited 2018 May 14];44(4):211–4. Available from: http://www.scielo.br/pdf/rb/v44n4/en_v44n4a04.pdf

159. Shariff F, Kiely E, Curry J, Drake D, Pierro A, McHoney M. Outcome After Laparoscopic Fundoplication in Children Under 1 Year. J Laparoendosc Adv Surg Tech [Internet]. 2010 Sep [cited 2018 Mar 13];20(7):661–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20822420

160. Shah SR, Jegapragasan M, Fox MD, Prince JM, Segura BJ, Kane TD. A review of laparoscopic Nissen fundoplication in children weighing less than 5 kg. J Pediatr Surg [Internet]. 2010 Jun [cited 2016 Feb 8];45(6):1165–8. Available from: http://www.sciencedirect.com/science/article/pii/S0022346810001971

161. Arasu TS, Wyllie R, Fitzgerald JF, Franken EA, Siddiqui AR, Lehman GA, et al. Gastroesophageal reflux in infants and children comparative accuracy of diagnostic methods. J Pediatr [Internet]. 1980 May [cited 2018 May 14];96(5):798–803. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7365577

162. Salvatore S, Hauser B, Vandemaele K, Novario R, Vandenplas Y. Gastroesophageal Reflux Disease in Infants: How Much is Predictable with Questionnaires, pH-metry, Endoscopy and Histology? J Pediatr Gastroenterol Nutr [Internet]. 2005 Feb [cited 2016 Aug 2];40(2):210–5. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00005176-200502000-00024

163. Goldani HA, Nunes D La, Ferreira CT. Managing gastroesophageal reflux disease in children: The role of endoscopy. World J Gastrointest Endosc [Internet]. 2012 Aug 16 [cited 2018 Apr 26];4(8):339–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22912907

164. Vieira MC, Pisani JC, Mulinari RA. Diagnosis of reflux esophagitis in infants: histology of the distal esophagus must complement upper gastrointestinal endoscopy. J Pediatr (Rio J) [Internet]. 2004 May 15 [cited 2018 Apr 26];80(3):197–202. Available from: http://www.jped.com.br/Redirect.aspx?varArtigo=1183

165. Rosen R, Fritz J, Nurko A, Simon D, Nurko S. Lipid-laden macrophage index is not an indicator of gastroesophageal reflux-related respiratory disease in children. Pediatrics [Internet]. 2008 Apr [cited 2018 May 3];121(4):e879-84. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/18362101

166. Krishnan U1, Mitchell JD, Tobias V, Day AS BT. Fat laden macrophages in tracheal aspirates as a marker of reflux aspiration: a negative report. J Pediatr Gastroenterol Nutr. 2002;35(3):309–13.

167. Tobin JM, McCloud P, Cameron DJ. Posture and gastro-oesophageal reflux: a case for left lateral positioning. Arch Dis Child [Internet]. 1997 Mar 1 [cited 2018 Mar 29];76(3):254–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9135268

168. Ewer AK, James ME, Tobin JM. Prone and left lateral positioning reduce gastro-oesophageal reflux in preterm infants. Arch Dis Child Fetal Neonatal Ed [Internet]. 1999 Nov [cited 2018 Mar 29];81(3):F201-5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10525024

169. Omari TI, Rommel N, Staunton E, Lontis R, Goodchild L, Haslam RR, et al. Paradoxical impact of body positioning on gastroesophageal reflux and gastric emptying in the premature neonate. J Pediatr [Internet]. 2004 Aug [cited 2018 Mar 29];145(2):194–200. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15289766

170. Corvaglia L, Martini S, Aceti A, Arcuri S, Rossini R, Faldella G. Nonpharmacological management of gastroesophageal reflux in preterm infants. Biomed Res Int [Internet]. 2013 [cited 2018 Mar 29];2013:141967. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24073393

171. Meyers WF, Herbst JJ. Effectiveness of Positioning Therapy for Gastroesophageal Reflux. Pediatrics [Internet]. 1982 [cited 2018 Mar 29];69(6). Available from: http://pediatrics.aappublications.org/content/69/6/768?ijkey=c720a5e802c7ede7824af9875ecbc865e0023560&keytype2=tf_ipsecsha

172. Orenstein SR1 MJ. Efficacy of conservative therapy as taught in the primary care setting for symptoms suggesting infant gastroesophageal reflux. - PubMed - NCBI. J Pediatr [Internet]. 2008;152(3):310–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/?term=orenstein+macgowan++conservative

173. Horvath A, Dziechciarz P, Szajewska H. The effect of thickened-feed interventions on gastroesophageal reflux in infants: systematic review and meta-analysis of randomized, controlled trials. Pediatrics [Internet]. 2008 Dec 1 [cited 2018 Mar 29];122(6):e1268-77. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19001038

174. Woods CW, Oliver T, Lewis K, Yang Q. Development of necrotizing enterocolitis in premature infants receiving thickened feeds using SimplyThick®. J Perinatol [Internet]. 2012 Feb [cited 2018 Mar 29];32(2):150–2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22289705

175. Beal J, Silverman B, Bellant J, Young TE, Klontz K. Late onset necrotizing enterocolitis in infants following use of a xanthan gum-containing thickening agent. J Pediatr [Internet]. 2012 Aug 1 [cited 2018 Mar 29];161(2):354–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22575248

176. Keady S. Update on drugs for gastro-oesophageal reflux disease. Arch Dis Child Educ Pract Ed [Internet]. 2007 Aug 1 [cited 2018 Apr 25];92(4):ep114-8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17644666

177. Terrin G, Passariello A, De Curtis M, Manguso F, Salvia G, Lega L, et al. Ranitidine is associated with infections, necrotizing enterocolitis, and fatal outcome in newborns. Pediatrics [Internet]. 2012 Jan [cited 2016 Jul 31];129(1):e40-5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22157140

178. Moore DJ, Tao BS-K, Lines DR, Hirte C, Heddle ML, Davidson GP. Double-blind placebo-controlled trial of omeprazole in irritable infants with gastroesophageal reflux. J Pediatr [Internet]. 2003 Aug [cited 2018 Apr 25];143(2):219–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12970637

179. Orenstein SR, Hassall E, Furmaga-Jablonska W, Atkinson S, Raanan M. Multicenter, double-blind, randomized, placebo-controlled trial assessing the efficacy and safety of proton pump inhibitor lansoprazole in infants with symptoms of gastroesophageal reflux disease. J Pediatr [Internet]. 2009 Apr 1 [cited 2016 Mar 17];154(4):514–520.e4. Available from: http://www.jpeds.com/article/S0022347608008640/fulltext

180. MacLennan S, Augood C, Cash-Gibson L, Logan S, Gilbert RE. Cisapride treatment for gastro-oesophageal reflux in children. Cochrane Database Syst Rev [Internet]. 2010 Apr 14 [cited 2018 Apr 25];(4):CD002300. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20393933

181. Levin DN, Diamond IR, Langer JC. Complete vs partial fundoplication in children with esophageal atresia. J Pediatr Surg [Internet]. 2011 May [cited 2018 May 18];46(5):854–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21616240

182. Kubiak R, Andrews J, Grant HW. Long-Term Outcome of Laparoscopic Nissen Fundoplication Compared with Laparoscopic Thal Fundoplication in Children. Ann Surg [Internet]. 2011 Jan [cited 2018 Mar 13];253(1):44–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21233605

183. LaRiviere CA, Parimi C, Huaco JC, Acierno SA, Garrison MM, Goldin AB. Variations in preoperative decision making for antireflux procedures in pediatric gastroesophageal reflux disease: a survey of pediatric surgeons. J Pediatr Surg. 2011;46(6):1093–8.

184. Richards CA, Smith V V, Milla PJ, Spitz L, Andrews PLR. The histological appearances of Nissen-type fundoplication in the ferret. Neurogastroenterol Motil [Internet]. 2003 Apr [cited 2018 May 14];15(2):121–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12680911

185. Richards CA, Carr D, Spitz L, Milla PJ, Andrews PL. Nissen-type fundoplication and its effects on the emetic reflex and gastric motility in the ferret. Neurogastroenterol Motil [Internet]. 2000 Feb [cited 2018 May 14];12(1):65–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10652115

186. Di Lorenzo C, Orenstein S. Fundoplication: friend or foe? J Pediatr Gastroenterol Nutr [Internet]. 2002 Feb [cited 2016 Aug 5];34(2):117–24. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11840028

187. Sifrim D, Holloway RH, Tack J, Zelter A, Missotten T, Coulie B, et al. Effect of sumatriptan, a 5HT1 agonist, on the frequency of transient lower esophageal sphincter relaxations and gastroesophageal reflux in healthy subjects. Am J Gastroenterol [Internet]. 1999 Nov [cited 2016 Aug 5];94(11):3158–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10566707

188. Baerg J, Thorpe D, Bultron G, Vannix R, Knott EM, Gasior AC, et al. A multicenter study of the incidence and factors associated with redo Nissen fundoplication in children. J Pediatr Surg [Internet]. 2013 Jun [cited 2015 Sep 5];48(6):1306–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23845623

189. Rothenberg SS1. The first decade's experience with laparoscopic Nissen fundoplication in infants and children. - PubMed - NCBI. J Pediatr Surg [Internet]. 2005;40(1):142–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15868575

190. Rothenberg SS. Two Decades of Experience with Laparoscopic Nissen Fundoplication in Infants and Children: A Critical Evaluation of Indications, Technique, and Results. J Laparoendosc Adv Surg Tech [Internet]. 2013 Sep 3 [cited 2018 May 15];23(9):791–4. Available from: http://online.liebertpub.com/doi/abs/10.1089/lap.2013.0299

191. Mauritz FA, van Herwaarden-Lindeboom MYA, Stomp W, Zwaveling S, Fischer K, Houwen RHJ, et al. The effects and efficacy of antireflux surgery in children with gastroesophageal reflux disease: a systematic review. J Gastrointest Surg [Internet]. 2011 Oct [cited 2016 Aug 11];15(10):1872–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21800225

192. Cheung KM, Tse HW, Tse PWT, Chan KH. Nissen fundoplication and gastrostomy in severely neurologically impaired children with gastroesophageal reflux. Hong Kong Med J. 2006;12(4):282–8.

193. Tovar JA1, Luis AL, Encinas JL, Burgos L, Pederiva F, Martinez L OP. Pediatric surgeons and gastroesophageal reflux. - PubMed - NCBI. J Pediatr Surg [Internet]. 2007;42(2):277–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17270535

194. Diamond IR, Mah K, Kim PCWW, Bohn D, Gerstle JT, Wales PW. Predicting the need for fundoplication at the time of congenital diaphragmatic hernia repair. J Pediatr Surg [Internet]. 2007 Jun [cited 2016 Mar 14];42(6):1066–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17560222

195. Fasching G, Huber A, Uray E, Sorantin E, Lindbichler F, Mayr J. Gastroesophageal reflux and diaphragmatic motility after repair of congenital diaphragmatic hernia. Eur J Pediatr Surg Off J Austrian Assoc Pediatr Surg . [et al] = Zeitschrift für Kinderchirurgie [Internet]. 2000 Dec [cited 2016 Mar 14];10(6):360–4. Available from: https://www-thieme-connect-de.libproxy.ucl.ac.uk/products/ejournals/abstract/10.1055/s-2008-1072391

196. Muratore CS, Utter S, Jaksic T, Lund DP, Wilson JM. Nutritional morbidity in survivors of congenital diaphragmatic hernia. J Pediatr Surg [Internet]. 2001 Aug [cited 2015 Aug 27];36(8):1171–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11479850

197. Arena F, Romeo C, Baldari S, Arena S, Antonuccio P, Campennì A, et al. Gastrointestinal sequelae in survivors of congenital diaphragmatic hernia. Pediatr Int [Internet]. 2008 Feb [cited 2016 Mar 14];50(1):76–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18279210

198. Maier S, Zahn K, Wessel LM, Schaible T, Brade J, Reinshagen K. Preventive antireflux surgery in neonates with

congenital diaphragmatic hernia: a single-blinded prospective study. J Pediatr Surg [Internet]. 2011 Aug [cited 2016 Mar 14];46(8):1510–5. Available from: http://www.sciencedirect.com/science/article/pii/S0022346811003198

199. Levy MS, Sorrels CW, Wagner CW, Jackson RJ, Barnes RW, Smith SD. Evolution of the modified Rossetti fundoplication in children: surgical technique and results. Ann Surg [Internet]. 1999 Jun [cited 2018 Apr 25];229(6):774-9; discussion 779-80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10363890

200. Wockenforth R, Gillespie CS, Jaffray B. Survival of children following Nissen fundoplication. Br J Surg [Internet]. 2011 May [cited 2015 Sep 5];98(5):680–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21351077

201. St Peter SD, Valusek PA, Calkins CM, Shew SB, Ostlie DJ, Holcomb GW. Use of esophagocrural sutures and minimal esophageal dissection reduces the incidence of postoperative transmigration of laparoscopic Nissen fundoplication wrap. J Pediatr Surg. 2007;42(1):25–30.

202. Lobe TE, Schropp KP, Lunsford K. Laparoscopic Nissen fundoplication in childhood. J Pediatr Surg [Internet]. 1993 Mar [cited 2018 May 21];28(3):358-60; discussion 360-1. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8468646

203. Franzén T, Anderberg B, Wirén M, Johansson K-E. Long-term outcome is worse after laparoscopic than after conventional Nissen fundoplication. Scand J Gastroenterol [Internet]. 2005 Nov [cited 2016 Aug 9];40(11):1261–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16334434

204. Pacilli M, Eaton S, McHoney M, Kiely EM, Drake DP, Curry JI, et al. Four year follow-up of a randomised controlled trial comparing open and laparoscopic Nissen fundoplication in children. Arch Dis Child [Internet]. 2014 Jun 1 [cited 2018 May 16];99(6):516–21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24532685

205. Fox D, Barnard J, Campagna EJ, Dickinson LM, Bruny J, Kempe A. Fundoplication and the Pediatric Surgeon: Implications for Shared Decision-Making and the Medical Home. Acad Pediatr. 2012;12(6):558–66.

206. Mattioli G, Repetto P, Carlini C, Torre M, PiniPrato A, Mazzola C, et al. Laparoscopic vs open approach for the treatment of gastroesophageal reflux in children. Surg Endosc [Internet]. 2002 May 1 [cited 2018 May 21];16(5):750–2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11997815

207. Somme S, Rodriguez JA, Kirsch DG, Liu DC. Laparoscopic versus open fundoplication in infants. Surg Endosc [Internet]. 2002 Jan 5 [cited 2018 May 21];16(1):54–6. Available from: http://link.springer.com/10.1007/s00464-001-8147-1

208. Chang AB, Lasserson TJ, Gaffney J, Connor FL, Garske LA. Gastro-oesophageal reflux treatment for prolonged non-specific cough in children and adults. Cochrane database Syst Rev [Internet]. 2011 Jan [cited 2016 Mar 17];(1):CD004823. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21249664

209. McHoney M, Wade AM, Eaton S, Howard RF, Kiely EM, Drake DP, et al. Clinical Outcome of a Randomized Controlled Blinded Trial of Open Versus Laparoscopic Nissen Fundoplication in Infants and Children. Ann Surg [Internet]. 2011 Aug [cited 2018 May 18];254(2):209–16. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00000658-201108000-00006

210. Fyhn TJ, Knatten CK, Edwin B, Schistad O, Aabakken L, Kjosbakken H, et al. Randomized Controlled Trial of Laparoscopic and Open Nissen Fundoplication in Children. Ann Surg [Internet]. 2015 Jun [cited 2016 Feb 1];261(6):1061–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26291953

211. Launay V, Gottrand F, Turck D, Michaud L, Ategbo S, Farriaux JP. Percutaneous endoscopic gastrostomy in children: influence on gastroesophageal reflux. Pediatrics [Internet]. 1996 May [cited 2016 Aug 2];97(5):726–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8628615

212. Toporowska-Kowalska E, Gębora-Kowalska B, Jabłoński J, Fendler W, Wąsowska-Królikowska K. Influence of percutaneous endoscopic gastrostomy on gastro-oesophageal reflux evaluated by multiple intraluminal impedance in children with neurological impairment. Dev Med Child Neurol [Internet]. 2011 Oct [cited 2016 Aug 2];53(10):938–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21752017

213. St Peter SD, Ostlie DJ, Holcomb GW. The use of biosynthetic mesh to enhance hiatal repair at the time of redo Nissen fundoplication. J Pediatr Surg. 2007;42(7):1298–301.

214. Peter SDS, Barnhart DC, Ostlie DJ, Tsao K, Leys CM, Sharp SW, et al. Minimal vs extensive esophageal mobilization during laparoscopic fundoplication: a prospective randomized trial. 2011;

215. Kimber C, Kiely EM, Spitz L. The failure rate of surgery for gastro-oesophageal reflux. J Pediatr Surg [Internet]. 1998

Jan [cited 2018 May 23];33(1):64–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9473102

216. Smith CD, Othersen HB, Gogan NJ, Walker JD. Nissen fundoplication in children with profound neurologic disability. High risks and unmet goals. Ann Surg [Internet]. 1992 Jun [cited 2016 Aug 3];215(6):654-8; discussion 658-9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1632687

217. Wales PW1, Diamond IR, Dutta S, Muraca S, Chait P, Connolly B LJ. Fundoplication and gastrostomy versus image-guided gastrojejunal tube for enteral feeding in neurologically impaired children with gastroesophageal... - PubMed - NCBI. J Pediatr Surg [Internet]. 2002;37(3):407–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11877658

218. Srivastava R, Downey EC, O'Gorman M, Feola P, Samore M, Holubkov R, et al. Impact of fundoplication versus gastrojejunal feeding tubes on mortality and in preventing aspiration pneumonia in young children with neurologic impairment who have gastroesophageal reflux disease. Pediatrics [Internet]. 2009 Jan [cited 2016 Aug 4];123(1):338–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19117901

219. Morabito A, Lall A, Lo Piccolo R, McCarthy H, Kauffmann L, Ahmed S, et al. Total esophagogastric dissociation: 10 years' review. J Pediatr Surg [Internet]. 2006 May [cited 2018 May 15];41(5):919–22. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022346806000194

220. Peters RT, Goh YL, Veitch JM, Khalil BA, Morabito A. Morbidity and mortality in total esophagogastric dissociation: A systematic review. J Pediatr Surg [Internet]. 2013 Apr [cited 2018 May 15];48(4):707–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23583122

221. de Lagausie P, Bonnard A, Schultz A, Van Den Abbeel T, Bellaiche M, Hartmann JF, et al. Reflux in esophageal atresia, tracheoesophageal cleft, and esophagocoloplasty: Bianchi's procedure as an alternative approach. J Pediatr Surg [Internet]. 2005 Apr [cited 2018 May 15];40(4):666–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15852275

222. MinkyYu Lee (MKLabs Co. L(. StarUML [Internet]. 2011. Available from: http://staruml.io/download

223. Al-Awar, J., Chapanis, A., Ford R. Tutorials for the first-time computer user. IEEE Trans Prof Commun. 1981;PC-24(1).

224. P. Bourque and R.E. Fairley E. Guide to the Software Engineering Body of Knowledge, Version 3.0. IEEE Comput Soc [Internet]. 2014; Available from: www.swebok.org.

225. Brooke J. SUS-A quick and dirty usability scale [Internet]. [cited 2018 Oct 16]. Available from: https://hell.meiert.org/core/pdf/sus.pdf

226. Jakob Nielsen TKL. A mathematical model of the finding of usability problems. In: CHI '93 Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. 1993. p. 206–13.

227. ISO. Ergonomic requirements for office work with visual display terminals – Part 11: Guidance on usability. 1998;9241–11.

228. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. Int J Hum Comput Interact [Internet]. 2008 Jul 29 [cited 2018 Oct 16];24(6):574–94. Available from: http://www.tandfonline.com/doi/abs/10.1080/10447310802205776

229. OfCom. Ofcom. The Market in Context. 2011;

230. YouGov. Smartphone growth in the UK: Challenges in the handset market. 2012;

231. Venables WN, Smith DM. An Introduction to R Notes on R: A Programming Environment for Data Analysis and Graphics Version 3.4.3 (2017-11-30). R Gentlem R Ihaka Copyr c [Internet]. 1990 [cited 2018 Mar 13]; Available from: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf

232. Groh F. Gamification: State of the art definition and utilization. Inst Media Informatics Ulm Univ. 2012;39.

233. Hintsa Performance. HeiaHeia. 2012.

234. Virzi R. Refining the Test phase of Usability Evaluation: How many subjects is enough?". Hum Factors. 1992;34:457–68.

235. Lewis J. Sample Sizes for Usability Studies: Additional Considerations. Hum Factors. 1994;36(2):368–78.

236. Nielsen, J., and Molich R. Heuristic evaluation of user interfaces. In: ACM CHI'90 Conf. 1990.

237. M. K. Data Mining: Concepts, Models, Methods, and Algorithms, Second Edition.

238. Jeremy Howick, Iain Chalmers, Paul Glasziou, Trish Greenhalgh, Carl Heneghan, Alessandro Liberati, Ivan Moschetti, Bob Phillips and HT. Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence (Background Document). Oxford Cent Evidence-Based Med. 2011;

239. Checklist for Retrospective Database Studies. [cited 2017 May 9]; Available from: https://www.ispor.org/workpaper/research_practices/A_Checklist_for_Retroactive_Database_Studies-Retrospective_Database_Studies.pdf

240. CHEERS Checklist Items to include when reporting economic evaluations of health interventions. [cited 2017 May 9]; Available from: http://www.equator-network.org/wp-content/uploads/2013/04/Revised-CHEERS-Checklist-Oct13.pdf

241. Lantz B. Machine Learning with R [Internet]. 2nd ed. 2015 [cited 2017 Oct 9]. Available from: http://pdf.th7.cn/down/files/1602/Machine Learning with R, 2nd Edition.pdf

242. Guyon I, Elisseeff A, De AM. An Introduction to Variable and Feature Selection. J Mach Learn Res [Internet]. 2003 [cited 2017 Jun 9];3:1157–82. Available from: http://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf

243. Cole TJ. Too many digits: the presentation of numerical data. [cited 2017 Jun 2]; Available from: http://adc.bmj.com/content/archdischild/100/7/608.full.pdf

244. ONS. Live Births in England and Wales by Characteristics of Birth. 2010 [cited 2017 Jun 11]; Available from: http://webarchive.nationalarchives.gov.uk/20130701212709/http://www.ons.gov.uk/ons/dcp171778_241936.pdf

245. ONS. Deaths registered in England and Wales - Office for National Statistics [Internet]. 2015 [cited 2017 Jun 11]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregistrationsummarytables/2015

246. Mathisen B, Worrall L, Masel J, Wall C, Shepherd RW. Feeding problems in infants with gastro-oesophageal reflux disease: a controlled study. J Paediatr Child Health [Internet]. 1999 Apr [cited 2018 Mar 13];35(2):163–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10365354

247. Trinick R, Johnston N, Dalzell AM, McNamara PS. Reflux aspiration in children with neurodisability--a significant problem, but can we measure it? J Pediatr Surg [Internet]. 2012 Feb 1 [cited 2018 Mar 13];47(2):291–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22325378

248. DICK AC, COULTER P, HAINSWORTH AM, BOSTON VE, POTTS SR. A Comparative Study of the Analgesia Requirements Following Laparoscopic and Open Fundoplication in Children. J Laparoendosc Adv Surg Tech [Internet]. 1998 Dec 30 [cited 2018 Mar 13];8(6):425–9. Available from: http://www.liebertonline.com/doi/abs/10.1089/lap.1998.8.425

249. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Minind, Inference and Prediction [Internet]. 2008 [cited 2017 Jun 24]. Available from: https://statweb.stanford.edu/~tibs/ElemStatLearn/printings/ESLII_print10.pdf

250. The Comprehensive R Archive Network.

251. Perlich C, Provost F, Simonoff JS. Tree Induction vs. Logistic Regression: A Learning-Curve Analysis. J Mach Learn Res [Internet]. 2003 [cited 2018 Mar 13];4(Jun):211–55. Available from: http://www.jmlr.org/papers/v4/perlich03a.html

252. Delorme A. Statistical Methods. In: Encyclopedia of Medical Devices and Instrumentation [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2006 [cited 2018 Mar 13]. Available from: http://doi.wiley.com/10.1002/0471732877.emd318

253. Guyon I, Guyon I. A Scaling Law for the Validation-Set Training-Set Size Ratio. T BELL Lab [Internet]. 1997 [cited 2018 Mar 13]; Available from: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.33.1337

254. Burnham KP, Anderson DR. Multimodel Inference: Understanding AIC and BIC in Model Selection Multimodel Inference Understanding AIC and BIC in Model Selection. 2004 [cited 2017 Jun 24];33(261). Available from: http://smr.sagepub.com

255. Firth D. Bias reduction of Maximum Likelihood Estimates. Biometrika [Internet]. 1993 [cited 2017 Jul 3];80(1):27–38. Available from: http://www.jstor.org

256. Kosmidis I, Firth D. Bias reduction in exponential family nonlinear models. Biometrika [Internet]. 2009 Dec 1 [cited 2017 Jul 3];96(4):793–804. Available from: https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/asp055

257. Kosmidis I, Firth D. A generic algorithm for reducing bias in parametric estimation. Electron J Stat [Internet]. 2010 [cited

2017 Jul 3];4(0):1097–112. Available from: http://projecteuclid.org/euclid.ejs/1287147913

258. Kosmidis I. Improved estimation in cumulative link models. J R Stat Soc Ser B (Statistical Methodol [Internet]. 2014 Jan [cited 2017 Jul 3];76(1):169–96. Available from: http://doi.wiley.com/10.1111/rssb.12025

259. Sing T, Sander O, Beerenwinkel N, T L. ROCR: visualizing classifier performance in R. 2005; Available from: http://rocr.bioinf.mpi-sb.mpg.de

260. Quinlan JR. Simplifying decision trees. Int J Man Mach Stud [Internet]. 1987 Sep 1 [cited 2017 Oct 9];27(3):221–34. Available from: http://www.sciencedirect.com/science/article/pii/S0020737387800536

261. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics [Internet]. 2005 Oct 15 [cited 2017 Jun 24];21(20):3940–1. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16096348

262. Liaw A, Wiener M, Andy Liaw M. Breiman and Cutler's Random Forests for Classification and Regression Description Classification and regression based on a forest of trees using random inputs. 2015 [cited 2017 Oct 27]; Available from: https://www.stat.berkeley.edu/~breiman/RandomForests/

263. Agrawal R, Imieliński T, Swami A, Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93 [Internet]. New York, New York, USA: ACM Press; 1993 [cited 2018 Feb 27]. p. 207–16. Available from: http://portal.acm.org/citation.cfm?doid=170035.170072

264. Brin S, Motwani R, Ullman JD, Tsur S. Dynamic Itemset Counting and Implication Rules for Market Basket Data. [cited 2018 Feb 27]; Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.41.6476&rep=rep1&type=pdf

265. Hahsler M, Grün B, Hornik K, Wien W, Buchta C. Introduction to arules – A computational environment for mining association rules and frequent item sets. [cited 2018 Feb 27]; Available from: https://cran.r-project.org/web/packages/arules/vignettes/arules.pdf

