Variations in GP decision making in the diagnosis of lung cancer

Rachel Elizabeth Sequeira

Department of Applied Health Research, UCL

SUBMISSION FOR PHD IN APPLIED HEALTH RESEARCH

Supervisors

Professor Rosalind Raine

Professor Stephen Duffy

Dr Jessica Sheringham

Funder

Department of Health Policy Research Unit for
Cancer Awareness, Screening and Early Diagnosis

I, Rachel Elizabeth Sequeira confirm that the work presented in this thesis is my own. Where information has been derived from other sources or conducted by other individuals/groups, I confirm that I have indicated this in the thesis.

Abstract

Background

The United Kingdom's lung cancer patients have lower survival than patients in comparable countries. Delays in diagnosis may contribute to this. There are significant socio-demographic variations in the interval between cancer patients first presenting to their general practitioner (GP) and referral, but it is unclear why these exist.

Aim

To examine patient and GP characteristics associated with GPs' referral decisions, focusing on patients with symptoms indicative of lung cancer.

Methods

Study 1: Systematic literature review considering non-clinical patient, GP and practice characteristics associated with variations in GPs' referral of patients for investigations or to secondary care.

Study 2: GP decision making study: a factorial experiment using interactive multimedia vignettes to examine GPs' decisions to refer patients with symptoms indicative of lung cancer, and a survey to examine factors influencing decision making.

Results

Study 1: 11,791 titles were screened; 47 were of sufficient quality and relevance for inclusion. There was strong evidence that patients over 75 were less likely to be investigated or referred, and of variations by patient gender. However few higher quality studies examined associations with patient ethnicity and GP or practice characteristics, or considered why socio-demographic variations occurred.

Study 2: 227 GPs completed the study. GPs were less likely to investigate older than younger patients, and black patients than white. The survey identified several factors that GPs believe affect their referral decisions (such as patients' lifestyles), some of which may explain the observed differences in GPs' referral decisions.

Conclusions

My thesis identified socio-demographic variations in GP decision making that are independent of clinical characteristics (for lung cancer and more widely) and factors that may underlie these. Further research addressing the extent to which these factors contribute to socio-demographic variations, and the development of primary care interventions which address these findings, could reduce delays in lung cancer diagnosis.

Table of Contents

Index of Figures

Index of Tables

Glossary of abbreviations used in the thesis

A&E	accident and emergency department at a hospital
BMI	body mass index
CA-125	blood test for a cancer protein produced by some ovarian cancers
CAMHS	child and adolescent mental health services
CAPER studies	Cancer Prediction in Exeter studies: a group of primary care case-control research studies
CASP	the Critical Appraisal Skills Programme: providing tools to assist in the critical appraisal of research studies
CBT	cognitive behavioural therapy: a talking therapy used in treatment of some mental health conditions
CCG	clinical commissioning group: these commission most of the hospital and community healthcare services in the local area that they are responsible for (they replaced primary care trusts in 2013)
CHD	coronary heart disease
CI	confidence intervals: a range of values likely to include a certain population parameter
CONDUIT	Cutting Out Needless Deaths Using IT programme: a South-West London database programme used for research
COPD	chronic obstructive pulmonary disease

CT	computerised tomography: a type of X-ray imaging producing detailed images
CXR	chest X-ray
DM	diabetes mellitus
DSM	Diagnostic and Statistical Manual of Mental Disorders: a standard classification of mental health conditions
ECG	electrocardiogram: a test recording the heart's electrical activity
ENT	ear, nose and throat: a hospital speciality
EOI	an expression of interest from a general practitioner interested in taking part in our research study
F1	foundation year 1 doctor (newly qualified)
GP	general practitioner: a doctor specialising in general practice
GPRD	General Practice Research Database (now known as Clinical Practice Research Datalink, CPRD)
HADS	Hospital Anxiety and Depression Scale: used to determine the levels of depression and anxiety a patient is experiencing
HbA1c	glycosylated haemoglobin: a blood test used to assess the average blood glucose concentration over several weeks
HDL	high-density lipoprotein cholesterol: a blood test forming part of the cholesterol panel of tests
HIV	human immunodeficiency virus, the cause of HIV infection

IAPT	Improving Access to Psychological Therapies programme: a National Health Service programme offering interventions for people with depression and anxiety disorders
IBS	irritable bowel syndrome
ICBP	International Cancer Benchmarking Partnership: a partnership of clinicians, academics and policy-makers studying variations in cancer survival
IMD	Index of Multiple Deprivation: a United Kingdom government study of deprivation in England
IT	information technology and the use of computer software
LVD	left ventricular dysfunction (of the heart)
MeSH	Medical Subject Headings: a vocabulary used to index journal articles and books
MRI	magnetic resonance imaging: a scan using magnetic fields to produce detailed images
NACDPC	National Audit of Cancer Diagnosis in Primary Care: undertaken in 2009/2010 in England
NAEDI	National Awareness and Early Diagnosis Initiative: an initiative to co-ordinate and support research and projects to improve early cancer diagnosis in England
NHS	National Health Service: the United Kingdom's publically funded healthcare system
NICE	National Institute for Health and Care Excellence: provides national guidance and healthcare advice

NLST	National Lung Screening Trial: a research trial in the United States of America
OOH	out-of-hours services: provide healthcare services outside of normal general practice surgery hours
OR	estimated odds ratio: the odds that an outcome will occur given a particular exposure
p value	the probability of finding the observed results of an analysis if the null hypothesis is true
PCRN	primary care research network: regional networks that provide infrastructure for primary care research
PCT	primary care trust: administrative bodies responsible for commissioning health services within a local area (replaced in 2013 by clinical commissioning groups)
PDF	portable document format: a file format used to present documents on a computer
PHQ-9	patient health questionnaire, 9 item version: a questionnaire used for diagnosing, monitoring and measuring the severity of depression
PMB	post-menopausal bleeding
PPV	positive predictive value: the probability that an individual has a disease
PRU	Policy Research Unit in Cancer Awareness, Screening and Early Diagnosis: a programme of studies to inform and evaluate policies to improve cancer outcomes in the United Kingdom

QMAS	Quality Management and Analysis System: a computer system previously used by the National Health Service
QOF	Quality and Outcomes Framework: a voluntary annual incentive programme for general practices in England
QRESEARCH	a general practice research database
RAT	risk assessment tool: tools developed to assist general practitioners in selecting which patients to send for investigation
SEC	socio-economic circumstance: a combined sociological and economic measure of an individual or household's social and economic position in relation to others, often based on a number of factors
SHO	senior house officer doctor (a level of junior doctor)
SIGN	Scottish Intercollegiate Guidelines Network: develop evidence based clinical practice guidelines
STI	sexually transmitted infection
TB	tuberculosis
TWW	two week wait referral pathway: an urgent referral route for patients with suspected cancer
UCL	University College London
UK	United Kingdom
USA	United States of America

1 : Background

1.1 : Introduction

"To achieve our ambition that cancer mortality and survival rates should match the best, it will be essential to prevent more cancers developing in the first place and to ensure they are diagnosed while the cancer is at an earlier stage. Tackling inequalities will be fundamental to this." - *Improving outcomes: a strategy for cancer (Department of Health, January 2011)*[1]

Cancer is the leading cause of mortality in the United Kingdom (UK),[2] and was responsible for 29% of all deaths in England and Wales in 2014.[3] Cancer incidence is also increasing, such that one in two people in the UK born after 1960 will be diagnosed with cancer during their lifetime.[4]

With both cancer incidence and mortality rising, improving outcomes for patients with cancer is a key initiative for the UK government and the National Health Service (NHS). The importance of research in this field is widely recognised: the National Clinical Research Institute (a UK-wide partnership between cancer research funders) recorded that £498 million was spent on cancer research funding in 2015.[5]

A number of research programmes have been established to address how to improve cancer outcomes:

- *The Department of Health's Policy Research Unit in Cancer Awareness, Screening and Early Diagnosis (PRU)*
  Established in 2011 by the Department of Health. A programme of studies to inform and evaluate policies to improve UK cancer outcomes, focusing on the initial part of the cancer pathway (pre-treatment). Both the general practitioner (GP) decision making study and my PhD are funded by the PRU programme, within the 'early diagnosis' strand.
- *International Cancer Benchmarking Partnership (ICBP)*
  A global programme of clinicians, academics and policy-makers from six countries, established to consider how and why cancer outcomes vary between countries. It is funded by a number of partners, including the Department of Health and the National Cancer Action Team.

- *National Awareness and Early Diagnosis Initiative (NAEDI)*
  A joint initiative between the Department of Health and Cancer Research UK established in response to the Department of Health's 2007 Cancer Reform Strategy, and funded by a consortium of partners brought together by the National Cancer Research Institute.
- *The Accelerate, Coordinate, Evaluate (ACE) Programme*
  A more recent early diagnosis initiative between Cancer Research UK and Macmillan Cancer Support, established in 2014 to support the NHS outcome of 'preventing people from dying prematurely'.

The aim of my PhD is to provide a small piece in the jigsaw of 'improving cancer outcomes'. I focus on variations in the early diagnosis of lung cancer, specifically the role that GP decision making plays in this.

There are a number of reasons to focus on lung cancer.

- It is the second most commonly diagnosed cancer in the UK,[6] and the leading cause of cancer mortality (responsible for over 21% of the UK's cancer deaths in 2014).[7]
- Many cancers' survival rates have improved dramatically over the last 25 years, but the survival rate for lung cancer has remained low.[8] One reason for this is that surgery is the only curative treatment for most lung cancers, but can only be performed in early stage disease. Data from the National Cancer Intelligence Network (NCIN) shows that in 2015 just 23% of lung cancers were diagnosed at an early stage (defined here as stage 1 or 2 disease), compared to 54% of all cancers.[9] Improving early diagnosis of lung cancer therefore has the potential to increase survival.
- Lung cancer survival rates vary across the UK population (see Section 1.2). There is also some evidence of variation in both diagnosing and treating UK lung cancer patients (see Section 1.5) although the evidence is not consistent. It remains unclear where in the diagnostic and treatment pathway the variation that leads to these survival differences occurs.

I will now give a brief overview of the current knowledge about early diagnosis of cancer, in particular regarding the GP's role in this. This will highlight the relevance and importance of the two studies that make up my PhD. I will also describe the scope and parameters of my PhD.

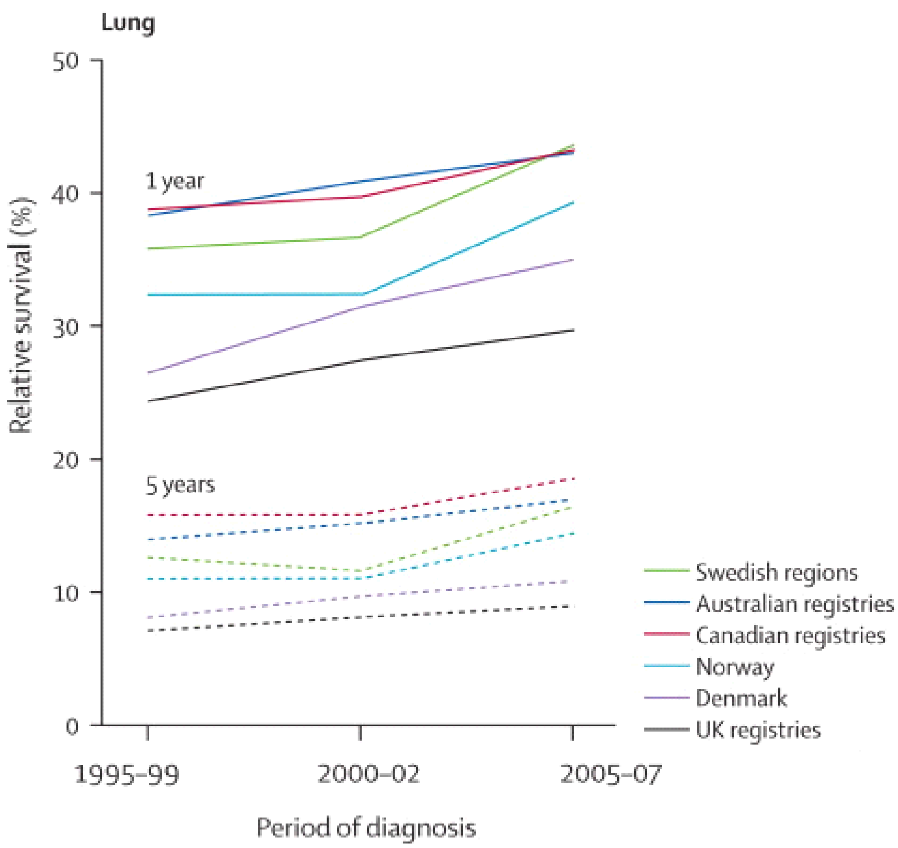
1.2 : Disparities in cancer survival

There is significant variation in cancer survival internationally and within the UK.

1.2.1 : International variation in survival

The UK's cancer survival rates lag significantly behind those in comparable countries.[10] Coleman et al (2011)[10] reported six different countries' trends in one and five year survival rates[i] for four of the most common cancers (including lung, see Figure 1) between 1995 and 2007. Whilst relative survival improved over time in all countries, the UK consistently performed worst for all cancers. This inter-country variation was particularly marked for one year survival, and the inequalities were greatest for patients 65 years and older.

Figure 1 : Age-standardised one and five year relative survival trends for lung cancer between 1995 and 2007 by country *(Coleman et al 2011)*[10]



[i] Where 'survival rate' is a measure of those patients in a group who survive for a defined period of time, expressed as a proportion of all those in the group alive at the beginning of the time period.

Abdel-Rahman et al (2009)[11] estimated that at least 6,500 cancer-related deaths could be avoided each year in the UK if our survival rates matched the mean in Europe. Survival in the UK matching those countries with the highest rates of survival would equate to avoidance of over 11,000 excess premature deaths in the UK annually, representing over 10% of cancer-related mortality.

Results from the ICBP studies (2013)[12] suggest that in recent years the 'survival gap' between the UK and the best-performing countries may have begun to reduce in breast cancer. However there is no evidence that the proportion of excess deaths is decreasing in lung cancer,[12] and for the oldest patients there is evidence that the survival gap between the UK and the best-performing countries is actually increasing.[13]

1.2.2 : UK variation in survival

NHS England, NHS Scotland, NHS Wales and Health and Social Care Northern Ireland (HSCNI) together provide healthcare for the whole UK population, based on the following principles:[14]

- to provide a comprehensive service available to all;
- that access to services is based on an individual's clinical need;
- to aspire to the highest standards of excellence.

Given these principles, it is therefore perhaps surprising that variation in cancer survival exists within the UK. However there is evidence of significant variation between different population groups. Examples include:

Gender	The effect of gender on survival varies between cancer types.[15;16] For lung cancer, men have a lower survival rate than women.[17]
Socio-economic circumstance	For the majority of cancers, relative survival rates are lower for the most socially disadvantaged patients - even when their higher rates of all-cause mortality are accounted for.[18] In lung cancer the survival gap between the most and least disadvantaged patients has actually increased with time.[19]

Age	As age at diagnosis increases, relative survival rates decrease for almost all cancers, including lung - even when accounting for higher rates of all-cause mortality in older age groups.[18;20]
Ethnicity	There is some evidence that patients of Asian ethnicity have higher survival rates than patients of white ethnicity, but there is no significant difference in survival rates between black or white ethnic groups.[21] Ethnicity information is not available for a significant proportion of patients with cancer, so it is important to exercise caution in interpreting these data.
Region	There are wide geographic disparities in survival rates across England for the eight most common cancers.[16] Patients with lung cancer in London have a substantially higher one year survival rate than patients in the North West and East Anglia.[17]
Clinical Commissioning Groups	Survival rates also differ at a more local level. In 2012 there was a 26% range between the highest and lowest one year cancer survival rates estimates for individual Clinical Commissioning Groups (CCGs) in England.[22]

1.3 : The effects of reducing diagnostic delay

The cancer survival gap between the UK and comparable countries appears very soon after diagnosis and is greatest between one and three months post-diagnosis. Poor one year survival rates are generally considered to indicate more advanced disease at diagnosis, since therapeutic options are more likely to influence long-term rather than short-term survival in cancer. This suggests that diagnostic delays may contribute significantly to the UK's excess premature, cancer-related deaths.[23] Several researchers therefore make a case that *earlier* diagnosis of cancer will improve survival.

There is strong evidence to support this for breast cancer (two systematic reviews),[24;25] and some evidence beginning to emerge for colorectal cancer.[25;26] This evidence relates to symptomatic diagnosis; there is also evidence that screening for pre-symptomatic disease reduces mortality for both breast and colorectal cancer.[27;28] For lung cancer the picture relating to earlier symptomatic diagnosis is more uncertain; however we know the following:

Neal et al's systematic reviews (a scoping review published in 2009 and full review published in 2015)[25;29] examining the effect of diagnostic delays on lung cancer survival reported equivocal results.

Neal noted significant challenges in comparing studies as a result of their differing definitions of delay and varying outcome measures. In addition he highlights that the majority of studies had substantial methodological limitations (e.g. they do not consider lead time bias[ii] or account for variations in the speed of tumour growth) making it impossible to assess the 'true' effect of diagnostic delay. However he reports that one of the studies, Tørring et al (2013),[30] that did report a positive association between mortality and longer diagnostic intervals addressed the key sources of bias, including a 'waiting time paradox' (which he describes as the issue of patients with very aggressive disease presenting early but having poor outcomes). As a result, whilst we cannot be certain from the literature reviewed by Neal that improving timeliness of diagnosis has an effect in lung cancer survival, it is a realistic possibility.

[ii] Where 'lead time' is the period between the early detection of a cancer (usually the result of screening or other early testing) and its usual clinical symptomatic presentation, and 'lead-time bias' is when, as a result, survival time appears prolonged, even if the earlier detection has actually had no effect on the course of the disease and overall survival time.

Studies considering proxy outcomes for early diagnosis show promise for improving survival

It seems plausible that diagnosing cancers earlier may result in diagnosis at an earlier stage of disease. A number of clinical studies have shown that a reduction in lung cancer stage at diagnosis leads to improved outcomes. Hamilton et al (2013)[31] evaluated the effects of his lung cancer risk assessment tool (RAT) and found that GPs diagnosed significantly more lung cancers using the tool (compared to prior to its introduction), some of which were early stage disease. Introduction of the RAT was associated with increased rates of curative treatment and therefore potentially an increased survival rate, although long-term data are required to confirm this. In addition, the National Lung Screening Trial of low dose computerised tomography (CT) screening for lung cancer (NLST, 2011)[32] conducted in the United States reported that (on average) cancers were detected at an earlier stage amongst screened patients than controls, and that there was a subsequent increase in resection rate and decrease in mortality rates in the group randomised to CT screening (although it is important to note that screening is designed to identify asymptomatic cancers rather than the symptomatic cancers that would present to the GP).

There is also an indication that factors other than early diagnosis might affect survival. For example whilst older patients with lung cancer may be more likely to be diagnosed with early stage disease,[33] they still have a poorer rate of survival than younger patients.[20] Both patient-related factors (e.g. delayed presentation to the GP) and health service-related factors (e.g. underuse of appropriate treatment) could account for this. It could also reflect some patients' decisions not to undergo potentially curative surgery.

While the link between early diagnosis and improved survival for lung cancer is not fully established, there is consensus within both the scientific and medical communities that avoidable delays and non-clinical variations in diagnosis of cancer are not acceptable. There is clear evidence that variations exist, and a strong likelihood that they have implications for differences in survival. It is therefore very important to understand what these variations are for lung cancer, and why they might be occurring, so that they can be addressed.

1.4 : The diagnostic pathway for a cancer patient

Richards (2009)[23] proposes that differences in patient pathways to receiving a cancer diagnosis and treatment are likely to contribute to disparities in survival.