266. Douglas Bates A, Maechler Maintainer Martin Maechler M. Package "Matrix" Title Sparse and Dense Matrix Classes and Methods. 2017 [cited 2017 Oct 27]; Available from: https://cran.r-project.org/web/packages/Matrix/Matrix.pdf

267. Aggarwal CC, Procopiuc C, Yu PS. Finding localized associations in market basket data. IEEE Trans Knowl Data Eng [Internet]. 2002 [cited 2018 Feb 27];14(1):51–62. Available from: http://ieeexplore.ieee.org/document/979972/

268. Gauderer MWL, Ponsky JL, Izant RJ. Gastrostomy without laparotomy: a percutaenous endoscopic technique. 1980 Dec [cited 2016 Aug 2];15(6). Available from: http://linkinghub.elsevier.com/retrieve/pii/S002234688080296X

269. Jones VS, La Hei ER, Shun A. Laparoscopic gastrostomy: the preferred method of gastrostomy in children. Pediatr Surg Int [Internet]. 2007 Oct 15 [cited 2018 May 24];23(11):1085–9. Available from: http://link.springer.com/10.1007/s00383-007-2015-6

270. Sullivan PB. Gastrostomy feeding in the disabled child: when is an antireflux procedure required? Arch Dis Child [Internet]. 1999 [cited 2018 Jun 7];81:463–4. Available from: http://adc.bmj.com/content/archdischild/81/6/463.full.pdf

271. Fonkalsrud EW, Ament ME. Gastroesophageal reflux in childhood. Curr Probl Surg [Internet]. 1996 Jan [cited 2018 Jun 3];33(1):3–70. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0011384096800128

272. Suksamanapun N, Mauritz F, Franken J, van der Zee D, van Herwaarden-Lindeboom M. Laparoscopic versus percutaneous endoscopic gastrostomy placement in children: Results of a systematic review and meta-analysis. J Minim Access Surg [Internet]. 2016 [cited 2018 Jul 10];0(0):0. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27251841

273. Akay B, Capizzani TR, Lee AM, Drongowski RA, Geiger JD, Hirschl RB, et al. Gastrostomy tube placement in infants and children: is there a preferred technique? J Pediatr Surg [Internet]. 2010 Jun [cited 2018 Jul 10];45(6):1147–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20620310

274. Fox D, Campagna EJ, Finnell SME, Partrick DA, Kempe A. Concomitant Fundoplication With Gastrostomy. J Pediatr Gastroenterol Nutr [Internet]. 2016 Dec [cited 2018 May 23];63(6):e163–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27070655

275. Peters RT, Balduyck B, Nour S. Gastrostomy complications in infants and children: a comparative study. Pediatr Surg Int [Internet]. 2010 Jul 8 [cited 2016 Aug 2];26(7):707–9. Available from: http://link.springer.com/10.1007/s00383-010-2612-7

276. Wilson GJP, van der Zee DC, Bax NMA. Endoscopic gastrostomy placement in the child with gastroesophageal reflux:

is concomitant antireflux surgery indicated? J Pediatr Surg [Internet]. 2006 Aug [cited 2016 Aug 2];41(8):1441–5. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022346806003204

277. Viswanath N, Wong D, Channappa D, Kukkady A, Brown S, Samarakkody U. Is Prophylactic Fundoplication Necessary in Neurologically Impaired Children? Eur J Pediatr Surg [Internet]. 2010 Jul 21 [cited 2018 Jun 3];20(04):226–9. Available from: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0030-1249677

278. Wadie GM, Lobe TE. Gastroesophageal reflux disease in neurologically impaired children: the role of the gastrostomy tube. Semin Laparosc Surg [Internet]. 2002 Sep 1 [cited 2015 Aug 27];9(3):180–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12407528

279. W G van der Merwe, R A Brown, J D Ireland EG. Percutaneous endoscopic gastrostomy in children. Pediatr Infect Dis J Geneva WHO [Internet]. 1998 [cited 2018 Jun 7];17(10):581–6. Available from: https://pdfs.semanticscholar.org/de8c/f3c47c1d31ff13bb428d1719a3068a36b5f9.pdf

280. Wheatley MJ, Wesley JR, Tkach DM, Coran AG. Long-term follow-up of brain-damaged children requiring feeding gastrostomy: Should an antireflux procedure always be performed? J Pediatr Surg [Internet]. 1991 Mar [cited 2016 Aug 2];26(3):301–5. Available from: http://linkinghub.elsevier.com/retrieve/pii/002234689190506O

281. Barnhart DC, Hall M, Mahant S, Goldin AB, Berry JG, Faix RG, et al. Effectiveness of fundoplication at the time of gastrostomy in infants with neurological impairment. JAMA Pediatr [Internet]. 2013 Oct 1 [cited 2016 Feb 19];167(10):911–8. Available from: http://archpedi.jamanetwork.com.libproxy.ucl.ac.uk/article.aspx?articleid=1723815

282. DC, Hall M MS. Effectiveness of Fundoplication at the Time of Gastrostomy in Infants With Neurological Impairment. JAMA Pediatr. 2013;167(10):911–8.

283. Yap BKY, Nah SA, Chen Y, Low Y. Fundoplication with gastrostomy vs gastrostomy alone: a systematic review and meta-analysis of outcomes and complications. Pediatr Surg Int [Internet]. 2017 Feb 26 [cited 2018 Jun 5];33(2):217–28. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27889821

284. Pacilli M, Eaton S, Despoina A, Ae M, Lopez PJ, Spitz L, et al. Factors predicting failure of redo Nissen fundoplication in children. 2007;

285. Noble LJ, Dalzell AM, El-Matary W. The relationship between percutaneous endoscopic gastrostomy and gastro-oesophageal reflux disease in children: a systematic review. Surg Endosc [Internet]. 2012 Sep 22 [cited 2016 Aug 2];26(9):2504–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22437953

286. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ [Internet]. 2010 Mar 23 [cited 2018 May 24];340:c332. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20332509

287. World Health Organisation. Nutritional anaemias: Tools for effective prevention and control. [Internet]. Geneva; 2017 [cited 2018 May 24]. Available from: http://apps.who.int/iris/bitstream/handle/10665/259425/9789241513067-eng.pdf?sequence=1

288. Wade A, Pan H, Eaton S, Pierro A, Ong E. An investigation of minimisation criteria. BMC Med Res Methodol [Internet]. 2006 Dec 15 [cited 2018 Jun 7];6(1):11. Available from: http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-6-11

289. Neill JKO', Neill PJO', Goth-Owens T, Horn B, Cobb LM. Care-Giver Evaluation of Anti-Gastroesophageal Reflux Procedures in Neurologically Impaired Children: What Is the Real-Life Outcome? [cited 2018 May 24]; Available from: https://ac.els-cdn.com/S0022346896907411/1-s2.0-S0022346896907411-main.pdf?_tid=12373c38-e8bc-4962-b258-6c0cb70871a9&acdnat=1527200110_fc8b077fd12b392b490367168b163401

290. McHoney M, Wade AM, Eaton S, Howard RF, Kiely EM, Drake DP, et al. Clinical outcome of a randomized controlled blinded trial of open versus laparoscopic nissen fundoplication in infants and children. Ann Surg. 2011;

291. Pierro, Agostino, Eaton S. REMOS Trial Protocol V2. London; 2008.

292. Getz K, Getz K. The Impact of Protocol Amendments on Clinical Trial Performance and Cost. [cited 2018 Jun 21]; Available from: https://www.mdsol.com/sites/default/files/Abstract_The Impact of Protocol Amendments on Clinical Trial Performance and Cost_2016.pdf

293. McHoney M, Eaton S, Wade A, Carnielli V, Kiely E, Drake D, et al. Effect of Laparoscopy and Laparotomy on Energy

and Protein Metabolism in Children: A Randomized Controlled Trial. J Pediatr [Internet]. 2010 Sep 1 [cited 2018 Jun 12];157(3):439–444.e2. Available from: https://www-sciencedirect-com.libproxy.ucl.ac.uk/science/article/pii/S0022347610002118

294. O'Neill JK, O'Neill PJ, Goth-Owens T, Horn B, Cobb LM. Care-giver evaluation of anti—Gastroesophageal reflux procedures in neurologically impaired children: What is the real-life outcome? J Pediatr Surg [Internet]. 1996 Mar 1 [cited 2018 Jun 21];31(3):375–80. Available from: https://www-sciencedirect-com.libproxy.ucl.ac.uk/science/article/pii/S0022346896907411

295. Loyd BH, Abidin RR. Revision of the Parenting Stress Index. J Pediatr Psychol [Internet]. 1985 Jun 1 [cited 2018 Jun 21];10(2):169–77. Available from: https://academic.oup.com/jpepsy/article-lookup/doi/10.1093/jpepsy/10.2.169

296. Kubiak R1, Spitz L, Kiely EM, Drake D PA. Effectiveness of fundoplication in early infancy. - PubMed - NCBI. J Pediatr Surg [Internet]. 1999;34(2):295–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10052808

297. Knatten CK, Fyhn TJ, Edwin B, Schistad O, Emblem R, Bjørnland K. Thirty-day outcome in children randomized to open and laparoscopic Nissen fundoplication. J Pediatr Surg [Internet]. 2012 Nov [cited 2016 Feb 1];47(11):1990–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23163988

298. McAteer J, Larison C, LaRiviere C, Garrison MM, Goldin AB. Antireflux procedures for gastroesophageal reflux disease in children: influence of patient age on surgical management. JAMA Surg [Internet]. 2014 Jan [cited 2016 Aug 1];149(1):56–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24197138

299. Hogan WJ1 SR. Life after antireflux surgery. Am J Med. 2000;108(Suppl 4a):181S–191S.

300. Lee JH1, Zhang J2, Wei L3 YS. Neurodevelopmental implications of the general anesthesia in neonate and infants. - PubMed - NCBI. Exp Neurol [Internet]. 2015;272:50–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25862287

301. Network S intercollegiate guidelines. British guideline on management of asthma [Internet]. 2014. Available from: www.brit-thoracic.org.uk

302. 2016 ICD-10-CM Diagnosis Codes K21.* : Gastro-esophageal reflux disease [Internet]. [cited 2016 Apr 26]. Available from: http://www.icd10data.com/ICD10CM/Codes/K00-K95/K20-K31/K21-

303. Macharia, E. , Groves, R., Coates A. TARDIS:REFLUX [Internet]. 2013. Available from: https://itunes.apple.com/us/app/tardis-reflux/id792659458?mt=8

304. Macharia EW. Focus group testing of the TARDIS:REFLUX app.

305. Marshall B, Warren JR. UNIDENTIFIED CURVED BACILLI IN THE STOMACH OF PATIENTS WITH GASTRITIS AND PEPTIC ULCERATION. Lancet. 1984;323(8390):1311–5.

306. Spechler SJ, Sharma P, Souza RF, Inadomi JM, Shaheen NJ, American Gastroenterological Association. American Gastroenterological Association technical review on the management of Barrett's esophagus. Gastroenterology [Internet]. 2011 Mar [cited 2016 Nov 21];140(3):e18–52; quiz e13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21376939

307. Loots CM1, van Wijk MP, Blondeau K, Dalby K, Peeters L, Rosen R, Salvatore S, Wenzl TG, Vandenplas Y, Benninga MA OT. Interobserver and intraobserver variability in pH-impedance analysis between 10 experts and automated analysis. - PubMed - NCBI. J Pediatr [Internet]. 2012;160(3):441–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21924738

308. Gastro-oesophageal reflux disease in children and young people: diagnosis and management | Guidance and guidelines | NICE. [cited 2018 May 14]; Available from: https://www.nice.org.uk/guidance/ng1/chapter/1-Recommendations#diagnosing-and-investigating-gord

309. Zimbric G, Bonkowsky JL, Jackson WD, Maloney CG, Srivastava R. Adverse outcomes associated with gastroesophageal reflux disease are rare following an apparent life-threatening event. J Hosp Med [Internet]. [cited 2016 Aug 11];7(6):476–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22532496

310. Goldin AB, Sawin R, Seidel KD, Flum DR. Do antireflux operations decrease the rate of reflux-related hospitalizations in children? Pediatrics [Internet]. 2006 Dec 1 [cited 2016 Feb 19];118(6):2326–33. Available from: http://pediatrics.aappublications.org.libproxy.ucl.ac.uk/content/118/6/2326

311. Van Howe RS, Storms MR, Mittal R, Holloway R, Penagini R, Sherman P, et al. Gastroesophageal reflux symptoms in infants in a rural population: longitudinal data over the first six months. BMC Pediatr [Internet]. 2010 Dec 11 [cited 2016 Aug 11];10(1):7. Available from: http://bmcpediatr.biomedcentral.com/articles/10.1186/1471-2431-10-7

312. Mathei J, Coosemans W, Nafteux P, Decker G, De Leyn P, Van Raemdonck D, et al. Laparoscopic Nissen fundoplication in infants and children: analysis of 106 consecutive patients with special emphasis in neurologically impaired vs. neurologically normal patients. Surg Endosc [Internet]. 2008 Apr [cited 2015 Sep 5];22(4):1054–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17943378

313. Böhmer C. The prevalence of gastroesophageal reflux disease in institutionalized intellectually disabled individuals. Am J Gastroenterol [Internet]. 1999 Mar [cited 2016 Feb 1];94(3):804–10. Available from: http://www.sciencedirect.com/science/article/pii/S0002927098007357

314. Kawahara H1, Okuyama H, Kubota A, Oue T, Tazuke Y, Yagi M OA. Can laparoscopic antireflux surgery improve the quality of life in children with neurologic and neuromuscular handicaps? - PubMed - NCBI. J Pediatr Surg [Internet]. 2004;39(12):1761–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15616922

315. Kawahara H, Nakajima K, Yagi M, Okuyama H, Kubota A, Okada A. Mechanisms responsible for recurrent gastroesophageal reflux in neurologically impaired children who underwent laparoscopic Nissen fundoplication. Surg Endosc [Internet]. 2002 May [cited 2016 Aug 11];16(5):767–71. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11997818

316. Kawahara H1, Imura K, Nakajima K, Yagi M, Kamata S OA. Motor function of the esophagus and the lower esophageal sphincter in children who undergo laparoscopic nissen fundoplication. - PubMed - NCBI. J Pediatr Surg [Internet]. 2000;35(11):166–71. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11083449

317. Farhath S, Aghai ZH, Nakhla T, Saslow J, He Z, Soundar S, et al. Pepsin, a reliable marker of gastric aspiration, is frequently detected in tracheal aspirates from premature ventilated neonates: relationship with feeding and methylxanthine therapy. J Pediatr Gastroenterol Nutr [Internet]. 2006 Sep [cited 2015 Sep 22];43(3):336–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16954956

318. Garland JS, Alex CP, Johnston N, Yan JC, Werlin SL. Association between tracheal pepsin, a reliable marker of gastric aspiration, and head of bed elevation among ventilated neonates. J Neonatal Perinatal Med [Internet]. 2014 Jan 1 [cited 2015 Sep 22];7(3):185–92. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25318633

319. da Silva O, Stevens D. Complications of airway management in very-low-birth-weight infants. Biol Neonate [Internet]. 1999 Jan [cited 2015 Sep 5];75(1):40–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9831682

320. Johnson S, Evans TA, Draper ES, Field DJ, Manktelow BN, Marlow N, et al. Neurodevelopmental outcomes following late and moderate prematurity: a population-based cohort study. Arch Dis Child Fetal Neonatal Ed [Internet]. 2015 Jul [cited 2015 Jul 27];100(4):F301-8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4484499&tool=pmcentrez&rendertype=abstract

321. Van Meurs KP, Robbins ST, Reed VL, Karr SS, Wagner AE, Glass P, et al. Congenital diaphragmatic hernia: Long-term outcome in neonates treated with extracorporeal membrane oxygenation. J Pediatr [Internet]. 1993 Jun [cited 2016 Mar 14];122(6):893–9. Available from: http://www.sciencedirect.com/science/article/pii/S0022347609900130

322. Kieffer J, Sapin E, Berg A, Beaudoin S, Bargy F, Helardot P. Gastroesophageal reflux after repair of congenital diaphragmatic hernia. J Pediatr Surg [Internet]. 1995 Sep [cited 2016 Mar 14];30(9):1330–3. Available from: http://www.sciencedirect.com/science/article/pii/0022346895904972

323. Sigalet DL, Nguyen LT, Adolph V, Laberge J-M, Hong AR, Guttman FM. Gastroesophageal reflux associated with large diaphragmatic hernias. J Pediatr Surg [Internet]. 1994 Sep [cited 2016 Mar 14];29(9):1262–5. Available from: http://www.sciencedirect.com/science/article/pii/0022346894908192

324. Pederiva F, Martinez L, Tovar JA. Enteric nervous system and esophageal-gastrointestinal motility in experimental congenital diaphragmatic hernia. Eur J Pediatr Surg Off J Austrian Assoc Pediatr Surg . [et al] = Zeitschrift für Kinderchirurgie [Internet]. 2014 Apr 17 [cited 2016 Mar 14];24(2):141–9. Available from: https://www-thieme-connect-de.libproxy.ucl.ac.uk/products/ejournals/html/10.1055/s-0033-1345106

325. Pederiva F, Rodriguez JI, Ruiz-Bravo E, Martinez L, Tovar JA. Abnormal intrinsic esophageal innervation in congenital diaphragmatic hernia: a likely cause of motor dysfunction. J Pediatr Surg [Internet]. 2009 Mar [cited 2016 Mar 14];44(3):496–9. Available from: http://www.sciencedirect.com/science/article/pii/S0022346808007082

326. Chrispin AR, Friedland GW, Waterston DJ. Aspiration pneumonia and dysphagia after technically successful repair of oesophageal atresia. Thorax [Internet]. 1966 Mar [cited 2016 Aug 10];21(2):104–10. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/5935835

327. Peetsold MG, Heij HA, Nagelkerke AF, Deurloo JA, Gemke RJBJ. Pulmonary function impairment after trachea-esophageal fistula: a minor role for gastro-esophageal reflux disease. Pediatr Pulmonol [Internet]. 2011 Apr [cited 2016 Aug 10];46(4):348–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20967841

328. Rosen R1, Hu L2, Amirault J3, Khatwa U4, Ward DV5 OA. 16S community profiling identifies proton pump inhibitor related differences in gastric, lung, and oropharyngeal microflora. - PubMed - NCBI. J Pediatr [Internet]. 2015;166(4):917–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25661411

329. Varal IG, Köksal N, Ozkan H, Doğan P, Bağcı O, Doğruyol H, et al. Very Low Birth Weight Infant Necessitating Nissen Fundoplication for Weaning off the Mechanical Ventilator. APSP J Case Rep [Internet]. 2014 May [cited 2016 Mar 17];5(2):15. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4090813&tool=pmcentrez&rendertype=abstract

330. Duncan DR, Amirault J, Johnston N, Mitchell P, Larson K, Rosen RL. Gastroesophageal Reflux Burden, Even in Children That Aspirate, Does Not Increase Pediatric Hospitalization. J Pediatr Gastroenterol Nutr [Internet]. 2016 Aug [cited 2016 Aug 10];63(2):210–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26794490

331. Djeddi D-D, Kongolo G, Stéphan-Blanchard E, Ammari M, Léké A, Delanaud S, et al. Involvement of autonomic nervous activity changes in gastroesophageal reflux in neonates during sleep and wakefulness. PLoS One [Internet]. 2013 Jan [cited 2015 Sep 22];8(12):e83464. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3862728&tool=pmcentrez&rendertype=abstract

332. Havemann BD, Henderson CA, El-Serag HB. The association between gastro-oesophageal reflux disease and asthma: a systematic review. Gut [Internet]. 2007 Dec [cited 2015 Sep 5];56(12):1654–64. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2095717&tool=pmcentrez&rendertype=abstract

333. Greifer M, Ng K, Levine J. Impedance and extraesophageal manifestations of reflux in pediatrics. Laryngoscope [Internet]. 2012 Jun [cited 2015 Aug 27];122(6):1397–400. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22447689

334. Drucker DE1, Michna BA, Krummel TM, Hunt PK, Hartenberg MA SA. Transthoracic Nissen fundoplication for gastroesophageal reflux in patients with severe kypho-roto-scoliosis. - PubMed - NCBI. J Pediatr Surg [Internet]. 1989;24(1):46–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2723993

335. van der Zee DC, Bax NM. Laparoscopic Thal fundoplication in severely scoliotic children. Surg Endosc [Internet]. 1995 Nov [cited 2016 Aug 10];9(11):1197–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8553233

336. Tatekawa Y1, Kanehiro H NY. Laparoscopic modified Thal fundoplication for gastroesophageal reflux in a patient with severe scoliosis and sliding esophageal hiatal hernia. - PubMed - NCBI. J Pediatr Surg [Internet]. 2006;41(10):E15-18. Available from: http://www.ncbi.nlm.nih.gov/pubmed/?term=tatekawa+laparoscopic+thal

337. Ledermann SE, Spitz L, Moloney J, Rees L, Trompeter RS. Gastrostomy feeding in infants and children on peritoneal dialysis. Pediatr Nephrol [Internet]. 2002 Apr [cited 2016 Aug 10];17(4):246–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11956875

338. Mekahli D, Shaw V, Ledermann SE, Rees L. Long-term outcome of infants with severe chronic kidney disease. Clin J Am Soc Nephrol [Internet]. 2010 Jan [cited 2016 Aug 10];5(1):10–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19965538

339. Humphrey GM, Najmaldin A. Laparoscopic gastrostomy in children. Pediatr Surg Int [Internet]. 1997 Sep [cited 2016 Aug 10];12(7):501–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9238116

340. Coleman JE, Watson AR, Rance CH, Moore E. Gastrostomy buttons for nutritional support on chronic dialysis. Nephrol Dial Transplant [Internet]. 1998 Aug [cited 2016 Nov 21];13(8):2041–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9719162

341. Luzzani S, Macchini F, Valadè A, Milani D, Selicorni A. Gastroesophageal reflux and Cornelia de Lange syndrome: typical and atypical symptoms. Am J Med Genet A [Internet]. 2003 Jun 15 [cited 2015 Sep 16];119A(3):283–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12784293

342. Dobbelsteyn C, Peacocke SD, Blake K, Crist W, Rashid M. Feeding difficulties in children with CHARGE syndrome: prevalence, risk factors, and prognosis. Dysphagia [Internet]. 2008 Jun [cited 2015 Sep 16];23(2):127–35. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18027028

343. Maggadottir SM, Sullivan KE. The diverse clinical features of chromosome 22q11.2 deletion syndrome (DiGeorge syndrome). J allergy Clin Immunol Pract [Internet]. Jan [cited 2015 Sep 21];1(6):589–94. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24565705

344. Gulack BCH, Adibe OO. Laparoscopic antireflux surgery in infants with single ventricle physiology: a review. J Laparoendosc Adv Surg Tech A [Internet]. 2013 Aug [cited 2015 Sep 22];23(8):733–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23859741

345. Macharia E, Eaton S, DeCoppi P, Curry J, Kiely E, Yates R, et al. Fundoplication improves growth failure in patients with single ventricle cardiac anomalies. In: British Association of Paediatric Surgery- European Association of Paediatric Surgery Joint Conference. 2012.

346. Mercado-Deane MG, Burton EM, Harlow SA, Glover AS, Deane DA, Guill MF, et al. Swallowing dysfunction in infants less than 1 year of age. Pediatr Radiol [Internet]. 2001 Jun [cited 2015 Sep 22];31(6):423–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11436889

347. Ludemann JP, Hughes CA, Noah Z, Holinger LD. Complications of pediatric laryngotracheal reconstruction: prevention strategies. Ann Otol Rhinol Laryngol [Internet]. 1999 Nov [cited 2015 Sep 22];108(11 Pt 1):1019–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10579227

348. Faner R, Gutiérrez-Sacristán A, Castro-Acosta A, Grosdidier S, Gan W, Sánchez-Mayor M, et al. Molecular and clinical diseasome of comorbidities in exacerbated COPD patients. Eur Respir J [Internet]. 2015 Aug 6 [cited 2015 Sep 15]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/26250499

349. Samuel M, Holmes K. Quantitative and qualitative analysis of gastroesophageal reflux after percutaneous endoscopic gastrostomy. J Pediatr Surg [Internet]. 2002 Feb [cited 2016 Aug 2];37(2):256–61. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022346802914703

350. Isch JA, Rescorla FJ, Tres Scherer L., West KW, Grosfeld JL. The development of gastroesophageal reflux after percutaneous endoscopic gastrostomy. J Pediatr Surg [Internet]. 1997 Feb [cited 2016 Aug 2];32(2):321–3. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022346897902025

351. Toporowska-Kowalska E, Gębora-Kowalska B, Fendler W, Popińska K, Szlagatys-Sidorkiewicz A, Grzybowska-Chlebowczyk U, et al. Diagnosis of gastroesophageal reflux and anti-reflux procedures among Polish children with gastrostomies: a 10-year nationwide analysis. Eur J Clin Nutr [Internet]. 2013 Nov [cited 2016 Aug 2];67(11):1169–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24065062

352. Catto-Smith AG, Jimenez S. Morbidity and mortality after percutaneous endoscopic gastrostomy in children with neurological disability. J Gastroenterol Hepatol [Internet]. 2006 Apr [cited 2015 Aug 27];21(4):734–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16677161

353. Kimber C, Beasley S. Limitations of percutaneous endoscopic gastrostomy in facilitating enteral nutrition in children: Review of the shortcomings of a new technique. J Paediatr Child Health [Internet]. 1999 Oct 5 [cited 2016 Aug 2];35(5):427–31. Available from: http://doi.wiley.com/10.1046/j.1440-1754.1999.355418.x

354. Lee SL1, Shabatian H, Hsu JW, Applebaum H HP. Hospital admissions for respiratory symptoms and failure to thrive before and after Nissen fundoplication. J Pediatr Surg. 2008;43(1):63–5.

355. Hebson CL1, Oster ME, Kirshbom PM, Clabby ML, Wulkan ML SJ. Association of feeding modality with interstage mortality after single-ventricle palliation. J Thorac Cardiovasc Surg [Internet]. 2012;144(1):173–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22244571

356. Hutton JL, Colver AF, Mackie PC. Effect of severity of disability on survival in north east England cerebral palsy cohort. Arch Dis Child [Internet]. 2000 Dec [cited 2016 Aug 2];83(6):468–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11087278

357. Strauss D, Shavelle R. Life expectancy in cerebral palsy. Arch Dis Child [Internet]. 2001 Nov [cited 2016 Aug 2];85(5):442. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11676432

358. Brooks JC, Strauss DJ, Shavelle RM, Tran LM, Rosenbloom L, Wu YW. Recent trends in cerebral palsy survival. Part I: period and cohort effects. Dev Med Child Neurol [Internet]. 2014 Nov [cited 2016 Aug 2];56(11):1059–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24966011

359. Sleigh G, Brocklehurst P. Gastrostomy feeding in cerebral palsy: a systematic review. Arch Dis Child [Internet]. 2004 Jun [cited 2016 Aug 3];89(6):534–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15155398

360. Knott EM, Fraser JD, Alemayahu H, Drake WB, St Peter SD, Perryman KM, et al. Evaluation of cardiac function during laparoscopic gastrostomy in pediatric patients with hypoplastic left heart syndrome using intraoperative transesophageal echocardiography. J Laparoendosc Adv Surg Tech A [Internet]. 2014 Oct [cited 2016 Mar 10];24(10):738–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25295635

361. Davies RR, Carver SW, Schmidt R, Keskeny H, Hoch J, Pizarro C. Gastrointestinal complications after stage I Norwood versus hybrid procedures. Ann Thorac Surg [Internet]. 2013 Jan [cited 2016 Mar 10];95(1):189-95; discussion 195-6. Available from: http://www.sciencedirect.com/science/article/pii/S0003497512014518

362. Goldin AB, Sawin R, Seidel KD, Flum DR. Do antireflux operations decrease the rate of reflux-related hospitalizations in children? Pediatrics [Internet]. 2006 Dec [cited 2016 Aug 1];118(6):2326–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17142515

363. Freedman B. Equipoise and the Ethics of Clinical Research. N Engl J Med [Internet]. 1987 Jul 16 [cited 2016 Nov 16];317(3):141–5. Available from: http://www.nejm.org/doi/abs/10.1056/NEJM198707163170304

364. Chard JA, Lilford RJ. The use of equipoise in clinical trials. Soc Sci Med [Internet]. 1998 Oct [cited 2016 Nov 17];47(7):891–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9722109

365. Vakili C. A previously nondescribed dominant branch from the left splanchnic nerve innervating the gastroesophageal junction: the significance of its preservation during fundoplication for gastroesophageal reflux disease warrants clarification. Surg Endosc [Internet]. 2009 Dec 29 [cited 2016 Nov 17];23(12):2862–3. Available from: http://link.springer.com/10.1007/s00464-009-0524-1

366. Shin MS, Shim JO, Moon JS, Kim HS, Ko JS, Choi J-H, et al. Impedance-pH monitoring and conventional pH monitoring are complementary methods to detect association between gastroesophageal reflux and apnea-related symptoms in preterm infants and neonates. J Matern Fetal Neonatal Med [Internet]. 2012 Nov [cited 2016 Jul 31];25(11):2406–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22694343

367. Malcolm WF, Gantz M, Martin RJ, Goldstein RF, Goldberg RN, Cotten CM, et al. Use of medications for gastroesophageal reflux at discharge among extremely low birth weight infants. Pediatrics [Internet]. 2008 Jan [cited 2016 Jul 31];121(1):22–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18166553

368. Guillet R, Stoll BJ, Cotten CM, Gantz M, McDonald S, Poole WK, et al. Association of H2-blocker therapy and higher incidence of necrotizing enterocolitis in very low birth weight infants. Pediatrics [Internet]. 2006 Feb [cited 2016 Jul 31];117(2):e137-42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16390920

369. Bilali A1, Galanis P, Bartsocas C, Sparos L VE. H2-blocker therapy and incidence of necrotizing enterocolitis in preterm infants: a case-control study. Pediatr Neonatol. 2013;54(2):141–2.

370. Beck-Sague CM, Azimi P, Fonseca SN, Baltimore RS, Powell DA, Bland LA, et al. Bloodstream infections in neonatal intensive care unit patients: results of a multicenter study. Pediatr Infect Dis J [Internet]. 1994 Dec [cited 2016 Jul 31];13(12):1110–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7892080

371. Bianconi S, Gudavalli M, Sutija VG, Lopez AL, Barillas-Arias L, Ron N. Ranitidine and late-onset sepsis in the neonatal intensive care unit. J Perinat Med [Internet]. 2007 [cited 2016 Jul 31];35(2):147–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17302510

372. Gupta RW, Tran L, Norori J, Ferris MJ, Eren AM, Taylor CM, et al. Histamine-2 Receptor Blockers Alter the Fecal Microbiota in Premature Infants. J Pediatr Gastroenterol Nutr [Internet]. 2013 Apr [cited 2016 Jul 31];56(4):397–400. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00005176-201304000-00011

373. Chandrasekaran M, Fleming P. Question 1: does the use of ranitidine increase the risk of NEC in preterm infants? Arch Dis Child [Internet]. 2014 Apr 1 [cited 2016 Feb 22];99(4):390–2. Available from: http://adc.bmj.com/content/99/4/390.1.full

374. Goldin AB, Garrison M, Christakis D. Variations between hospitals in antireflux procedures in children. Arch Pediatr Adolesc Med [Internet]. 2009 Jul [cited 2016 Aug 1];163(7):658–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19581550

375. Bruns NE1, Shah MA2, Dorsey AN1, Ponsky TA1 SO. Pediatric surgery - a changing field: national trends in pediatric

surgical practice. - PubMed - NCBI. J Pediatr Surg [Internet]. 2016;51(6):1034–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26987709

376. Gundeti M. Abdominal Procedures- Fundoplication. In: Pediatric Robotic and Reconstructive Urology: A Comprehensive Guide [Internet]. Wiley; 2011. p. 204. Available from: https://books.google.com.au/books?id=YeYKcVqrUMkC&pg=PA204&lpg=PA204&dq=fewer+fundoplications&source=bl&ots=yPGu6zke4p&sig=XwXdlL2b43om_K6OpP4mnWXoHLI&hl=en&sa=X&ved=0ahUKEwiHtOGAgJzOAhWIm5QKHSvxC9MQ6AEINDAD#v=onepage&q=fewer fundoplications&f=false

377. Fortunato JE1 CC. Outcomes of percutaneous endoscopic gastrostomy in children. Curr Gastroenterol Rep. 2011;13(3):293–9.

378. Kubiak R, Andrews J, Grant HW. Laparoscopic Nissen Fundoplication Versus Thal Fundoplication in Children: Comparison of Short-Term Outcomes. J Laparoendosc Adv Surg Tech [Internet]. 2010 Sep [cited 2018 May 18];20(7):665–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20822421

379. Livingston MH, Shawyer AC, Rosenbaum PL, Jones SA, Walton JM. Fundoplication and gastrostomy versus percutaneous gastrojejunostomy for gastroesophageal reflux in children with neurologic impairment: A systematic review and meta-analysis. J Pediatr Surg [Internet]. 2015 May [cited 2016 Aug 5];50(5):707–14. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022346815001001

380. Slavin JL. Position of the American Dietetic Association: health implications of dietary fiber. J Am Diet Assoc [Internet]. 2008 Oct [cited 2016 Nov 18];108(10):1716–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18953766

381. Pentiuk S, O'Flaherty T, Santoro K, Willging P, Kaul A. Pureed by Gastrostomy Tube Diet Improves Gagging and Retching in Children With Fundoplication. J Parenter Enter Nutr [Internet]. 2011 May 1 [cited 2016 Nov 18];35(3):375–9. Available from: http://pen.sagepub.com/cgi/doi/10.1177/0148607110377797

382. O'Gorman EA. Complete tubefeeding : everything you need to know about tubefeeding, tube nutrition and blended diets. The Author; 2012. 222 p.

383. Barnhart DC, Berry JG, Srivastava R. Does Nissen fundoplication decrease reflux-related problems including respiratory illness in children with neurological impairment?--reply. JAMA Pediatr [Internet]. 2014 Feb 1 [cited 2016 Feb 19];168(2):188–9. Available from: http://archpedi.jamanetwork.com.libproxy.ucl.ac.uk/article.aspx?articleid=1819104

384. Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for communicating results of medical investigations. Cochrane database Syst Rev [Internet]. 2012 [cited 2016 Aug 8];(6):CD007456. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22696369

385. Smith C, Gold J, Ngo TD, Sumpter C, Free C. Mobile phone-based interventions for improving contraception use. Cochrane database Syst Rev [Internet]. 2015 [cited 2016 Aug 8];(6):CD011159. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26115146

386. Semper HM, Povey R, Clark-Carter D. A systematic review of the effectiveness of smartphone applications that encourage dietary self-regulatory strategies for weight loss in overweight and obese adults. Obes Rev [Internet]. 2016 [cited 2016 Aug 8]; Available from: http://doi.wiley.com/10.1111/obr.12428

387. Phillips KA, Epstein DH, Mezghanni M, Vahabzadeh M, Reamer D, Agage D, et al. Smartphone Delivery of Mobile HIV Risk Reduction Education. AIDS Res Treat [Internet]. 2013 [cited 2016 Aug 8];2013:231956. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24159383

388. Jenny JY1. Measurement of the knee flexion angle with a Smartphone-application is precise and accurate. - PubMed - NCBI. J Arthroplast [Internet]. 2013;28(5):784–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23499407

389. Moves - Activity Diary for iPhone and Android [Internet]. Available from: https://www.moves-app.com/

390. Kwapisz JR, Weiss GM, Moore SA. Activity Recognition using Cell Phone Accelerometers.

391. Behar J, Roebuck A, Domingos JS, Gederi E, Clifford GD. A review of current sleep screening applications for smartphones. Physiol Meas [Internet]. 2013 Jul [cited 2016 Aug 2];34(7):R29-46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23771214

392. Wu W, Dasgupta S, Ramirez EE, Peterson C, Norman GJ. Classification Accuracies of Physical Activities Using Smartphone Motion Sensors. J Med Internet Res [Internet]. 2012 Oct 5 [cited 2016 Aug 11];14(5):e130. Available from:

http://www.jmir.org/2012/5/e130/

393. Norgall T, Schmidt R, von der Grün T. Body area network--a key infrastructure element for patient-centered telemedicine. Stud Health Technol Inform [Internet]. 2004 [cited 2016 Aug 11];108:142–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15718640

394. Houston TK, Cooper LA, Ford DE. Internet Support Groups for Depression: A 1-Year Prospective Cohort Study. Am J Psychiatry [Internet]. 2002 Dec [cited 2016 Nov 18];159(12):2062–8. Available from: http://psychiatryonline.org/doi/abs/10.1176/appi.ajp.159.12.2062

395. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ [Internet]. 2004 May 15 [cited 2016 Aug 11];328(7449):1166. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15142921

396. Smart NJ. A survey of smartphone and tablet computer use by colorectal surgeons in the UK and Continental Europe. Colorectal Dis [Internet]. 2012 Sep [cited 2016 Aug 11];14(9):e535-8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22747977

397. Data Protection Act 1998.

398. NHS Code of Practice. 2003;

399. Warnes GR. gmodels: Various R programming tools for model fitting [Internet]. 2013. Available from: http://cran.r-project.org/package=gmodels

400. Kuhn M, Team C from JW and SW and AW and CK and AE and TC and ZM and the RC. caret: Classification and Regression Training [Internet]. 2014. Available from: http://cran.r-project.org/package=caret

401. Konis K, Konis MK. Package “safeBinaryRegression” Title Safe Binary Regression. 2015 [cited 2018 Mar 2]; Available from: https://cran.r-project.org/web/packages/safeBinaryRegression/safeBinaryRegression.pdf

402. Daniel E. Ho, Kosuke Imai, Gary King EAS. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. J Stat Softw [Internet]. 2011 [cited 2017 Oct 27];42(8):1–28. Available from: https://cran.r-project.org/web/packages/MatchIt/citation.html

403. Therneau TM, Atkinson EJ, Foundation M. An Introduction to Recursive Partitioning Using the RPART Routines. 2017 [cited 2017 Oct 27]; Available from: https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf

404. Milborrow S. Plotting rpart trees with the rpart.plot package. 2016 [cited 2017 Oct 27]; Available from: http://www.milbo.org/rpart-plot/prp.pdf

405. Package “party” Title A Laboratory for Recursive Partytioning. 2017 [cited 2017 Oct 27]; Available from: https://cran.r-project.org/web/packages/party/party.pdf

406. Zeileis A, Hothorn T. partykit: A Toolkit for Recursive Partytioning. [cited 2017 Oct 27]; Available from: https://cran.r-project.org/web/packages/partykit/vignettes/partykit.pdf

COMPARISON OF UPPER GASTROINTESTINAL CONTRAST STUDIES VERSUS PH IMPEDANCE

Pediatr Radiol (2012) 42:946–951

DOI 10.1007/s00247-012-2405-3

ORIGINAL ARTICLE

Comparison of upper gastrointestinal contrast studies and pH/impedance tests for the diagnosis of childhood gastro-oesophageal reflux

Eva W. Macharia

Received: 9 October 2011 /Revised: 9 January 2012 /Accepted: 12 February 2012 /Published online: 4 August 2012

Abstract

Background The upper gastrointestinal (UGI) contrast study is used in the assessment of children with gastro-oesophageal reflux (GOR) and for detection of structural anomalies. The pH study is more sensitive than the UGI study for the diagnosis of GOR. The pH study has been replaced by the pH/ impedance test, which detects both acid and nonacid reflux. Objective To compare the UGI contrast study with the pH/ impedance test for the diagnosis of GOR in children. Materials and methods We retrospectively reviewed consecutive records of children investigated for GOR from October 2008 to February 2010, and compared the findings of UGI studies with those of pH/impedance tests.

Results The UGI studies revealed GOR in 116 of 579 children (20%). Of the children undergoing a UGI study, 66 also underwent a pH/impedance test. Using the pH/impedance tests as the reference for GOR, UGI had a sensitivity of 42.8% and a negative predictive value of 24%. There was no significant correlation ($P>0.05$) between the reflux index and the number of reflux episodes in the pH/impedance tests and height of reflux in the UGI study. There were low incidences of malrotation (0.9%), hiatus hernia (1%) and delayed gastric emptying (0.4%).

Conclusion The UGI study had low sensitivity for the diagnosis of GOR and low yield for the diagnosis of structural anomalies.

Keywords Gastro-oesophageal reflux . Upper gastrointestinal contrast study . pH/impedance . Malrotation

FUNDOPLICATION IN VENTILATOR DEPENDENT INFANTS WITH GORD

Macharia, E. Eaton, S. De Coppi P, Pierro A. "Fundoplication in ventilator-dependent infants with gastro-esophageal reflux". European Journal of Pediatric Surgery -2012; 22: 91-96.

Abstract

Aim: In ventilator-dependent infants and complex co-morbidities, severe gastro-oesophageal reflux (GOR) may lead to recurrent chest infections and acute life threatening events, therefore contributing to ventilator dependence. It is often difficult to predict whether anti-reflux surgery will improve the respiratory status of an infant and allow weaning off the ventilator. The aim of this study is to assess the clinical outcomes in a cohort of ventilator-dependent infants who underwent fundoplication to help wean them off ventilation.

Methods: Between January 2006 and December 2010, 26 ventilator-dependent infants underwent a fundoplication for symptoms of GOR . 7 infants had suffered acute life threatening events (ALTE). Three patients had episdoes of aspiration. The rest of the cohort (n=16) suffered an acute or chronic deterioration in respiratory status. The median age at surgery was 5.8 months (range 0.8-19.4 months). The median weight at surgery was 6.3kg (range 4-15.1 kg. For each infant, data were collected on significant co-morbidities, pre- and post-operative ventilation status, pre- and post-operative GOR symptoms and survival.

Results: All infants underwent a Nissen fundoplication with no intra-operative morbidity or mortality. 12 infants had a laparoscopic fundoplication. 14 infants had an open fundoplication. Post-operatively, all infants received invasive positive pressure ventilation on the intensive care unit. All infants were successfully weaned from ventilation. The median time to extubation was in 4 days (range 2-18 days). The median postoperative ICU stay was 9 days (range 3-52 days).

Ten patients (38%) were deceased within the study capture period. In patient Z, mortality was directly related to operative morbidity (missed trachea-oesophageal fistula). Patient Y died following several episodes of pneumonia following fundoplication. In this patient, a post-operative upper GI contrast study confirmed an intact wrap without herniation. In the other 8 patients, mortality was related to underlying their serious underlying co-morbidities and not directly related to recurrence of reflux symptoms.

Conclusion: In infants with severe3 GOR and signiciant co-morbidities, fundoplication may be a usefulk procedure to assist ventilator dependence.

CASE STUDY DESCRIBING THE PILLARS, PERSONNEL AND PROCESS OF DEVELOPING THE TARDIS:REFLUX SMARTPHONE APP

THELANCET-D-13-04906R1

Title: Case study describing the pillars, personnel and process of developing the TARDIS:REFLUX smartphone app.

Article Type: Conference Abstracts

Authors: E W Macharia, BA, MA Oxon, MbBChir Cantab, MRCS; Richard Groves, BSc. MSc. ; Anthony Coates, BSc Hons (1); Joseph I Curry, MBBS, FRCS.; Rashmi R Singh, MBBS, MRCS; Tim J Cole, BA, BPhil MA Cantab, PHD DSC Manuscript Region of Origin: UNITED KINGDOM

Abstract: Background Many smartphone apps for healthcare are commercially available to track conditions such as asthma, diabetes, etc. Regulations governing data protection for NHS patients and limited expertise within the healthcare professions restrict development of apps as clinical and research adjuncts. Our aim was to demonstrate the feasibility of, and define the roadmap for developing apps as clinical and research tools within an NHS environment. For this case study we focused on gastroeosophageal reflux (GOR) in children, as this is a relatively common condition. Symptoms occur daily yet are reviewed infrequently by clinicians, often at 3- monthly intervals. Where questionnaires are used, they are applied retrospectively, require interval recall and are delivered on paper. Epidemiological understanding of GOR is limited by the paucity of symptom data. To address this challenge we developed an iPhone app to Track Activity in Relation to reflux DISease(TARDIS: REFLUX).

Methods We developed a questionnaire for parents of GOR patients to prospectively capture symptom frequency and key events. A project proposal describing data collection requirements and security arrangements drove project specification. Funding, ethical and institutional Caldicott Guardian approvals were obtained. Following tender, app developers were identified and the app built. A software application was designed to transfer data from users' phones to a central, secure database. A web interface for reviewing transmitted data was designed.

Findings The four pillars of this app are a data collector, a data transfer protocol, a central database and a data viewer. The roadmap for healthcare app development in an NHS environment is characterised as follows:

* Data specification: clinical data collection requirements defined (task duration: 2 weeks)

* Wire-framing the data collector: Mock-up diagrams used to translate the data collection questionnaire into app views (2 weeks)

* User interface design: data collector and native data store coded (4 months)

* User management: user registration, login protocols and level of access coded (1 month)

* Data security and ethical approval: Systems Level Security Protocol and IRAS application to obtain NHS REC approval (NREC 12/NW/0837) (4 months)

* Data management, transfer and storage: an application program interface(API) to provide data transfer from users' iPhones to the central database (3 months)

* Data viewer: online dashboard enabling secure access and review of data in central database.

The app was subjected to beta-testing through a 2-step iterative process involving app target users (3 months).

* A randomised cross-over study, in which users compare a paper questionnaire against the TARDIS:REFLUX app, is on-going.

* A focus group (n=8), was convened to test usability. Seven participants successfully downloaded the app with no crashes during the session. Users required 6.8±2.8 seconds to login and 6.9±3.0 seconds to record a symptom. The app received a "good" mean usability rating of 84 out of 100 (Systems Usability Scale®).

Interpretation The TARDIS: REFLUX app illustrates feasibility and provides a roadmap for developing apps as clinical and research adjuncts. It allows real-time remote tracking of GOR symptoms in children to enhance clinical care. The data architecture is robust and secure, and meets NHS data protection standards. Developing software for use on patient-held smartphones reduces the cost of development and allows large-scale use.

IMPACT OF COMBINED MULTI-CHANNEL INTRA-LUMINAL IMPEDANCE STUDY ON DIAGNOSIS OF GASTRO-OESOPHAGEAL REFLUX DISEASE

Purpose: Combined multi-channel intra-luminal impedance and pH (CIP) monitoring is increasingly being used to diagnose gastro-oesophageal reflux (GOR) in the paediatric population. It may replace 24-hour pH recording as the investigation of choice prior to fundoplication. Our aim is to evaluate diagnostic sensitivity of each modality and the potential impact on rates of fundoplication.

Methods: We reviewed records of 121 patients with GOR symptoms assessed with 24-hour CIP studies in our centre (September 2008 - February 2010). Nasogastric pH/impedance probes were positioned at the level of thoracic vertebrae 8-10 (~3 cm above gastro-oesophageal junction). Position was confirmed on chest X-Ray. The diagnosis of GOR was based on quantitative and qualitative analysis of CIP study. Reflux index (percentage of time pH <4) was extracted from the oesophageal acid exposure recording during CIP study. Data were analysed, descriptive statistics and bi-variate correlations generated using Microsoft Excel and SPSS.

Results: Mean recording time was 16.7 ± 7.6 hours. On the basis of CIP study, GOR was diagnosed in 86 children (71%). Mean age of patients with GOR (6.01 ± 5.36 years) was similar to those without GOR (5.4 ± 4.6 years, p= n.s.). Diagnostic sensitivity of CIP versus a pH recording only was evaluated (Table 1,2).

Table 157: Diagnosis of GOR based on pH recording versus combined pH/impedance study (CIP)

pH recording (reflux index)		CIP study	
		GOR	No GOR
No GOR (<5%)	n=86	43	33
Mild GOR (5-10%)	n=24	23	1
Severe GOR (>10%)	n=21	20	1
Total		86	35

Table 158: Predictive value of pH recording compared to combined pH/impedance study

Correlation between significant CIP parameters and a positive CIP test was analysed. Oesophageal exposure time (r^2 = 0.28, p=0.002), reflux episodes over 5 minutes (r^2= 0.3, p=0.001), total reflux episodes (r^2= 0.39, p=0.000 and Boix-Ochoa Composite score (r^2= 0.3, p=0.003) were significantly correlated to a CIP diagnosis of GOR. Median acid /bolus clearance time and reflux time were not significantly with CIP diagnosis of GOR. Twelve (14%) of 86 patients had fundoplication following CIP study; only 7 (58%) were positively diagnosed on pH recording.

Conclusion: Our study indicates that GOR is diagnosed more frequently on CIP study compared to pH only study. This difference in diagnostic sensitivity may be accounted for by non-acid GOR events. This difference may potentially impact patient selection for fundoplication.

APPLE JUICE PH RECODING VERSUS PH IMPEDANCE FOR THE ASSESSMENT OF GASTRO-OESOPHAGEAL REFLUX IN INFANTS

Aim of study: In neonates and infants on frequent or continuous feeds, pH studies are an unreliable measure of gastro-oesophageal reflux. The apple juice pH recording (AJ), by creating an acid medium in the stomach, avoids the buffering effect of milk feeds. The pH/Impedance test (PI) is able to measure both acid and non-acid reflux.

Method: We reviewed all AJ and PI studies in infants (n=52) requested by paediatric surgeons. Data are median (range), and compared by Mann-Whitney test.

Main results: Between Jan 2002-Jun 2008, AJ was used in 32 infants; 20 infants had PI performed between Aug 2008-Dec 2010 (Table). Both groups had similar age and symptoms of gastro-oesophageal reflux.

Table 159: Patient characteristics and main results of investigations

	AJ (n=32)	**PI (n=20)**	**p**
Age at investigation (months)	5.4 (1.1-17.8)	5.9 (1.5-23.1)	0.3
Neurological impairment	18 (56%)	16/20 (80%)	0.4
Oesophageal Atresia	3 (9%)	1 (5%)	0.4
Congenital diaphragmatic hernia	1	0	0.5
Duration of study	23.5 hours	23.3 hours	0.4
Required fundoplication	21(65%)	11(55%)	0.1
Number of reflux episodes	183 (23-728)	84 (16-466)	0.014
Number of reflux episodes/hour	8.1 (0.95-31.9)	3.8 (0.8-20.3)	0.021
Number of reflux episodes >5min	7 (0-36)	1 (0-26)	0.021
Duration of longest reflux (min)	29(1.9-551)	8.5 (1.2 -172)	0.041
Reflux index (%)	13.9 (0.6-56)	6 (0.3-43)	0.014

Conclusion: The apple juice pH recording is associated with longer acid exposure and more reflux episodes per hour compared to PI. This may be an artefact due to the non-physiological environment generated by the apple juice. pH/impedance records both acid and non-acid reflux episodes in a physiological environment and is useful in infants on milk feeds.

FUNDOPLICATION IS EFFECTIVE IN WEANING INFANTS AND CHILDREN WITH REFLUX FROM VENTILATION AND ICU DEPENDENCY

Aim of study: to determine effectiveness of fundoplication in infants and children requiring pre-operative intensive care unit (ICU) support with ventilation. **Methods**: Retrospective review, following institutional approval, of Paediatric, Neonatal and Cardiac ICU patients admitted before or after fundoplication for gastro-oesophageal reflux (GOR). Data are median (range).

Main Results: Of 596 children undergoing fundoplication at our institution between 2006 and 2010, 101 (16.9%) required peri-operative ICU support. Excluding children who had previous fundoplication, 26 of these infants and children could not be weaned from the ventilator and ICU treatment before fundoplication. There were 7 children with neurological impairment (NI) plus chronic lung disease (CLD), 5 with NI alone, 8 with CLD alone, 4 with complex cardiac anomalies, 1 with oesophageal atresia and trachea-oesophageal fistula and 1 with tracheal malformation. 4 children (15.3%) had tracheotomies. Age at operation was 5.8 (0.8-19.4) months; weight was 6.3 (4-15.1) kg. The fundoplication was performed successfully laparoscopically in 12 (46%). Post-operatively, all patients were weaned from positive pressure support in 4 (2-18) days, and had a postoperative ICU stay of 9 (3-52) days. Symptoms recurred in 4 patients, of whom 3 (all with NI), required re-do fundoplication after 1.4 (1.1-1.6) years.

Conclusions: Nissen fundoplication helps weaning infants and children with GOR from the ventilator and can be performed laparoscopically independently of ventilator dependency, age and size of the child. Poorer outcomes are seen with children with neurological impairment.

DEFINING THE UTILITY OF THE UPPER GI CONTRAST STUDY IN ASSESSMENT OF GASTRO-OESOPHAGEAL REFLUX

Introduction: The upper gastro-intestinal contrast study (UGI) is used in assessment of gastro-oesophageal reflux (GOR) and detection of structural anomalies e.g. malrotation, hiatus hernia. For reflux detection, the pH study is more sensitive than the UGI. The pH study has been eclipsed by the combined pH/impedance (P-I) study, which detects both acid and non-acid reflux.

Aim: To compare the utility of the UGI study against the P-I study, and assess its utility in detection of structural anomalies.

Method: We reviewed records of children undergoing investigation for GOR (October 2008 - February 2010).

Results: In 579 patients undergoing UGI, GOR was detected in 116 (20%). The incidence of malrotation (0.9%), hiatus hernia (1%) and delayed gastric emptying (0.4%) was low. In 66 patients undergoing both UGI and P-I studies, the sensitivity of UGI was low (42.8%). The negative predictive value was 24%. There was no correlation between reflux index and reflux episodes on P-I, and height of reflux on UGI.

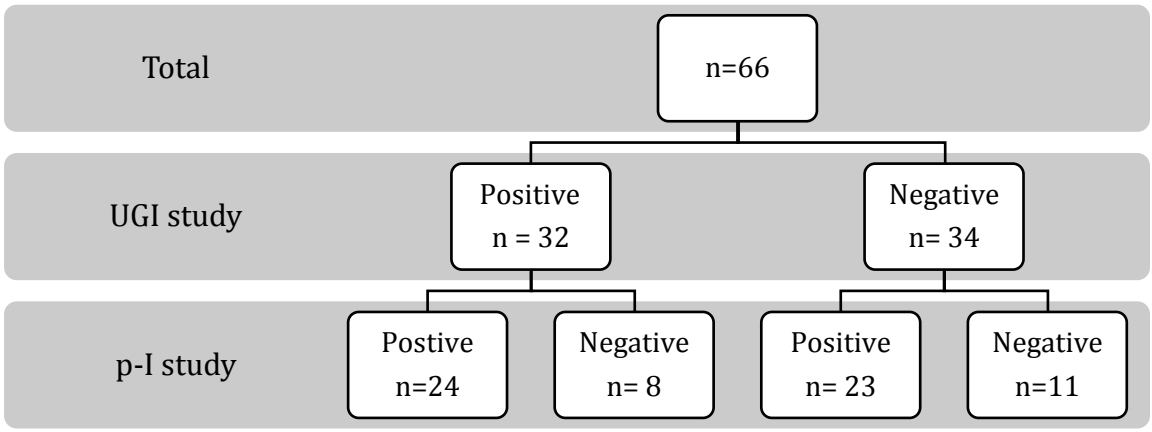


Figure 143: Comparing frequency of positive and negative diagnosis on UGI and pH/impedance study.