1.4.1 : The diagnostic pathway when lung cancer is suspected

The majority of patients with lung cancer present symptomatically.[34] NICE guidelines recommend urgent chest X-ray (that is chest X-ray to be performed within two weeks) as the first line investigation for almost all patients with suspected lung cancer - direct referral to a specialist is only recommended for patients aged 40 years or older who present with unexplained haemoptysis (1.1.1: 'refer people using a suspected cancer pathway referral (for an appointment within 2 weeks) for lung cancer if they are aged 40 and over with unexplained haemoptysis').[35] If the findings of the chest X-ray suggest lung cancer, the next step recommended is referral to a specialist for an appointment within two weeks (1.1.1: 'refer people using a suspected cancer pathway referral (for an appointment within 2 weeks) for lung cancer if they have chest X-ray findings that suggest lung cancer').[35] More expensive and invasive diagnostic tests or procedures (e.g. computerised tomography (CT) scan or bronchoscopy) are, in general, only performed as second line investigations where a chest X-ray has identified abnormalities, and thus usually under guidance from a specialist - although 'open access' CT is currently being piloted in several English CCGs.

Chest X-ray is a readily available and reasonably cheap diagnostic test, capable of identifying lung cancer.[34] It is also quite accurate: there are relatively few falsely positive chest X-ray reports, whilst false negative chest X-rays only occur in a quarter of cancers (these are either due to the cancer not being visible, or it being missed by the radiologist producing the report).[36] Doctors are therefore able to have a relatively low threshold for requesting a chest X-ray in a patient presenting with symptoms of lung cancer; indeed about 20% of all chest X-rays requested by primary care are investigating a suspected lung cancer.[37]

1.4.2 : Key intervals in the diagnostic pathway for cancer

For any cancer where patients present symptomatically, a number of events occur between the time that a patient first notices their symptoms and the point at which they receive a diagnosis of cancer (and start treatment). Patient, healthcare and disease factors all contribute to the length of the intervals between these events. Several researchers have sought to use theoretical models to present the events that mark a patient's progress through the diagnostic pathway. These include Walter et al's (2012)[38] model of pathways to treatment shown in Figure 2, and the model developed by Oleson et al (2009)[39] shown in Figure 3.

Figure 2 : Model of pathways to treatment *(Walter et al, 2012)[38]*

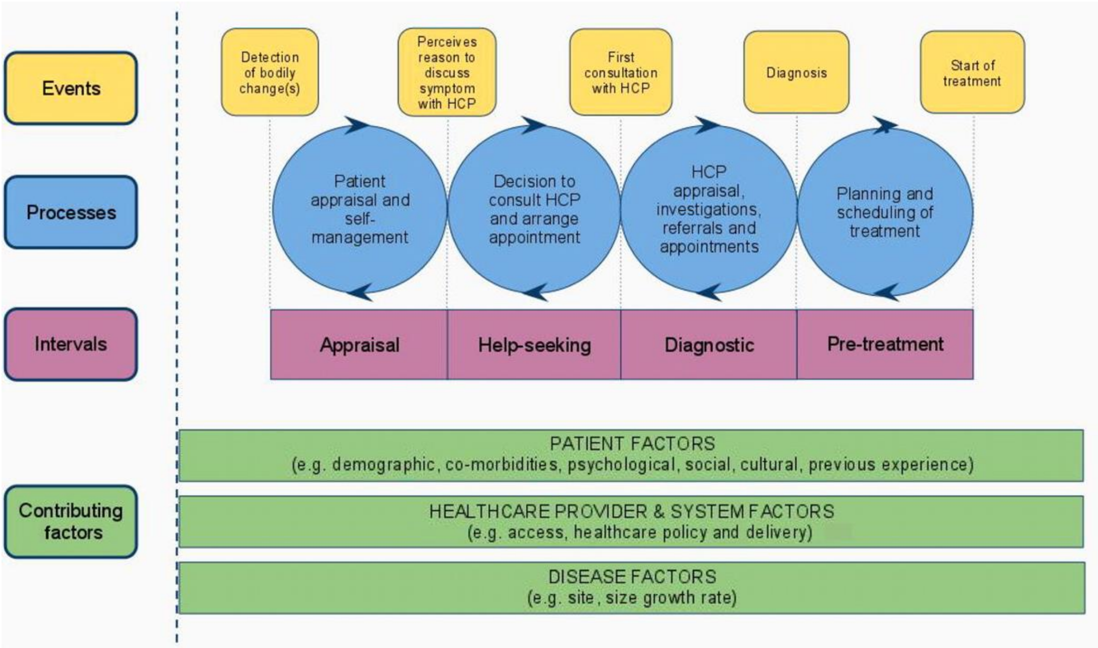
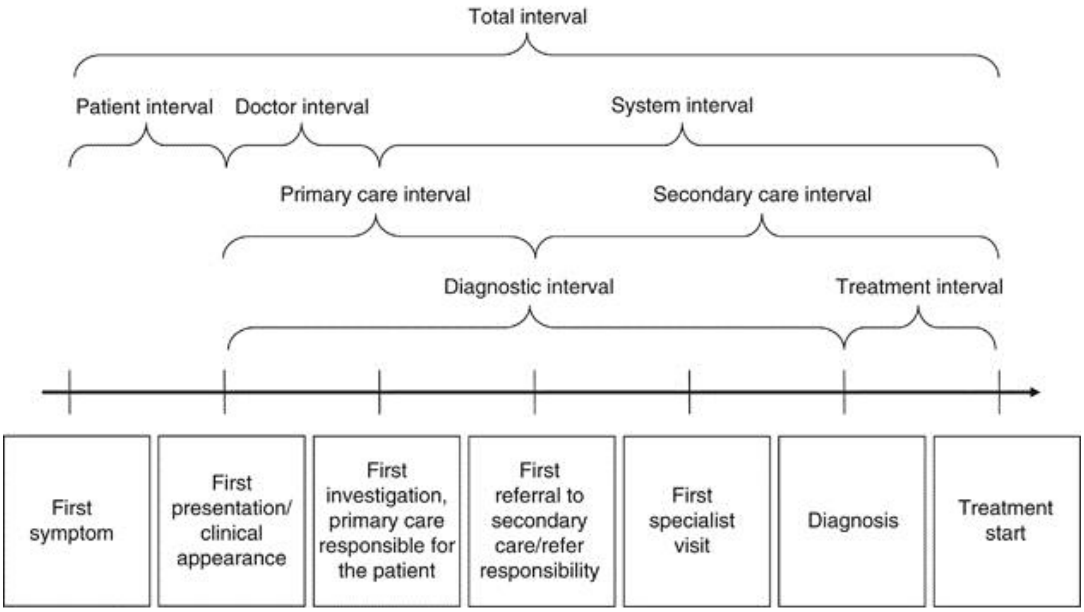


Figure 3 : A summary of key events and associated intervals in the cancer diagnostic pathway *(Oleson et al, 2009)*[39]



For the purposes of my PhD, I will describe the intervals in the cancer diagnostic pathway as follows (based on the 2012 Aarhus statement):[40]

- *Patient interval*
  The time between the appearance of the first symptom(s) and the patient's first presentation to a health professional.
- *Primary care interval*
  The time between the patient's first presentation to a GP and their being referred to secondary care.
  (NB: there will be no primary care interval if the patient first presents as an emergency)
- *Secondary care interval*
  The time between the patient's first presentation to secondary care (typically via a GP referral, but in some incidences as an emergency) and the initiation of treatment.

There is potential for delay to occur in any of these intervals: if this happens then a patient's diagnosis will be delayed.

Both the length of these intervals, and their relative importance, can vary between cancers. In cancers where most patients present with typical and/or visible symptoms and signs (e.g. melanoma or breast cancer) the length of the patient interval tends to account for a large portion of the overall time to referral.[41;42] As Lyratzopoulos et al (2015)[43] reflect, this suggests the importance of the patient interval in these cancers and thus (if one is seeking to improve timeliness of diagnosis) a need to focus on increasing patients' symptom awareness and encouraging appropriate help-seeking behaviour. In other cancers the relative contribution of the primary care interval to the overall time to referral is larger; this is particularly true for cancers where patients commonly present with symptoms of low specificity (e.g. lung cancer or myeloma).[44] For these cancers optimising the effectiveness and timeliness of the diagnostic process in primary care is therefore a key priority.[43]

As discussed in Section 1.4.1, chest X-ray is the recommended first line investigation for patients with symptoms that could be due to a lung cancer. Since GPs are able to request chest X-rays from primary care (indeed NICE guidelines recommend that for most patients GPs wait for the findings of an urgent chest X-ray before referring to secondary care),[35] and because a chest X-ray has reasonable accuracy as a diagnostic test for lung cancer, for many patients evidence suggestive of lung cancer will be produced in primary care before a specialist referral and histological examination. This differs from many other cancers. As a result the relative importance in the diagnostic pathway of the primary care interval compared to the secondary care interval is greater in lung cancer than some other cancers (e.g. colorectal, where NICE guidelines suggest that patients with symptoms suggestive of cancer are referred urgently to a secondary care specialist who will then perform diagnostic tests such as colonoscopy).

The primary care interval is therefore a particularly important part of the diagnostic pathway for lung cancer, and thus a research priority when seeking to improve patient outcomes.

1.5 : Evidence of inequalities in the cancer diagnostic pathway in the UK

There is evidence of socio-demographic variation in cancer survival within the UK (Section 1.2.2). Lyratzopoulos et al (2012)[44] propose that by increasing our understanding of the socio-demographic variation in the process and timeliness of cancer diagnosis, greater improvement in UK cancer survival will be achievable.

Excluding those cancers with current population screening programmes in the UK (breast, cervical and colorectal), the diagnostic pathway typically begins with a patient noticing a symptom and seeking medical help. Whilst there is significant potential for inequalities in the patient interval, this is outside the scope of my PhD. Inequalities may also occur in the secondary care interval: once diagnosed with cancer there are significant socio-demographic differences in the likelihood that particular groups of patients will receive optimal treatment, including for lung cancer.[45;46] This is also outside the scope of my PhD.

My PhD focuses on the primary care interval, specifically GPs' role in the cancer diagnostic pathway. In this section I will therefore focus on our current understanding of how delays can manifest in the primary care interval and current evidence of socio-demographic inequalities here, as well as highlighting gaps in our knowledge.

1.5.1 : Timeliness of GPs' decision to refer

Most patients with cancer who present to their GP are diagnosed relatively promptly – the 2011 National Audit of Cancer Diagnosis in Primary Care (NACDPC)[47] revealed that on average 82% were referred for specialist assessment within two visits to their GP. Nevertheless, some patients with certain cancers require significantly more visits before referral: for example the 2011 NACDPC recorded that 31% of lung cancer patients visited their GP three or more times before referral,[47] and Lyratzopoulos et al (2012)[44] using data from the National Cancer Patient Experience Survey observed a very similar percentage (33%). The 2011 NACDPC showed little evidence of socio-demographic variation in the number of GP visits preceding diagnosis for cancer patients as a whole, except younger adults having a greater number of visits.[47] However since only 14% of GP practices in England supplied data, we cannot be certain how representative these results are.

Lyratzopoulos et al (2012)[44] observed that younger patients and women were more likely to have attended the GP three or more times before their lung cancer diagnosis, although they found no significant variation by socio-economic circumstance. It is not clear why some patients visit the GP more than others before a referral is made. Lyratzopoulos et al (2013)[42] propose that this variation may reflect differences in GPs' symptom awareness or their access to/use of diagnostic tests; however this has not yet been subject to empirical examination.

MacLeod et al's systematic review (2009)[48] found that both socio-demographic (age, education, ethnicity, gender, socio-economic status) and clinical (presenting symptom, medical history) characteristics were associated with timeliness of GPs' referral of patients with a number of cancers. For lung cancer they concluded that older or more socially disadvantaged patients were more likely to experience delayed referral (and therefore a longer primary care interval) but there was not enough evidence to assess the impact of the other characteristics.

1.5.2 : GPs' referral process

There is also significant variation in referral behaviour between GPs and between GP practices: Meechan et al (2012)[49] examined GPs' use of the two week wait (TWW) referral pathway for suspected cancer and found considerable between-practice variation in both the proportion of patients diagnosed via the TWW (detection rate) and the proportion of TWW referrals found to have cancer (conversion rate).

GPs' choice of specialty to refer to, and the appropriateness of this specialty, may also have an impact on the length of a patient's diagnostic interval. Barrett et al (2008)[50] found that only 73% of lung cancer patients diagnosed via the GP were initially referred to respiratory specialists, and that those patients initially referred elsewhere also had a lower rate of chest X-ray investigation prior to referral.

1.5.3 : No primary care interval

The 2011 NACDPC reported that 20.3% of patients with lung cancer present as an emergency, higher than the average for all cancers combined (12.9%).[47] It also found that housebound patients or those over 80 years were more likely to present as an emergency.[47] Patients with cancer admitted as an emergency typically have lower rates of survival,[45;47;51] and those with lung cancer have a lower resection rate (which likely reflects a later stage of disease at presentation).[52] Both a systematic review by Mitchell et al (2015),[53] and a study of cancer-specific variation in emergency presentation by Abel et al (2015),[54] found that older patients, women and those with higher levels of deprivation were more likely to have an emergency presentation of lung cancer.

Emergency presentation may reflect patients not having visited their GP, for example due to difficulties in access or unwillingness to seek help: Mitchell et al's (2015)[53] review also observed that patients with lung cancer who had a lower primary care use or who lacked a regular source of primary care were also more likely to present as emergencies. However it may also reflect patients who have previously attended their GP for the same or related symptoms but who were not referred to secondary care at that point (in a study of colorectal cancer patients, Sheringham et al (2014)[55] observed that 84% of those presenting as an emergency had seen their GP in the 6 months prior to their diagnosis),[55] or patients who were advised to attend accident and emergency directly (MacLeod reports that a third of patients who present as emergencies to Accident and Emergency departments (A&E) have been referred there by their GP).[56]

1.5.4 : Summary

Studies therefore show evidence of significant socio-demographic variations in the length of the primary care interval. What is not yet clear is why these variations occur: whilst these studies consider the 'output' of the primary care interval (referral to secondary care), they do not provide information about what is happening *during* the primary care interval. To address this we need to increase our understanding of how GPs decide which patients to refer, and the extent to which this is affected by patient socio-demographic characteristics or by GP characteristics.

1.6 : GPs' role in the early diagnosis of cancer

The length of a patient's primary care interval is influenced by both patient and health-service factors, but is primarily determined by the GP. In the UK GPs act as gate-keepers to secondary care. The management decisions that GPs make can therefore have significant implications: both for patient outcomes (e.g. how early a patient with cancer is diagnosed) and healthcare costs. Understanding GPs' decision making processes in the diagnosis and referral of patients with symptoms that might indicate cancer is therefore key to increasing our understanding of early diagnosis and for the development of strategies for change.

1.6.1 : Role in the cancer diagnostic pathway

1.6.1.1 : Eliciting symptoms

When patients become aware of a new symptom, most will initially visit their GP.[50] Their progression along the cancer diagnostic pathway therefore relies on the GP identifying any symptoms of concern. There are numerous reasons why both the presence, or the full extent, of a patient's symptoms might not be elicited during a GP consultation.

Some of these reasons may reflect how a GP takes a patient's history: they may not pick up on patients mentioning (or alluding to) symptoms, they may not ask about relevant symptoms, or they may not ask questions in such a way that patients understand and provide the necessary information. A GP successfully eliciting a patient's symptoms also relies on that patient recognising them as important and worthy of mentioning to the doctor. As work by Walabyeki looking at understanding of cancer symptoms in smokers highlights,[57] not everyone is aware of symptoms that can indicate lung cancer: many identified cough or weight loss as warning signs, but the significance of shortness of breath or chest pain was much less recognised. If patients are unaware of the potential significance of their symptoms they may be less likely to mention them to their GP.

As Lyratzopoulos et al (2015)[58] discuss, structural factors in general practice in the UK may also negatively influence the elicitation of symptoms. Many patients attending general practices may feel (through either explicit or implicit suggestion of 'consultation norms') that they should only consult their GP about one problem per appointment,[59] which is likely to reduce the likelihood that GPs elicit the presence of symptoms that a patient believes are unrelated to their presenting complaint. The increasing workload in general practice is also likely to have an effect. Most general practices book 10 minute appointment slots during which a GP is often required to elicit the presence of (and important details about) symptoms, perform examinations, identify and explain the diagnosis, discuss management options, explore any concerns the patient has and write a record of the consultation. GPs therefore face considerable time pressures in their practice,[60] which may reduce their capacity to elicit symptoms.[61] It has also been noted that in countries with publically funded health systems, such as the UK, many patients worry about consulting the GP for symptoms that may 'waste the doctor's time'.[62] In addition to potentially leading to some symptomatic patients not attending their GP in the first place,[58] this also suggests that those who do attend may be hesitant to declare the full breadth or complexity of their symptoms.

1.6.1.2 : Evaluating the level of risk

Patients' progression along the cancer diagnostic pathway is also dependent on GPs' evaluation of their level of risk. On average, a full-time GP will see only one new diagnosis of lung cancer each year.[63] By contrast, they are likely to see patients with the most common presenting symptoms of lung cancer (cough and shortness of breath)[47] almost daily. Because the symptoms of lung cancer are non-specific,[64] GPs need to distinguish those patients with a high risk of serious disease from those with mild, self-limiting or acute illness. If GPs do not recognise a patient's presenting symptoms as being of sufficiently high risk to merit further investigation then their progress along the diagnostic pathway will be delayed.

There is some evidence that guidelines may influence GPs' decision making: McBride et al (2010)[65] observed that there was greater socio-demographic variation in GPs' referral decisions when patients presented with a symptom that did not have clear national guidelines about which patients to refer/not refer.[65] However even though there are national guidelines for investigation and referral of many suspected cancers (including for lung cancer), no threshold level of risk has been published. GPs therefore face a challenge in determining what level of risk of lung cancer justifies investigation or referral.

We do not know the average risk of cancer in patients who are referred for investigation, although Meechan et al (2012)[49] found that 11% of TWW referrals resulted in a cancer diagnosis, so this can be used as a rough estimate. However there is evidence that most patients would choose to be tested at a much lower level of risk, even below the threshold of national guidelines: Banks et al's (2014)[66] vignette study looking at preferences for investigation in primary care attendees found that 92% wanted to be investigated for symptoms that had just a 1% risk of being due to lung cancer (low risk).

The GP's role is highly challenging: if they investigate/refer too few patients there is the risk of delayed diagnosis and poorer patient outcomes; whilst over-investigation and over-referral have implications for resource use and NHS costs.

1.6.2 : Summary

GPs make their management decisions based on patients' presenting symptoms; the focus of this research is therefore on GPs' decision making process when presented with symptoms indicative of lung cancer. Increasing our understanding of what management decisions GPs make, as well as *why* they make them, is key if we are seeking to reduce the primary care interval and improve early diagnosis of lung cancer.

1.7 : Implications for my PhD

From the evidence discussed in this introduction so far, it is clear that GPs manage patients in different ways, their aim being to keep patients with acute or self-limiting illness within primary care, and refer those at higher risk for further investigation or to secondary care. We also know that there is clinical and non-clinical variation in GPs' management decisions and it is likely that this contributes to the inequalities seen in the diagnosis of lung cancer within primary care, which may in turn contribute to the UK's socio-demographic variation in lung cancer survival rates.

It is possible to take a number of different approaches when exploring how to improve lung cancer diagnosis. Researchers may choose to focus on the diagnostic tests themselves - for example comparing the efficacy, cost-effectiveness and acceptability of existing investigations, or seeking to develop novel tests that aim to identify cancers at an earlier stage. Other approaches to research involve exploring how clinicians use diagnostic tests: perhaps by examining factors (clinical or non-clinical) influencing their management decisions, or evaluating the value of guidelines or decision aids. Also, as discussed in Section 1.4.2, research may focus on specific intervals in the diagnostic pathway for lung cancer. Each of these approaches contribute to our understanding of how to improve lung cancer diagnosis.

The importance of the primary care interval in the diagnostic pathway for lung cancer, together with the consideration that avoidable delays in diagnosis have the potential to affect patient outcomes, indicates that addressing the issue of the potential for delay in the primary care interval may provide useful insights that contribute to the earlier diagnosis of (and potentially subsequent improved outcomes in) lung cancer. I therefore address this in my PhD.

My starting point is that there may be socio-demographic variation in the length of the primary care interval, and that it is likely that this contributes to differences in survival. I therefore believe that reducing socio-demographic variations in the diagnosis of lung cancer is an important priority. As a result, I have chosen to examine the extent to which GPs' decisions to refer patients to secondary care or for investigation vary with patient or GP characteristics. I have decided to focus particularly on the impact of non-clinical factors on GPs' decision making, since it is reasonable to hypothesise that these could contribute to much of the socio-demographic variation in the primary care interval.