Conclusion: The UGI study is poorly sensitive for detection of GOR. It is useful in the assessment of malrotation and hiatus hernia but the incidence of these conditions is low. The utility of UGI for the assessment of GOR is questionable.

IMPACT OF FUNDOPLICATION ON GROWTH IN PATIENTS WITH SINGLE VENTRICLE CARDIAC ANOMALIES

Aim: Single ventricle cardiac anomalies (SVCA) e.g. hypoplastic left heart syndrome (HLHS), tricuspid and mitral atresia, is a complex cardiac anomaly. Patients often need multiple cardiac surgeries and prolonged ventilation in the neonatal period. Poor feeding and failure to thrive is a feature of SVCA. Thirty per cent of children with SVCA will have gastro-oesophageal reflux (GOR).

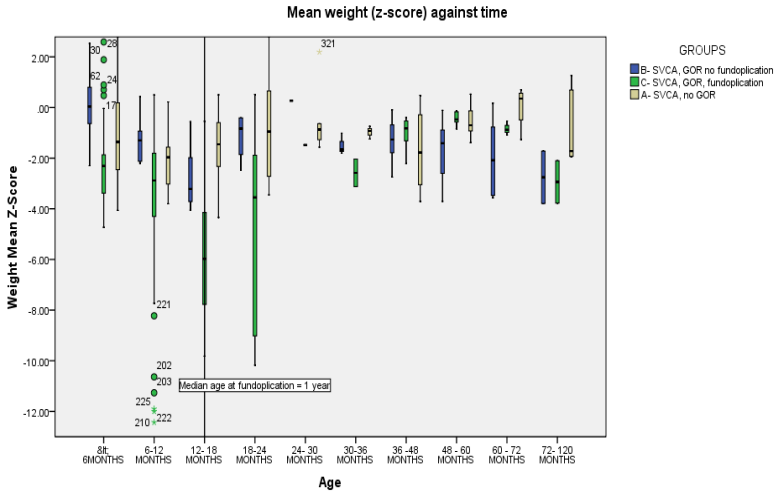
Methods: We retrospectively reviewed inpatient, clinic and operative records to identify patients managed for SVCA at our institution between January 2000 and December 2010. We reviewed patient notes for symptoms of reflux, operative details and serial weight measurements.

Results: We identified 47 patients who had Norwood stage I procedures for hypoplastic left heart over the past 10 years at our institution. Twenty patients had no GOR (Group A). Fourteen patients received medical treatment for GOR (Group B). Thirteen patients underwent Nissen fundoplication for GOR (Group C).

Table 160: Characteristics of patients who underwent Norwood procedures for hypoplastic left heart.

Mean ± standard deviation, *p<0.05	Group A	Group B	Group C
Number	20 (42%)	14 (30%)	13 (27%)
Gestation (weeks)*	38.94 ±3.65	36.17 ± 1.92	36.86 ± 0.818
Birth weight *(kilograms)	3.24 ± 0.53	3.05 ± 0.25	2.65 ± 0.52
Mean age at Norwood 1* (days)	5.21 ± 1.53	8.09 ± 4.75	9.45 ± 6.36
Median age at Norwood 1 (days)	6 (4-8)	6 (3-26)	7 (4-24)
Mortality	0	2	4
Transplant	2	0	1

Figure 144: Trend in weight z score for patients with a single ventricle cardiac anomaly



Conclusion: This study indicates that, in children with SVCA and GOR fundoplication results in catch-up growth.

SECTION IX: APPENDICES

SECTION I APPENDIX ITEMS

Summary of abbreviations

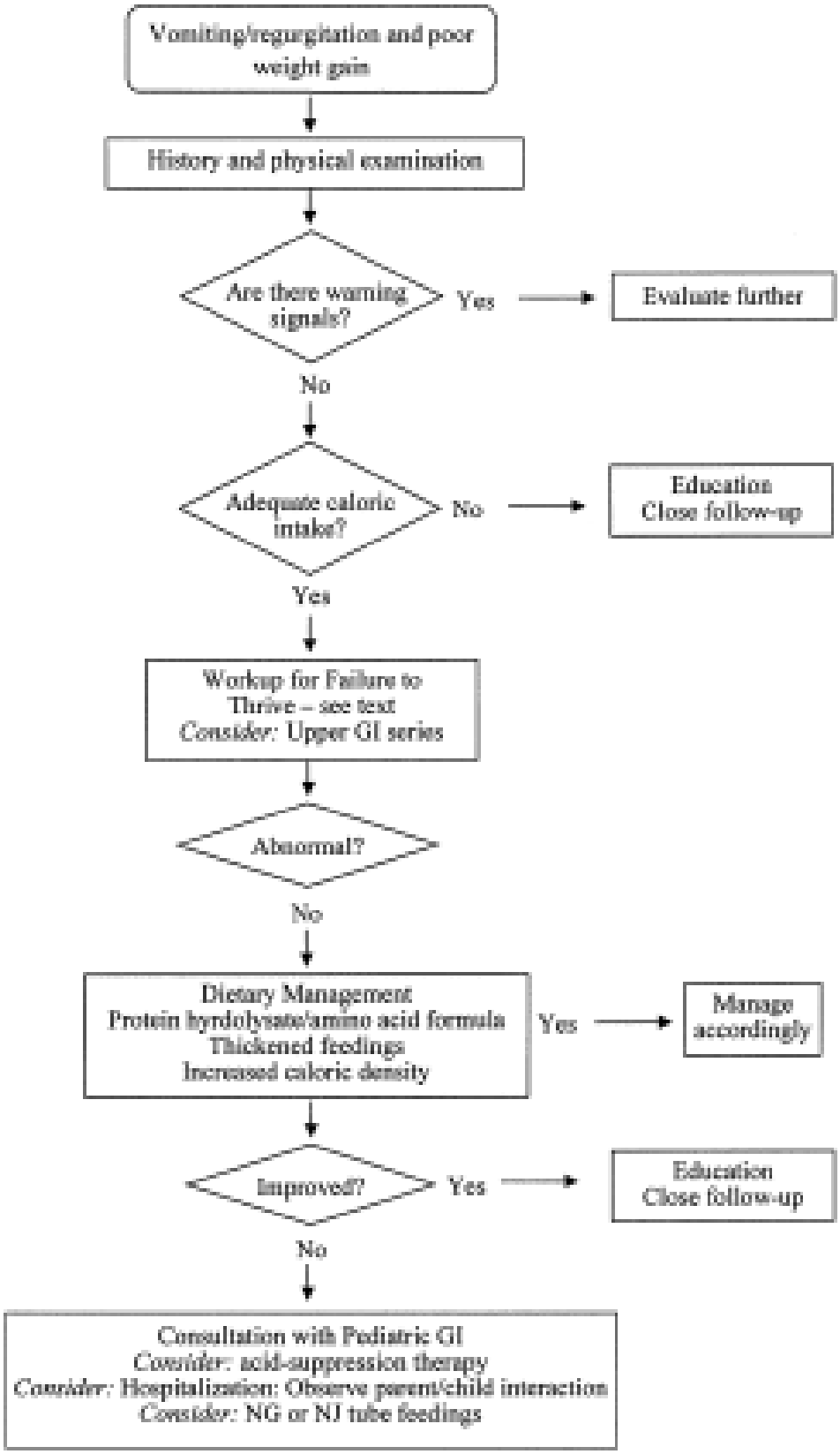
Acronym	Disambiguation
95%CI	95 per cent confidence interval
ALTE	Apparent life-threatening event
ANOVA	Analysis of Variance
AR	Acid reflux
ASM	Acid suppression medication
AUC	Area under the curve
CDD	Clinical documents database
CDH	Congenital diaphragmatic hernia
CENTRAL	Cochrane Central Register of Controlled trials
CP	Cerebral palsy
Df	Degrees of freedom
DGH	District general hospital
DT	Decision tree
ESPGHAN	European Society for Paediatric Gastroenterology, Hepatology, and Nutrition
FPR	False positive rate
GOJ	Gastro-oesophageal junction
GORD	Gastro-oesophageal reflux disease
GOSH	Great Ormond Street Hospital
GP	General Practitioner
GT	Gastrostomy tube
HCW	Healthcare workers
ICD	International Classification of Diseases
IPP	International Private Patients
IUGR	Intra-uterine growth restriction
LGT	Laparoscopic gastrostomy tube
LOS	Lower oesophageal sphincter
NAR	Non-acid reflux
NASPGHAN	North American Society for Paediatric Gastroenterology, Hepatology, and Nutrition
NGT	Nasogastric tube
NI	Neurological impairment
OA±TOF	Oesophageal atresia with or without tracheo-oesophageal fistula
OGD	Oesophago-gastro-duodenoscoppy
P/CG	Parents and Caregivers
PACS	Picture Archiving and Communication System
PATHLAB	Pathology Laboratory database

PEG	Percutaneous Endoscopic Gastrostomy tube
PGA	Physician Global Assessment
PIMS	Patient information management system
PIMS-OR	Patient information management system for operating theatre
R(DB)CT	Randomised double blind controlled trial
RCT	Randomised controlled trial
SQL	Standard Query Language
TRLOS	Transient relaxation of the lower oesophageal sphincter
UGI	Upper gastro-intestinal contrast study

Key definitions in understanding pH impedance

1. Reflux: a retrograde propagation of drop in current (impedance)
2. Acid reflux: reflux episode associated with a drop in pH to <4, or a drop in impedance during a period when oesophageal pH was already <4.0.
3. Non-acid reflux: a retrograde drop in impedance to >50% of the baseline impedance value in at least two most-distal channels. PH remains above 4.
4. Bolus clearance time: drop of impedance to >50% baseline) to recovery to baseline
5. Acid clearance time: drop of pH< 4 to time of recovery of pH >4.
6. Bolus reflux: a retrograde drop in resistance ≥50% of the baseline in at least two distal impedance channels.
7. End of a reflux episode: when resistance returns to at least 50% of the initial value.

ESPGHAN approach to the infant with recurrent regurgitation and weight loss



SECTION II APPENDIX ITEMS

TARDIS:REFLUX database

We obtained a MySQL database is hosted on a remote and secure database server i.e. mysql-server.ucl.ac.uk. The mysql-server.ucl.ac.uk service comprises a MySQL relational database package, running on a server maintained and supported by UCL Information Services Division (UCL ISD). The URL address for our database was http://mysql-server.ucl.ac.uk/sejjewm/. This address denotes an SQL database granted to the user (researcher). Access is limited to the user through password protection.

The database can be accessed using custom desktop database viewer software (MySQL Workbench 5.2). The database is hosted on a firewall-protected (Cisco Catalyst 6500 FWSM) wired local area network (WLAN). Therefore, the database can only be accessed from a desktop computer within the UCL network. Data is stored in an encrypted form within an encrypted drive. Back-up copies are also encrypted through disk-to-disk replication. This data environment meets ISO-27001 security standards. In the event that data files need to be removed from this environment, these files will be encrypted (Sophos Free Encryption) before secure file transfer (FTP).

Access to the database is password protected and limited to the database owner (the researcher). Privileged access to user accounts is granted, audited and monitored by UCL ISD.

Root account privileges available to the researcher. The researcher is therefore able to:

- CREATE (records)
- READ
- UPDATE
- DELETE

TARDIS:REFLUX Application Programming Interface (API)

The API was built to communicate between the app and the central database. There were two operations/tasks:

- Authenticate app users
- Transmit symptom and event data for each user from the app to the central database.

We commissioned an API for this purpose. The software developer (ABC) worked within the parameters described in the data specification phase. This API was designed using REST (Representational State Transfer) methods. The aim of REST-ful architecture is to enable communication between an application and remote server using a pre-defined set of methods or instructions that are performed on the remote server. The standard REST-ful instructions are:

- Create (a record)
- Read
- Update
- Delete

According to the specifications, data were to be transmitted in one direction only, i.e. from the app user's phone to the database. No app user, symptom or event data could be passed from the central database back to the app user's device. This design specification was to ensure that person-identifiable data protection requirements were met. When the method is carried out, the API can return a message which carries information on the success or failure of the operation. The API was designed to return messages if the authentication of users (task 1) or the data transmission (task 2) did not occur successfully.

As the data travelled in one direction only, only one "method" was required. A method is an instruction that is performed by a program at a specified resource. The resource for the TARDIS system was a central database. The methods available included:

- GET: This method retrieves data from the resource. It is synonymous with the READ instruction used in database design.
- POST: This method inserts new data to the resource. This is synonymous with the CREATE instruction used in database design.
- PUT: This method updates the resource with data and is synonymous with both CREATE and UPDATE.
- DELETE: This method deletes data from the resource

Data were transmitted in one direction only, i.e. from the app, to the central database. Therefore, the only methods utilised were POST and DELETE. These methods would act on particular data objects which, in RESTful terms are called parameters. The parameters for the TARDIS API are detailed in the table below.

Parameter name	Description	Type
username	The app user's username. This usually denotes the child whose symptoms are being tracked.	string
password	An 8-digit string of letters and numbers. This also corresponds to the Unique Study Number assigned to participants in the Pilot Study described in Chapter 0 below.	string
email	The app user's email address	string
datetime	This is a timestamp i.e. the data and time that API operation occurs	time
symptom_value	This is the symptom reported. This is one of the 26 symptoms enumerated in the symptom data object.	string
event_class	This is the event category i.e. Medication, Investigation, Diagnosis or Procedure	string
event_name:	This is the selected event e.g. Investigation > Weight	string

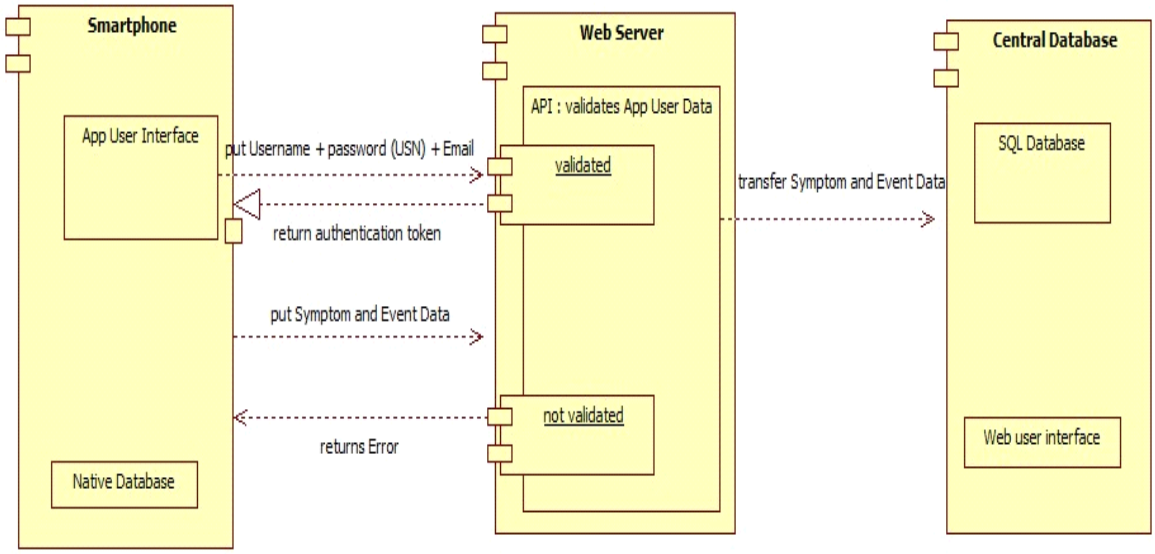
event_value	Weight events have a further enumeration i.e. an integer is inputted representing the actual weight measurement in kilograms.	integer
date	This is the date (+/- time) when a symptom or event occurs. It is a property of the symptom and event data objects. The app user inputs this data.	time

As described above, the API tasks are defined as

- Authenticate app users
- Transmit symptom and event data for each user from the app to the central database

The data architecture of app/API interactions is illustrated in the figure below.

Figure 145: Component diagram demonstrating data architecture for app/ API for user login and authentication



The method for authentication was designed as follows:

POST: /TARDIS/ios/v1/user_login

Use login wrapped the three parameters (i.e. email, username, password) into one identifier. Once submitted, the 3-point identifier was set to return a response in form of a 12-digit number. This 12-digit number was a session id (user_session_id). A unique session id was generated for each new login interaction between the app and the API. The id was denoted with a positive number if authentication was successful. This positive number was the session id. The session id was used in all further POST operations to identify the user sending the data within that session. If authentication was unsuccessful, a negative session id was generated. A negative session id triggered an Error Alert object on the app.

The session id was then used to tag all further inputs within the session. For example, the method for transmission of symptom data was as follows:

POST: /TARDIS/ios/v1/symptom_input

The symptom_input object wraps three parameters. These are:

- user_session_id
- symptom_name
- date: date+time of the symptom

This method returns either of the following results:

- OK
- Error

Summary of pilot study participants

TARDISID	SEX	AGE AT RECRUITMENT	SMARTPHONE OWNERSHIP	SMARTPHONE TYPE	DEVICE USED FOR STUDY	GROUP	RECRUITMENT DATE	DATE OF RANDOMISATION	COMPLETED STUDY
Ljwm7r3u	FEMALE	8.70	YES	IPHONE	IPHONE	A-APP FIRST	19/09/2013	19-Sep-13	NO
S2MJQ4c6	MALE	3.40	NO	-	IPAD	A-APP FIRST	09/07/2013	09-Jul-13	NO
BpTphZN9	MALE	0.57	YES	IPHONE	IPHONE	A-APP FIRST	11/07/2013	11-Jul-13	YES
tZf7T3X6	FEMALE	5.15	YES	IPHONE	IPHONE	A-APP FIRST	19/08/2013	19-Aug-13	YES
4JjPzNAc	FEMALE	14.80	YES	IPHONE	IPHONE	A-APP FIRST	20/08/2013	20-Aug-13	YES
scaXTRh9	FEMALE	4.89	YES	IPHONE	IPHONE	A-APP FIRST	29/05/2013	29-May-13	NO
8QFpTjmJ	MALE	8.75	YES	IPHONE	IPHONE	A-APP FIRST	04/06/2013	04-Jun-13	NO
fgEXLEVu	MALE	9.64	YES	IPHONE	IPHONE	A-APP FIRST	19/09/2013	19-Sep-13	NO
ykMnRbEz	MALE	1.74	YES	IPHONE	IPHONE	A-APP FIRST	23/09/2013	03-Nov-13	NO
kwdaD9mz	MALE	7.20	NO	-	IPAD	A-APP FIRST	02/07/2013	02-Jul-13	NO
2QVqvnmJ	FEMALE	0.64	YES	IPHONE	IPHONE	A-APP FIRST	23/07/2013	23-Jul-13	NO
ZLKhNzhq	FEMALE	0.95	YES	IPHONE	IPHONE	A-APP FIRST	04/07/2013	04-Jul-13	NO
v9PZMKYq	FEMALE	7.20	YES	IPHONE	IPHONE	A-APP FIRST	09/07/2013	09-Jul-13	NO
vaMwBdSZ	FEMALE	7.29	YES	IPHONE	IPHONE	A-APP FIRST	29/05/2013	29-May-13	NO
UyeVwtKm	FEMALE	11.63	YES	IPHONE	IPHONE	A-APP FIRST	22/05/2013	14-May-13	YES
Aku9ahCk	MALE	1.22	YES	IPHONE	IPHONE	A-APP FIRST	28/05/2013	28-May-13	YES
wgdb5exJ	MALE	1.31	YES	IPHONE	IPHONE	A-APP FIRST	30/05/2013	30-May-13	YES
dNrHQYKj	FEMALE	1.24	YES	IPHONE	IPHONE	A-APP FIRST	11/06/2013	11-Jun-13	YES
sNFBAHud	FEMALE	0.49	YES	IPHONE	IPHONE	A-APP FIRST	25/07/2013	25-Jul-13	YES
FKyPabp7	MALE	1.12	YES	IPHONE	IPHONE	A-APP FIRST	16/09/2013	16-Sep-13	YES
9VaLdE9x	MALE	1.55	YES	IPHONE	IPHONE	B-PAPER FIRST	23/05/2013	23-May-13	NO
nPjxJ3Wk	FEMALE	3.51	YES	IPHONE	IPHONE	B-PAPER FIRST	17/06/2013	13-Jun-13	NO
wnD6fDmD	MALE	0.90	YES	IPHONE	IPHONE	B-PAPER FIRST	13/06/2013	13-Jun-13	YES
n3zT9Em8	FEMALE	0.99	YES	IPHONE	IPHONE	B-PAPER FIRST	11/07/2013	11-Jul-13	YES
SzJHqhTK	FEMALE	5.45	NO	-	IPAD	B-PAPER FIRST	16/08/2013	16-Sep-13	YES

TARDISID	SEX	AGE AT RECRUITMENT	SMARTPHONE OWNERSHIP	SMARTPHONE TYPE	DEVICE USED FOR STUDY	GROUP	RECRUITMENT DATE	DATE OF RANDOMISATION	COMPLETED STUDY
vbMpUk6F	FEMALE	0.71	YES	IPHONE	IPHONE	B-PAPER FIRST	30/05/2013	30-May-13	NO
2EK6vsWQ	FEMALE	0.79	YES	IPHONE	IPHONE	B-PAPER FIRST	19/09/2013	19-Sep-13	yes
MNGHYEbC	MALE	0.56	YES	IPHONE	IPHONE	B-PAPER FIRST	23/05/2013	23-May-13	NO
j6p24hqu	FEMALE	2.14	YES	IPHONE	IPHONE	B-PAPER FIRST	03/06/2013	03-Jun-13	NO
9mWXjvL7	MALE	3.11	YES	IPHONE	IPHONE	B-PAPER FIRST	20/08/2013	20-Aug-13	NO
VBBweMuU	MALE	8.78	YES	IPHONE	IPHONE	B-PAPER FIRST	19/09/2013	19-Sep-13	NO
X7GT76CX	FEMALE	3.13	YES	IPHONE	IPHONE	B-PAPER FIRST	23/05/2013	23-May-13	NO
rDjzwF6D	MALE	2.89	YES	IPHONE	IPHONE	B-PAPER FIRST	04/06/2013	04-Jun-13	NO
6H5srcEj	MALE	13.10	NO	-	IPAD	B-PAPER FIRST	11/07/2013	11-Jul-13	YES
5GDWD62j	MALE	17.64	YES	IPHONE	IPHONE	B-PAPER FIRST	04/07/2013	04-Jul-13	NO
JcERZnba	MALE	1.94	YES	IPHONE	IPHONE	B-PAPER FIRST	20/06/2013	20-Jun-13	YES
8xrkRKaG	MALE	1.85	NO		IPAD	B-PAPER FIRST	11/07/2013	11-Jul-13	YES
wXjtgMDN	FEMALE	1.13	YES	IPHONE	IPHONE	B-PAPER FIRST	25/07/2013	25-Jul-13	YES
flmzNyPn	MALE	2.99	NO	-	IPAD	B-PAPER FIRST	20/08/2013	20-Aug-13	YES
MV73WDyc	MALE	5.77	NO	-		CONTROL	18/06/2013		NO
ZNbbKeM2	FEMALE	4.36	YES	IPHONE, BUT DAMAGED		CONTROL	30/05/2013		YES
vdVGEyL6	MALE	0.22	NO	NOKIA		CONTROL	21/05/2013		YES
xQRH5jpX	MALE	8.74	YES	SAMSUNG GALAXY		CONTROL	02/07/2012		NO
8aCXPxYH	FEMALE	2.30	YES	IPHONE, BUT DAMAGED		CONTROL	04/06/2013		NO
u5KdwsQx	FEMALE	0.66	YES	SAMSUNG GALAXY		CONTROL	29/05/2013		NO
mcVPY6Sx	FEMALE	7.36	YES	SAMSUNG GALAXY		CONTROL	04/07/2013		NO
uN2yHPKV	FEMALE	5.41	YES	SAMSUNG S4		CONTROL	15/08/2013		NO
QAE32ae2	FEMALE	8.96	NO	-		CONTROL	23/07/2013		NO
66wF9BDg	MALE	5.44	NO	-		CONTROL	14/06/2013		NO
c6DfTGKQ	MALE	4.63	NO	-		CONTROL	09/07/2013		NO
tRK7zbEq	MALE	2.16	YES	BLACKBERRY		CONTROL	21/05/2013		YES

TARDISID	SEX	AGE AT RECRUITMENT	SMARTPHONE OWNERSHIP	SMARTPHONE TYPE	DEVICE USED FOR STUDY	GROUP	RECRUITMENT DATE	DATE OF RANDOMISATION	COMPLETED STUDY
TBhqcL8j	MALE	12.62	YES	BLACKBERRY		CONTROL	25/07/2013		YES
EyZtwkrR	MALE	1.01	YES	GALAXY		CONTROL	28/05/2013		YES
5DwWKDkh	FEMALE	0.19	NO	NOKIA		CONTROL	29/05/2013		YES
N6H8kpYs	MALE	15.19	NO	NOKIA		CONTROL	17/06/2013		YES
mKGHyd9S	FEMALE	0.16	YES	SAMSUNG GALAXY		CONTROL	28/05/2013		YES
HVaqn3rd	FEMALE	1.91	YES	SAMSUNG GALAXY		CONTROL	03/03/2013		YES
Mg3dQt5D	MALE	13.18	YES	SAMSUNG GALAXY		CONTROL	08/07/2013		YES
c2ggJFue	MALE	4.38	NO	TESCO MOBILE		CONTROL	25/07/2013		YES
YSTGppRf	FEMALE	8.50	NO	-		CONTROL	25/07/2013		YES
4asctzB3	MALE	8.38	NO	-		CONTROL	23/07/2013		YES

Reflux UX: Analysis of interviews

Usability

Participants were asked specifically if they found the app easy to use. Most users agreed that the app was indeed “easy to use”, “user friendly”, and “straightforward to use”. Other adjectives applied were:

- Simple (2.41% coverage)
- Easy (1.67% coverage)

There was praise for the app as an idea, with phrases like:

- great idea, awesome idea (1.41% coverage)
- very useful, handy (0.93% coverage)
- great data collecting tool
- fast, quick
- fun

Not all users had praise for the app. Some negative adjectives used were:

- Too simple
- Tedious
- Serious

DESIGN EFFICIENCY

Improvements to design contributed to 32% of coverage. Participants suggested improvements to the symptom menu. Sorting the symptom menu in alphabetical order would make individual symptoms easier to locate. Event reporting was also found to be particularly problematic. Participants were expecting more investigations and not just the option of ‘weight’. Participant 7 (pharmacist) identified a potential source of error. There were three entries used to remark on the status of a single drug i.e. ‘Started Omeprazole’, ‘Increased/Decreased Omeprazole’, ‘Stopped Omeprazole’. A user would need to read to the end of the phrase to understand which option best reported the change in medication.

When there is more than one entry and you have a drug in your head... you may enter it "Oh, Omeprazole" and then you realize that there is more than one entry for a drug…

Source: Interviews 12:32.3 - 12:50.3, Participant 7

Improvements could also be made to the reporting functions and there were many modifications to the charting functions suggested. Participant 5 suggested a line graph in place of a density map to visualise common symptoms and their trends over time. An eagle –eyed participant (6) spotted the use of half-percentage points in the symptom frequency graph. This was deemed to be superfluous detail. Participant 2 suggested that the pie chart demonstrated the distribution of common symptoms should be annotated with percentages. In its current form, the user can visualise and estimate the proportions.

Participant 3 suggested substituting pictures for words as this might make the choices easier and faster to select. For example, picture of a baby crying may be used to instead of the symptom ‘crying’. The use of colour to aid navigation was also suggested.

SUBJECTIVELY PLEASING

We asked users to comment on the look and feel and, and specifically, the colour palette applied to the app. Examples of open questions used include:

What do you think about the colours that we have used? Do you think that these would appeal?

Source: Interviews 5:30.6 - 5:35.9, Interviewer: EMW

Can you think of ways that we night improve? For example, the look and feel of it?

Source: Interviews 37:18.4 - 37:19.2, Interviewer: EWM

Participants described the colours as “bright” and “nice”. Participants agreed that the colours would appeal to the parent of a young child. The colour coding of the home screen to differentiate the symptom, event and report choices was assessed as “really good”. However, some participants suggested that we should go even further.

Maybe some colours to make it different when you change the page. For example, if you record the symptoms, it should be blue. For the parents, if the children are crying or something, they should go quickly to the right page. If a busy mom has to look more than once to use the app, it might make her tired or she might find it boring.

Source: Interviews 18:34.2 - 19:46.0, Participant 3

There appeared to be a tension between palletes used and target audience. Participants asserted that, if the app was being targeted at children, there needed to be more done to appeal to this audience. Participant 4 considered that the app may not be taken seriously by parents if it was too playful. On the contrary, participant 6 felt that making the app more child-friendly, yet still targeted at the parent was desirable. The parent may use these features to engage the child.

Yes, but also if a parent has a young child, it’s nice to have a playful way of displaying everything. Because it could be a reason for the child to get involved. "That's a small book, let’s look, how did you do today, oh look, you have improved, so keep taking your medication."

Source: Interviews 7:56.1 - 8:24.2, Participant 6

LEARNABILITY

Participants commented on the learnability of the app. Participant 4 said of the app:

I thought elements of it were a little bit tricky to use. But nothing that, if you used a few times, you couldn't get used to.

Source: Interviews 29:12.9 - 29:45.7, Participant 4

This comment demonstrates that the app was not uniformly easy to use for all participants. It also suggests the participant felt it was accessible with more use, hence learnable.

CONTENT VALIDITY

Users highlighted words and phrases which were confusing, or did not capture their intended meaning. When participants navigated to the ‘events’ tab and selected ‘investigations’, they expected more investigations to select from and not just the option of ‘weight’. Furthermore, as participant 5 said,

‘Investigations’ suggested invasive procedures e.g. endoscopy. This participant suggested the use of a ‘softer’ term.

Participant 7 introduced the idea of user-generated data. The participant suggested that, when users change the dose of a medication, the user should have the option of submitting a free –text comment to explain the change. This view was re-iterated by participant 5 who elaborated on the benefit of adding user comments to a tracking app.

… A lot of things can be picked up from other apps. Things like Nike plus. That's an app I use for recording my runs. Although it’s a different criteria, it is still an app that records data with many variables. It records the time of your run, how you feel after your run. And there is a space there to put comments. So I put things like- had 5 traffic stops. That would affect my time on my run…

Source: Interviews 37:44.4 - 38:26.8, Participant 5

Utility

TRACKING

Several participants commented on the utility of apps for tracking health conditions. A participant who was also a parent (2) commented on the utility of an app to capture information that she could use to decide whether to make a GP appointment. A participant who was also a research coordinator felt that an app could anticipate a change in a health condition and therefore trigger an appointment.

Every so often, we could get information from the children and the parents on how many joints are affected at the present time. If ever a child is having a flare up, it would be great to have an app to report that the child is having a flare up before they have an appointment, which means waiting for the condition to be treated.

Source: Interviews 22:15.3 - 22:57.3, Participant 1

A contrasting view was offered by participant 5, a male student with chronic disease. He opposed the idea that apps should be used to continuously track symptoms. In fact, he observed that – as a person with chronic disease- he enjoyed periods of wellness. He would consider tracking symptoms during these periods as an unwanted reminder of his disease.

Yes, that is something from my experience, and other people I have talked to with chronic medical conditions. When they are in a period of remission, they tend to want to forget about the illness and move on with their lives.

Source: Interviews 34:04.3 - 34:25.6, Participant 5

SHARING

We discussed the role of apps in sharing data. Generally, the idea of apps being used to share a patient’s health information was seen in a positive light. Apps are seen to “make it easy” to share information. The role of apps in enabling fast information sharing was also reflected on. One participant observed that such an app could be used to reduce time spent waiting for a hospital appointment. By enabling the health professional to gauge the severity of symptoms occurring in the community, the app could be used to ‘triage’ patients and prioritise those who having exacerbations.

A case that is not too complex would mean a child is only seen here once a year. I think that even when a case is not too complex, they don't get priority to be seen. It is hard for consultants to gauge what has happened in 12 months. It's a really long time.

Source: Interviews 23:09.9 - 23:41.8, Participant 1

Willingness to share data using an app was contingent on information about the organisation or the people receiving the data. Some participants felt happy to share their data if it was being used anonymously in a study for the "greater good". Other participants indicated they would require understanding of the data destination.

I would need to know who is getting it. Maybe the nurse or the doctor. What they are going to use that information for, especially if there are details of my child e.g. name, surname, personal information.

Source: Interviews 27:38.9 - 27:54.9, Participant 2

The app was felt to have data sharing utility from a research perspective. One participant felt that such an app would help solve "missing data problems".

It would definitely be useful for the clinician, but useful for research too. It would solve the problem of missing data because no one is recording symptoms when the patient is not here

Source: Interview 24:00.9 - 24:19.7, Participant 1

OTHER CONDITIONS

The app as a medium was considered useful for tracking other conditions. Suggestions made included diabetes, constipation with soiling, rheumatoid arthritis and chronic abdominal pain. One participant with chronic illness considered the app a tool that her younger self might have appreciated.

I've been a diabetic since I was nine. Going back to when I was a child, if there was something there for me to monitor, and I my parents allowed me too, I would definitely go for it.

Source: Interviews 31:45.1 - 32:24.6, Participant 4

Participant 6 observed that app utility depended on the type of data being collected. In his experience as a researcher into HIV, data of interest for both patients and researchers were lab values. Therefore, an app that transmitted these data from the lab to the patient would be of greater utility than one which required the patient to enter these values from another source.

DATA COLLECTION TOOL

Participants agreed that the app had utility as a data collecting tool. However, the idea of effortful versus effortless data collection emerged. While participants agreed on the desirability of collecting data, the ease with which the data was collected was a factor. One participant contrasted the app with an electronic scale that automatically and remotely sent sensed data to a desktop application.

One thing I have is a scale. It measures body fat perspective and your weight. All you do is stand on a scale and it sends the data to a computer. So you get a graph of your weight and how it changes over time.

Source: Interviews 35:01.2 - 35:28.3, Participant 5

Participant 8 observed the importance of such a tool in digitising data. Once data has been digitised, it is then quick and easy to share.

ENABLING PATIENTS AND PARENTS

The role of the app in enabling parents to manage their child's condition was a common idea, accounting for 12% of comment coverage. The app would allow parents to "see for themselves". One participant (7) suggested that the app should contain "targets" that could be used to motivate health behaviour.

Another participant reinforced the idea that parents were already invested in the outcome i.e. improving their child's health.

Parents want to do as much as they can to help their child get better. They also want to do what they can to help their clinician help their child.

Source: Interviews 24:19.7 - 25:20.0, Participant 1

The engagement between the app user and the person receiving the data was seen as a two-way street. Some participants felt that busy parents would need incentives in order to invest the time and effort. It was considered fitting that the researcher or clinician should give "something back" to the user in return for completing the data entry tasks. The ability for parents to visualise data and obtain a health record was considered as an incentive.

DATA VISUALISATION

Participants felt that the ability to visualise data was valuable. Participant 8 suggested that the data could be transferred into a computer spreadsheet and then visualisations applied. However, he pointed out that the ability to visualise the data directly on one's device was a useful feature. Data visualisations were considered as part of the "reward" given to the user.

The fact that you can get a graph at the end of it... the fact that you can actually have a records, that's nice. Even though you might be using it, it's nice that you can give the person something back. If they are going to take the time to use it, it's good that they have some sort of feedback. It's good to have a record there so they can see their own record of what's happening. They can take more of an initiative on their condition when they monitor it themselves.

Source: Interviews 13:06.9 - 13:51.3, Participant 7

CLINICAL UTILITY

Participant 1 introduced the idea of utility for clinicians. The data received by the app could be used by clinicians to triage patient remotely and prioritise those with more severe symptoms.

AUDIENCE

There was some debate on the topic of whether the app should be targeted at the parent or the child (16% coverage). Some participants (1, 6) felt that children "who are old enough to use an iPad and iPhone" ought to be able to use the app. However, no agreement on the age of iOS achievement could be reached. Other participants (5) felt the app would be "difficult for a child".

The argument of the use of the app to promote positive health behaviour in the child had some mileage. The parent might the app to be a useful tool with which to engage the child on the health condition:

"…let's look, how you did today, oh look, you have improved, so keep taking your medication."

Source: Interviews 7:56.1 - 8:24.2, Participant 6

Some participants clearly had mixed feelings. While it might be fun to develop the app for a child, treating the condition was an important and serious matter that could not be left to the child.

Yes, like having an animal that is increasing or decreasing the levels? I don't know. Maybe for the child it is more difficult (to develop). They might start to think that this is a game. But it's not clear if they are playing or doing something serious. So you have to be careful. I think that it is better for a parent to use.

Source: Interviews 7:20.9 - 7:48.5, Participant 6

Most participants agreed that the app should be directed towards parents. This might be because the app was "too medical for a child". Parents are invested in getting their child better. However, for a child to complete recording health related data might be "too much". Furthermore, for an adult with chronic disease completing the diary for themselves would be unappealing. However, completing it for their child would be easy to do.

I think it's a bit different for adults. I am diabetic and I wouldn't necessarily find an app like that for me. That's something that I probably wouldn't use. But for children, from a parent's perspective, it would be a lot easier.

Source: Interviews 29:49.8 - 30:17.1, Participant 4

I think for your children, it is a little bit different. The time investment is more for them. However, with yourself, you always take a little bit of a back seat. But I do think with children, you do.

Source: Interviews 30:24.1 - 30:42.2, Participant 4

SECTION III APPENDIX ITEMS

Search strategy and data processing

Clinical documents database

CDD is Microsoft SharePoint customised for GOSH. It allows verified users to upload, access and download a variety of document formats. It is accessed via a secure server (https://cdd.moss.sharepoint.gosh.nhs). Documents created by clinicians are stored on CDD as part of a patient's record. As a Microsoft product, SharePoint applications have a high level of integration with Microsoft Office applications.

Construct and content

CDD was initialised at GOSH in September 2005. The strategy for document management was prospectively upload on-going clinical data to the SharePoint site. Letters are uploaded in batches by administrative staff supporting clinicians. For example, following a clinical consultation, the clinician dictates a letter to the patient and GP. These letters are typed and formatted by the clinician's secretary, who then uploads them to CDD. Records on CDD therefore reflect a contemporaneous account of the patient's symptoms and clinician's findings.

Each record is a row of data in a flat table. The 'Filename' field contains a hyperlink which, when clicked, opens the record.

Figure 146: An example of a CDD table

Filename	Document type	PatientID	Date of Birth	Department	Document Date	Author Name	Created	Last modified
Exxx DS Clinic 05.12	Reports 939900 29/05/	939900	13/07/2009	Cardiothoraci	18/10/2011	SquirA	23/11/2011	
Nxxx 22 02 2006.do	Reports 781453 11/09/	781453	11/09/2010	Anaesthetics	17/02/2011	FrancR2	04/03/2011	
Lxxx 2011.06.22.DO	Reports 913660 16/11/	913660	13/07/2009	Anaesthetics	22/02/2011	Rose	22/02/2011	04/03/2011
Sxxx 22.10.2010.do	Clinic letter 917299 01	917299	08/02/2009	Anaesthetics	22/02/2011	Rose	22/02/2011	29/03/2011
Sxxx 06.12.2010.do	Outgoing General Cor	917299	08/02/2009	Anaesthetics	23/06/2011	Rose	23/06/2011	04/07/2011
Axxx PX322401 cl 8	Clinic letter PX322401	PX322401	22/09/2010	BMT	28/05/2012	FanniM	21/06/2012	
931102_2011_04_07	Clinic letter 931102 03	931102	19/10/2011	Cardiology	21/11/2011	GillaC2	28/11/2011	
Sxxx2011.03.15.DO	Reports 937814 20/04/	937814	02/01/2011	Cardiology	23/09/2011	SchutK	29/09/2011	
Kxxx 939249 - Lette	Clinic letter 939249 15	939249	15/09/2010	Cardiology	13/10/2011	SchutK	31/10/2011	
Sxxx 94028113-05-1	Outgoing Referral Lett	940281	22/09/2010	Cardiology	09/03/2011	McDonC4	16/03/2011	
967424_2012_04_17	Outgoing General Cor	967424	20/08/2010	CICU	04/04/2012	Lisa	04/04/2012	04/04/2012
Cxx2011.05.19.DOC	Reports 924217 14/05/	924217	11/09/2002	Cleft	06/01/2011	KingD1	07/01/2011	

The header of each table contains other fields used for identification and indexing.

Utility

CDD can be searched via the index fields. Due to integration of SharePoint with Microsoft Office applications, search capability mirrors the workings of Office applications. Significantly, document-level text searching is possible for the whole database. The document formats utilised are listed in **Table 161** below.

Table 161: CDD document formats

Document type	Microsoft Office format

Clinic Letter	Word
Discharge Summary	Word, Rich text format, Excel
Reports	Rich Text Format, Portable Document Format (PDF)
Multiple patient documents	Word
Correspondence	Email/Outlook

Text search terms/ strings are chosen to be as broad as possible to capture all patients with GORD at GOSH. Below is a summary of the search strings utilised.

Table 162: Summary of keyword searches

Category	Rationale	Search string
Diagnosis	GOR or GORD has been mentioned	“G* reflux” ,“GOR” , “GORD” , “*oesophageal reflux”
Medications	Typical medications for GORD	omepr*,lansopr*, pantop*, ranit*, gavis*, PPI , H2.
Interventions	Surgery for GORD	Fundoplication, fundopl*, Nisse*
Comorbidities	comorbidities associated with GORD	*sophageal atresia, trache*esophagel fistula, tracheoesophageal fistula, OA+/-TOF, OA*TOF, *, H*type “C* D* H*”, CDH Achalasia, achalas* prematurity, premature, exprematurity, ex-premature, ex-prem, exprem

To ensure that only paediatric patients (<16 years) were included in the cohort, we restricted searches to the capture period i.e. patients born between 01/01/1994 and 31/12/2010.

Data quality

1. Referential integrity: All CDD records are indexed using PatientID. Records in all other electronic databases were also indexed using PatientID. Therefore, we have a high degree of confidence that pooled records indexed using a PatientID correspond to the same patient.

2. Consistency: Indexing fields e.g. DOB and PatientID, Author and Department are required. Therefore, a similar degree of information is available for each record.

3. Accuracy: During a clinic, the clinician makes notes which are stored in paper records. These notes can be illegible. The clinician also dictates a clinic letter which is subsequently transcribed by a medical secretary. Transcribed letters are checked and signed by the clinician before they are posted. Accuracy is this aided by multiple stages where errors can be detected and corrected.

Despite these failsafe, typographic and keystroke errors were detected. Boolean search modifier e.g. wildcard/ *, query/?, quotations "*" were useful in creating word searches accounting for errors and variation in spelling. The researcher used clinical experience to anticipate word use and phraseology.

4. Timeliness: Most CDD records are links to documents created prospectively. Where records link to retrospective documents, they are indexed as 'Multiple patient documents'. Therefore it was possible to ascertain a chronology of events. This was important for time-critical variables e.g. mortality, fundoplication.

5. Verification: CDD was useful in verifying search results from other databases. Document meta-data e.g. date of creation, date last modified is automatically created when the record is uploaded or modified. These metadata also provided useful verification points.

Post-processing and cleaning

CDD data were highly redundant. Search queries yielded all records containing the search term e.g. 'tracheoesophageal fistula', This necessitated limited searches, manual review and data cleaning procedures. To remove duplicate records, we pasted records into Microsoft Excel and ran the 'remove duplicates' macro. Metadata e.g. documents stored in 'Draft' versus 'Live' areas could be distinguished, thus separating unverified draft clinic letters from verified letters.

Searching Strategy

Search by Diagnosis

We search clinic letters for records where GOR or GORD has been stated. Clinic letters are saved as Word documents therefore the search is limited by document type i.e. Word document. This excludes other records with these key words e.g. hand-over sheets in .xls format (Microsoft Excel).

Serial searches are conducted for the following terms:

"G* reflux" ,"GOR" , "GORD" , "*oesophageal reflux"

Data post-processing

We summarise post-processing procedures required and the results of the search and detail (Figure 147).

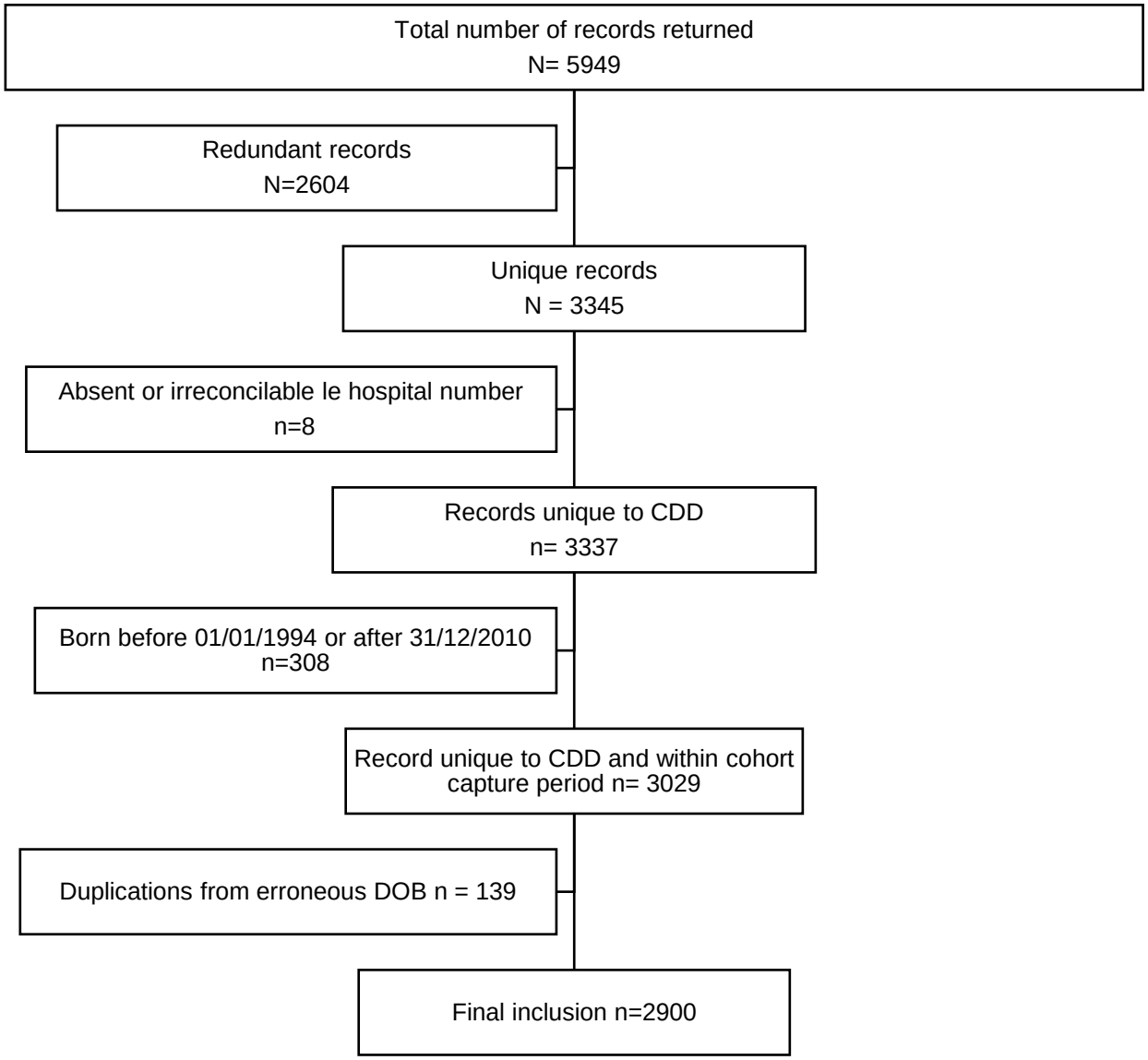
1. PatientID redundancy: Out of 5949 records in rows, 2604 were duplicates. Therefore, 3345 unique records identified and preserved.

2. PatientID errors: We found 8 records without a hospital number or with an irreconcilable error i.e. cannot be validated by DOB, firstname or surname. Subsequently, 3337 records were retained.

3. Capture period: There were 86 patients born before 01/01/1994 and 218 patients are born after the 31/12/2010. A total of 308 records are excluded leaving 3029 records.

4. DOB errors: There were 139 duplicated records based on erroneous DOB. These were identified and rationalised, leaving 2900 unique records.

Figure 147: Inclusions and exclusions following search based on GORD diagnosis



Following data pre-processing, records were uploaded to the core database [Retrospective GOR.db]. This search by GOR diagnosis identified most patients included in master list of cohort patients. This master list is was contained in a table named 'Inclusion List'. This table was indexed by PatientID, which was set as a primary key. In a table, a primary key field can only contain unique values i.e. no duplications. Setting this requirement means that each record, representing a patient, can only be included or counted once.

Search by Comorbidity/ Specialist team

Comorbidity data were searched for in CDD. Clinic letters are uploaded by doctors working in different specialist teams. In CDD, letters are indexed by specialist team (**Table 163**).

Table 163: Illustrating comorbidity information from a CDD keyword search. All PatientID and DOB couplets are fictitious, for the purpose of anonymising the patient.

PatientID	DOB	Author	Department
92M737	10/04/2010	MendoP	Palliative
9516L5	03/04/2010	Martina	SLT
9229P1	25/03/2010	HillsM	Palliative
919L58	16/03/2010	WimalN	Neuro-disability
9B2010	13/03/2010	Alex	Respiratory

GOSH patients are referred for tertiary and sometimes quaternary review. It is a safe assumption that a patient seen by a cardiothoracic surgeon has a cardiothoracic comorbidity. By reviewing the specialist teams, different comorbidity categories can be identified. These are summarised in the table below.

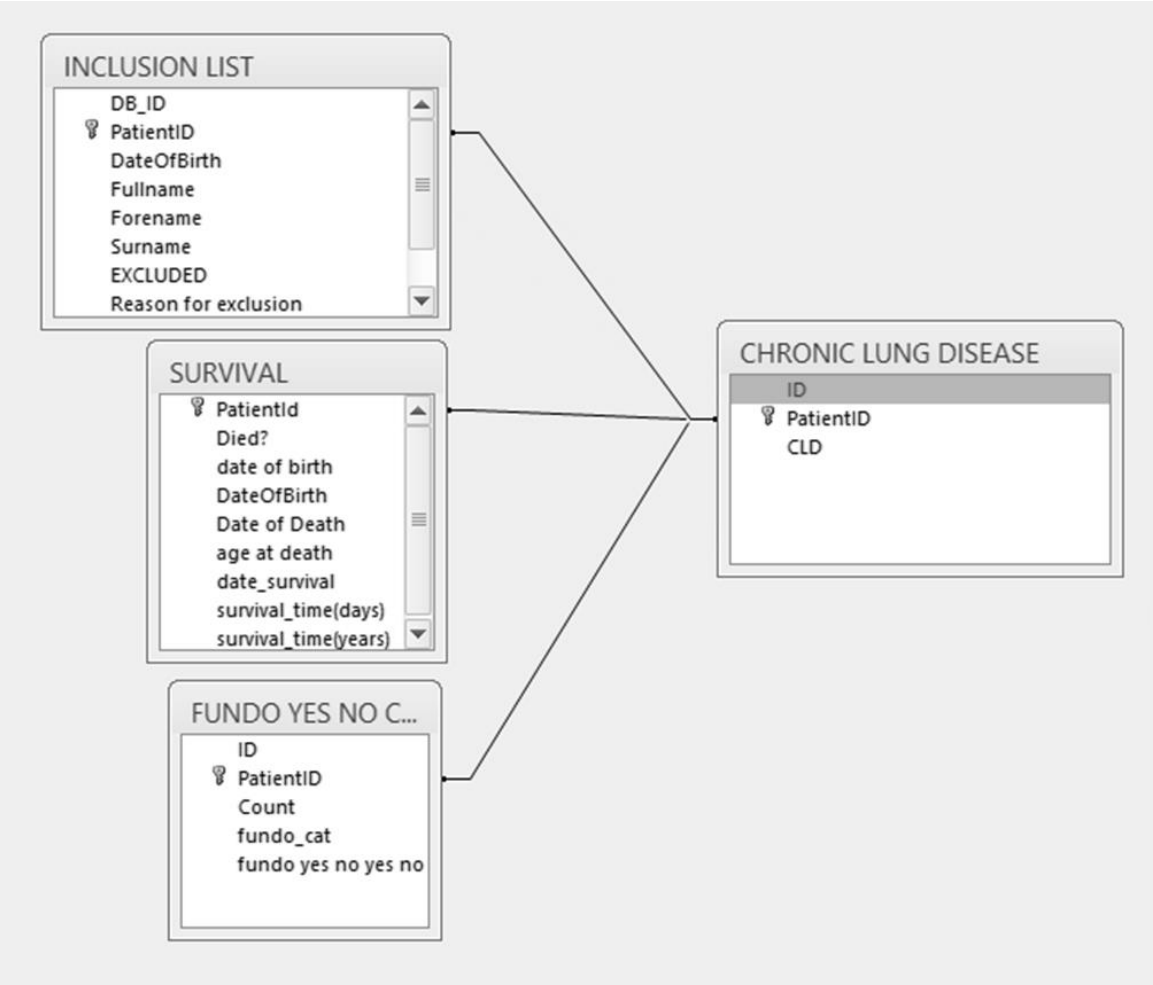
Table 164: Specialist teams and associated comorbidities

Specialist team	Comorbidity category
Asthma	Asthma
Bone Marrow Transplant	Bone marrow transplant
Cardiac	Cardiac comorbidities
Cardiothoracic	Cardiothoracic surgery
Craniofacial	Cleft and craniofacial anomalies
Cystic Fibrosis	Cystic Fibrosis
Dental	Dental disease
Endocrinology	Endocrine disease
Epidermolysis Bullosa	Epidermolysis Bullosa
Genetics	Genetic and chromosomal anomalies
Immunology	Immune disorders
Intensive Care	Aspiration
Metabolic	Metabolic disease
Musculoskeletal	Skeletal anomalies
Neurology, Neuro-disability	Neurological impairment
Oncology	Oncological disease
Renal Transplant	Chronic renal failure
Respiratory Medicine	Chronic Lung Disease
Respiratory Medicine (Sleep Lab)	Sleep apnoea
Speech and Language (Videoflouroscopy)	Feeding and swallowing disorders
Tracheal	Tracheal and laryngeal anomalies
Tracheostomy	Tracheostomy
PICU	Acute respiratory distress/ALTE
Urology	Renal dysplasia

Records were sorted by specialist team. We summarised all unique clinical teams involved in the care of these children and identified where relevant, corresponding comorbidities e.g. cardiothoracic team corresponded to cardiothoracic comorbidity. Patients with a particular comorbidity were included in the corresponding table. The comorbidity field value was 1 = comorbidity present and 0 = comorbidity absent. For each comorbidity, a table was created. Comorbidity tables had only 2 fields: a primary key (PatientID) and a categorical variable (co-morbidity true/false).