I begin by discussing a systematic review that I have conducted examining the evidence for associations between non-clinical patient and GP characteristics and variations in GPs' referral for further investigation or to secondary care. Two published reviews have examined similar literature but do not address my research question specifically.

Hajjaj et al (2010)[67] reviewed the literature on non-clinical causes of variation in clinical decision making. However this review is subject to a number of limitations: its search methods were not systematic, it used few search terms, and it is not clear how studies were selected for inclusion. In addition, it was not focused solely on *GPs'* decision making, and the results were not reported consistently or comprehensively, making it difficult to make comparisons between studies. As discussed in Section 1.5, MacLeod et al (2009)[48] reviewed socio-demographic characteristics associated with delays in diagnosis (including within primary care) in studies of patients with cancer. However patients do not present complaining of 'cancer', but with symptoms of varying specificity and likelihood of an underlying diagnosis of cancer. In order to achieve a more complete understanding of GP decision making it is therefore important to consider how GPs' management varies by presenting symptom. This is particularly key for lung cancer where the presenting symptoms may be non-specific and common, and new presentations are rare (and therefore lung cancer is less likely to be the cause than many other diagnoses, such as chronic obstructive pulmonary disease).

I then go on to specifically consider how GPs manage patients presenting with symptoms of lung cancer, whether this varies by patient and/or GP characteristics, and which factors may underlie the decisions they make.

2 : Research overview

2.1 : Research aim

To examine the patient and GP characteristics associated with GPs' decisions to refer patients with symptoms indicative of lung cancer for chest X-ray or to a respiratory specialist.

2.2 : Research design

I addressed this aim through two studies:

- Study 1: A systematic literature review
- Study 2: An online factorial study examining variations in GP decision making

2.2.1 : Study 1

A systematic literature review of the non-clinical patient, GP and practice characteristics associated with variations in UK GPs' decisions to refer patients for investigations (including diagnostic tests) or to secondary care

This review:

- *Had UK focus*
  The importance of primary care and the role of GPs differs between countries, which may have implications for GPs' decision making behaviour. My PhD focuses on UK GPs; therefore I only included studies conducted (solely or partially) in a UK population.
- *Was not symptom or disease specific*
  Whilst my PhD focuses on lung cancer, my systematic review considers variations in GP referral behaviour more widely. This is for two reasons: firstly patients with lung cancer present with symptoms not a disease; secondly some factors underlying non-clinical variations in GP decision making may be independent of patients' symptoms or diagnosis.

The findings of Study 1 informed Study 2, in particular:

- the content of the post-consultation survey;
- the in-depth analysis performed on data collected in Study 2.

I report the methods and findings of Study 1 in Chapter 3 of this thesis.

2.2.2 : Study 2

Examining variations in GPs' decision making for patients presenting with symptoms of lung cancer: a factorial study using interactive, web-based patient vignettes

Study 2, ***'the GP decision making study'***, focuses on the behaviour of GPs practising in England. It has two parts:

a) The ***'vignette study'*** (Study 2a) examines the management decisions that GPs make in response to patients presenting with symptoms that could indicate lung cancer. It explores whether these decisions vary by patient or GP characteristics, or any combination of these. The vignette study has a factorial design[iii] and used a novel methodological approach. I report the development and methods of the vignette study in Chapter 4 of this thesis, and the results in Chapter 5.

[iii] A 'factorial design' experiment involves examining two or more experimental factors, each of which have a number of discrete possible values (e.g. gender or ethnicity). A series of experimental units are generated by creating all possible combinations of these values across all the experimental factors:

e.g. Experimental factor A has three possible values (Ai, Aii, Aiii)
Experimental factor B has two possible values (Bi, Bii)
Experimental factor C has two possible values (Ci, Cii)

Combining all these values across the three experimental factors therefore generates twelve experimental units:

Experimental unit number	*Factor A*	*Factor B*	*Factor C*
1	Ai	Bi	Ci
2	Ai	Bi	Cii
3	Ai	Bii	Ci
4	Ai	Bii	Cii
5	Aii	Bi	Ci
6	Aii	Bi	Cii
7	Aii	Bii	Ci
8	Aii	Bii	Cii
9	Aiii	Bi	Ci
10	Aiii	Bi	Cii
11	Aiii	Bii	Ci
12	Aiii	Bii	Cii

A factorial design therefore enables the effect of each experimental factor (and interactions between experimental factors) on the outcome measure to be studied.

b) The ***'post-consultation survey'*** (Study 2b) was completed by all participating GPs immediately after the vignette study. It explores the extent to which GPs believe certain factors influence their referral decisions for real patients who present in a similar manner to those in the vignette study.

   I report the methods and findings of the post-consultation survey in Chapter 6.

In Chapter 7 of this thesis I consider the data from the GP decision making study (Study 2) as a whole.

The findings of Study 2 are intended to inform the design and development of interventions to improve GP decision making when patients present with symptoms that could indicate lung cancer.

The outline for the GP decision making study was originally designed by my supervisor, Professor Raine (RR), and formed a component of the successful application to become a Policy Research Unit.

3 : Systematic literature review (Study 1)

3.1 : Introduction

As detailed in Section 2.1, the aim of my PhD was to examine non-clinical characteristics associated with GPs' decisions to refer patients with symptoms indicative of lung cancer for appropriate diagnostic investigation (chest X-ray) or to secondary care. I began by conducting a systematic review (Study 1) to explore what evidence of non-clinical variations in GPs' referral behaviour there is in the existing research literature. As well as enhancing our knowledge of this field, one purpose of my systematic literature review was to inform the research questions and study design of Study 2, which formed the rest of my PhD.

Whilst the overall focus of my PhD (and of Study 2) is on the diagnosis of lung cancer, I chose to review any literature that had examined GPs' referral for investigations or to secondary care, regardless of the presenting symptoms or underlying condition of patients in the study. This was for two reasons:

- My PhD focuses on decision making within primary care, where patients typically present with symptoms rather than a disease; furthermore (as discussed in Section 1.6) lung cancer may present with non-specific symptoms. Simply reviewing studies of patients either with lung-related symptoms, or who went on to receive a diagnosis of lung cancer, might not capture all relevant aspects of GPs' decision making.
- Many of the factors underlying non-clinical variations in GPs' decision making may be independent of patients' symptoms or diagnoses, and thus even studies of patients with conditions very distinct from lung cancer have the potential to provide useful insight into GPs' referral behaviour.

It is also of note that a scoping review I conducted of the literature revealed very few studies which specifically examined referral of patients with either symptoms of lung cancer, or who went on to receive a diagnosis of lung cancer.

In this chapter I discuss the aim and methods of my systematic literature review, details of the relevant literature identified, and discuss both the associations and the gaps in the literature that my review has demonstrated.

3.1.1 : Aim

To identify the non-clinical patient, GP and practice related characteristics significantly associated with variation in UK GPs':

- referral of patients for investigations, including diagnostic tests;
- referral of patients to secondary care services.

3.1.2 : Objectives

To conduct a systematic review to identify and critically appraise all relevant literature on the determinants of referral for GPs working in the UK in order to:

- determine if there are any clear associations between patient, GP or practice characteristics, or a combination of these, and GPs' decisions to refer patients;
- identify areas of uncertainty or inconsistency;
- identify possible explanations for any areas of uncertainty or inconsistency, and propose ways to address these.

3.2 : Method

3.2.1 : Search strategy

I initially performed a brief scoping review to select databases to search and to develop search terms and synonyms for the systematic review. I also sought advice from the systematic review librarian at the Royal Free Hospital.

My systematic review's search combined four principles:

- '**patient**': to ensure that studies identified related to the consultation and management of patients;
- '**decision making/outcome**': search terms related to either the decision making process or the specific outcomes I was considering in this study - referral or diagnostic investigation;
- '**general practice**': to restrict the search to studies in the primary care setting;
- '**socio-demographic characteristics**': search terms related to the four most studied characteristics - age, gender, ethnicity and socio-economic circumstance.

Each principle consisted of a variety of appropriate phrases, synonyms and Medical Subject Headings (MeSH) terms combined with an ‘or’ clause. I then combined the four principles with an ‘and’ clause to create the final search. I applied publication year and language limits, and because the review is only of UK studies set an exclusion of 'United States'. My search strategy is included in Appendix 1.

I searched the following databases: Medline, Embase, Web of Science, PsycInfo and Social Policy and Practice. I performed citation searching on the reference lists of the papers selected for full review. In addition Professor Willie Hamilton, an expert in primary care diagnosis and member of the Policy Research Unit, has had sight of the papers included in this review.

I exported all search results into Reference Manager and removed duplicates.

3.2.2 : Screening

I used the findings of my scoping review to develop selection criteria for the inclusion or exclusion of records. I refined these in consultation with my supervisors and Joe McDonnell (JMc), a senior public health trainee on academic secondment to the department where I am based. The final selection criteria, and my basis for these, are detailed in Appendix 2.

I used a three stage approach to assess all records retrieved by the search and to determine whether they met the selection criteria, at each stage excluding studies which clearly did not meet the criteria. I initially screened records by title, then screened the abstracts. Finally I screened the full papers of the records still remaining: initially to exclude non-UK studies, and then to exclude studies that did not meet the other selection criteria.

For quality assurance a second reviewer (JMc) independently screened a proportion of the records at each stage: 4% of the titles, 20% of the abstracts and 23% of the full papers. I then calculated the kappa statistic, which can be used to measure the agreement and reliability between two raters.[68] The kappa statistic for abstract screening was 0.74 (the upper limit of good agreement), and 0.88 for full paper screening (very good agreement), indicating that we reached a consensus for most of the records.

Where JMc and I reached conflicting decisions about inclusion/exclusion of a study we read the paper in more detail together and discussed our application of the selection criteria. Where disagreements still remained I brought the title or abstract through to the next stage of screening; there were no unresolved disagreements at the full paper stage.

3.2.3 : Quality assessment and data extraction

I then assessed the quality of the papers selected for full review using a critical appraisal tool which I developed.

I considered the suitability of the following validated tools for my review:

- CASP (Critical Appraisal Skills Programme)[69]
- SIGN (Scottish Intercollegiate Guidelines Network)[70]
- Heller et al (2008)[71] - Critical appraisal for public health: A new checklist.

I chose to adapt Heller et al's (2008)[71] checklist.

There are a number of reasons why I chose to base my checklist on Heller et al's (2008)[71] checklist (as opposed to the CASP or SIGN tools):

- Heller et al's (2008)[71] checklist was designed specifically for evaluation of public health studies, which are predominantly observational (as were the majority of studies in my systematic review).
- Because the CASP and SIGN tools vary by study design it would have been harder to make comparisons between studies, which was necessary in this review. Whilst Heller et al's (2008)[71] checklist did contain some questions that did not apply to all studies I appraised, the majority were applicable regardless of study design.
- The CASP and SIGN tools do not specifically address several of the issues that Sanderson et al (2007)[72] highlighted as being important for assessing quality and susceptibility to bias, for example whether the study clearly defined its inclusion/exclusion criteria (not addressed in the SIGN tool), if the potential for design-specific sources of bias is assessed (not addressed in either the CASP or SIGN tools) or whether statistical methods are used appropriately (not addressed in the SIGN tool).
- SIGN tools require evaluation of how effectively issues are addressed by the studies. This does give more information than the Yes/No in Heller et al's (2008)[71] checklist; however when evaluating the tools I found it was often difficult to make a judgement about when something was 'well' covered and when it was simply 'adequate' - it often seemed more effective to make an overall assessment of quality.

I made adaptations to Heller et al's (2008)[71] checklist to ensure that I could assess the quality of papers included in this systematic review, and to allow me to undertake data extraction at the same time as quality assessment. I created a database to enter and store the critical appraisal and data extraction information.

My supervisor Dr Jessica Sheringham (JS), JMc and I piloted my adapted checklist to ensure that it was an appropriate tool for critical appraisal and data extraction for this review. Appendix 3 shows the final version of the collection tool.

JS acted as a second reviewer for quality assessment and data extraction. We both conducted independent critical appraisal and quality assessment of all studies selected for full review using the data collection tool. This included rating each study according to both its quality and its relevance to this systematic review. JS and I then met afterwards to compare our appraisals and ratings for each paper, discussing discrepancies and reaching a consensus where possible. All our disagreements in rating were minimal and did not result in a different analysis outcome for the paper: we both agreed completely on which papers were rated medium or higher and should be examined in depth, which were rated lower than medium but still met my inclusion criteria for the review, and which should be excluded. We reached a consensus for the majority of papers, and in the few cases (4 out of 68, 5.9%) where we had a minor disagreement on rating (e.g. between 'low' or 'low/medium') decided that I would use my own rating assessment.

Where abstracts were selected for full paper screening but the full paper was not available online, I searched for the paper in the British Library. I then screened those papers which could be sourced in the same manner as those papers available online.

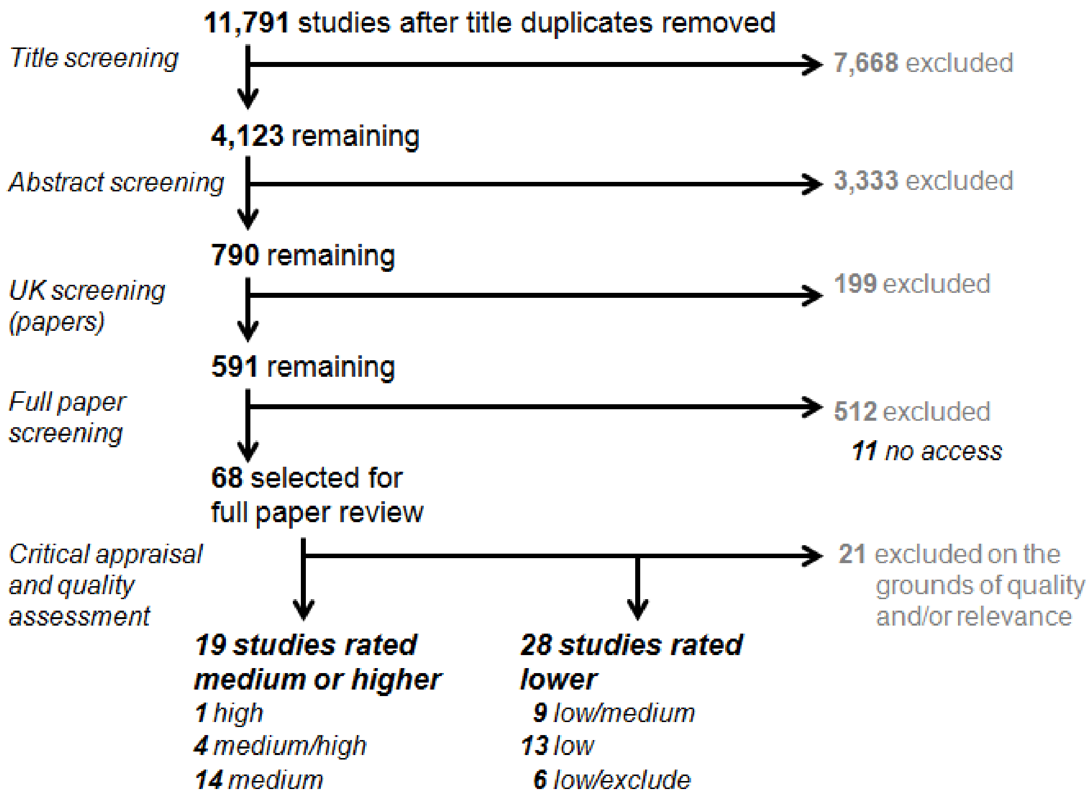
3.3 : Results

3.3.1 : *The literature search, screening and appraisal process*

3.3.1.1 : A summary of study inclusion in the review

Figure 4 summarises the results of the systematic literature search, screening process, critical appraisal and quality assessment. From the 11,791 unique studies identified, 68 were selected for full paper review.

Figure 4 : Flowchart of study inclusion



3.3.1.2 : Papers with no online access

I was able to source 54 of the studies that were not available online in the British Library. I discussed six of these with JS; however we concluded that none of the six met the criteria for full paper review. 11 studies could not be screened - six were missing from British Library records, whilst five were in journals or issues not held by the library. Since only 11 out of the 11,791 studies that my original searches identified could not be screened, I feel confident that my review is likely to be fairly comprehensive.

3.3.1.3 : Quality assessment and rating decisions

JS and I critically appraised and assessed the quality of the 68 papers selected for full paper review. After discussion we allocated 47 studies a rating (ranging from high to low/exclude), and excluded 21 studies on the grounds of quality and/or relevance. This involved evaluating how effectively each study tackled a number issues that have been identified as key when evaluating study quality or relevance such as: how well the study population reflected the UK population, study response rate, how well a study addressed potential sources of bias, how relevant a study's aims were to the aim of my systematic review, and whether a study's findings were presented clearly and in a useable format.[69;70;73]

Whilst some systematic reviews allocate studies a rating by scoring them in categories such as these and then totalling to create a summary score for each study, I chose not to use this approach. The main reason for this was that since my systematic review included studies with a variety of study designs, different categories were of different importance for different studies. It was therefore not appropriate to use a single common measure to evaluate each study's quality. Whilst the checklist I developed to critically appraise studies (Appendix 3) ensured that the same information was collected for all studies, the relative importance of some of these factors potentially differed with study design (e.g. the risk of certain sources of bias, or the necessity to account for possible confounding factors). Furthermore, the Cochrane Handbook (many people see Cochrane reviews as the gold standard of systematic reviews)[74] states that calculating a summary score may be an unreliable assessment of validity, and less likely to be transparent for readers of the review.[73]

Although our rating of studies in this systematic review did not use a quantitative scoring system, it was still based on how effectively each study tackled these key issues of quality and relevance. Before allocating any studies a rating, both JS and I independently read several studies with a wide range of both quality and relevance. Whilst the aims, designs and content of all 47 studies rated were extremely heterogeneous, JS and I identified the key features of studies that we believed should be rated highly, as well as the significant shortcomings that meant a study would have to be allocated a low rating for this review. As Table 1 describes, if a study had a clear design, a large and generalisable population, addressed sources of bias well (or acknowledged the potential limitations of its finding) and was relevant to my systematic review's aim, then it was rated highly. If a study fell short in most of these areas it was rated low. Studies allocated a medium rating generally had shortcomings in one or two of these areas, but were sufficiently relevant and had methods of sufficient quality that it was possible to have confidence in their results.

Table 1 : Typical features of studies assessed as low quality and high quality

Low quality studies	**High** quality studies
- poor generalisability *small scale study (often single site) or a sample very different to the general population* - lack of relevance to the aims of my review - unaddressed potential sources of bias *e.g. response bias, selection bias, reporting bias* - poor/unclear reporting of study findings	- sample size large enough to answer the study's question - multi-site across at least two different regions - considered key potential confounders in the analysis (where applicable) - clear attention to potential sources of bias *methods to address these and/or significant transparency about limitations due to bias* - results reported clearly, generally using raw data

In consultation with my supervisors, I decided to only report the findings of the 19 studies rated medium or higher. The 28 studies rated lower were either low quality and/or not very relevant or generalisable to the question of my systematic review and I therefore had concerns about the confidence that could be placed in their findings, with the result that my confidence in their findings was not on par with my confidence in the higher rated studies.

All 47 studies rated are very heterogeneous, with many of the low rated studies having examined different diseases, characteristics and outcomes to the higher rated studies. Therefore whilst I do not focus on the results of the lower rated studies here, I have briefly considered their designs, methods and outcome measures (and how these compared to the higher rated studies) as there are some marked differences between studies in each of these groups. I discuss this further in Section 3.3.3.2.

3.3.2: The 19 medium and high rated studies

3.3.2.1 : Study settings, populations and design

Table 2 (see pages 58-62) summarises the 19 studies rated medium and high.

Seven of the studies comprised national data from across the United Kingdom (UK). Three studies used data from both English and Scottish GP practices and patients. Seven studies were based solely in England, two were based solely in Scotland.

An extremely heterogeneous range of symptoms or medical conditions were examined across the 19 studies. Six studies involved patient diagnosis, considering presenting symptoms ranging from hip pain to dyspepsia to depression. Ten studies focused on the management of patients with pre-existing medical conditions, in particular diabetes (five studies) and coronary heart disease (three studies). One study examined renal function testing as a whole, and therefore included both diagnostic tests and follow up tests for long term management. Finally two studies considered referral for prevention - smoking cessation courses and an exercise scheme.

Study sample sizes ranged from 128 to 1,852,762, reflecting in large part the studies' methods and sources of data.

18 of the 19 studies were retrospective observational: 13 were cross-sectional and five were cohort studies. One group of researchers (Bonte et al, 2008)[75] conducted a factorial study using vignettes.

Table 2 : Summary of study settings, populations, designs and methods for the 19 studies rated medium or higher

Paper	Study setting	Type of consultation	Study population	Characteristics studied	Outcome measure(s)	Data source	Rating
Studies focusing on diagnosis of symptoms							
Macfarlane, 2012 [76]	GP practices in one Scottish city and one English county	*Diagnostic:* - back pain	14,680 patients	*Patient:* age	*Referral:* specialist, physiotherapy, exercise referral, cognitive behavioural therapy (CBT),	Patient postal questionnaire	Medium
de Lusignan, 2011 [77]	29 GP practices across south west London (England)	*Diagnostic or management:* - renal function testing	220,721 patients from 29 GP practices	*Patient:* age, gender, ethnicity, co-morbidity (diabetes)	*Investigation:* creatinine, micro-albuminuria, proteinuria	Local database (CONDUIT network)	Medium
Juni, 2010 [78]	40 GP practices across 2 south west counties (England)	*Diagnostic:* - hip pain	1,302 patients from 40 GP practices	*Patient:* gender	*Referral:* specialist care	Screening questionnaire and patient interview	Medium/high
McBride, 2010 [65]	326 GP practices across England and Scotland	*Diagnostic:* - hip pain - dyspepsia - post-menopausal bleeding (PMB)	- hip pain: 23,121 patients - dyspepsia:101,212 patients - PMB: 5,492 patients	*Patient*: age, gender, socio-economic circumstance (SEC), co-morbidity	*Referral:* secondary care	National database (health improvement network)	High

Abbreviations: CBT = cognitive behavioural therapy, PMB = post-menopausal bleeding, SEC = socio-economic circumstance

Table 2 *(continued)*- Summary of study settings, populations, designs and methods for the 19 studies rated medium or higher

Paper	Study setting	Type of consultation	Study population	Characteristics studied	Outcome measure(s)	Data source	Rating
Studies focusing on diagnosis of symptoms (continued)							
Tate, 2010 [79]	488 GP practices across the UK	*Diagnostic:* - symptoms of ovarian cancer	1,107 patients from ~488 GP practices	*Patient:* age	*Investigation:* CA-125, ultrasound scan, CT scan *Referral:* gynaecology	National database (GPRD)	Medium/high
Kendrick, 2009 [80]	38 GP practices across 3 regions of England	*Diagnostic:* - symptoms of depression	2,294 patients from 38 GP practices	*Patient:* age, gender, co-morbidities, symptom severity	*Referral:* mental health services, social services	Patient notes	Medium/high
Bonte, 2008 [75]	GP practices across 3 regions of England	*Diagnostic:* - symptoms of coronary heart disease (CHD)	128 GPs	*Patient:* gender	*Investigation:* number of CHD tests ordered *Referral:* cardiology, other medical professional	Factorial video vignette study (of GPs)	Medium

Abbreviations: CHD = coronary heart disease, CT = computerised tomography

Table 2 *(continued)*- Summary of study settings, populations, designs and methods for the 19 studies rated medium or higher

Paper	Study setting	Type of consultation	Study population	Characteristics studied	Outcome measure(s)	Data source	Rating
Studies focusing on the management of already diagnosed disease							
Vamos, 2011 [81]	GP practices across the UK	*Management:* - diabetes	422 GP practices	*Practice:* size	*Investigation:* cholesterol, HbA1c	National database (GPRD)	Medium
Coleman, 2010 [82]	GP practices across the UK	*Management:* - commencing anti-hypertensive treatment	74,096 patients	*Patient:* age, gender, SEC, co-morbidity (diabetes), smoking, symptom severity	*Investigation:* baseline and follow up investigations	National database (GPRD)	Medium
Hamilton, 2010 [83]	GP practices across the UK	*Management:* - diabetes	154,945 patients	*Patient:* age, gender, SEC	*Investigation:* cholesterol, HbA1c	National database (GPRD)	Medium/high
Verma, 2010 [84]	26 GP practices across 1 London borough (England)	*Management:* - diabetes	4,309 patients from 26 GP practices	*Patient:* ethnicity	*Investigation:* cholesterol, HbA1c	Patient notes	Medium
McGovern, 2008 (CHD)[85]	GP practices across Scotland	*Management:* - coronary heart disease (CHD)	75,495 patients	*Patient:* age, gender, SEC	*Investigation:* exercise testing, cholesterol	English and Scottish database	Medium
McGovern, 2008 (DM)[86]	GP practices across Scotland	*Management:* - diabetes (DM)	310 GP practices	*Patient:* age, gender, SEC	*Investigation:* creatinine, cholesterol, HbA1c	English and Scottish database	Medium

Abbreviations: CHD = coronary heart disease, DM = diabetes mellitus, HbA1c = glycosylated haemoglobin, SEC = socio-economic circumstance

Paper	Study setting	Type of consultation	Study population	Characteristics studied	Outcome measure(s)	Data source	Rating
Studies focusing on the management of already diagnosed disease (continued)							
Millett, 2008 [87]	32 GP practices in south London (England)	*Management:* - CHD	3,101 patients from 32 GP practices	*Patient:* ethnicity	*Investigation:* cholesterol	Patient notes	Medium
Phatak, 2008 [88]	GP practices across the UK	*Management:* - commencing statin treatment	57,296 patients	*Patient:* age, gender, smoking, BMI, risk of CHD	*Investigation:* lipid testing	National database (GPRD)	Medium
Millett, 2007 [89]	GP practices across England and Scotland	*Management:* - diabetes	1,852,762 patients from 8,970 GP practices	*Practice:* size, diabetes caseload, deprivation	*Referral:* retinal screening *Investigation*: creatinine, micro-albuminuria, cholesterol, HbA1c	National database (QMAS)	Medium
Saxena, 2007 [90]	GP practices across England and Scotland	*Management:* - CHD, - stroke	8,970 GP practices	*Practice:* size, caseload, deprivation	*Investigation:* exercise testing, cholesterol, CT/MRI	National database (QMAS)	Medium

Abbreviations: BMI = body mass index, CHD = coronary heart disease, CT = computerised tomography, HbA1c = glycosylated haemoglobin, MRI = magnetic resonance imaging, SEC = socio-economic circumstance

Table 2 *(continued)*- Summary of study settings, populations, designs and methods for the 19 studies rated medium or higher

Paper	Study setting	Type of consultation	Study population	Characteristics studied	Outcome measure(s)	Data source	Rating
Studies focusing on prevention							
Simpson, 2010 [91]	GP practices across the UK	*Prevention:* - smoking cessation	483,239 patients (2006/7 data) from 525 GP practices	*Patient:* age, gender, SEC	*Referral:* smoking cessation services	National database (QRESEARCH)	Medium
Sowden, 2008 [92]	317 GP practices across 6 London PCTs (England)	*Prevention:* - exercise referral	7,985 patients from 317 GP practices	*Patient:* age, gender *Practice:* deprivation, primary care trust (PCT)	*Referral:* exercise scheme	Patient notes	Medium

Abbreviations: PCT = primary care trust, SEC = socio-economic circumstance

3.3.2.2 : Associations between non-clinical characteristics and referral

Table 3 (see pages 71-75) gives details of the association between a number of different non-clinical patient, GP and practice characteristics, and referral for investigations or to secondary care. I will discuss the results for each characteristic in turn; for each looking first at papers that considered investigations, and then those that considered secondary care referral.

3.3.2.2.1 : Patient age

12 of the included studies (that is those rated medium or higher) examined the association between patient age and referral for investigations or to secondary care. Older patients were significantly less likely to be referred to secondary care, and also frequently less likely to have blood test investigation.

Referral for investigations, including diagnostic tests

Seven of these 12 studies looked at the association between patient age and the likelihood of their receiving investigations. All seven studies considered blood tests, one study (Tate et al, 2010)[79] also studied the association between age and more invasive investigations, and another (McGovern et al, 2008, 1)[85] also considered referral for exercise testing or specialist assessment.

Four of these studies examined the association between patient age and the blood test monitoring of patients with chronic conditions: diabetes, hypercholesterolaemia and coronary heart disease (CHD). The overall pattern was that older patients, in these studies typically defined as those aged 75 years or older, were statistically less likely to receive cholesterol testing, and HbA1c testing where applicable (the two studies of diabetics). This pattern was observed in three studies; the fourth study looking at this association, McGovern et al (2008, 2),[86] did not observe a statistically significant difference.

Another study, Coleman et al (2010),[82] examined the association between patient age and monitoring of patients with hypertension commencing anti-hypertensive treatment. They observed that whilst older patients were less likely to undergo baseline blood tests before commencing treatment, they were more likely than younger patients to have follow up blood testing within six months of starting therapy.

Older patients were also statistically more likely to undergo renal function testing than those of younger age, both in the context of managing those with known diabetes (McGovern et al, 2008, 2)[86] and renal function testing in general (de Lusignan et al, 2011).[77]

Tate et al (2010)[79] examined the association between patient age and referral for blood tests, scans and invasive investigations for patients with symptoms of ovarian cancer. They did not observe any statistical difference in rates of CA-125 marker blood tests or computerised tomography (CT) scans by patient age, but did observe a negative association between patient age and rates of both ultrasound scanning and invasive investigations (laparoscopy, laparotomy and/or oopherectomy).

Referral to secondary care

Seven of the 12 medium and high rated studies that considered patient age examined referral of patients (with a wide variety of symptoms) to out of practice services, including to secondary care. The overwhelming pattern was that older patient age decreased the likelihood of referral: in six of these studies older patients, typically defined as those aged 75 years or older, were less likely to be referred than those who were younger, and in the seventh there was no statistical significance.

This association was observed for a wide variety of symptoms and conditions, from referral of patients with depression to mental health or social services (Kendrick et al, 2009)[80] to referral of those with symptoms that could indicate cancer, such as post-menopausal bleeding and dyspepsia (McBride et al, 2010)[65] or symptoms of ovarian cancer (Tate et al, 2010).[79]

Two studies considered the impact of patient age on referral for joint pain, the prevalence of which increases significantly with patient age. McBride et al (2010)[65] considered hip pain, and observed a statistically significant decrease in the likelihood of referral to secondary care for the oldest group of patients, those aged 85 years and older. The authors adjusted for a range of patient factors, including patient gender and co-morbidity. Macfarlane et al (2012)[76] examined the effect of patients' age on referral for back pain. Whilst they observed that patients aged over 70 years were slightly less likely to be referred to secondary care specialists, this was not statistically significant. This study did not adjust for patient co-morbidity, but did take into account patient gender and the disease severity and impact as potential confounding factors.

The oldest group of patients also appear less likely to be referred for preventative healthcare opportunities. Although they did not consider patient co-morbidity or capability, Sowden et al (2008)[92] noted that patients aged 75 years or older were less likely to be referred for exercise referral schemes. Simpson et al (2010)[91] identified a non-linear relationship between patient age and referral for smoking cessation services with older patients more likely to be referred than the youngest patients but, following the pattern seen in the majority of these studies, the oldest patient group in the study (those aged 75 years or older) being the least likely to be referred.

3.3.2.2.2 : Patient gender

12 of the medium and high rated studies examined the association between patient gender and referral for investigations or to secondary care. 11 of these studies observed a difference in rates of investigation and referral between men and women; however which gender was more likely to be referred varied depending on the patient's underlying symptoms or condition.

Referral for investigations, including diagnostic tests

Seven of the 12 studies looked at the association between patient gender and the likelihood of their receiving investigations.

Four of these studies examined the association between patient gender and the blood test monitoring of patients with chronic conditions: hypercholesterolaemia, CHD and diabetes. In three of these studies women were statistically less likely to receive cholesterol testing, whilst the fourth (Phatak et al, 2008)[88] did not find significant difference by gender for cholesterol testing of patients starting statin therapy.

Coleman et al (2010)[82] examined the association between patient gender and the monitoring of patients with hypertension commencing anti-hypertensive treatment. They observed that women were less likely to undergo baseline blood tests before commencing treatment, although there was not a significant difference by gender for the likelihood of receiving follow up blood testing.

Two studies considered the association between patient gender and renal function testing. McGovern et al (2008, 2)[86] observed that women with diabetes were less likely to have creatinine testing than men. By contrast, de Lusignan et al (2011)[77] found that women in general practice were overall (that is not limited to diabetic patients) more likely to undergo creatinine testing, although whilst they adjust for patient age with gender, and patient ethnicity with gender, the potential confounder of diabetes as a co-morbidity was not considered alongside gender.

Two studies examined the association between patient gender and more complex investigations for CHD. Bonte et al (2008)[75] did not observe any significant difference by gender in rates of request for cardiac investigation for patients with symptoms of CHD; however McGovern et al (2008, 1)[85] found that women with CHD were less likely than men to be referred for exercise testing or specialist assessment.

Referral to secondary care

Six of the 12 medium and high rated studies considered referral of patients to out of practice services, including to secondary care.

Three of these studies observed that female patients were less likely to be referred to particular secondary care services than male patients. Two studies (Juni et al, 2010 and McBride et al, 2010)[65;78] considered the impact of patient gender on referral for hip pain. Despite hip pain being more prevalent in women,[78] both studies observed that women were significantly less likely to be referred to secondary care. Both studies' authors adjusted for patient age, and Juni et al (2010)[78] also adjusted for disease severity. Bonte et al (2008)[75] examined the association between gender and referral for patients with symptoms of CHD. They observed that women were less likely to be referred to cardiology than men with the same symptoms, although equally likely to be referred to other specialities.

Kendrick et al (2009)[80] explored the referral of those with depression to mental health or social services. The study's data comprised patients whose depression severity had been classified using two different scoring systems (both of which are widely used): the Hospital Anxiety and Depression Scale (HADS), and the patient health questionnaire, 9 item version (PHQ-9). Perhaps rather surprisingly, Kendrick et al (2009)[80] observed that where patients were assessed using PHQ-9 women were less likely to be referred than men, but where HADS had been used women were more likely than men to be referred.

Two studies did not identify any significant relationship between gender and referral to out of practice services: McBride et al (2010)[65] considered the management of patients with dyspepsia, whilst Simpson et al (2010)[91] studied referral for smoking cessation.

Only one study found that women were more likely to be referred for an out of practice service: Sowden et al (2008)[92] observed that women were more likely to be referred for exercise referral schemes.

3.3.2.2.3 : Patient ethnicity

Only three of the medium and high rated studies explored the impact of patient ethnicity, and all of these studies considered its association with referral for simple investigations (blood or urine testing).

Referral for investigations, including diagnostic tests

de Lusignan et al (2011)[77] examined the effect of patients' ethnicity on renal function testing, measured by blood test (creatinine) or urine test (proteinuria). No details were available for why each individual was having their renal function tested. The authors noted that both South Asian and black patients were more likely to have had their creatinine tested than white patients. This is not unexpected given the high risk and burden of diabetes, and subsequent diabetic nephropathy, in patients of these ethnicities compared to white patients. However diabetes co-morbidity does not appear to entirely explain this effect. There is no statistically significant difference in proteinuria testing between the three ethnic groups.

Two of the studies examined whether the likelihood of patients with known chronic conditions receiving blood test monitoring varied by patients' ethnicity. Verma et al (2010)[84] considered the frequency of cholesterol and HbA1c monitoring of diabetic patients, whilst Millett et al (2008)[87] examined cholesterol testing of patients with CHD. Both of these studies found no difference in testing between white and non-white patients (South Asian or black) when adjusting for age and gender (Millett et al, 2008 also adjusted for socio-economic circumstance and GP practice level clustering).[87]

Referral to secondary care

This systematic review did not identify any medium or high rated studies which considered the impact of patients' ethnicity on the likelihood of referral to secondary care.

3.3.2.2.4 : Patient socio-economic circumstance

Six of the medium and high rated studies examined the association between patients' socio-economic circumstance (SEC) and referral for investigations or to secondary care. All these studies defined patients' SEC using area based measures (Townsend, Carstairs, or the Index of Multiple Deprivation). Three of these studies considered the association between patients' SEC and their referral for blood tests, whilst three looked at referral to more complex services (one study examined both of these). Many of the studies did not find a significant association between patient socio-economic circumstance and referral; where it was found to have an impact it was generally the most deprived patients who were the least likely to be investigated or referred to secondary care.

Referral for investigations, including diagnostic tests

Coleman et al (2010)[82] examined the association between SEC and the likelihood that patients commencing anti-hypertensive drugs received both baseline blood tests and follow up blood tests (within six months of commencing treatment). The authors observed that patients in the intermediate deprivation quintiles were most likely to receive monitoring, even after a wide variety of potential confounding factors (such as co-morbidities, smoking status, age, gender and blood pressure) were adjusted for.

McGovern et al (2008, 2)[86] and Hamilton et al (2010)[83] explored the association between patients' SEC and blood test monitoring of known diabetic patients. McGovern et al (2008, 2)[86] did not observe any statistically significant relationship between patient deprivation and cholesterol, HbA1c or creatinine testing. By contrast, Hamilton et al (2010)[83] observed that more deprived patients were less likely to have cholesterol and HbA1c testing (although of note is that Hamilton et al, 2010 did not adjust for potential confounding factors whereas McGovern et al, 2008 did).[83;86]

Referral to secondary care

McBride et al (2010)[65] examined the association between SEC and referral to secondary care for patients with three distinct symptom groups, adjusting for several key potential confounding factors. Patients with both hip pain and dyspepsia were less likely to be referred when they were from a more deprived area. However for patients with post-menopausal bleeding the authors observed no statistically significant relationship between patients' SEC and their referral.
McGovern et al (2008, 1)[85] considered the association between CHD patients' SEC and both cholesterol blood testing and referral for exercise testing or to specialist assessment (no distinction is made between these in this study, as they are a combined Quality Outcomes Framework target). They did not observe any effect of patient deprivation on either of these outcomes.

Only one study observed that patients from more deprived areas were statistically more likely to be referred, Simpson et al's (2010)[91] examination of referral to smoking cessation services. This study was a descriptive analysis and did not adjust for potential confounding factors. However rates of referral are counted as percentages of the total number of patients in each category, so this effect is not simply reflecting higher rates of smoking among the lower SEC population.

3.3.2.2.5 : GP characteristics

This systematic review did not identify any high or medium rated studies which considered the impact of individual GPs' personal characteristics on patients' likelihood of referral for investigations or to secondary care.

3.3.2.2.6 : Practice characteristics

Four of the 19 medium and highly rated studies examined practice characteristics.

Three of these studied the association between practice size and rates of referral for investigations or to secondary care. Two studies (Vamos et al, 2011 and Millett et al, 2007)[81;89] considered the impact of practice size on blood test investigations for patients with known diabetes. Vamos et al (2011)[81] did not observe any significant difference in either cholesterol and HbA1c testing by practice size, but Millett et al (2007)[89] found that larger GP practices had higher rates for both these investigations. Millett et al (2007)[89] also examined the effect of practice size on other outcomes for diabetes patients and found that larger practices were also more likely to perform creatinine and proteinuria testing, and retinal screening. A third study, Saxena et al (2007)[90] considered the association between practice size and referral for patients with CHD or stroke, considering a range of simple investigations (cholesterol testing), more complex diagnostic tests (echocardiogram or CT and MRI scans), and referral for exercise testing and specialist assessment. The authors observed that for all these outcome measures, referral was more likely in larger practices.

Two of the medium and high rated studies considered the association between the size of a practice's caseload for a particular disease, and referral of patients with (or suspected of having that disease) for investigations or to secondary care. For both diabetes (Millett et al, 2007)[89] and CHD/stroke (Saxena et al, 2007)[90] practices with a higher caseload were more likely to refer patients. This was the case for all investigations and referral outcomes considered: from simple blood tests to referral for scans or specialist assessment.