Comorbidity tables were then linked by PatientID to the primary index table i.e. –'INCLUSION LIST'. The table structure for comorbidity data is illustrated in **Figure 148** below.

Figure 148: CHRONIC LUNG DISEASE is linked to INCLUSION LIST, SURVIVAL and FUNDO by the primary key i.e. PatientID.



Comorbidities that are known to be associated with GORD were also investigated. CDD was searched to identify patients with CDH, OATOF and prematurity.

Table 165: Search terms for comorbidities associated with GORD

Comorbidity	Term	Identified
OATOF	"*sophageal atresia" OR "trachea-oesophageal fistula" OR "tracheoesophageal fistula" OR "OA+/-TOF" OR "OA*TOF" OR "H*type".	302
CDH	"C* D* H*", CDH	134
Prematurity	"prematurity" OR "*premature*" OR "exprematur*" OR "ex-prematur*" OR "exe-prem*"OR "exprem*"	1417
Achalasia	Achalas*	40

Patients with achalasia required special consideration. Achalasia is a condition in which the lower oesophageal sphincter muscles fail to relax appropriately. This leads to progressive narrowing and

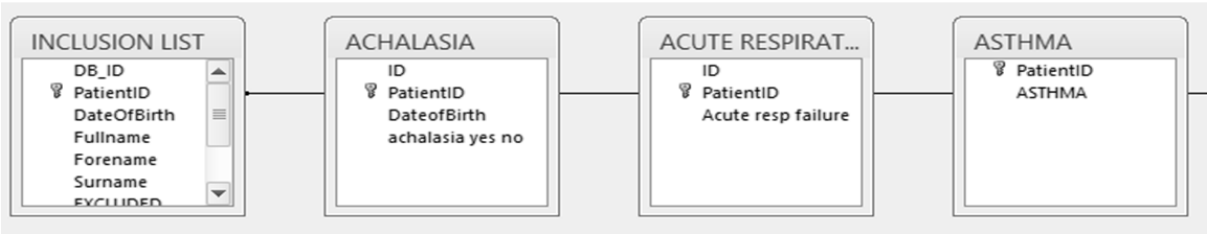
funnelling of the distal oesophagus. This narrowing obstructs passage of food and fluids and can cause dysphagia and regurgitation. Regurgitation in particular can mimic GORD. Some patients with achalasia will have both achalasia and GORD depending on the degree of distal oesophageal junction stricture. Like patients with GORD, patients with suspected achalasia will have upper GI contrast studies and endoscopy. Given the co-incidence of achalasia and GORD, a keyword search was also performed for patients with achalasia. The rationale for this search was to identify and exclude a subset of patients with achalasia and no GORD.

There were 101 records from the PICU department in the GORD search yield. These records were manually reviewed to identify relevant comorbidities. Respiratory events were characterised as acute respiratory failure or aspiration ALTE.

We also identified 290 patient records arising from the palliative care team. We reviewed these 290 individual patient records in PIMS to identify date of death. These records were cross-referenced with mortality data generated from PIMS (See page 502). There were 204 patients known to the palliative care team whose mortality was confirmed on PIMS. There were 86 records of patients known to the palliative care team but not confirmed as mortalities on PIMS. Manual review of the corresponding CDD records detected 39 patients who had died, but whose PIMS record had not been updated with this information.

Comorbidity tables were linked to each other, as well as the index table 'INCLUSION LIST' using a left join.

Figure 149: The primary key PatientID sets referential integrity between the INCLUSION LIST, ACHALASIA, ACUTE RESPIRATORY DISTRESS and ASTHMA



Search by Medications

The medications keyword search identifies patients being treated for GORD who might otherwise be missed. The GORD search string identifies records in which the condition is explicitly mentioned. However, when GORD is well controlled, it may not be included in the list of active problems. For example, GORD may be a secondary problem not central to admission episode. Another example is patients with a history of OATOF. These patients have a higher incidence of GORD. GORD is anticipated and acid suppression medications (ASM) prescribed after primary oesophageal atresia repair soon after birth.

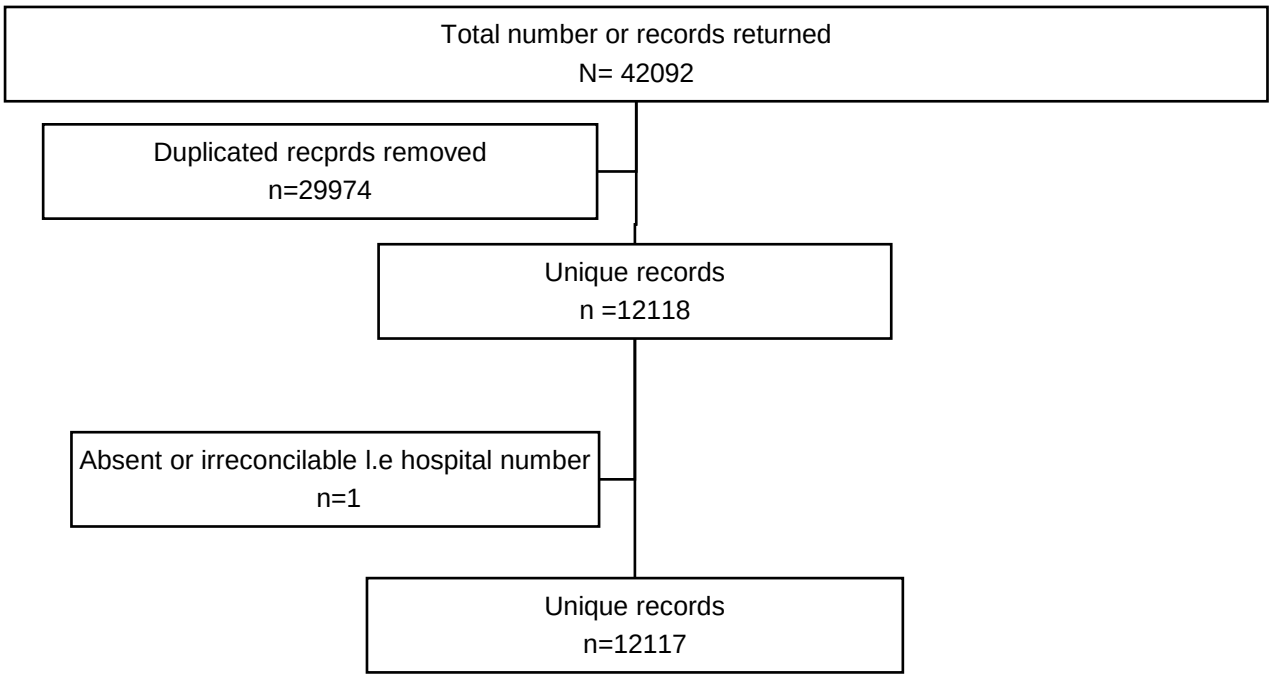
We performed a keyword search to identify patients on the following ASM: omeprazole, lansoprazole, pantoprazole, ranitidine, PPI, H2 antagonists/ blockers. The motility agent domperidone is sometimes used for patients with GORD. However, it is also used for children with chronic abdominal pain, nausea and vomiting. We chose not to search for this term for this reason. Furthermore, it would not be in keeping with guidelines to treat GORD with domperidone but no acid suppression medications.

Gaviscon is also commonly prescribed as a first line medication in infants and children with reflux. Therefore, Gaviscon was included in the search string.

Strings : omepr*,lansopr*, pantop*, ranit*, gavis*, PPI , H2.

Search yield and necessary post-processing procedures are summarised in the flowchart below (Figure 150).

Figure 150: Yield of keyword search by medication



Search by 'Fundoplication'

The fundoplication keyword was designed to identify records containing the word fundoplication. The strings utilised were: "fundoplication" OR fundopl* OR Nisse*.

However, we found that fundoplication was mentioned in many contexts. Furthermore, mention did not correspond to the operation having been carried out. For example, fundoplication was mentioned as a treatment option.

"*We may consider a fundoplication, but I feel at the moment her reflux is well managed on medication." (779809)*

Fundoplication was mentioned incidentally:

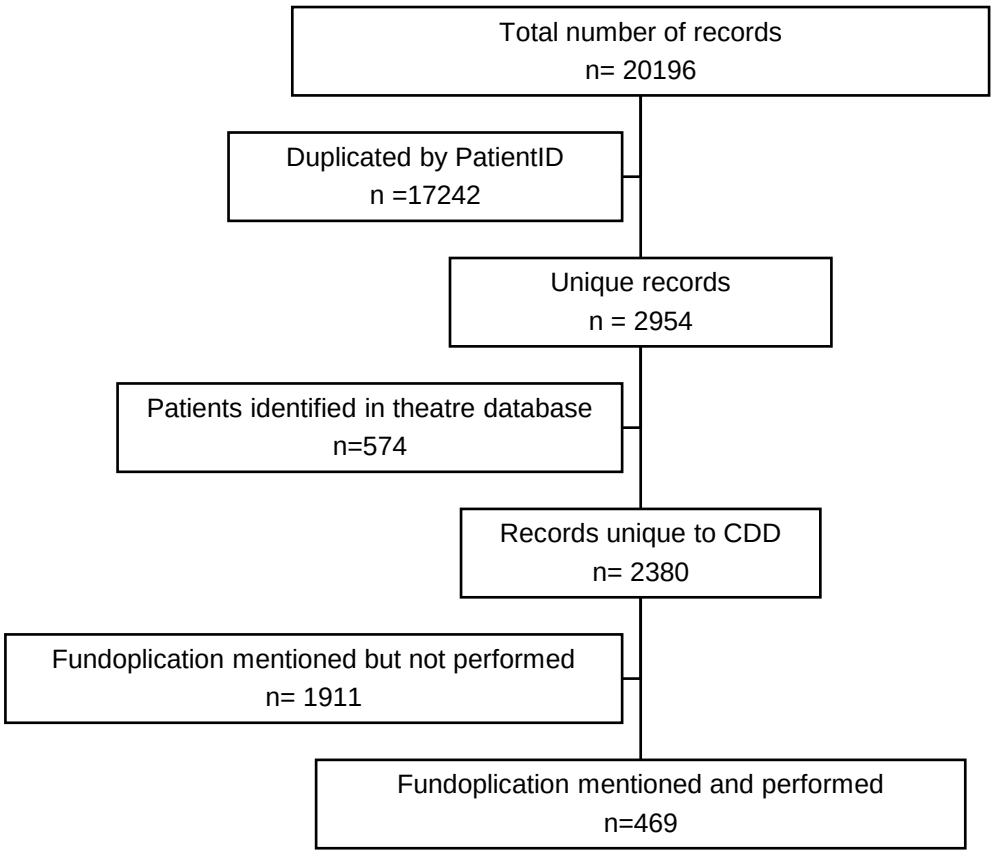
"We did discuss gastric pacing , which has now found a niche in treating post-fundoplication nausea." (638876)

Fundoplication was mentioned as a negative finding.

"…(the patient) has never had Nissen Fundoplication." (645116)

Fundoplication was also included as an option in videoflouroscopy questionnaires.

Figure 151: Search yield and post-processing for fundoplication keyword search of CDD



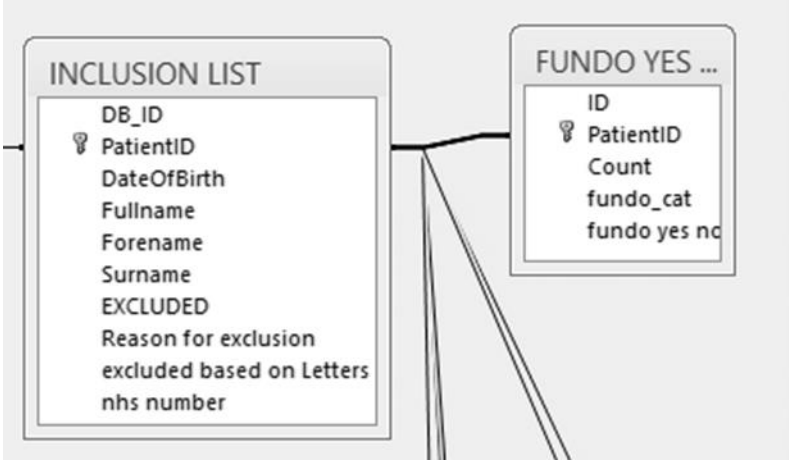
To identify records where mention of fundoplication does not correspond to an operated patient, an analysis of context was necessary. Therefore, we manually reviewed the clinic letters to identify keyword context.

CDD records were also cross-referenced with operating theatre records (PIMS-OR). This was done to validate CDD records of fundoplication. It also enabled identification of patients who had fundoplication at GOSH. It is important to note that those who had fundoplication elsewhere would not be captured by operating theatre records. Therefore, the CDD search was crucial in identifying this subset of patients.

The fundoplication search yield and post-processing is detailed in the flowchart below (Figure 151). Most of the 469 operated patients had a primary fundoplication at GOSH (n=360, 77%). .In 71 instances (12%), the year of operation was recorded in the clinic letter e.g. “Nissen fundoplication 2000”. In these instances, the midpoint of the year (i.e. 2nd of July) was assumed as the operation date. Age at operation was calculated using this date.

Records were included in the [Retrospective GOR.db] as a table ’FUNDO’. Fundoplication status was a categorical field i.e. Yes / No. The table also included a numerical ‘Count’ field to record the number of fundoplications a patient had. Lastly, a categorical field was introduced to recode fundoplication count as a categorical variable i.e. one fundoplication/ greater than one fundoplication’. The ‘FUNDO’ table was linked to the ‘INCLUSION LIST’ by joining PatientID.

Figure 152: FUNDO is linked to the INCLUSION LIST table by the PatientID



Patient information management system (PIMS)

Construct and content

PIMS is an electronic relational database that is used to store administrative data. PIMS is hosted on the GOSH virtual private network and can only be accessed from a network computer by authorised personnel. PIMS is encounter-based database i.e. data are structured in silos based on date and type of encounter. These encounters include:

1. Registration: A patient's first admission or clinic visit. A record is created on the Master Patients Index. A PatientID is created at this first encounter.

2. Outpatient appointments: Appointments are booked using PIMS. When a patient attends the appointment, the system is updated.

3. Inpatient episodes: Admission, discharge and transfer date are recorded. Transfers are patient movement e.g. from ward to theatre.

4. Updating patient mortality status

PIMS is also used for bed management, referral management (referring patients between specialities) and document tracking.

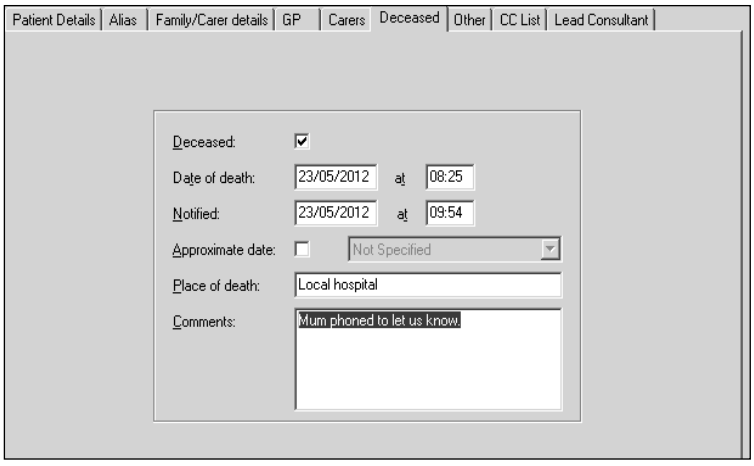
Utility

PIMS was used primarily to gather mortality data. PIMS was also invaluable for verification of DOB and PatientID.

Mortality data arise from various sources:

1. Demographics Batch Service (DBS): The DBS is a central NHS service that enables users to trace demographic data. Users (Hospitals and their administrators) submit a batch of NHS numbers and obtain and update on demographic information i.e. name, address, DOB, date of death (DOD). This process is called a DBS trace. Notification of death can arise from clinical (e.g. palliative care team), clerical staff, and sometimes from parents (**153**).

Figure 153: An example of a PIMS recorded updated with a date of death following a phone call from a parent



Data quality

When searching the CDD database, we identified patients with GORD keywords in their records known to the palliative care team (n=86). Within this subset 39 had died. However, none of these had their PIMS recorded updated with Date of Death. This indicates that there PIMS database is incomplete and underestimates mortality.

Given the non-trivial nature of a DOD update, mortality data entered on PIMS is likely to be accurate.

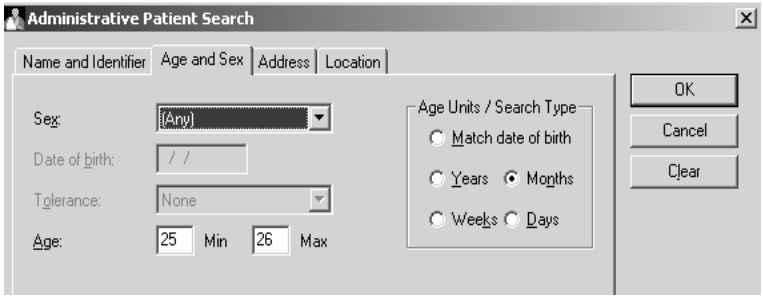
However, we found that there was often a lag between the date of death and its notification. Some occurrences of mortality may be missed due to this reporting-recording lag. An important limitation of these data is absence of cause of death information. To systematically gather this information access to civil registrations of death (death certificates) would be required. Access to this data can be obtained from the Office of National Statistics(ONS) Health episodes data. Requesting access to death registration records specific to our cohort and institution would require informed patient consent. Obtaining this consent was beyond the scope of this project.

Search strategy

This PIMS database has limited search capability for bulk records. It is possible to search for individual patients by their PatientID, DOB or name. However, where these parameters are unknown, PIMS can be searched by age bracket.

To overcome this, we searched for mortality within the capture period i.e. 1/1/1994 and 31/12/2012. In the example below, we search PIMS for all patients between 25 to 26 months old at the time of the search.

Figure 154: Age bracket (25-26 months) search of PIMS database for mortality data



The search yields a list of patients of that age. One of the fields returned is 'Date of Death. Results are then sorted by Date of Death. Records with a 'Date of Death' are selected and cross-matched with [RetrospectiveGOR.db] The search was conducted between 24th and 30th of January 2013.Therefore, the search found only patients who had their DOD updated to PIMS by this date.

We identified 925 cases of mortality from this PIMS search. Prior review of CDD had also identified 39 palliative care patients who had died (see page 491). Therefore, in total, we identified 964 cases of mortality. Mortality records were amalgamated into the Retrospective GOR.db in a table labelled 'SURVIVAL'. A left join was used to create linkage with other records in the database. Unmatched patients in the list of mortalities were excluded from the database. Therefore, unmatched patients in [RetrospectiveGOR.db] were assumed to be alive.

Radiology database

Radiological images and reports are stored, manipulated and distributed through the Picture Archive and Communication System (PACS). PACS is an electronic database hosted on the NHS N3 cloud server. It is used at GOSH and many other NHS hospitals. At GOSH, it is also used as an administrative portal for tracking encounters with patients.

Construct and content

The database is accessed through a "thick-client" network set-up. Desktop computers on the GOSH local area network (LAN) are equipped with a software application to directly access the PACS server. The database can be accessed in three main ways. The researcher was granted PACS Viewer access allowing searching by patient-specific identifiers e.g. PatientID, DOB, or by searching for episodes e.g. Chest X-rays done on 19/9/2010. However, it is not possible for general users to run batch queries of PACS

Utility

The PACS system at GOSH was installed in March 2006. Unfortunately, images from the previous image storage application were not transferred to the new system. Database review of images prior to March 2006 was not possible.

As we required a batch query of the database, we approached the PACS administrator with a query request. Institutional approval for this request was incorporated in the ethical approval for the REMOS study. Transfer was via a .csv file type which translates easily into an excel spreadsheet for collation and manipulation.

Data quality

- Accuracy: As the radiology reports are reviewed by radiological consultants before being released to patients, we expect a high degree of accuracy.
- Consistency: UGIC studies were reported by different consultants. There was inconsistency in the application of provocation procedures for reflux. There was also inconsistency in the description of reflux e.g. a 'wisp of reflux' into the fundoplication wrap versus a 'beak of refluxate'. The researcher used clinical experience and standardise these reports into the reflux categories **Table 46Table 166**

Table 166: Data transformation: recoding the level of reflux

LEVEL OF REFLUX	Code
Zero	0
Proximal	1
Mid	2
Distal	3
Oropharynx	4
Aspiration	5

Inadequate study	6
No UGI done	7

- Completeness: The PACS database is incomplete and only contains records from its implementation in March 2006.

Search strategy

The UGI contrast study is used to investigate patients with GORD. Historically, it has been used to demonstrate reflux. However, it is a low-yield and inaccurate reflux diagnostic(156). Clinicians request this investigation to diagnose/exclude malrotation and hiatus hernia.

We requested records of patients who underwent an UGI contrast study from the radiology records administrator. The query parameters were:

Table 167: Radiology database query parameters

Field	Parameter
Image type	Upper GI Contrast Study
From	01/03/2006
To	31/12/2010
Reported fields (Results)	• PatientID • DOB • Name • Investigation type • Date of investigation • Text of investigation report

Post-processing and cleaning

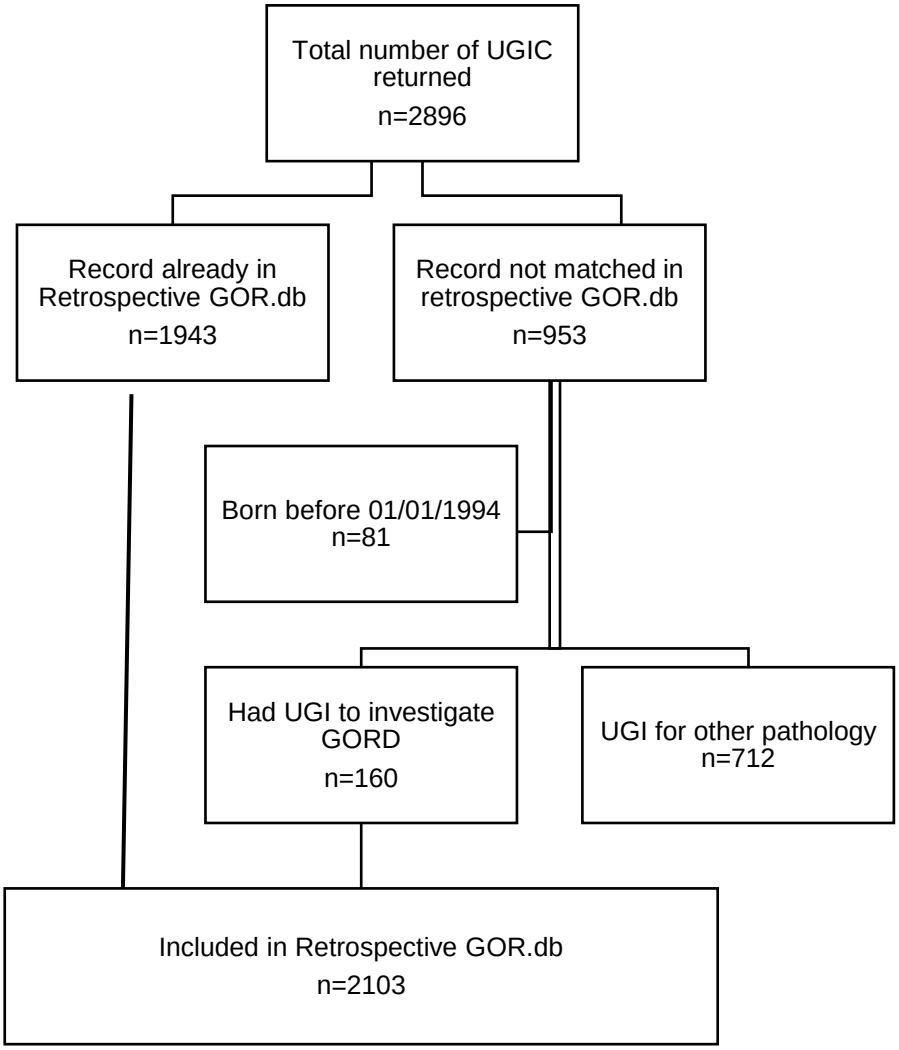
Not all reporting radiologists included the indication for the UGIC in their study report. To ensure inclusion only patients having an UGIC for GORD indications, we cross-referenced search results with those of the CDD keyword search. There was a 67% match (1943 /2896 records = 67%) between patients having an UGIC and patients with already identified in the CDD keyword search. This suggests that our search strategy was reasonable sensitive in identifying patients with GORD. Those records not cross-referenced in CDD were reviewed manually. This exercise identified patients who had UGIC for indications other than GORD(Figure 155) e.g.

- Oesophageal stricture or leak following caustic ingestion
- Lower gastrointestinal contrast study or Contrast enema wrongly coded as UGIC

- UGIC for achalasia / dysphasia

The UGI report for each patient record manually reviewed and the height of reflux was coded according to the information given in the radiologist's report.

Figure 155: Radiology records were reviewed to identify patients investigated for GORD



Radiology records were amalgamated into [RetrospectiveGOR.db] in a table titled 'UGI FINAL'. Using a left join, this table was linked to INCLUSION LIST by PatientID.

Theatre database

At GOSH, activity in the operating theatres is tracked using an extension of the PIMS database for operating room (PIMS-OR). PIMS-OR extension is an admissions, discharge and transfer (ADT) tracker. Clinical and administrative staff book theatre appointments, admit and discharge patients into the OR. The PIMS-OR patient record is updated with the procedure performed.

Construct and content

PIMS-OR was implemented in January 2006. Therefore, prospective records are only available from January 2006. Operating theatre records prior to this date were not migrated to the new system. Neither are these records accessible for searching.

PIMS-OR is accessible as a desktop application in each theatre. Theatre personnel use the PatientID to select and update patient records. Batch searches of the database are not supported by this desktop application and can only be carried out by the PIMS-OR administrator.