Three studies examined the association between practice deprivation and referral to secondary care or for investigations. Two of these studies observed that practices in more deprived areas had lower rates of referral to both investigations and services outside primary care: Millett et al (2007)[89] found that patients with diabetes were less likely to have cholesterol, HbA1c, creatinine or proteinuria tests if they were part of practices in more deprived areas, whilst Saxena et al (2007)[90] found that in practices in more deprived areas patients with CHD were less likely to have blood tests for cholesterol or to be referred for echocardiogram, exercise testing or specialist assessment, and that patients with strokes were less likely to have had cholesterol testing or CT/MRI scanning. By contrast Sowden et al (2008)[92] observed that patients from practices in more deprived areas were more likely to be referred to an exercise scheme.

Table 3 (on the following pages) summarises all the associations between patient and practice characteristics and referral for investigations or to secondary care for all 19 medium and high rated studies. I discuss these findings in Section 3.4.2.

Table 3 : Associations between patient and practice characteristics and referral for investigations or to secondary care for the 19 studies rated medium or higher

Key: ↑ = more likely, ↓= less likely, ↕ = variable pattern (explained more in notes), ↔ = no difference (at a statistical significance of $p<0.05$), - = not examined

Abbreviations: BMI = body mass index, HbA1c = glycosylated haemoglobin, SEC = socio-economic circumstance

Paper	Disease	Outcome	Patient characteristic				Practice characteristic			Adjusted for	Notes
			Age *older vs. younger*	**Gender** *female vs. male*	**Ethnicity** *non-white vs. white*	**Socio-economic circumstance** *more deprived vs. less deprived*	**Deprivation** *more deprived vs. less deprived*	**Size** *large vs. small*	**Caseload** *higher vs. lower*		
Macfarlane, 2012 [76]	Back pain	- referral to any specialist (e.g. rheumatologist, psychologist)	↔	-	-	-	-	-	-	gender, severity/impact (chronic pain grade score)	referral of those age >70 less likely reported in text but not significant
de Lusignan, 2011 [77]	Renal function testing	- creatinine tested	↑	↑	↑	-	-	-	-	models for: age+gender, age+ethnicity, gender+ethnicity, diabetes+ethnicity	
		- proteinuria tested	↑	↔	↔	-	-	-	-		
Vamos, 2011 [81]	Diabetes	- cholesterol tested	-	-	-	-	-	↔	-	practice level clustering	
		- HbA1c tested	-	-	-	-	-	↔	-		
Coleman, 2010 [82]	Anti-hypertensive therapy	- baseline blood tests	↓	↓	-	↕*	-	-	-	age, gender, SEC, co-morbidity (diabetes), BMI, smoking, blood pressure, drug class prescribed	* intermediate deprivation quintiles most likely to have monitoring
		- follow up blood tests (up to 6 months)	↑	↔	-	↕*	-	-	-		
Hamilton, 2010 [83]	Diabetes	- cholesterol tested	↓	↓	-	↓	-	-	-	none	older age defined as ≥ 75yr NB: SEC disparities very small
		- HbA1c tested	↓	↓	-	↓	-	-	-		

Key: ↑ = more likely, ↓= less likely, ↕ = variable pattern (explained more in notes), ↔ = no difference (at a statistical significance of $p<0.05$), - = not examined

Abbreviations: CT = computerised tomography, HbA1c = glycosylated haemoglobin, PMB = post-menopausal bleeding, SEC = socio-economic circumstance

Paper	Disease	Outcome	Patient characteristic				Practice characteristic			Adjusted for	Notes
			Age *older vs. younger*	**Gender** *female vs. male*	**Ethnicity** *non-white vs. white*	**Socio-economic circumstance** *more deprived vs. less deprived*	**Deprivation** *more deprived vs. less deprived*	**Size** *large vs. small*	**Caseload** *higher vs. lower*		
Juni, 2010 [78]	Hip pain	- referral to specialist (rheumatology or orthopaedics)	-	↓	-	-	-	-	-	age, disease severity	
McBride, 2010 [65]	PMB	- referral to secondary care	↓	-	-	↔	-	-	-	age, gender, SEC, co-morbidity (number of drug classes prescribed), practice level clustering	negative gradient in referral for PMB by age, for hip pain only those aged ≥85 yr
	Hip pain	- referral to secondary care	↓	↓	-	↓	-	-	-		
	Dyspepsia	- referral to secondary care	↕*	↔	-	↓	-	-	-		* for dyspepsia those aged 55-64 most referred
Simpson, 2010 [91]	Smoking cessation	- referral to smoking cessation services	↕*	↔	-	↑	-	-	-	none	* tendency for older patients to be referred more, except those ≥75
Tate, 2010 [79]	Ovarian cancer symptoms	- CA-125 blood test	↔	-	-	-	-	-	-	none	negative association with age (between 40 and 84 years)
		- CT scan	↔	-	-	-	-	-	-		
		- ultrasound scan	↓	-	-	-	-	-	-		
		- invasive investigation	↓	-	-	-	-	-	-		
		- gynaecology referral	↓	-	-	-	-	-	-		

Key: ↑ = more likely, ↓= less likely, ↕ = variable pattern (explained more in notes), ↔ = no difference (at a statistical significance of $p<0.05$), - = not examined

Abbreviations: CHD = coronary heart disease, HbA1c = glycosylated haemoglobin, SEC = socio-economic circumstance

Paper	Disease	Outcome	**Patient characteristic**				**Practice characteristic**			Adjusted for	Notes
			Age *older vs. younger*	**Gender** *female vs. male*	**Ethnicity** *non-white vs. white*	**Socio-economic circumstance** *more deprived vs. less deprived*	**Deprivation** *more deprived vs. less deprived*	**Size** *large vs. small*	**Caseload** *higher vs. lower*		
Verma, 2010 [84]	Diabetes	- cholesterol tested - HbA1c tested	- -	- -	↔ ↔	- -	- -	- -	- -	age, gender	
Kendrick, 2009 [80]	Depression	- referral to mental health or social services	↓	↕*	-	-	-	-	-	age, gender, co-morbidity (diabetes, CHD, other physical), severity, past history, region	older age defined as ≥ 65yr * men assessed with PHQ-9 more likely to be referred, those assessed with HADS less likely
Bonte, 2008 [75]	CHD	- investigation - referral to cardiology - referral to other specialties	- - -	↔ ↓ ↔	- - -	- - -	- - -	- - -	- - -	none (but controlled factorial design)	
McGovern, 2008 (1) [85]	CHD	- cholesterol tested - referral for exercise testing/ specialist assessment	↓ ↓	↓ ↓	- -	↔ ↔	- -	- -	- -	age, gender, SEC, co-morbidity (CHD related), practice size, practice level clustering	

Key: ↑ = more likely, ↓= less likely, ↕ = variable pattern (explained more in notes), ↔ = no difference (at a statistical significance of $p<0.05$), - = not examined

Abbreviations: CHD = coronary heart disease, HbA1c = glycosylated haemoglobin, HDL = high-density lipoprotein, SEC = socio-economic circumstance

Paper	**Disease**	**Outcome**	**Patient characteristic**				**Practice characteristic**			**Adjusted for**	**Notes**
			Age *older vs. younger*	**Gender** *female vs. male*	**Ethnicity** *non-white vs. white*	**Socio-economic circumstance** *more deprived vs. less deprived*	**Deprivation** *more deprived vs. less deprived*	**Size** *large vs. small*	**Caseload** *higher vs. lower*		
McGovern, 2008 (2) [86]	Diabetes	- cholesterol tested - HbA1c tested - creatinine tested	↔ ↔ ↑	↓ ↓ ↓	- - -	↔ ↔ ↔	- - -	- - -	- - -	age, gender, SEC, co-morbidity (diabetes related), practice level clustering	
Millett, 2008 [87]	CHD	- cholesterol tested	-	-	↔	-	-	-	-	age, gender, SEC, practice level clustering	
Phatak, 2008 [88]	Statin therapy	- cholesterol testing *total* *HDLs* *triglycerides*	 ↓ ↓ ↓	 ↔ ↑ ↑	 - - -	 - - -	 - - -	 - - -	 - - -	age, gender, region, co-morbidity (hypertension), risk of CHD, smoking, baseline cholesterol	older age defined as ≥ 75 years
Sowden, 2008 [92]	Exercise referral scheme	- referral to exercise referral scheme	↓	↑	-	-	↑	-	-	patient age and gender; practice area, deprivation, training status, distance to scheme	older age defined as ≥ 75 years

Key: ↑ = more likely, ↓= less likely, ↕ = variable pattern (explained more in notes), ↔ = no difference (at a statistical significance of $p<0.05$), - = not examined

Abbreviations: CHD = coronary heart disease, CT = computerised tomography, HbA1c = glycosylated haemoglobin, MRI = magnetic resonance imaging

Paper	Disease	Outcome	Patient characteristic				Practice characteristic			Adjusted for	Notes
			Age *older vs. younger*	**Gender** *female vs. male*	**Ethnicity** *non-white vs. white*	**Socio-economic circumstance** *more deprived vs. less deprived*	**Deprivation** *more deprived vs. less deprived*	**Size** *large vs. small*	**Caseload** *higher vs. lower*		
Millett, 2007 [89]	Diabetes	- cholesterol tested	-	--	-	-	↓	↑	↑	none	
		- HbA1c tested	-	-	-	-	↓	↑	↑		
		- creatinine tested	-	-	-	-	↓	↑	↑		
		- proteinuria tested	-	-	-	-	↓	↑	↑		
		- retinal screening performed	-	-	-	-	↓	↑	↑		
Saxena, 2007 [90]	CHD	- cholesterol tested	-	-	-	-	↓	↑	↑	none	
		- referral for exercise testing/ specialist assessment	-	-	-	-	↓	↑	↑		
		- referral for echocardiogram	-	-	-	-	↓	↑	↑		
	Stroke	- cholesterol tested	-	-	-	-	↓	↑	↑		
		- CT or MRI scan	-	-	-	-	↓	↑	↑		

3.3.3 : The 28 lower rated studies

28 studies were rated lower than medium, of which nine were rated low/medium, 13 rated low and six rated low/exclude.

Table 4 (pages 77-83) summarises the characteristics and outcome measures examined in each of these studies, as well as the key reasons for each study's lower rating.

Table 4 : Summary of study design and reasons for low rating of the 28 studies rated lower than medium

Abbreviations: COPD = chronic obstructive pulmonary disease, PCT = primary care trust, SEC = socio-economic circumstance

Paper	Disease	Outcomes	Characteristics studied			Reasons for low rating				
			Patient	*GP*	*Practice*	*Low/biased response rate*	*Not representative of UK population*	*No denominator group*	*No adjustment for confounding*	*Results not usable for my review*
Martin, 2012 [93]	COPD	Referral to pulmonary rehabilitation	ethnicity	-	-	-	**Yes** three London PCTs	-	-	-
Baughan, 2011 [94]	Suspected cancer	Urgent referrals for suspected cancer	age	-	-	-	-	-	**Yes**	**Yes** reported results not in usable format for me
de Lusignan, 2011 [95]	Depression and/or anxiety disorders	Referral to IAPT services (Improving Access to Psychological Therapies)	age, gender, ethnicity, SEC	-	-	**Yes** risk of respondent bias could not be assessed	-	**Yes**	**Yes**	-
Grimshaw, 2011 [96]	Low back pain	Referral for lumbar spine X-ray	-	beliefs	-	**Yes** low response rate	-	-	-	**Yes** study aim different to my review aim
Hammouche, 2011 [97]	Hypertension	Cholesterol, blood glucose, creatinine, electrolytes and proteinuria tested	age, gender, SEC	-	size, deprivation	**Yes** risk of recruitment bias	**Yes** 18 Norfolk GP practices	-	**Yes** considered some but no accounting for clustering	-

Table 4 *(continued)* - Summary of study design and reasons for low rating of the 28 studies rated lower than medium

Abbreviations: PCT = primary care trust, SEC = socio-economic circumstance

Paper	Disease	Outcomes	Characteristics studied			Reasons for low rating				
			Patient	*GP*	*Practice*	*Low/biased response rate*	*Not representative of UK population*	*No denominator group*	*No adjustment for confounding*	*Results not usable for my review*
Hu, 2011 [98]	Parkinson's Disease	Seen a specialist, sub-optimal care (delay in referral)	age, gender, SEC	-	-	**Yes** poor reporting of response rate	-	-	-	-
Jinks, 2011 [99]	Knee pain	Referral to rheumatology or orthopaedics	age, gender	-	-	**Yes** risk of selection bias	**Yes** three GP practices in North Staffordshire	-	-	-
Raymond, 2011 [100]	Prevention	Cholesterol and blood glucose tested	self efficacy	-	-	-	**Yes** three GP practices in London	-	-	**Yes** self efficacy is not in the scope of my review
Wagg, 2011 [101]	Urinary incontinence	Referral for specialist opinion, cystometry testing	age	-	-	**Yes** potential for recruitment bias	**Yes** only one practice per PCT, may not be representative	**Yes**	-	-
McGorm, 2010 [102]	Unspecified (patients with medically unexplained symptoms)	Referral to specialist services	age, gender, SEC	-	-	**Yes** potential for respondent bias	**Yes** five GP practices in one Scottish city	-	-	**Yes** study does not make a distinction between different symptoms

Table 4 *(continued)* - Summary of study design and reasons for low rating of the 28 studies rated lower than medium

Abbreviations: HbA1c = glycosylated haemoglobin, HIV = human immunodeficiency virus

Paper	Disease	Outcomes	Characteristics studied			Reasons for low rating				
			Patient	*GP*	*Practice*	*Low/biased response rate*	*Not representative of UK population*	*No denominator group*	*No adjustment for confounding*	*Results not usable for my review*
Murray, 2010 [103]	Coronary heart disease	Cholesterol tested	ethnicity	-	-	-	**Yes** one (ethnically diverse) London borough only	-	**Yes** no statistical tests (not even significance)	**Yes** study's primary aim different to my review
Kumar, 2010 [104]	Rheumatoid arthritis	'GP delay' (weeks from primary to secondary care)	ethnicity	-	-	**Yes** no data on response rates	**Yes** one hospital trust	**Yes** no information on those remaining in primary care	-	-
Nicholson, 2010 [105]	Epididymo-orchitis	Chlamydia, Gonorrhoea, microbial and urine testing	age	-	-	-	**Yes** some patients managed outside GP practice	-	**Yes**	**Yes** good quality but study focus was not on investigation
Sadler, 2010 [106]	Sexually transmitted infections	Chlamydia and HIV testing	-	age	size, deprivation	-	**Yes** combines two very different populations	-	**Yes**	-
Fischbacher, 2009 [107]	Diabetes	Cholesterol and HbA1c tested, retinal screening	ethnicity	-	-	-	**Yes** one area of Scotland	-	-	**Yes** unclear which results were ordered in secondary care

Table 4 *(continued)* - Summary of study design and reasons for low rating of the 28 studies rated lower than medium

Abbreviations: HbA1c = glycosylated haemoglobin

Paper	Disease	Outcomes	Characteristics studied			Reasons for low rating				
			Patient	*GP*	*Practice*	*Low/biased response rate*	*Not representative of UK population*	*No denominator group*	*No adjustment for confounding*	*Results not usable for my review*
Ingram, 2009 [108]	Presentation to out-of-hours GPs	Referral to hospital (as an emergency)	-	gender, attitudes (including tolerance of risk)	-	**Yes** poor response rate, quite small sample size	-	-	-	**Yes** no data to compare referred and not referred clinically
Loo, 2009 [109]	Atrial fibrillation	Echo-cardiography	age	-	-	**Yes** very small sample size	**Yes** small region (South Devon)	-	**Yes** no statistical tests (not even significance)	-
Green, 2008 [110]	Disordered eating	Intention to refer to eating disorder services	-	age, gender, training status, years practising, attitudes	size	**Yes** low response rate, poor sample size	**Yes** one county	-	-	**Yes** results and analysis not presented clearly
Tahrani, 2008 [111]	Diabetes	Cholesterol, HbA1c, creatinine, micro-albuminuria tested, retinal screening	-	-	size	-	-	-	**Yes**	**Yes** crude examination of practice characteristic examined

Table 4 *(continued)* - Summary of study design and reasons for low rating of the 28 studies rated lower than medium

Abbreviations: CHD = coronary heart disease, CT = computerised tomography, ECG = electrocardiogram, HbA1c = glycosylated haemoglobin, LV D = left ventricular dysfunction, MRI = magnetic resonance imaging

Paper	**Disease**	**Outcomes**	**Characteristics studied**			**Reasons for low rating**				
			Patient	*GP*	*Practice*	*Low/biased response rate*	*Not representative of UK population*	*No denominator group*	*No adjustment for confounding*	*Results not usable for my review*
Ashworth, 2007 [112]	Mental health (on lithium) Coronary heart disease (CHD) Left ventricular disease (LVD) Stroke	Lithium, creatinine, thyroid testing Exercise ECG testing Echo-cardiography Referral for CT/MRI	-	-	deprivation	-	-	-	**Yes**	**Yes** unclear paper - many results only reported in text not clear tables
Crilly, 2007 [113]	Angina	Cholesterol testing, exercise ECG, coronary angiography, thallium scan	gender	-	-	-	**Yes** 15% Liverpool population only	-	-	-
Gray, 2007 [114]	Diabetes	Cholesterol, HbA1c, creatinine, micro-albuminuria testing and retinal screening	ethnicity	-	-	-	**Yes** 32 practices in one (ethnically diverse) London PCT	-	-	-

Table 4 *(continued)* - Summary of study design and reasons for low rating of the 28 studies rated lower than medium

Abbreviations: CHD = coronary heart disease, OOH = out-of-hours, PMB = post-menopausal bleeding

Paper	**Disease**	**Outcomes**	**Characteristics studied**			**Reasons for low rating**				
			Patient	*GP*	*Practice*	*Low/biased response rate*	*Not representative of UK population*	*No denominator group*	*No adjustment for confounding*	*Results not usable for my review*
McLean, 2007 [115]	CHD Stroke Diabetes	Cholesterol testing (for all diseases) and retinal screening (diabetes only)	-	-	rurality	-	**Yes** unclear if rural Scotland generalisable to rural rest of UK	-	**Yes**	-
Parker, 2007 [116]	Colorectal bleeding PMB	Relevant referral or investigation (for both symptoms)	age	-	-	-	-	-	**Yes** does not appear to consider for referral results	**Yes** analyses focus on co-morbidity not non-clinical characteristics
Ridsdale, 2007 [117]	Headache	Referral to neurologist	age, gender	-	-	**Yes** low response rate, risk of participation bias	**Yes** 18 GP practices in one London region only	-	-	-
Roberts, 2007 [118]	Health checks in patients with schizophrenia	Cholesterol tested	co-morbidity	-		**Yes** low response rate, small sample size	-	-	-	**Yes** considers co-morbidity but no demographic
Rossdale, 2007 [119]	Presentation to out-of-hours (OOH) GPs	Referral to hospital (as an emergency)	-	gender, role, years practising number of OOH consults	-	-	**Yes** one out-of-hours centre in Bristol	-	-	**Yes** no data to compare referred and not referred clinically

Table 4 *(continued)* - Summary of study design and reasons for low rating of the 28 studies rated lower than medium

Abbreviations: SEC = socio-economic circumstance

Paper	Disease	Outcomes	Characteristics studied			Reasons for low rating				
			Patient	*GP*	*Practice*	*Low/biased response rate*	*Not representative of UK population*	*No denominator group*	*No adjustment for confounding*	*Results not usable for my review*
Weich, 2007 [120]	Depression	Referral for psychological treatment	age, gender, SEC	-	-	**Yes** high attrition rate, effects of bias uncertain	-	-	-	**Yes** cannot be certain if referral was via GP or other route

3.3.3.1 : Quality assessment and critical appraisal

As with the higher rated studies, the 28 lower rated studies differed considerably in their aims, study design and methods, and therefore there were also differences in the reasons they were allocated a low rating during quality assessment and critical appraisal (studies were allocated a low rating if deemed low quality and/or not very relevant or generalisable to my systematic review). However there were some commonly occurring features in those studies given a low rating.

3.3.3.1.1 : Low or potentially biased response rate

Low or biased response rate was an issue in 14 of the lower rated studies.