Procedures are updated using Office of Population Censuses and Surveys (OPCS) codes for interventions and procedures. OPSC is a coding system used in many NHS hospitals for standardising reporting of procedures performed. Standardisation is necessary for audit, governance and billing activity. Code categories apply to a body region e.g. G codes apply to the upper digestive system and H codes apply to the lower digestive system. Three or four numerals follow the letter to further categorise the procedure e.g.:

- Q07.4 Total abdominal hysterectomy not elsewhere classified
- Q07.5 Subtotal abdominal hysterectomy

Data quality

- *Consistency*: PIMS-OR data entries are selected from pre-populated lists. Therefore, descriptions are standardised and no free text is used.
- *Referential integrity:* In PIMS-OR the primary index key for each record is restricted to PatientID only. No records are indexed by private patient aliases or 10-digit NHS identifiers. Therefore, PIMS-OR records can be linked to CDD records with a high degree of confidence in referential integrity.
- *Accuracy*: Theatre coding is often done by the theatre nurses or operating department technicians. Data entry into the system is rarely done by the operating surgeons. There may be errors of omission for complex procedures e.g. fundoplication with pyloroplasty may be recorded simply as fundoplication. To mitigate, a keyword search of CDD was used to cross-reference PIMS-OR data.
- *Completeness*: PIMS –OR only contained prospective data from January 2006. Therefore PIMS-OR was an incomplete source of data particularly for patients born before January 2006. The CDD search was valuable for identifying these patients.
- *Loss of information*: OPSC codes preclude the use of surgical eponyms. However, there is a difference between Nissen and Toupet fundoplication. The difference in these procedures may have an influence on outcomes. There was no unique code for oesophago-gastric dissociation. Although this is a salvage procedure that is rarely performed, it is an important outcome to consider. A keyword search of the CDD was performed to address this data gap.

Search Strategy

We requested the PIMS-OR database administrator to run a batch query to identify patients who had fundoplication at GOSH between 1/1/2006 and 31/12/2010. The batch database query is detailed below (**Table 168**).

Table 168: PIMS-OR database query parameters

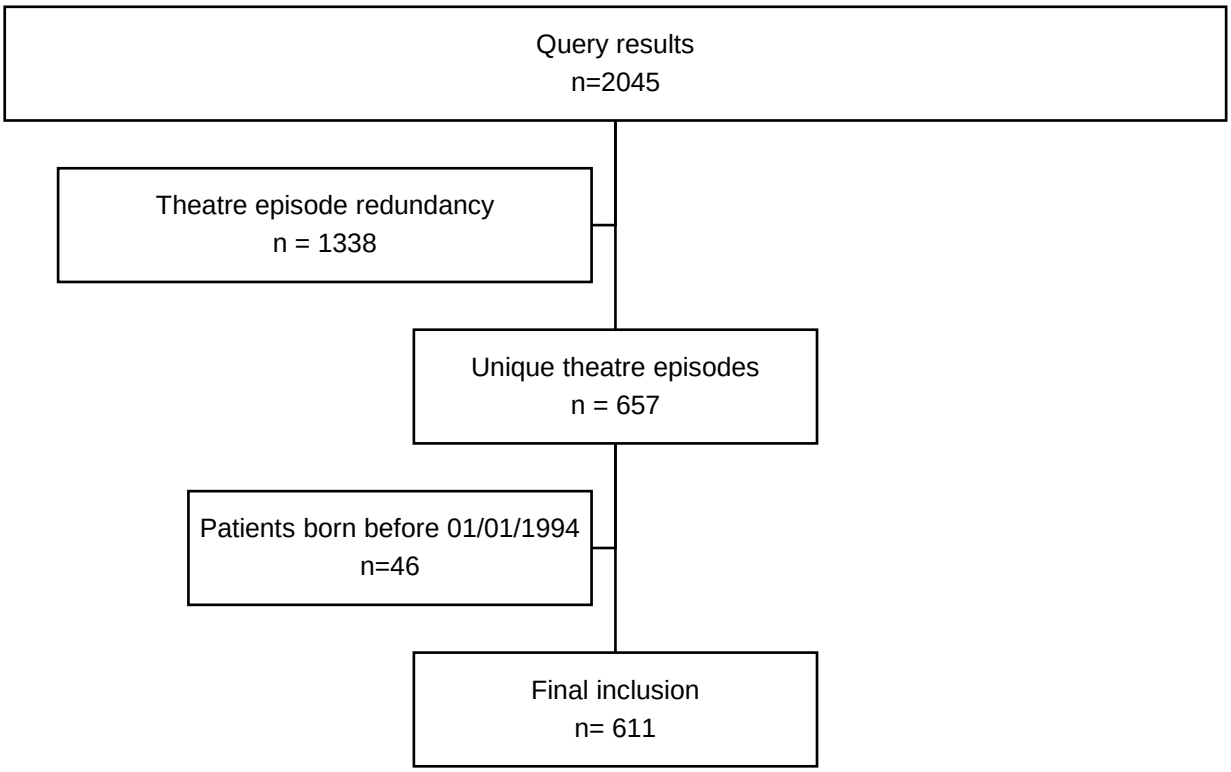
Query	**Field**	**Parameter**
Search parameters	Code	G241: Anti-reflux fundoplication using thoracic approach G243: Anti-reflux fundoplication using abdominal approach G251: Revision of fundoplication of stomach
Limit	From	01/01/2006
Limit	To	31/12/2010
Output fields	• PatientID • Patient Name • DOB • Gender • Date of operation • Operation done • Consultant • Diagnosis	

Results were returned in form of a .csv document. Data were processed in Microsoft Excel, then imported to the Microsoft Access database.

Post-processing and cleaning

The query returned 2045 records. As the admission, discharge and transfer episodes were recorded separately, it was necessary manually sort and delete redundant episodes.

Figure 156: Data cleaning procedures for theatre records



Therefore, between January 2006 and December 2010, 611 anti-reflux procedures were conducted. On reviewing PatientID, we found that these 611 procedures were carried out in 544 unique patients. Therefore, some patients had more than one fundoplication.

These data were imported into the database [RetrospectiveGOR.db] as a table titled 'FUNDO'. This database already contained 469 records of patients who had fundoplication identified through CDD search. In total, CDD and PIMS-OR searches identified 1080 episodes of fundoplication.

Histopathology database: PATHLAB

When a patient is investigated using an oesophago-gastro-duodenoscopy (OGD), the oesophagus is visually inspected for macroscopic signs of GORD. Biopsies from the stomach, duodenum and oesophagus may be taken. These biopsy samples are sent to the pathology laboratory where histopathologists examine them for microscopic signs of GORD i.e. oesophagitis.

Construct and Content

Results of the histopathological analysis are documented and archived in .pdf (Portable Document Format) reports. These reports are uploaded to an electronic database – PATHLAB. The PATHLAB was launched in 2003. Biopsy data prior to cannot be accessed digitally and can only be identified by searching individual patient records. Records in the database are indexed according to the following fields:

Field	Description (Required*)
Episode	A 7 –digit code serial code which is a unique identifier for each biopsy sample. A typical code has a 2-digit prefix representing the year of the biopsy e.g. 08S1320.
Unit Number	PatientID
DOB	DOB
Age	Age (years)
Date received	Date the sample was received in the lab
Name	Patient's full name
Investigation	Categorises the tissue sample by body region. For gastric and oesophageal tissue, this is 'Upper GI'.
Report	A link to a PDF document containing the report.

Search strategy

PATHLAB reports can be accessed through using a dedicated web viewer - OMNIWEB. Clinicians and researchers can view individual patient records by searching for PatientID or DOB. However, it is not possible to perform a whole database searches using OMNIWEB. Database queries can only be searched by administrative users e.g. histopathology scientists and technicians. We requested a histopathology technician to search PATHLAB all records of Upper GI biopsies taken between 1st January 2003 and 31st December 2010. Query results were returned in .pdf format. Results were copied and pasted this data into a Microsoft Excel spreadsheet for manipulation. Results were indexed by PatientID allowing linkage to the other data in [RetrospectiveGOR.db]. However, reports did not contain information on the indication for biopsy and histopathology finding, it was necessary to manually review individual patient reports on OMNIWEB.

Data quality

- Consistency: There was noticeable inconsistency in reporting of oesophageal biopsy findings. Many reports did not include the indication for endoscopic biopsy. It was necessary to clarify the indication by cross-referencing with the patient record in CDD. Biopsies not related to GORD e.g. endoscopy for Crohn's disease were excluded.
- Information loss: A standardised grading system e.g. Los Angeles classification, was not applied for the majority of patients. Subjective statements e.g. "very mild oesophagitis", "negligible inflammation" were evident. To mitigate this inconsistency, reports were coded into a binary variable i.e. oesophagitis Yes/No. In doing so, information about severity of oesophagitis was lost.
- Completeness: Not all patients who have an OGD had an oesophageal biopsy taken. Gastric and duodenal biopsies may be taken where inflammatory bowel disease is considered. By rationalising the episode number, date of sample receipt and PatientID, multiple biopsies were merged into one patient record. .

Pre-processing and cleaning

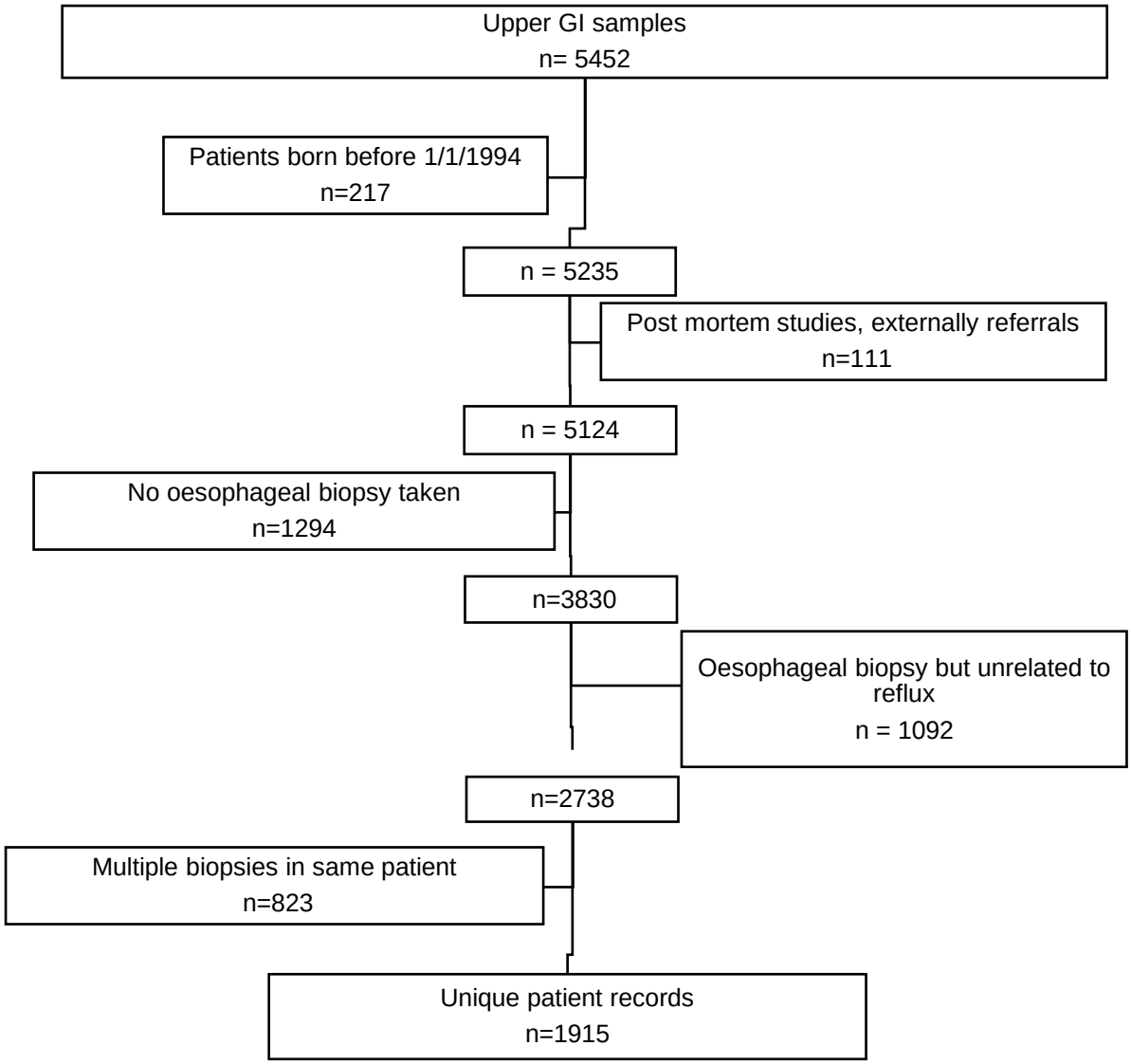
Between 1st January 2003 and 31st December 2010, 5453 oesophageal biopsy samples were received in the histology laboratory. On inspecting the query results, there were some anomalous episode numbers e.g.

G prefix e.g. G06S3306 (n=107)

T prefix e.g. T09S1606 (n=3)

G- and T-prefixed samples are those sent from other hospitals for histopathology review. Typically, these are sent to GOSH histopathologists for a second, expert opinion. We also found one sample coded with a 6-digit episode number starting with 1 e.g. 10S1320. Episode numbers prefixed with 1 are taken during post-mortem examinations. These samples were also excluded from further analysis.

Figure 157: Pre-processing operations for the PATHLAB OGD and biopsy data



In summary, between January 2003 and December 2010, 1915 biopsies to investigate GORD were carried out. These records were included in [Retrospective GOR.db] as a table 'OGD 2003 to 2010' and linked to 'INCLUSION LIST' using PatientID and a left join.

Gastroenterology databases: pH impedance reports

Combined pH-impedance studies have available at GOSH March 2008. Prior to this date, patients underwent pH studies at GOSH. The data from the pH studies were stored on the ambulatory recorder (Digi Trapper). On completing the study, data were transferred via SD card to one of two GOSH computers containing the software for data analysis. Following analysis, the report was available on screen, but was also printed out for the patient's notes. A digital copy of the report was stored in a .coo format.

A .coo file (a.k.a. COOkie file) is a method of data storage developed for printing images. Cookie files contain the coordinates for an image to be printed in a Cartesian space. Cookie files are text editable. Therefore, they are useful for printing reproducible reports where report parameters are fixed but parameter values are valuable. Cookie files cannot be converted into Word or Rich Text Format documents. Therefore, the only method available for reviewing .coo pH study reports was to review individual patient digital or paper notes.

We were unable to find an accurate or comprehensive master list of patients who had pH studies. Therefore, it was decided not to review pH studies as looking through individual patient notes to identify those who had a study was simply not feasible.

When pH-impedance studies were first introduced in March 2008, reports were stored on a shared Gastroenterology Department drive on the GOSH network. However, since October 2009, all pH impedance study reports are uploaded onto CDD as part of the patient record. pH-impedance reports generated prior to this were also uploaded retrospectively. Therefore, pH-impedance data was available on CDD for review.

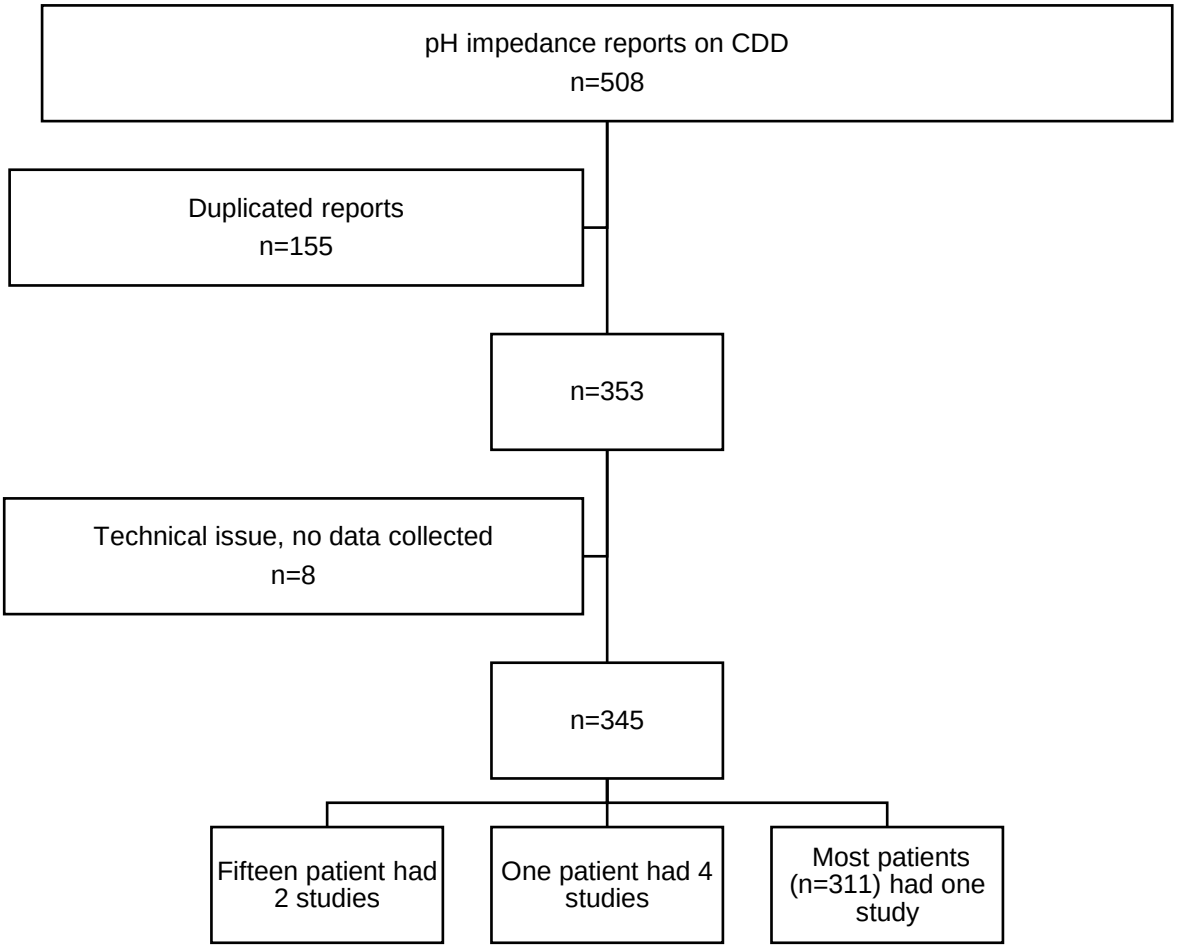
Search strategy

All pH impedance reports are stored as .rtf (Rich Text Format) files on CDD. Text search is possible on .rtf files. pH impedance reports are standardised i.e. same header and field structure. They all contain the text "pH impedance". Therefore, a search for "pH impedance" yielded all pH impedance studies as well as other documents containing this phrase. We then restricted the yield to .rtf format to exclude other records containing this phrase and isolate pH impedance studies.

Data post-processing

Between 2008 and 2010, there were 508 pH impedance reports of pH impedance studies filed on CDD. There was some redundancy i.e. reports loaded onto CDD more than once. Manual review of reports led to further exclusions and detection of patients who had more than one study. We identified 345 pH impedance reports performed in 327 unique patients.

Figure 158: Data post-processing for pH impedance reports



Data from the pH impedance records were included in [RetrospectiveGOR.db] in a table titled 'pH IMPEDANCE'. Data were linked to 'INCLUSION LIST' by PatientID. Data were categorised based on reflux index (continuous variable) and height of reflux (categorical).

SECTION IV APPENDIX ITEMS

R packages used in logistic regression analysis

The packages utilised to automate this analysis were those used and described in the logistic regression analysis for first fundoplication.

1. gmodels(399): contains various programming tools for model fitting. It was applied here for graphical presentation and tabulation.
2. caret(400): was used to automate the data partitioning process.
3. safeBinaryRegression(401): was used to identify linear separation terms.
4. ROCR(261): was used to generate visualisations for performance metrics e.g. ROC curves.

R packages used in decision tree analysis

1. Caret (400): Subsets data into training and testing sets.

2. Matchit (402): Automates selection of matched samples from treated and control groups.

3. Matrix(266): generates extensions of the basic matrix data type to enable data manipulation.

4. Rpart (403): Automates recursive partitioning for classification, regression and survival trees.

5. Rpart.plot (404): An extension package for Rpart for plotting classification and regression trees.

6. Party(405)and Partykit(406): Extension packages for Rpart, utilised for visualisation of DTs.

7. ROCR (261): Used to automate plotting of receiver operating characteristic curves

8. Gmodels(399): applied here for graphical presentation and tabulation

9. Randomforests(262): generates classification and regression ensembles/forests based on random splits.

Student's T-test table

Explain one and two tails briefly here

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

SECTION V APPENDIX ITEMS

REMOS Trial protocol version 2

Gastrostomy with medical treatment *versus* gastrostomy with fundoplication in children with neurological impairment

1. Background

Neurological impairment (NI) is unfortunately an increasing problem in paediatrics. The prevalence of NI is around 0.2% of live births, rising to ~5% of low birthweight infants and to 25% of extreme prematurity survivors Each paediatrician is therefore likely to encounter several children with NI and it is estimated that there are 6000 UK children with significant feeding related health problems due to NI.

Feeding difficulties and gastroesophageal reflux Children with NI frequently experience profound feeding difficulties leading to malnutrition and growth failure. Much of these difficulties can be bypassed by giving nutrition via a gastrostomy. However, the feeding problems can be compounded by pathological gastro-oesophageal reflux (GOR), a well-known problem present in 14-75% of NI children. Complications of GOR include failure to thrive, aspiration pneumonia, oesophageal stricture and anaemia. In these patients, recurrent chest infections are due to aspiration, but are exacerbated by malnutrition and consequent poor immunological status. GORD treatment is based on medical therapy or surgical procedures such as fundoplication. The aims of GOR treatment are to achieve relief of symptoms, prevent complications such as aspiration and pneumonia and achieve adequate nutrition. However, it is not clear whether, in NI children with GOR, a gastrostomy should be associated with medical therapy or with anti-reflux surgery.

Medical therapies for GOR in NI children These include feed thickeners, H_2 receptor antagonists, antacids and prokinetics, and more recently proton pump inhibitors (e.g. omeprazole). These therapies have poor response rates with studies showing that only 13% of NI children respond completely to medical management. Proton pump inhibitors (PPIs) are said to work by decreasing the refluxate acidity and decreasing acid secretion volume, thereby improving gastric emptying. Uncontrolled studies of omeprazole have reported good tolerability and efficacy suggesting high rates of symptom relief and reduction of vomiting in up to 90% of participants. PPIs have been found to have a good long-term safety profile in adults, but only limited information on long-term use in children is available.

Surgical treatment of GOR Historically, medical management of GOR in the child with NI has been characterized by a high failure rate, which has led to the frequent use of surgical anti-reflux procedures. Nissen fundoplication is the commonest surgical procedure, which is designed to mechanically prevent GOR and now frequently performed laparoscopically. Complications include gas bloat, 'dumping syndrome', retching and dysphagia. Moreover, fundoplication is associated with a high recurrence rate and significant morbidity and mortality in this group of children, with a 40% surgical failure rate, and 59% post-operative complications with a 1 to 3% mortality rate.

Controversies There is agreement that NI children with feeding difficulties can benefit from a gastrostomy. However, it is still controversial whether this should be associated with medical therapy of GOR or with fundoplication. Recently, some authors, based on retrospective case-note review, have suggested that a gastrostomy in combination with medical therapy prevents reflux in the majority of

cases. In a prospective study, Sullivan et al. suggested that gastrostomy placement results in a decrease in respiratory tract infection and hospitalization. In our prospective follow-up of a randomized controlled trial we have shown that the recurrence rate of GOR in NI children who underwent laparoscopic or open fundoplication was 7% at 1 year and 18% at 4 years. Vernon-Roberts and Sullivan recently performed a systematic review to compare the effectiveness of anti-reflux surgery and anti-reflux medication for children with NI and GOR who are undergoing gastrostomy placement. They found no randomised controlled trial and concluded there is need for robust scientific evidence to compare risks or benefits of the two interventions.

2. Hypothesis and purpose of the research

The hypothesis to be tested in this randomised controlled trial is that fundoplication is superior to medical treatment of GOR in NI children requiring a gastrostomy. We will measure the effectiveness of surgery or medical treatment on the primary end point of quality of life. In addition, we will measure the effectiveness of the two procedures on secondary end points such as nutritional parameters, percentage of time the oesophageal pH is pathologically acid, non-acid reflux, complications, failure of interventions, and cost of treatment.

3. Plan of investigation

Study design

A randomised controlled trial will be conducted in 60 children with NI requiring gastrostomy and treatment for severe GOR. The patients will receive either: a) laparoscopic gastrostomy with maximal medical treatment for GOR (medical group) or b) laparoscopic gastrostomy and Nissen fundoplication (fundoplication group). The trial will take place at Great Ormond Street Hospital, London. Since both study groups receive a laparoscopic gastrostomy, the only difference between the two arms of the trial will be medical *versus* surgical management of GOR.

To be included in the trial, the child much satisfy each of the following three criteria: 1) neurological impairment with a referral for gastrostomy and/or fundoplication; 2) presence of GOR symptoms including one or more of: vomiting, recurrent aspiration pneumonia (>3 episodes in a year requiring antibiotic administration), anaemia (defined according to WHO) , failure to thrive (weight < 2nd percentile for age for at least 6 months); 3) documented GOR on 24h oesophageal pH monitoring (reflux index ≥10%).

Children with any of the following criteria will be excluded: 1) hiatus hernia (> one vertebral body), intestinal malrotation or gastric outlet obstruction or severely delayed gastric emptying on upper GI contrast study; 2) previous gastrostomy, fundoplication or other abdominal surgery; 3) acute life-threatening events and/or apnoea associated with GOR and requiring urgent operative treatment.

- At baseline (before trial inclusion), clinical assessment will include: a) 24 hour oesophageal pH monitoring, (the gold standard to diagnose GOR[12]); b) upper GI contrast study to document presence of gross GOR, and to exclude gastric outlet obstruction, intestinal malrotation and hiatus hernia. This will be reviewed by the same consultant paediatric radiologist (MH); c) intraluminal oesophageal impedance, carried concomitantly with pH

monitoring, to detect non-acid as well as acid reflux[10]; d) gastric emptying time measured by octanoate breath test; e) gastrointestinal symptoms and quality of life. In patients meeting the inclusion criteria we will also measure nutritional indices including dietary intake, weight and body composition.

- After intervention we will measure: a) 24 hour oesophageal pH monitoring and oesophageal impedance (12 months); b) nutritional indices including calorie intake, weight centile/velocity and body composition (6 and 12 months); c) gastric emptying time measured by octanoate breath test (6 months); d) gastrointestinal symptoms and quality of life (6 and 12 months).

The operations in the trial will be performed by 5 consultant surgeons with considerable experience in the laparoscopic procedures therefore excluding the effects of a learning curve[13].