Several of the low rated studies had a very small sample size: for example Loo et al (2009)[109] report data for 131 patients (managed by GPs from a single practice), whilst Green et al (2008)[110] surveyed 88 GPs. This affected their quality rating because of concerns that they might not be representative of the overall population.

A number of studies had a low response rate (in some less than a third of those approached participated),[96;110;118] which could lead to a risk of response bias. Other studies were rated low because their methods had significant potential for recruitment, participation or selection bias, or as a result of having a high attrition rate during the course of the study. Higher rated studies did not necessarily have completely unbiased methods, however in these studies the authors had taken steps to reduce potential sources of bias or, if this was not possible, were transparent about and aware of the risk of bias when drawing conclusions from their results.

Three studies either did not report a response rate, or did not report one that could be clearly interpreted. Since it was therefore not possible to evaluate whether these studies' response rates and consideration of bias were adequate, all three were also rated lower quality.

3.3.3.1.2 : Study sample population not representative of the UK population

In 18 of the 28 studies rated low quality the study's sample population could not be considered representative of the UK population overall, and therefore its generalisability was uncertain.

Table 4 gives more details why each of these study's sample populations was not considered representative. Typically this was either due to the study being small scale (e.g. a single hospital trust or out-of-hours practice) or confined to a specific area (e.g. one city, or a small region), or because the study population was of significantly different composition to the overall population of the UK (e.g. a single, very ethnically diverse primary care trust in London).

3.3.3.1.3 : No denominator group of all those eligible for referral

In three studies there was only information on the characteristics of patients who were referred, not those remaining in primary care. As a result it was not possible to compare those patients referred with a 'denominator group' to examine any non-clinical variation in referral in order to answer my systematic review's question.

3.3.3.1.4 : No adjustment for confounding

A number of studies did not adjust for any potential likely confounding factors, such as patient age and gender. Not adjusting for confounding did not mean that a study was automatically assigned a lower rating. This was because in some studies confounding was adjusted for within the study design itself rather than the analysis (e.g. a factorial design), whilst for others it was considered acceptable if the sample size was extremely large, or the rest of the study was of high quality. 11 of the lower rated studies did not adjust for confounding; in fact two of them had not reported any statistical analysis of their results, including significance testing.

3.3.3.1.5 : Study's results not usable for my systematic review

The primary aim of a number of the studies differed from the aim of my systematic review. This in itself was not an absolute reason for exclusion or low rating of a study: many of these studies contained information to answer the research question (e.g. a study evaluating how rates of cholesterol testing have changed over time by patients' socio-demographic characteristics still provided useful data on how cholesterol testing varies by non-clinical patient characteristics).[88] However several of the studies rated lower did not provide quantitative information about the characteristics or outcomes that I was interested in,[96;100;105;111;116;118] whilst in others the results or analyses were not presented clearly enough for me to use.[94;103;107;108;110;112;119]

Two studies were rated lower because it was not possible to determine whether all their data met the inclusion criteria. In the study by Weich et al (2007)[120] it was not possible to distinguish patients who had been referred from their GP from those who had been referred via another route. McGorm et al's (2010)[102] study of patients with medically unexplained symptoms grouped all patients together and, especially in light of the heterogeneity in referral patterns (often by disease type) seen between many of the higher rated studies, I felt it was not reasonable to consider as a whole the referral data for different, unknown symptoms.

Overall, 15 of the low rated studies did not have results that were usable (in content or format) to answer the question of my systematic review.

3.3.3.2 : Comparison of the low rated studies with those rated higher

I identified three main areas in which the 28 lower rated studies differ from the 19 medium and high quality studies already discussed in Section 3.3.2. These are the disease considered, the characteristics explored, and the study methods/designs.

3.3.3.2.1 : Disease considered

A wide variety of physical disease types were considered both within the higher quality studies and within the low quality studies. Both groups also studied patients presenting to their GP with symptoms (e.g. joint pain or symptoms suggestive of cancer) as well as those undergoing management or monitoring of a previously diagnosed condition (e.g. heart disease or diabetes). However certain diseases were less likely to be considered in the higher rated studies.

11 of the 68 studies critically appraised considered the management of mental health symptoms or conditions (including Parkinson's disease and dementia). However only one of these was rated medium/high and therefore considered in detail in this review: four studies were rated lower quality, whilst six were excluded on the grounds of quality. This meant it was therefore not possible for me to examine non-clinical variation in mental health referral and treatment in this systematic review, despite it being fairly frequently addressed in the literature.

Similarly, five of the critically appraised studies considered patients with sexually transmitted infections (STIs). Two of these were included in the review, classified as low quality; the other three were excluded after full paper review (either on the grounds of quality or relevance). This was primarily due to the complication that sexual health is frequently managed in specialised clinics (which do not require GP referral to access), and therefore those patients being investigated or managed in general practice are unlikely to be fully representative of the population as a whole.

There were also a small number of studies that examined referral but did not give details of the underlying symptoms or disease; these were typically studies examining referrals from out-of-hours GP services,[108;119] and therefore may reflect GPs' management of patients with acute medical situations. However since it was not possible to determine whether any non-clinical variation seen was appropriate variation (e.g. due to patients' differing underlying health needs) or not, these studies were rated lower.

3.3.3.2.2 : Characteristics explored

12 of the 68 papers critically appraised considered the effect of patient ethnicity, but only three of these were assessed as being high quality. This may reflect poor reporting of ethnicity in routine data (many of the studies included in this systematic review report on data from the 2000s, when reporting of ethnicity data was notoriously incomplete), as well as the variation in ethnic density and diversity across the United Kingdom. Studies exploring ethnicity were generally small single centre studies conducted in multi-ethnic populations; as a result their findings were often not generalisable to the rest of the country.

Few studies considered the effect of GP characteristics. Those that did tended to focus on GPs' psychological characteristics, such as their attitude to risk or their personal beliefs about illness. While these studies provided an interesting insight into additional deeper factors that may underlie GPs' decision making the majority were not of sufficient quality to meet the criteria for inclusion.

3.3.3.2.3 : Study methods/designs

Four of the studies evaluated used vignette methods;[75;110;121;122] however all but one of these were excluded on the grounds of quality. Generally this was due to potential priming of the GP participants (e.g. they were informed of the study subject at recruitment, or their management decisions were selected from a pre-determined multiple choice list) or significant potential for bias (e.g. a greater proportion of participants than expected had a specialist interest in the condition being studied). Vignettes were typically delivered on paper, although one study used a voice recording of a consultation.

3.3.3.2.4 : Findings of the lower rated studies

I have chosen not to report the findings of the 28 lower rated studies in depth, since these are studies which our critical appraisal and quality assessment did not rate highly - either due to their lack of relevance for my systematic review, or their quality. However it is worth noting that, in general, a similar pattern of association between non-clinical characteristics and GPs' referral of patients for investigations or to secondary care was seen in the lower rated studies as in the higher rated studies that I have reported in detail.

3.4 : Discussion

3.4.1 : Main findings

There has not been much research into non-clinical variation in GPs' referral of patients with symptoms that could indicate cancer. In this systematic review I have therefore included all papers that examine UK GPs' decisions to refer patients for any investigations (including diagnostic tests) or to secondary care, rather than using patients' symptoms or medical condition to determine a paper's inclusion.

This systematic review provides consistent evidence that patient age and gender are associated with variation in GPs' referral behaviour. The oldest patients (in particular those 75 years or older) were consistently less likely to be referred. This association was observed for patients with a wide variety of symptoms and conditions. In contrast, the effect of patients' gender does not have a consistent effect: for some symptoms or conditions women are more likely to be referred, whilst for others the situation is reversed.

There is more uncertainty regarding the association between other patient characteristics and GPs' referral behaviour. Higher levels of patient deprivation were associated with lower referral in some cases, but not all. Studies varied in the indicators they used to measure deprivation; furthermore all studies used area based measures of deprivation rather than personal indicators, so it is not possible to draw any firm conclusions about whether GP referral is associated with an individual patient's socio-economic circumstance. It is also not possible to form any conclusions about whether patient ethnicity has any effect, because despite several studies examining this there were not enough that were highly rated or of high enough quality methodologically (their sample populations were typically small and/or not representative of the UK population).

There were not enough studies examining the association between either individual GP or practice characteristics and GPs' referral behaviour to draw firm conclusions about their possible effect on GPs’ decisions to refer patients for investigations or to secondary care.

3.4.2 : Possible explanations for these findings

3.4.2.1 : Patient age

None of the studies in my systematic review provide empirical evidence for why older patients were less likely to be referred for investigations (including diagnostic tests) or to secondary care.

As has been proposed in the literature, some of the differences in referral for investigations or to secondary care by patient age could be explained by the effect of patient age on:

- *the likelihood of severe disease* [113;121;123]
  This might explain why we observed that older patients were more likely to receive particular blood test investigations: both de Lusignan et al (2011)[77] and McGovern et al (2008, 2)[86] found that renal function testing was more common in older patients (and rates of kidney disease are known to increase with age), whilst Coleman et al's (2010)[82] observation that older patients were more likely to have follow up blood testing after starting statin therapy could reflect a concern from GPs about increased risk of side-effects or medication interactions.[82]

- *risk of the investigation itself* [113;124]
  This could account for Tate et al's (2010)[79] finding that for patients with symptoms of ovarian cancer referral for invasive investigations was rarer for older patients, but that there were no differences by age for non-invasive tests.

- *the patient's underlying level of health and co-morbidities, which GPs could perceive as a contra-indication for treatment* [124;125]

  This might explain why Sowden et al (2008)[92] found that patients aged 75 years or older were less likely to be referred for exercise referral schemes, given that older patients are more likely to struggle with exercise and the authors did not adjust for patient co-morbidity or capability.

Only a few of the studies in my systematic review that examined patient age also adjusted for patient co-morbidity. This may explain some of the differences in findings seen between studies: for example McBride et al (2010)[65] (who did adjust for co-morbidity) found that older patients with hip pain were significantly less likely to be referred to secondary care, whilst Macfarlane et al (2012)[76] (who did not adjust for co-morbidity) reported no difference in referral by age for patients with back pain. The lack of consistency in studies' consideration of patients' co-morbidities also means that it is unclear whether older patients were indeed at greater risk of contra-indications, or less fit for investigations and treatment.

None of the studies in my systematic review that explored patient age reported examining patients' wishes for, or beliefs and concerns about, investigation and/or treatment.

It therefore remains unclear whether the differences in referral (for investigation and to secondary care) by patient age seen in my systematic review were appropriate and due to differences in patients' fitness for referral, or in accordance with patients' wishes, or if perhaps they reflect GPs' own expectations of patients' fitness or wishes for referral.

It is also possible that these differences are not in the best interests of patients: for example differences in GPs' referral behaviour might contribute to the particularly wide survival gap between the UK and the best-performing countries for the oldest patients with cancer.[10] Since overall life expectancy in the UK is increasing,[126] understanding the reasons behind this variation in referral by patient age (and whether it is detrimental) is likely to become an issue of increasing importance.

3.4.2.2 : Patient gender

The effect of patient gender on referral for investigations or to secondary care appears complex: for some symptoms and conditions women are more likely than men to be referred, whilst for other symptoms and conditions the situation is reversed.

As has been proposed in the literature, some of the differences in referral for investigations or to secondary care by patient gender may be explained by:

- *GPs' perception of the risk and/or likelihood of disease varying by gender* [75;113;127;128]

  The perception that women are at lower risk of CHD than men could explain why McGovern et al (2008, 1)[85] found that women with CHD were less likely than men to be referred for exercise testing or specialist assessment, and Bonte et al (2008)[75] observed that, even when presenting with identical symptoms to men, women were less likely to be referred to cardiology - but as likely to be referred to other secondary care specialities. It may also explain why several studies found that women were significantly less likely to receive cholesterol testing,[83;85;86] or to undergo baseline blood tests before commencing anti-hypertensive treatment:[82] hypercholesterolemia and hypertension are both known risk factors for CHD. However it is important to remember that GPs' perceptions of the risk of disease may not always be accurate: for example studies have shown rates of CHD in women are not dissimilar to those in men, but that the disease often presents with less typical symptoms.[113;127]

However a difference in perceived risk, or even prevalence, of disease cannot explain differences in referral by patient gender completely: both McBride et al (2010)[65] and Juni et al (2010)[78] observed that female patients with hip pain were less likely to be referred to secondary care than male patients, even though hip pain is more prevalent in women.[78] This difference in referral was seen despite both studies adjusting for patient age, and one adjusting for disease severity.

- *Women are less likely to be in full-time employment than men, so may have increased flexibility for appointments* [128;129]

  Whilst many studies observed that women were less likely to be referred to secondary care services than men, Sowden et al (2008)[92] found that women were consistently more likely than men to be referred to an exercise referral scheme. This study adjusted for a number of potential confounding factors, but not for employment, which could have contributed to this effect.

- *Women are more likely to visit their GP than men, and variation in referral could reflect different thresholds of symptom tolerance reached before a patient consults their GP* [123;128;130;131]

  Kendrick et al's (2009)[80] study of referral of patients with depression to mental health or social services had the surprising findings that women were less likely to be referred than men if assessed using one depression screening and severity score, but more likely to be referred when using a different score. This may reflect slight differences in the two scores: HADS (women assessed using this score were more likely to be referred than men) evaluates both depression and anxiety, whilst the PHQ-9 is based strictly on the Diagnostic and Statistical Manual of Mental Disorders (DSM). These findings could be explained by men being less likely than women to report mental health concerns to their GP, but when they do attend perhaps being more likely to meet the DSM threshold for diagnosis.

None of the higher rated studies that examined patient gender adjusted for any difference in the rates of GP attendance by gender. However given that the overall pattern appears that women are less likely to be referred in some situations, and this is despite the fact that we know women attend primary care more frequently, there may be other underlying influences that we are not yet aware of.

It therefore remains unclear what underlies the variation in referral (and the range of that variation) by patient gender seen in this systematic review. As with patient age it is also uncertain both whether this variation is intentional and, whether intentional or not, it is in the best interests of the patients.

3.4.2.3 : Deprivation

Whilst nearly half of the studies I examined in detail in my systematic review considered the association between the level of deprivation (patient or practice area) and GPs' referral of patients for investigation and to secondary care, their findings were inconsistent. This was even the case amongst studies considering the same outcome measures: for example between McGovern et al (2008, 2)[86] and Hamilton et al (2010)[83] both exploring the association between patients' socio-economic circumstance and blood test monitoring of known diabetic patients, and between Saxena et al (2007)[90] and McGovern et al (2008, 1)[85] examining the association between deprivation and referral for investigations for patients with CHD. These differences may reflect methodological differences between studies: for example different sample populations, or whether there was adjustment for potential confounding factors. It is also important to note that whilst all the medium or high-rated studies used area based measures of deprivation, they did not all use the same measure.

It is possible that, as with patient gender, the association between patient socio-economic circumstance and referral to secondary care varies according to a patient's symptoms: McBride et al (2010)[65] observed deprivation was significantly associated with referral of patients with hip pain or dyspepsia, but not those with post-menopausal bleeding. This may reflect the 'red flag' nature of post-menopausal bleeding (which should always be considered pathological) in contrast to dyspepsia and hip pain.

This systematic review does suggest that patients from more deprived areas may be more likely to be referred for preventative interventions than those who are more affluent: deprivation was shown to be associated with higher rates of referral to both smoking cessation services (Simpson et al, 2010)[91] and exercise referral schemes (Sowden et al, 2008).[92] This may reflect GPs having a greater awareness of unhealthy lifestyle behaviours in this population, although it must also be noted that, in respect of the exercise referral schemes considered in Sowden et al's (2008)[92] study, these were specifically targeted at deprived areas, which could explain their observation.

3.4.2.4 : Patient ethnicity

Only three of the included studies explored the impact of patient ethnicity on referral for investigations or to secondary care although, as discussed in Section 3.3.3.2.2, several of the lower rated studies also looked at ethnicity. The lack of high quality studies examining the association between ethnicity and referral is likely to reflect the difficulties of collecting ethnicity data, and the lack of complete data in several GP research databases due to poor reporting of ethnicity (particularly in studies using routine data from a few years ago, when ethnicity reporting was far less complete). It also reflects the ongoing challenge that researchers conducting observational studies examining patient ethnicity face: that of doing so in diverse enough populations, but also across large areas so that the findings are generalisable.

3.4.2.5 : GP and practice characteristics

Only four of the included studies examined the association between practice characteristics and referral for investigations or to secondary care, and no medium or high rated studies considered the impact of individual GPs' personal characteristics on these outcomes - although a number of the lower rated studies did consider either individual GP or practice characteristics. As with patient ethnicity, it is therefore not possible for my systematic review to effectively evaluate the impact of these characteristics on GPs' referral behaviour.

3.4.2.6 : Studies' consideration of underlying reasons for this variation

I considered whether any of the studies examining quantitative differences in referral and investigation went on to explore the possible reasons underlying these differences. I also looked at a number of the studies which met the criteria for inclusion in this systematic review but were excluded because they only contained qualitative data to see if they identified potential reasons for these differences in referral.

While a number of studies hypothesised about factors that *could* explain the non-clinical variations in GP decision making that my systematic review has identified, their findings did not provide concrete evidence for or against any of these suggestions.

3.4.3 : Strengths and limitations of my systematic literature review

3.4.3.1 : What this systematic review adds to our understanding of this field

In this systematic review I set out to identify and critically appraise all literature relevant to determining the non-clinical patient, GP and practice related characteristics significantly associated with variation in UK GPs' referral of patients for investigations (including diagnostic tests) or to secondary care specialists and services. I have conducted a wide study in which I have looked at the association between referral and several key patient socio-demographic characteristics, as well between referral and some practice characteristics.

Through this systematic review I have increased our understanding of this field: I have enhanced the existing literature by demonstrating that some socio-demographic characteristics (e.g. patient age and gender) are clearly associated with variation in GPs' referral behaviour, and have also identified areas of inconsistency and uncertainty where further research is required.

3.4.3.1.1 : Enhancing the existing literature

This is the first systematic review of this topic, and is more comprehensive than previous narrative reviews have been. I conducted this systematic review because (as I discussed in Section 1.7), whilst Hajjaj et al (2010)[67] reviewed the literature on non-clinical causes of variation in clinical decision making, the contribution of their review to my research is subject to a number of limitations: their search methods were not systematic or particularly thorough, it was unclear how studies were selected for inclusion, results were not reported consistently, and it was not focused solely on UK GPs' decision making.

In so far as they can be compared, the key findings from my systematic review are consistent with the findings of both the UK and worldwide studies that Hajjaj et al (2010)[67] report in their review: that the oldest patients are less likely to be referred to secondary care services and often also less likely to undergo investigations and diagnostic tests, and that patient gender has a clear but inconsistent association with referral (varying with both disease and outcome measure).

3.4.3.1.2 : Identifying remaining gaps in the literature

In conducting this systematic review I also set out to identify any areas of uncertainty or inconsistency in the association between non-clinical characteristics and GPs' decisions to refer patients, as well as to identify any possible explanations for these and considerations of how they could be addressed in further research.

Some gaps in the literature and our understanding that I have identified are:

- *Whether patient ethnicity is associated with variation in referral*
  As discussed in Section 3.4.2.4, my review identified a lack of high quality studies examining the association between ethnicity and referral. I proposed that this could reflect not just the lack of availability of ethnicity data and the challenge of conducting studies in populations with sufficient ethnic diversity, but also the need for data from across a large enough area to ensure that the findings are generalisable. The challenges of studying the effect of ethnicity is a well-recognised issue that researchers have been taking steps to improve.[132;133] However, reading more recent studies (published since 2012) that meet the inclusion criteria for my systematic review indicates that, whilst some studies using GP databases are now reporting a higher recording of patients' ethnicity data,[134] the majority of studies examining patient ethnicity continue to be local studies in multi-ethnic areas, so the issue of their findings not being generalisable remains.[135-138] In order to ensure a sufficiently diverse, yet generalisable, sample researchers may need to consider using different study methods to examine the association between patient ethnicity and referral.

- *Whether individual GPs' personal characteristics are associated with variation in referral*
  My systematic review also identified a lack of high quality studies examining the association between individual GPs' personal characteristics and their referral behaviour, despite a number of the lower rated studies seeking to address this. This is therefore an important topic for future high quality research to address.

- *The underlying reasons for non-clinical variation in referral*
  As discussed in Section 3.4.2.6, whilst a number of the studies in my systematic review hypothesised about factors that *could* explain the non-

clinical variations in GP decision making, my review could not assess whether these were true. There are three main reasons for this: that either the studies considering these characteristics had been conducted poorly and were therefore excluded from my review; or that no high quality studies considering these characteristics have been undertaken; or that any research considering these characteristics had been conducted on a small scale or qualitatively. This is an important gap in the literature because understanding the underlying reasons behind non-clinical variation in referral will be vital if we are to determine whether it has an impact on patient outcomes and, if so, how to resolve it.

3.4.3.2 : Limitations

3.4.3.2.1 : Unable to quantify the extent of the variation identified

There was significant heterogeneity in both diseases and outcomes in the 19 medium or higher rated studies included in this systematic review. This meant that I could not conduct a meta-analysis, so the extent to which it is possible to draw overall conclusions about these studies is limited. For example it is clear that aggregating the studies that examined the association between patient gender and referral for investigation or to secondary care could have resulted in an overall summary estimate that patient gender had no effect - whereas in reality my systematic review has identified that patient gender has a strong impact on referral, albeit in different directions for different symptoms and diseases.

3.4.3.2.2 : Not reflective of the full breadth of the literature

Another limitation of this review is that the medium and highly rated studies do not reflect the full breadth of the literature we critically appraised. As a result it was not possible for this systematic review to evaluate completely the extent to which non-clinical characteristics affect GPs' referral in all circumstances.

For example there is clearly substantial concern amongst researchers that there might be non-clinical variation in mental health referral and treatment, since a number of the 68 studies that met my review's inclusion criteria examined investigation and referral of patients with mental health issues.[80;95;98;110;118;120] However because the majority of studies are of poor quality, it is not possible to examine this effectively. This raises a potential question for future research in how either the study design, or the reporting, of studies about non-clinical variation in the diagnosis and management of mental health conditions can be improved.

Similarly, there is also a lack of high quality studies considering the impact of patient ethnicity and individual GPs' personal characteristics on investigation/referral. Developing research methods to study these topics that address some of the challenges and limitations identified in my systematic review must be a priority if we are to continue increasing our understanding of this field.

3.4.3.2.3 : Assessment of study rating was potentially subjective

As discussed in Section 3.3.1.3, I did not use a quantitative scoring system to allocate studies a rating of quality and relevance. While all scoring systems have an element of subjectivity, my approach did mean that there was a degree of subjectivity in the process of producing an overall rating that would not be present using a summary scoring approach. I sought to address this by having two reviewers (JS and I) independently rate each study and then discuss any discrepancies (which occurred for less than 6% of the studies rated). Since I only reported the findings of studies rated medium or higher, rating allocation has the potential to influence the findings of this systematic review. However it should be noted that whilst there were some studies which JS and I initially rated differently, our independent ratings always agreed on whether a study's rating was in the 'medium or higher' or 'low rated' category.

3.4.3.2.4 : Does not include more recent literature

This systematic review was thorough and extensive (covering just over 5 years of research) and I have confidence that it has captured the vast majority of the relevant literature from this period. However it only includes literature up to April 2012, which is a significant limitation, not least because of the huge volume of new literature which has been published since then. I looked extensively into how I might update this review, however this has proved challenging: using the same search strategy to search the literature published between April 2012 and December 2014 yielded an additional 21,445 studies once duplicates were removed. This was far too many studies to screen using the same method I used for my initial review, given the time available and that I would be working alone.

I subsequently tried a number of strategies to see whether they made updating the systematic review manageable practically:

- *Searching by target phrases to speed up screening*
  From my experience of the screening process when conducting my 5 year systematic review, I was aware that many of the studies identified would be able to be excluded very rapidly as their title would clearly demonstrate that they did not meet the review inclusion criteria: for example stating the country of their study, population group (e.g. children, or animals) or that

they used solely qualitative research methods. I therefore conducted specific targeted searches of the 21,445 study titles using key words such as country names, 'qualitative' and 'paediatric/children', and excluded those which did not meet the inclusion criteria. Whilst this did enable me to exclude a significant proportion of the studies, I was still left with 12,895 studies that would require further title, abstract and full paper screening (using the same process as described in Section 3.2.2).

- *Refining my search strategy*

  I was aware that whilst my initial search strategy was thorough, it yielded an enormous number of studies (11,791) of which very few (68, 0.6%) were finally selected for critical appraisal and quality assessment. This reflected the fact that my search strategy was designed to ensure as far as possible that I did not miss studies containing data on the influence of non-clinical characteristics (particularly patient socio-demographic characteristics) but whose research question and primary aim did not initially appear directly relevant to my systematic review's question. With the help of a systematic reviewer I therefore worked to refine and develop a more specific search strategy in order to try to reduce the number of studies to screen. However even with a tighter search strategy over 16,000 studies (with duplicates removed) were selected for title screening, again far too many to be feasible for me to screen manually.

- *Considering using text mining software*

  I considered the possibility of using text mining software (a software application called EPPI-Reviewer 4, developed by researchers at the UCL Institute of Education) which aims to streamline the process of screening titles and abstracts by using term recognition to identify key words in the titles and abstracts of papers the person screening selects for full paper review, then reordering the list of unscreened studies so that those most similar to the studies already selected for full paper review are viewed first.[139] Once the person screening reaches the threshold where a pre-agreed number of studies were consecutively excluded, the screening process is truncated.

  Whilst this approach has the potential to enable reviewers to practically conduct large scale systematic reviews, it does have limitations: in particular

> it will not identify literature that uses different language/key words to the studies already selected, and the truncation of the screening process may mean relevant studies are missed. In addition, a specific challenge that I faced when trialling using this software for my review update was that only a low proportion of studies met the inclusion criteria for my systematic review on both title and abstract and were therefore selected for full paper review. This may have limited the software's ability to effectively generate appropriate key words for the term recognition.

In the end, after discussion with my supervisors, we concluded that none of these options lowered the number of studies needing to be screened sufficiently for it to be viable for me to update this systematic review as part of my PhD (alongside the other work I needed to complete, such as Study 2).

3.4.4 : Implications for future research, policy and practice

3.4.4.1 : Future research

The literature contains empirical evidence that patient age and gender (and to some extent deprivation) are associated with differences in GPs' referral behaviour. Future research is needed to start to unpick why there is variation in GPs' referral for investigations (including diagnostic tests) or to secondary care by patients' age and gender, and whether this is conscious, intentional, and in patients' best interests. However we do not have enough information to be able to draw conclusions about the impact of patients' ethnicity, individual GP characteristics (socio-demographic, experience/work related, or psychological) or practice and organisational characteristics, on GPs' decisions to refer patients. Thus further research examining whether these characteristics in particular are associated with variation in GPs' referral is another important priority. However it is vital that the methods of future studies seek to address the methodological shortcomings my review has identified: for example considering potential relevant confounders (when examining the effect of patient age it is likely to be important to consider co-morbidity, whilst when examining the effect of patient gender it may be useful to also consider GP attendance rates).

This systematic review has also raised the issue of whether typical observational studies using retrospective data are able to provide the data we need to answer these future research questions. For example I have proposed patient ethnicity as a characteristic to investigate further; however it is likely that the methodological limitations of studying this, highlighted by my systematic review (i.e. either lack of recording of ethnicity data, or the under-representation of certain ethnic groups in many regions of the UK), may mean that producing high quality, national, observational studies examining patient ethnicity is virtually impossible. In this case, future researchers may need to consider novel ways to examine the effect of both this and other characteristics on GPs' decision making. It is also important to consider whether research questions exploring the underlying reasons why variation in referral occurs are best answered using retrospective data (used by the vast majority of studies in this systematic review) or whether other methods are needed.

3.4.4.2 : Policy and practice

If GPs are aware of the existence of non-clinical variation in referral it will enable them to consider the extent to which this is occurring in their own practice, as well as reflecting on why this could be occurring and whether it is due to conscious decision making, or unconscious tendencies.

In addition, this systematic review highlights the importance of recorded routine data for research and evaluation. Whilst studies using routine data are unlikely to be able to completely answer the research questions this review has highlighted, they remain a key source of information for understanding GPs' decision making. An increased focus on data recording would therefore be extremely valuable.

4 : The development and methods of the GP decision making study (Study 2)

4.1 : Introduction

Having conducted a systematic review (Study 1) to examine non-clinical characteristics associated with variation in UK GPs' decisions to refer patients for investigations (including diagnostic tests) and to secondary care, I then proceeded for the rest of my PhD to focus specifically on whether there is non-clinical variation in GPs' decisions to refer patients with symptoms indicative of lung cancer. I addressed this question in the GP decision making study (Study 2).

One purpose of my systematic literature review was to inform the research questions and study design of the GP decision making study. The review identified some clear areas of uncertainty and gaps in the existing literature for future research, as well as methodological shortcomings that will be important for future research studies to consider if they are to address these issues effectively. I therefore designed and developed the GP decision making study with these questions and considerations in mind, in order to ensure that the study would further increase our understanding of non-clinical variation in GPs' decision making - in particular for patients presenting with symptoms that could indicate a diagnosis of lung cancer.

In Chapter 4 I focus specifically on the design, development and methods of the GP decision making study, the development and delivery of which formed a significant part of my PhD. I then go on to report and discuss the results of the vignette study (Study 2a) in Chapter 5, and the post-consultation survey (Study 2b) in Chapter 6.

4.1.1 : Aim

To examine the constellation of patient and GP characteristics associated with the management decisions that GPs make for patients presenting with symptoms that could indicate lung cancer.

4.1.2 : Objectives

i) To develop an interactive study tool to examine GPs' decision making behaviour, and to evaluate its suitability to do this.

ii) To undertake a factorial vignette study to:

- identify factors associated with variation in GPs' management decisions for patients presenting with symptoms that could indicate lung cancer, in particular their decision to perform a chest X-ray;
- examine how these management decisions vary by non-clinical characteristics (patient or GP characteristics, or a combination of these).

iii) To conduct a questionnaire survey to understand some of the reasons behind GPs' management decisions.

4.2 : Methods

4.2.1 : GP decision making study team

The main work of the GP decision making study has been conducted by me and JS, overseen by RR. Many other people have contributed at various stages to the study's design and delivery. Table 5 lists them and how I reference them in this thesis.

Table 5 : The GP decision making study team members and their roles

Name	Role	Referred to as
Policy Research Unit (PRU) investigators		
Rosalind Raine	Project principal investigator, my supervisor	RR
Stephen Duffy	Director of PRU, project statistician, my supervisor	SD
Willie Hamilton	Academic GP consulted about content and analysis	WH
Una MacLeod	Academic GP also studying GP decision-making	UM
Spencer Robinson	UM's PhD student, researching similar aims qualitatively	SR
Greg Rubin	Academic GP consulted about content	GR
Study researchers		
Jessica Sheringham	Project manager, my supervisor	JS
Rachel Sequeira	Myself, PhD student working on project full-time	RS *(or I)*
Joe McDonnell	Public health trainee working on project 05/12 to 01/13	JMc
External collaborators		
Joseph Forrest	Atheneum Educational Technologies, designed software	JF
Matt Aucott	UCL Media Services, performed filming and clip cropping	MA
Phil Mason	UCL Media Services, performed filming and clip cropping	PM
Jonathan Myles	Statistician from Queen Mary, University of London	JM
Judith Offman	Researcher at Queen Mary, University of London	JO
Anjali Bajekal	GP consulted about content, usability and recruitment	AB
Mike Gocman	GP consulted about content and usability	MG
Dave Ardron	Lung cancer patient representative (father died from disease)	DA
Tom Haswell	Lung cancer patient	TH

Several GPs, members of our research department and the PRU also provided valuable feedback on study design and development, whilst a number of temporary administrative staff helped with recruitment.

4.2.2 : Choice of study design

There are three approaches that one can take in a study of GP decision making:

i. analysis of routine data (e.g. medical records or GP databases);

ii. non-participant observation of GP consultations;

iii. undertaking an experiment.

Methods (i) and (ii) have the advantage of using real life data. Routine data are readily available, but the mix of prevalent and incident events makes it difficult to disentangle the influence of specific consultations. Observation is a resource intensive method due to the low number of relevant events (in this case symptom constellations indicative of cancer). Furthermore, medical decision making is a complex process: as McKinlay et al (2002)[140] note, "disentangling (or unconfounding) the independent and combined contribution of physician preferences and prejudices on clinical decision-making presents a formidable methodological challenge" and, as my systematic review identified, it is something that much of the existing literature has not addressed adequately methodologically. Some studies using routine data or observation may attempt to overcome this limitation by using multilevel modelling, but this is unlikely to entirely eliminate confounding. It is important that the design and methods of future studies of non-clinical variation in GPs' decision making seek to address these methodological shortcomings and issues.

Using an experimental design enables patient characteristics to be controlled and manipulated so that both the independent and interactive contribution of patient and GP characteristics to decision making can be evaluated, without the effect of confounding. It also provides an opportunity to examine patient ethnicity in such a way that the results are generalisable, by providing a sufficiently diverse sample nationally and thus bypassing the issue of the geographical under-representation of certain ethnic groups in parts of the UK that makes it virtually impossible to produce high quality national observational studies examining patient ethnicity using routine data. Several researchers have used this approach, using a factorial design to estimate the effects of patient, GP or organisational characteristics (or combinations of these) on decision making for a range of conditions including coronary heart disease, eating disorders, depression and diabetes.[121;127;141;142]

In order to control experimental factors, many factorial studies present patient information as vignettes. Several studies of medical decision making, not just those with an underlying factorial design, use vignette methods: for example the International Cancer Benchmarking Partnership (2013)[143] used text-based vignettes to examine international variations in GP referral for lung cancer. However vignette studies may not provide the same findings as studies of real practice because of their artificiality.[144] Researchers have therefore used a variety of approaches to make the use of vignettes more like real life.

The simplest vignettes are text-based. They may include a highly realistic description, but physicians still do not receive information as they would if experiencing the situation first-hand: as Raine (2002)[145] notes, "written vignettes exclude a host of factors shown to affect physician response, including auditory and visual cues such as the patient's age, ethnicity, social class, physical appearance, non-verbal behaviour and voice quality, as well as organisational and structural features".[145] Some vignette studies address this limitation by using non-text media. A few have presented vignettes as video recordings, enabling auditory and visual cues to be incorporated.[127;141;146] Others have tried to improve authenticity by enabling physicians to interact with multimedia vignettes: Epstein et al's (2008)[141] study involved physicians speaking pre-scripted questions into a microphone which was then followed by a video clip of the actor's response, whilst Harries et al (2007)[125] used 'patient' photographs and gave physicians an option to select and view additional pieces of clinical information.

As Blumenthal-Barby et al (2015)[147] recognise, a key limitation of many vignette studies is that they do not simulate key features of real life consultations, in particular where the vignette design is not interactive and offers little or no opportunity for physicians to ask questions of the patient. This can risk priming, and thus potentially biasing, physicians' responses: either by cuing what they should notice about the patient, or by offering only a limited selection of response options.[148] Some studies have attempted to address this limitation: for example Kostopoulou et al (2014)[149] conducted a web-based vignette study where physicians received initial standardised information about the patient but could then request further information of their choice, whereupon a researcher selected the appropriate answer and it was displayed on the physician's screen. However this not only required a researcher to provide real time responses for each vignette, but also sometimes resulted in the vignette taking significantly longer than a typical real life consultation.

For the first time, to our knowledge, we have used the combination of interactive multimedia technology and non pre-scripted vignettes to present information to GPs in such a way that we captured the experience of a real life consultation as closely as possible. This was the 'virtual patient application'.

Vignette studies do have limitations: although a few studies using this design met the inclusion criteria for my systematic review all but two were excluded on the grounds of quality, and only one (Bonte et al, 2008)[75] was rated medium or higher. As I reported in Section 3.3.3.2.3 the low quality of these studies was primarily due to unsophisticated methods and/or a significant potential for bias. During the design and development of our vignette study we therefore carefully considered how to avoid replicating these same methodological shortcomings.

4.2.3 : The virtual patient application

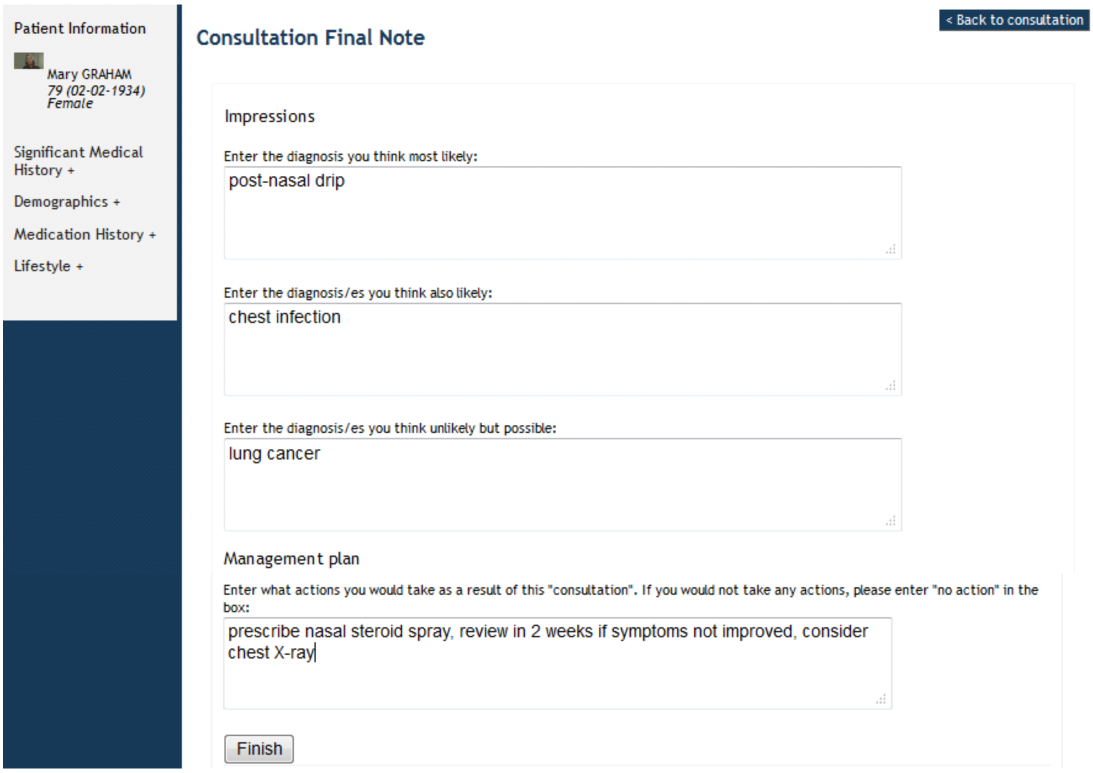
This was the online, interactive study tool which we developed and used to examine GP decision making in the vignette study.

GP participants used the application to undertake a series of six 'consultations' with virtual 'patients', each designed to take eight to ten minutes (reflecting the average length of a real life consultation). They initially watched a short video of the 'patient' (an actor) reporting a symptom. They were then able to seek further information by typing questions. The application interpreted each question and played an appropriate video clip of the 'patient's' response. GPs were also able to view each ‘patient’s’ medical records and receive findings of examinations or bedside tests they would have been able to perform if consulting an actual patient (Figure 5).

Figure 5 : The ‘consulting room’ and the actions GPs could perform to gain information

After GPs had collected all the information they wished, they typed their differential diagnosis and management plan (Figure 6). The application recorded GPs' questions, behaviour and decisions. This information provided the data for analysis.

Figure 6 : To complete the consultation GPs entered their differential diagnosis and management plan



Hosting the application online meant that GPs could complete the study in their consulting room during the working day, the same environment as their real life practice. They were encouraged to complete the study over three weeks, 'consulting' between one and three 'patients' per week.