Laparoscopic gastrostomy

Although percutaneous endoscopic gastrostomy is popular, we and others have described major complications including intestinal perforation and gastrocolic fistula. Laparoscopic gastrostomy has been suggested as a safer alternative and it is routinely performed in our centre. The procedure will be performed according to Jones et al., who found laparoscopic gastrostomy to be feasible, requiring minimal laparoscopic expertise and safe in 112 children over 6 years. Comparison with reports of percutaneous endoscopic gastrostomy indicates that laparoscopic gastrostomy should be the preferred method of gastrostomy placement in children. It may be a particular advantage in NI children, who are often scoliotic. Visualisation of the stomach allows correct gastrostomy placement. In addition, the two randomised groups will differ only for the treatment of reflux (medical vs surgical) and not for the type of gastrostomy performed.

Laparoscopic Nissen fundoplication with laparoscopic gastrostomy

The laparoscopically guided gastrostomy will be also performed as above, and laparoscopic Nissen fundoplication performed as described.

Medical treatment

The medical treatment for GOR will be optimized and follow a standard protocol including omeprazole (omeprazole up to 3mg/kg/d or a maximum of 80mg/d) plus domperidone (up to 2.4mg/kg/d up to 80mg/d). The patients randomised to fundoplication will not receive medical treatment postoperatively unless gastro-oesophageal reflux is re-occurring (see below).

Children in each treatment group may experience complications of gastro-oesophageal reflux, such as recurrent chest infections, vomiting, retching, or retrosternal pain. The group undergoing fundoplication in addition to the gastrostomy can also have gas bloating and blockage of the intestine. All the complications above can happen during routine treatment and are not caused by this study. As indicated above the study will determine which of the two treatments is more effective in reducing the above symptoms.

The **primary end-points** of the study are (i) quality of life using a score specifically designed for NI children undergoing anti-reflux procedures. This will be measured at 6 and 12 months after intervention (Appendix A). (ii) reflux index (percentage of time the oesophageal pH is below 4) on the 24 hour oesophageal pH monitoring at baseline (before randomisation) and 12 months after surgery (or at the time of failure as defined below).

The **secondary end points** of the trial will be:

1. Other parameters derived from the combined 24-hour oesophageal pH and impedance monitoring: number of reflux events (non-acid and acid); number of reflux events in the first two hours after the feed; average reflux height (based on the highest impedance electrode); average minimum proximal and distal pH; total acid clearance time per hour (total time with distal pH<4 relative to study duration) for acid reflux events; total reflux duration per hour (based on the impedance traces); DeMeester score.

2. A record of gastrointestinal symptoms (vomiting, retching, gas-bloating) measured at 6 and 12 months after intervention.

3. Daily antireflux medication (measured at 6 and 12 months after intervention).

4. Gastric emptying:: We have demonstrated a good correlation between the gold standard milk scan (radioactive isotope) and the non-invasive octanoate breath test which is based on a non-radioactive stable isotope (test available in our laboratory). In this trial we will measure gastric emptying time using the octanoate breath test before and after randomization (see above).

5. Complications: operative (i.e. conversion to open procedure, bleeding, intestinal perforation, pneumothorax) and postoperative (wound infection, bronchopneumonia, dysphagia, gas bloating, retching, vomiting, dumping syndrome).

6. Failure of intervention: will be defined as: *either* suspicion of ongoing severe GOR (at least 3 hospital admissions, or one ICU admission, for aspiration pneumonia requiring antibiotic administration), *or* falling weight centile category as defined in the minimization criteria at the 6 month evaluation *or* incapacitating vomiting after at least 6 months *or* herniation of the Nissen wrap (fundoplication group). Patients in both groups who have been defined as failing according to the criteria above will have a 24-hour pH study before considering further treatment. Patients failing in the medical gastrostomy group will be considered for laparoscopic Nissen fundoplication, whereas patients in the fundoplication group will be considered for either medical treatment or re-do fundoplication, at the discretion of the clinician responsible.

7. Nutritional indices (at baseline and at 6 and 12 months postoperatively): a) assessment of dietary intake (carbohydrate, fat and protein) using a 3-day diet diary ; b) weight centile and weight velocity; c) body composition (in routine pre-surgery and follow-up clinics) measured by 3 methods to increase precision: a) *skinfold thickness* which is a traditional method to rank individuals in terms of specific subcutaneous fat depots; b) *bioelectric impedance analysis* which measures impedance of the body to a small electric current to calculate fat free mass; c) *deuterium dilution* to measure total body water allowing estimation of fat free mass. We have normative data for all measures for the ages studied to allow Z-score calculation (JW) and we have expertise in these measurements.

8. Duration of the operation (hours), disposable instruments used (number and type), medications given for GOR and antibiotics during admission into hospital (number and dose), duration of hospital stay (days);

9. Medications for GOR (dose and number during the first postoperative year);

10. Episodes of chest infection (number requiring antibiotic treatment during the first postoperative year),

11. Admissions to hospital after randomisation (number during the first postoperative year);

12. Cost of care during the first year after randomisation (based on variables 8, 9, 10 and 11).

Randomisation

Allocation to groups will be made by weighted minimisation at enrolment into the study using the following criteria: 1) age [<1 year; 1-4 year; >4 year]; 2) severity of reflux index at pH study [5-10%, 10-20%; >20%]; 3) weight Z-score [<-2.2, -2.2 to -1.6 , >-1.6 to -0.5, >-0.5]); 4) triceps skinfold thickness Z-score [<-1.6, -1.6 to 0.5, >0.5]; 5) degree of neurological impairment based on GMFCS score [levels I-III, level IV, level V]; 6) presence of associated anomalies (yes/no); 7) medical treatment for GOR [none, <6 months before entry, >6 months before entry]. Minimisation and randomisation will be performed using our computerised randomisation program designed for randomised clinical trials.

Power calculation

We have previously performed a study following up children who had a laparoscopic Nissen fundoplication, and comparing quality of life before and after, using a published score for neurologically impaired children. In that study, we found a significant improvement of 32±19 points on the QOL scale. Assuming that children treated with gastrostomy plus medical therapy have only a 16 point improvement in QOL score, 30 patients in each arm, as originally proposed using pH study as the primary end point, would give us 90% power at the 0.05 level to detect this difference (Appendix B). In our hospital, we perform a large number of laparoscopic Nissen fundoplications and gastrostomies in NI children; approximately 90 fundoplications in NI children every year. We expect 70 of these children to be eligible for the trial each year and are therefore confident that 60 patients will be recruited in two years.

Organisation and Trial monitoring

In order to maximise compliance in the trial there will be a full time Research Fellow. To ensure that the trial progress is in accordance with Medical Research Council (U.K.) guidelines for good clinical practice in multicentre trials, the following Committees will be established: (1) Data Monitoring and Ethics Committee which will be independent of both the trial organisers and those providing therapy. This committee will perform interim analyses to: a) review assumptions underlying sample size considerations; b) modify or close intake to trial. The DMEC will review all available data after 1 year of recruitment has taken place. If there is overwhelming evidence of superiority of one treatment over the other (failure defined as: either suspicion of ongoing severe GOR, or falling weight centile at the 6 month evaluation or incapacitating vomiting after at least 6 months), we would stop the trial at that time.

(2) Trial Steering Committee which will This will include: i) independent Chairman (not involved in Trial); ii) two independent members (Paediatric Surgeon and Paediatrician); iii) nurse representative; iv)

parents' representative; v) trial co-ordinators (AP, PDC and SE); 6) representative of the Data Monitoring and Ethics Committee. A statistician will attend meetings as appropriate. The role of this Committee is to provide overall supervision of the trial and ensure that the trial is conducted to rigorous scientific, clinical and ethical standards.

4. Time frame to clinical application and number of patients likely to benefit

This trial addresses a contentious question concerning the treatment of NI children with GOR: medical or surgical? The clinical application of this study will be available as soon as the results of the trial will be analysed. GOR is one of the most common gastro-intestinal diseases in children and a condition encountered on a weekly basis in children's hospitals. It is one of the most frequent gastro-intestinal conditions requiring surgery in children, and is particularly common in neurologically impaired children. As described in the background above, around 6000 children in the UK have significant feeding related health problems due to NI, so it is expected that this trial will have immediate and profound impact on the treatment of many children in the UK and abroad.

5. Research team's ability to conduct the research

The UCL Institute of Child Health is an organisation well recognised nationally and internationally for its contribution to scientific research (gaining a 5∗A award in the last RAE). The Department of Paediatric Surgery has acquired an international reputation for translational research in surgical diseases of infants and children. We have created a collaborative group with the Departments of Neurology (Prof. F. Muntoni), gastroenterology (Dr N Shah), Nutrition (Dr J Wells), Radiology (Dr M Hiorns) and Dietetics (Ms V Shaw) to determine whether anti-reflux medical therapy or anti-reflux surgery is most efficacious in the treatment of NI children with significant GOR. We have established a clinical trial subunit responsible for randomised controlled trials, 3 multicentre and 2 based at Great Ormond Street Hospital and we have forged active collaboration with the Epidemiology and Biostatistics Unit. We therefore have the necessary experience of running trials and have written our own software for randomisation using minimisation.

6. Justification for the support requested

The proposed trial requires clinical expertise hence, we are requesting 3 year salary of a Specialist Registrar in Paediatric Surgery. The consumable costs are a realistic estimate based on the above power calculations. The equipment necessary for this project (oesophageal impedance and bioelectrical impedance for body composition) will be exclusively dedicated to the proposed trial.

Appendix

A: Quality of life: scoring based on the following factors (O'Neill JK, O'Neill PJ, Goth-Owens T, Horn B, Cobb LM: Care-giver evaluation of anti-gastroesophageal reflux procedures in neurologically impaired children: what is the real-life outcome? J Pediatr Surg 1996; 31:375-380**)**

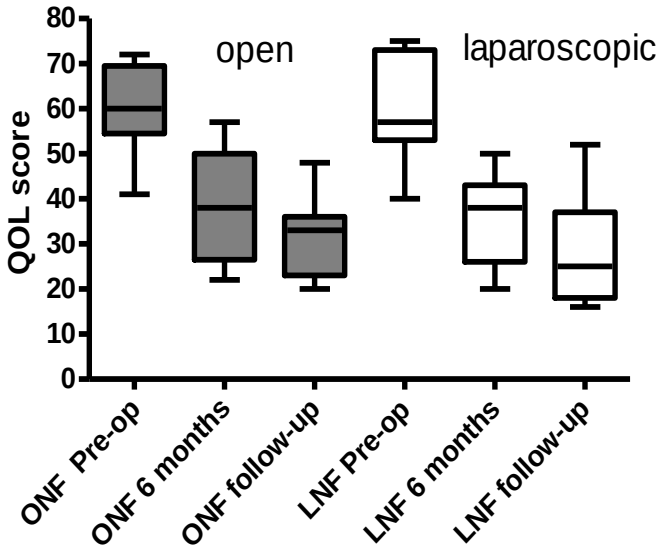
1. Parental attitude regarding daily care and the overall condition of their child (including: ease of feeding; physical comfort during feeding; problems with bowel conditions; child's problem with gas-bloat syndrome; pneumonia or other pulmonary condition; comfort of the child; child's ability to enjoy life; child developmental problem).

2. Parental attitudes regarding their child and the parents' overall quality of life (including: overall ease of caring for the child; overall enjoyment of the child; quality of time spent with the child; level of frustration in taking care of the child; level of concern about ability to properly take care of the child; overall quality of life).

3. Parental thoughts regarding their preparation for treatment (medical vs surgical) and outcome of treatment.

4. Parenting stress index-depression scale

B: Quality of life score in our randomised controlled trial comparing open and laparoscopic fundoplication



REMOS Trial Parent information leaflet

Great Ormond Street Hospital for Children NHS Trust

Gastrostomy with medical treatment versus gastrostomy with fundoplication in children with neurological impairment

Your child is being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the aim of the study

Children with neurological impairment often have eating difficulties and require surgery to place a feeding tube into their stomach (gastrostomy). These feeding difficulties can be caused by reflux of stomach acid up into the oesophagus leading to malnutrition and recurrent chest infections. The reflux can be treated with either antireflux surgery at the same time as the gastrostomy or with anti-reflux medication. All children in the study will have a surgical gastrostomy to help with feeding; the aim of this study is to compare the effectiveness of an additional anti-reflux operation (fundoplication) with that of anti-reflux medication (anti-acids).

Why has your child been chosen?

We aim to study 60 children with neurological impairment and eating difficulties who require treatment of reflux of stomach acid (vomit).

Does your child have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care your child will receive.

What will happen to your child during the trial?

This trial is a *randomised control trial*. Sometimes because we do not know which way of treating patients is best, we need to make comparisons. Infants selected will be put into two groups and then compared. We aim to give one group a surgical gastrostomy for feeding plus medical treatment for acid reflux, and another will receive a surgical gastrostomy for feeding plus an operation for the acid reflux. The groups are selected by a computer which has no information about the individual - i.e. by chance.

To investigate the effects of medical treatment or operation, we will record the data of the tests which are normally done before and after treatment independently of this trial. These include 24-hour oesophageal pH recording (measuring the acid in the gullet by a small tube), dietary intake, weight gain,

amount of body fat (measuring the fat under the skin), occurrence of respiratory tract infections, and failure of treatment. In addition, for this trial, we will record: 1) quality of life using a questionnaire after 6 and 12 months from intervention; 2) composition of the body (how much water, fat and muscle there is in the body) with two methods a) by giving some special and completely safe water and measuring the amount of water in the urine or saliva and b) measuring the water under the skin with an instrument that does not cause pain or discomfort and it is completely safe.

Are there any benefits to my child taking part in this trial?

There are no immediate benefits to your child from participating in the trial. The aim of this trial is to decide which is more effective to control acid reflux: surgical gastrostomy plus medical treatment or surgical gastrostomy plus fundoplication. We expect that this trial will help in the future hundreds of children with similar problems.

Are there any risks to my child?

Children in each treatment group may experience recurrent chest infections, vomiting, retching, or pain in the gullet. The group undergoing fundoplication in addition to the gastrostomy can also have gas bloating and blockage of the intestine. All the complications above can happen during the routine treatment of your child and are not caused by this study. As indicated above the study will determine which of the two treatment is more effective in reducing the above symptoms. This trial will look for any possible side effects of surgery or medical treatment, and there is an independent panel set up to monitor this.

What are the arrangements for compensation?

This research project has been approved by an independent Research Ethics Committee who believes that it is of minimal risk to your child. However, research can carry unforeseen risks and we want you to be informed of your rights in the unlikely event that any harm should occur as a result of taking part in this project.

This research is covered by a no-fault compensation scheme which may apply in the event of any significant harm resulting to your child from involvement in the project. Under this scheme it would not be necessary for you to prove fault. You also have the right to claim damages in a court of law. This would require you to prove fault on the part of the Hospital/Institute and/or any manufacturer involved.

Will any information collected be kept confidential?

Data collected for this project will be transferred to the Institute of Child Health, University College London. All information which is collected about your child during the course of the research will be kept strictly confidential. Any information about you which leaves the hospital will have your name and address removed so that your child cannot be recognised from it.

What will happen with the results of the trial?

The results will be analysed at the end of the study. We will aim to publish any findings in medical journals, and present at relevant medical conferences. None of the data presented will be identifiable as belonging to any child.

Who is funding this research?

The children's charity, SPARKS (Sport Aiding medical Research for Kids) is funding this trial. They provide money for running the tests as well as providing a salary for the researcher coordinating the trial.

Who has reviewed the study?

An independent panel have reviewed the study on behalf of SPARKS, and it has been subjected to rigorous ethical review by the Local Research Ethics Committee at Great Ormond Street and the Institute of Child Health

Who can I speak to for more information?

Chief Investigator:

Professor A Pierro MD FRCS(Eng) FRCS(Ed) FAAP

Nuffield Professor of Paediatric Surgery

Head of Surgery Unit

Institute of Child Health

30 Guilford Street

London WC1N 1EH

Telephone : 020 7905 2175/2641

Fax :020 7404 6181

email: a.pierro@ich.ucl.ac.uk

secretary:pierro.sec@ich.ucl.ac.uk

Thank you for your participation in this important trial.

REMOS Trial Parental Consent Form

Study Number:

Trial ID:

Great Ormond Street Hospital for Children NHS Trust

Gastrostomy with medical treatment versus gastrostomy with fundoplication in children with neurological impairment

Please initial box

1 I confirm that I have read and understand the information sheet dated 22/09/2008 version 1.0) for the above study and have had the opportunity to ask questions. ☐

2 I understand that my child's participation is voluntary and that I am free to withdraw my consent for my child's participation, without giving any reason, and without my medical care or legal rights being affected. ☐

3 I understand that sections of any of my child's medical notes may be looked at by responsible individuals from the Institute of Child Health or from regulatory authorities where it is relevant to taking part in research. I give permission for these individuals to have access to my child's records. ☐

4 I agree that my child can take part in the above study. ☐

5 I give permission for my child's General Practitioner to be contacted. ☐

6 I give my permission for the researchers to contact me in the future if there is an ethically approved long-term follow-up of this study. ☐

Name of Parent	**Date**	**Signature**
Name of Person taking consent (if different from researcher)	**Date**	**Signature**
Researcher	**Date**	**Signature**

Copies: 1 for patient; 1 for researcher; 1 to be kept with hospital notes

File Note- October 2010

TRIAL INFORMATION	
Trial	Gastrostomy with medical treatment versus gastrostomy with fundoplication in children with neurological impairment (REMOS)
R&D Number	R&D ref: 07SG34
Chief Investigator Principal Investigator	Professor Pierro Miss E Macharia

Issue 1: Measuring reflux

The study protocol describes use of 24 hr pH monitoring to measure reflux

The inclusion criteria are:

1. Neurological impairment with a referral for gastrostomy and/or fundoplication
2. Presence of GOR symptoms
3. Documented GOR on 24h oesophageal pH monitoring (reflux index ≥10%).

Oesophageal impedance recording is superior to 24 hour oesophageal pH monitoring as a measure of GOR. pH monitoring alone records only acid reflux events. Non-acid reflux comprises 40-50% of reflux events in milk-fed children, and children with neurological impairment. Therefore, oesophageal pH monitoring under-estimates reflux events by 50%. Oesophageal impedance recording is superior to 24 hour oesophageal pH monitoring as a measure of GOR. It provides standard pH study information (reflux index) as well as superior information on reflux events, non-acid reflux and symptom association. Therefore, it shall be used to quantify reflux on this study.

Preventative & Corrective Actions

Therefore, the inclusion criteria will be amended to read;

1. Neurological impairment with a referral for gastrostomy and/or fundoplication
2. Presence of GOR symptoms
3. Documented GOR on 24h pH **or** oesophageal pH/impedance monitoring.

A diagnosis of GOR on pH/impedance monitoring will be based on a skilled, qualitative and quantitative assessment of the investigation. This will take into account key parameters including reflux events, longest acid exposure events, oesophageal acid exposure time and symptom association.

To reflect our ability to measure non-acid reflux events, the minimization criteria will more accurately read

1. reflux index [<10%, 10-20%, > 20%]

Issue 2: Clarifying the use of triceps fold thickness

Triceps skin fold thickness, is detailed in the protocol as a minimization criteria.

4) triceps skinfold thickness Z-score [<-1.6, -1.6 to 0.5, >0.5];

However, on reviewing the minimization algorithm, this variable was never included in the minimization software criteria. Furthermore, as a distributive factor, weight is a better parameter than triceps thickness as there is less operator variability. Lastly, there is a measure of redundancy between triceps skinfold thickness and weight.
Preventative & Corrective Actions
Triceps skin fold thickness z-score will not be utilized as a minimization criteria. Triceps thickness will remain a secondary outcome measure.
Issue 3: Combined oesophageal pH+ impedance+manometry recording
Transient, inappropriate relaxation of the lower oesophageal sphincter is thought to be a mechanism underlying gastro-oesophageal reflux. Inappropriate relaxation may be triggered e.g. by air distension of the stomach, or may be secondary to underlying neurological impairment. The skeletal muscle of the proximal oesophagus is innervated by motor neurons arising from the nucleus ambiguous. The distal oesophagus is innervated by the nerves myenteric plexus within the smooth muscle layers. The myenteric plexus is , in turn, innervated by the dorsal motor nucleus of the vagus nerve. Recently, it has been demonstrated that the oesophageal peristalsis is actually comprised of two separate proximal and distal waves, separated temporally and spatially by the transition zone. The transition zone is the area where, on histology, oesophageal musculature changes from striated (proximal) to smooth (distal) muscle. The transition in musculature is also marked by a transition in innervations. There is evidence that the transition zone, as well as the lower oesophageal sphincter is implicated in reflux and dysphagia. A long (> 2cm) transition zone, or a long gap (>1cm) in proximal and distal peristalsis is strongly associated with dysphagia. In children with neurological impairment, reflux may be due to abnormal oesophageal motility arising from disco-ordinate peristalsis. Disco-ordinate peristalsis, therefore, may arise from disordered innervations or disordered musculature. Oesophageal manometry, will allow us to fully assess the role of oesophageal motility in GORD. Combined pH+ impedance+ manometry catheters are available. Therefore, **for the patient, the pathway will be identical, save for the additional measurement of pressures during swallowing.** A single catheter will be inserted. Catheter position will be confirmed on Xray as before, as well as on manometry. This catheter will initially be used to record pressure profiles. To do this, oesophageal pressures will be recorded for 10 swallowing events. The probe can then be left in situ for the standard 24 hour pH/impedance recording. There will be cost implications for the trial. However, these can be justified by the additional quantitative information gained about peristalsis, the lower oesophageal sphincter and the transition zone. Lastly, this addition presents an opportunity to be a pioneering institution in using impedance manometry to assess children with reflux.
Preventative & Corrective Actions
In addition to oesophageal impedance, we shall also perform oesophageal manometry.
Issue 4: Removing the impedance catheter after the study
From the feedback we have received so far, it is clear that it is expensive and inconvenient for parents to return after 24 hours to have the naso-gastric impedance catheter removed. In some cases, parents have had training

in passing and removing naso-gastric tubes, and will often manage their child's naso-gastric intubation independently.		
Preventative & Corrective Actions		
Where it is clear that a parent has been trained and is confident in managing their child's nasogastric tube, we shall allow them to remove the impedance catheter at the end of the study and make arrangements to have the catheter couriered back to the unit.		
SIGNATURES		
Author		
Print	Signature	Date
Authorised by (if required)		
Print	Signature	Date

Notice of substantial amendment (non-CTIMP)

For use in the case of all research other than clinical trials of investigational medicinal products (CTIMPs). For substantial amendments to CTIMPs, please use the EU-approved notice of amendment form (Annex 2 to ENTR/CT1) available in the Integrated Research Application System (IRAS) at http://www.myresearchproject.org.uk or on the EudraCT website at https://eudract.ema.europa.eu/document.html.

To be completed in typescript by the Chief Investigator in language comprehensible to a lay person and submitted to the Research Ethics Committee that gave a favourable opinion of the research ("the main REC"). In the case of multi-site studies, there is no need to send copies to other RECs unless specifically required by the main REC.

Further guidance is available at http://www.nres.nhs.uk/applications/after-ethical-review/notification-of-amendments/.

Details of Chief Investigator:	
Name:	Professor Agostino Pierro
Address:	Surgery Unit, UCL Institute of Child Health 30 Guilford Street, London
Postcode:	WC1N 1EH
Telephone:	0207 905 2641
Email:	Pierro.sec@ucl.ac.uk
Fax:	0207 4046181

Full title of study:	Gastrostomy with medical treatment versus gastrostomy with fundoplication in children with neurological impairment.
Lead sponsor:	SPARKS
Name of REC:	Institute of Child Health/ Great Ormond Street Hospital Research Ethics Committee.
REC reference number:	08/H0713/99
Name of lead R&D office:	Joint Institute of Child Health/ Great Ormond Street Research and Development Office
Date study commenced:	February 2010
Protocol reference *(if applicable),* current version and date:	Version 2 0 7/11/2008
Amendment number and date:	Amendment 1 10/12/2012

Type of amendment (indicate all that apply in bold)

(a) Amendment to information previously given on the REC Application Form

No

If yes, please refer to relevant sections of the REC application in the "summary of changes" below.

(b) Amendment to the protocol

Yes

If yes, please submit either the revised protocol with a new version number and date, highlighting changes in bold, or a document listing the changes and giving both the previous and revised text.

(c) Amendment to the information sheet(s) and consent form(s) for participants, or to any other supporting documentation for the study

No

If yes, please submit all revised documents with new version numbers and dates, highlighting new text in bold.

Is this a modified version of an amendment previously notified to the REC and given an unfavourable opinion?

No

Summary of changes

Briefly summarise the main changes proposed in this amendment using language comprehensible to a lay person. Explain the purpose of the changes and their significance for the study.

If this is a modified amendment, please explain how the modifications address concerns raised previously by the ethics committee.

If the amendment significantly alters the research design or methodology, or could otherwise affect the scientific value of the study, supporting scientific information should be given (or enclosed separately). Indicate whether or not additional scientific critique has been obtained.

…….

In the previous protocol (version 2), the following secondary end-points were to be measured:

1. Gastric emptying using octanoic acid
2. Nutritional indices: Using *deuterium dilution* to measure total body water and hence estimate fat free mass

Safety standards necessitate the administration of medical-grade octanoic acid and deuterium. Commissioning the provision of medical grade octanoic acid and deuterium is cost prohibitive. We have balanced the risks of administration of non-medical grade products with the benefits to the study and concluded that the best course of action is to omit these tests from the study. As these parameters are secondary outcomes, omission of these tests will not impair the methodology or alter the scientific value of this study.

……..

In the previous protocol (version 2), triceps skinfold thickness Z-score was to be used as a minimization criteria for randomization [<-1.6, -1.6 to 0.5, >0.5].

Triceps skin fold thickness, although included in the protocol, has not been utilized in the minimization software criteria. However, as a distributive factor, weight is a better parameter than triceps thickness as there is less operator variability. As there is a measure of redundancy between these two measures, triceps thickness will not be utilized as a criteria.

Triceps skin fold thickness z-score will not be utilised as a minimisation criteria.

……

In the previous protocol (version 2), severity of reflux index at pH study is used as a minimisation criteria. Severity is categorised as [5-10%, 10-20%; >20%].

Oesophageal impedance recording is superior to 24 hour oesophageal pH monitoring as a measure of GOR. pH monitoring alone records only acid reflux events. Non-acid reflux comprises 40-50% of reflux events in milk-fed children, and children with neurological impairment. Therefore, oesophageal pH monitoring under-estimates reflux events by 50% and is not a suitable measure of severity of reflux.

The reflux index criteria will more accurately read [<10%, 10-20%, > 20%]

Any other relevant information

3. Applicants may indicate any specific ethical issues relating to the amendment, on which the opinion of the REC is sought.

Cost of medical-grade octanoate

Cost of medical -grade deuterium

List of enclosed documents

Document	*Version*	*Date*
Protocol	Version 3	10/12/2012

Declaration by Chief Investigator

- I confirm that the information in this form is accurate to the best of my knowledge and I take full responsibility for it.

- I consider that it would be reasonable for the proposed amendment to be implemented.

Signature of Chief Investigator: ……

Print name: Professor Agostino Pierro

D*ate of submission:* …….

Declaration by the sponsor's representative

The sponsor of an approved study is responsible for all amendments made during its conduct.

The person authorising the declaration should be authorised to do so. There is no requirement for a particular level of seniority; the sponsor's rules about delegated authority should be adhered to.

- I confirm the sponsor's support for this substantial amendment.

Signature of sponsor's representative: …….……

Print name: *…….……*

Post: *…….……*

Organisation: *…….……*

Date: *……..*