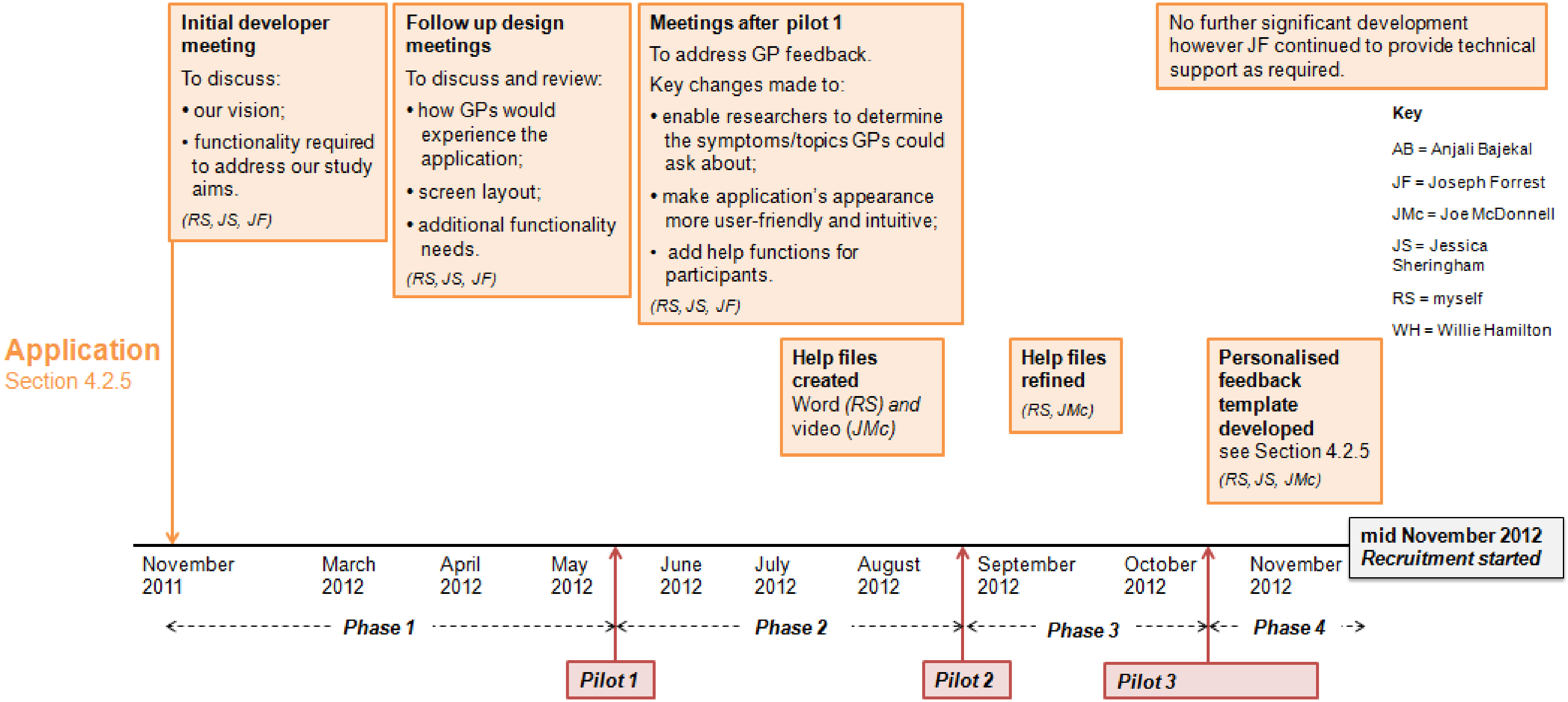
We collaborated with GPs to design the virtual 'consultations' so that they were as lifelike as possible; however we acknowledge that the experience was not completely true to life. In order to capture what factors GPs believed influence their real life decision making, I therefore also designed a post-consultation questionnaire survey (viewable at http://opinio.ucl.ac.uk/s?s=20054) which GPs completed at the end of the vignette study. I discuss the design, development and results of this post-consultation survey in Chapter 6.

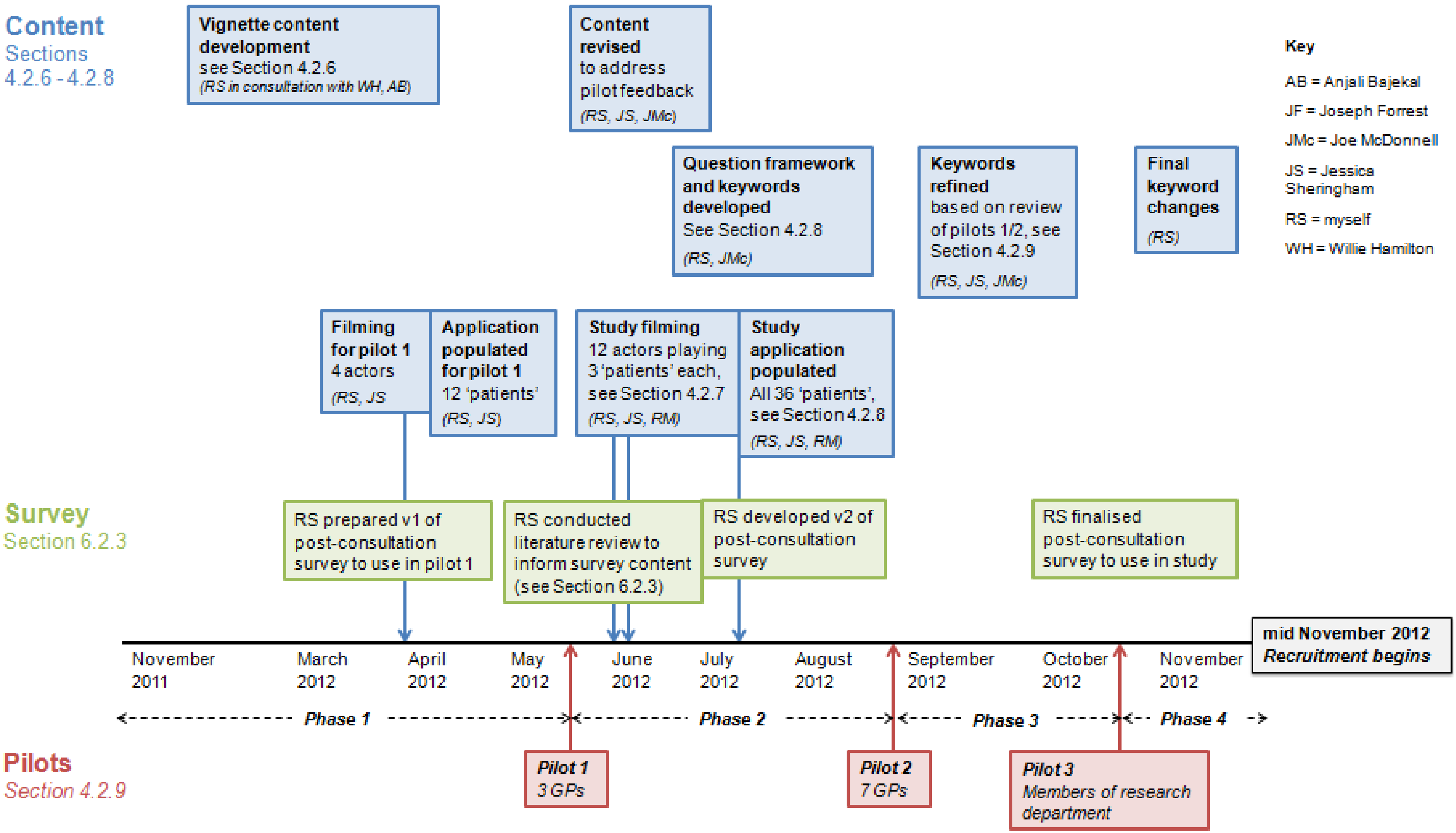
4.2.4 : An overview of the development of the virtual patient application and the post-consultation survey

Figure 7 illustrates the timing of key activities in the virtual patient application's development. Many activities happened concurrently: I will discuss these in Sections 4.2.5 to 4.2.9, and the development of the post-consultation survey in Section 6.2.3.

Figure 7 : Timeline of key activities in the development of the virtual patient application and the post-consultation survey

Box position and length correspond to the timing and duration of each activity. Where activities consisted of single events these are marked by arrows.





Content
Sections 4.2.6 - 4.2.8
Vignette content development
see Section 4.2.6
(RS in consultation with WH, AB)
Content revised
to address pilot feedback
(RS, JS, JMc)
Question framework and keywords developed
See Section 4.2.8
(RS, JMc)
Keywords refined
based on review of pilots 1/2, see Section 4.2.9
(RS, JS, JMc)
Final keyword changes
(RS)
Key
AB = Anjali Bajekal
JF = Joseph Forrest
JMc = Joe McDonnell
JS = Jessica Sheringham
RS = myself
WH = Willie Hamilton
Filming for pilot 1
4 actors
(RS, JS
Application populated for pilot 1
12 'patients'
(RS, JS)
Study filming
12 actors playing 3 'patients' each, see Section 4.2.7
(RS, JS, RM)
Study application populated
All 36 'patients', see Section 4.2.8
(RS, JS, RM)
Survey
Section 6.2.3
RS prepared v1 of post-consultation survey to use in pilot 1
RS conducted literature review to inform survey content (see Section 6.2.3)
RS developed v2 of post-consultation survey
RS finalised post-consultation survey to use in study
mid November 2012
Recruitment begins
November 2011
March 2012
April 2012
May 2012
June 2012
July 2012
August 2012
September 2012
October 2012
November 2012
Phase 1
Phase 2
Phase 3
Phase 4
Pilots
Section 4.2.9
Pilot 1
3 GPs
Pilot 2
7 GPs
Pilot 3
Members of research department

4.2.5 : Application development

The application software was produced by Joseph Forrest's (JF) team. RR secured JF's agreement to develop an interactive, video-based application for our GP decision making study in 2010. JF had experience of developing similar software for another study of physician decision making (Harries et al, 2007).[125]

JS and I met with JF in November 2011 to discuss the brief and aims of the study. The developers' role was to develop a 'shell' that we could use to insert the content we required. The final version of the application was ready in November 2012. During that time the application and its content went through a number of phases of development. After each of the first three phases JS and I ran a pilot to identify issues needing further refinement (see Section 4.2.9).

The development of an interactive application that could simulate a real life GP consultation and present content to GPs effectively and realistically was challenging. In this section I will briefly describe the main issues we faced and how we overcame them.

- We initially struggled to reach a shared understanding with the developers about the need for the application to be user-friendly for GPs. However once we fed back the pilot results and GPs' comments the developers started to understand the need for the application to be more intuitive.
- Making the application more user-friendly and intuitive was not straightforward, and some limitations could not be overcome - for example the application required GPs to repeat the name of the symptom they were asking about in all their questions (e.g. 'how long have you had chest pain' or 'what makes the breathlessness worse') which does not realistically mimic spoken conversation. We had to accept this and find ways to work around the application's constraints. I produced a PDF 'help' file (including trouble-shooting tips) that GPs could access whilst using the application, and JMc produced an introductory help video (viewable at http://www.ucl.ac.uk/stream/media/swatch?v=c22f1a2b58b8) which we asked all GPs to watch before starting the study. We also provided GPs with feedback after their first 'consultation' to reduce the likelihood that they missed key information in future 'consultations' because of repeated error. I ensured that

the feedback we provided was standardised for all GPs by producing a feedback template and list of key errors (Appendix 4).

- It became apparent that it would not be possible to develop all the functionality originally planned. JS and I therefore had to prioritise our requirements to ensure those that were key were met and, where possible, to find alternative solutions: for example we had originally planned for the post-consultation survey to be part of the virtual patient application, however I ended up building this myself using UCL Opinio software.[150]
- We could not expect our virtual 'consultations' to completely replicate real life. The pilots showed us the importance of managing GPs' expectations in the presentation of the application by acknowledging that it was simply a simulation.

4.2.6 : Content development

The interactive nature of our vignette presentation meant that developing the vignette study content was a complex process. In order to ensure that our vignettes were both authentic and suitable for answering our research questions we had to consider:

- **Study content**
  - our experimental factors ('patient' characteristics examined)
  - 'patient' profiles
  - GP characteristics examined
- **How this content was presented to the GP**
  - how 'patients' disclosed information about their presenting symptoms
  - examination and bedside test results
  - 'patient' medical records

4.2.6.1 : Study content

Our design used four experimental factors.

Three of these reflected 'patient' characteristics that are known to be associated with variation in lung cancer survival rates, but whose effect on inequalities in GPs' rates of referral for investigation or to secondary care is uncertain:

- **Ethnicity**:[21] three variations (white, black Caribbean, South Asian)
- **Gender**:[15] two variations (male, female)
- **Socio-economic circumstance**:[151] two variations (affluent, socio-economically disadvantaged)

Our fourth experimental factor was the **clinical risk of lung cancer**.[63] We used the following to derive risk level: age, smoking status, presenting symptoms and the duration of these symptoms.

We included three levels of risk, each of which we would expect GPs to manage differently:

Low 'watch and wait'

Medium *either* order a chest X-ray *or* 'watch and wait' with safety-netting [iv]

High order a chest X-ray

I created six clinically authentic 'patient' profiles (two at each level of risk), shown in Table 6. Each profile had two symptoms, one which the 'patient' would volunteer to the GP, the other which they would only disclose if questioned further and asked whether they had that symptom (e.g. "do you have a cough?")

[iv] Where 'safety-netting' involves the GP managing uncertainty, often by making contingency plans to review the 'patient' or adjust the management plan if symptoms worsen or continue, or something unexpected happens.

Table 6 : The six 'patient' profiles each GP saw

PPV = positive predictive value (the likelihood that someone with that combination of characteristics has lung cancer)

Low risk: 'watch and wait' appropriate

	Volunteered by 'patient' or available on screen as 'patient notes'			**Only available if GP asks**		**PPV**	**Notes**
Profile	**Age**	**Smoking status**	**Symptom 1**	**Symptom 2**	**Duration**		
1	58/59	Non-smoker	Breathlessness	Fatigue	10 days	0.4%	Also has swollen ankles *'Distracting vignette' – similar symptoms but history suggesting heart failure not lung cancer*
2	58/59	Smoker	Chest pain	Cough	10 days	1.1%	*Matched with profile 3 to examine effect of age*

Medium risk: <u>either</u> 'watch and wait' (with safety-netting) <u>or</u> refer for chest X-ray appropriate

	Volunteered by 'patient' or available on screen as 'patient notes'			**Only available if GP asks**		**PPV**	**Notes**
Profile	**Age**	**Smoking status**	**Symptom 1**	**Symptom 2**	**Duration**		
3	78/79	Smoker	Chest pain	Cough	Uncertain ~3 weeks	1.7%	*Matched with profile 2 to examine effect of age*
4	78/79	Non-smoker	Cough	Appetite loss	Uncertain ~3 weeks	2.5%	

High risk: immediate referral for chest X-ray appropriate

	Volunteered by 'patient' or available on screen as 'patient notes'			**Only available if GP asks**		**PPV**	**Notes**
Profile	**Age**	**Smoking status**	**Symptom 1**	**Symptom 2**	**Duration**		
5	58/59	Smoker	Breathlessness	Fatigue	>1 month	3-4%	COPD co-morbidity
6	78/79	Smoker	Chest pain	Weight loss	>1 month	14%	

All six profiles were presented to each participating GP; however the specific combinations of socio-demographic characteristics (gender, ethnicity and socio-economic circumstance) they viewed varied. We constructed a template of thirty-six 'patients' who together covered all combinations of our four experimental factors (Appendix 5). Each GP was randomly assigned six of these 'patients', one from each profile.

I developed the content of the six profiles in consultation with academic GPs (WH, UM, GR), patient representatives (DA, TH) and my supervisors.

With one exception, I aligned the risk profiles' content (and their expected management) to the NICE (National Institute for Health and Care Excellence) guidelines for the investigation of suspected lung cancer published in April 2011.[152] The exception was the inclusion of appetite loss, which I included on the advice of WH, and on the basis that a number of studies have shown a strong association between appetite loss and increased risk of lung cancer,[153;154] (this has since been reflected by the inclusion in the 2015 NICE guidelines of appetite loss as a symptom warranting urgent chest X-ray referral).[35] The risk profiles were also aligned with risk level using positive predictive values (PPVs) provided by WH based on his analysis from the CAPER study (2009).[63;155]

I included the most commonly presenting symptoms of lung cancer in the risk profiles. Symptoms were both lung-related and non-specific. My aim was to generate presentations that GPs would frequently encounter: about 70% of patients with lung cancer present with lung-related symptoms, and over 90% with ‘typical’ (but not necessarily specific) symptoms.[47;152]

We initially planned that all 'patients' would be smokers, since nearly 90% of patients diagnosed with lung cancer are current or ex-smokers.[156] However feedback from patient representative DA (whose father's diagnosis was delayed despite numerous GP consultations, possibly because as a non-smoker cancer was not expected) confirmed the importance of including profiles with both smokers and non-smokers.

We also collected information about GP and practice characteristics. Appendix 6 lists these, and the source of the information (either routine data, or via the registration questionnaire or post-consultation survey).

4.2.6.2 : How study content was presented to GPs

In this study GPs interacted with the content. In contrast to text-based vignette studies the information they received about a 'patient' depended on their behaviour during the 'consultation'. We sought to:

- replicate a real life GP consultation as far as possible;
- define the content and format of 'patient' video responses to GPs' questions;
- provide GPs access to background information about 'patients'.

4.2.6.2.1 : Replicating a real life GP consultation

GPs have several sources of information in a real life consultation (e.g. asking the patient questions, performing examinations, consulting medical records). We sought to mimic this as far as possible, although it is impossible to do so fully via an online application and we had to present some content differently (e.g. providing examination findings as text).

4.2.6.2.2 : Defining the content and format of 'patient' video responses to GPs' questions

In consultation with AB, I developed a list of questions that GPs would be likely to ask a patient presenting with the symptoms in each of the six profiles (Appendix 7). I refined this through consultation with WH, DA and TH, as well as using the results of the first pilot study which involved three GPs. WH and I then discussed the information that typical patients with lung cancer would provide in answer to these questions. The one exception to this was profile 1, representing a low risk of lung cancer, which I designed to be a 'distracting vignette' (suggesting a potential diagnosis of heart failure).

When considering how symptom details were presented to GPs we were guided by discussions with both GPs and patients about what would be realistic - both in the length of 'patient' responses to questions and the level of detail they provided. For example for the first pilot we filmed 'patients' providing short answers to questions, including the GP's initial question, "What seems to be the trouble?". The 'patient' also provided significant information about features of their symptom without prompting. However GPs told us that it would be more realistic for 'patients' to give an initial answer that was longer, but conveyed less information! Our final 'consultations' therefore had a long initial answer where the 'patient' disclosed the volunteered symptom and discussed how it troubled them, but nothing more.

Additional features of this symptom (such as what exacerbated it, or how long it had been present), and the second symptom of the 'patient' profile were only provided when the GPs asked specific questions to elicit this information.

Another example of how we sought to ensure that symptom details were presented to GPs in a realistic manner was that GPs needed to specifically ask a patient whether they had a symptom in order to receive an answer - a general question such as "any other symptoms?" received a "could you rephrase that" video response. There were two reasons for this - firstly for our analysis (because it allowed us to consider if GPs ask about particular symptoms more than others) and secondly the suggestion from patient representative TH who said that it was unrealistic for patients to disclose all their symptoms in response to a general question, which is consistent with peer-reviewed literature.[157]

4.2.6.2.3 : Providing GPs access to background information about 'patients'

AB and I developed a comprehensive list of examinations and tests that GPs might perform, including tests unrelated to the risk profile symptoms to avoid priming GPs' behaviour. I prepared examination and test results findings for all tests for each of the six profiles. In most cases results were the same for all ‘patients’ with that profile, although some varied according to 'patient' gender. On the advice of WH, the respiratory and cardiovascular examinations were unremarkable for all six profiles; this was to ensure we were studying GPs' responses to the presence/absence of symptoms, rather than to positive examination findings.

I also created medical records for each of the ‘patients’. These included information on socio-demographic and lifestyle characteristics, details of past medical history and medication, and a recent consultation history. For authenticity many 'patients' had co-morbidities; however I ensured that (with the exception of profile 5, which had a co-morbidity of COPD that was reflected in the profile's PPV) these did not relate to their presenting symptoms, since this could alter the likelihood of lung cancer and interfere with the risk level calculation.

4.2.7 : Filming the video content

Our key requirements for the video content were:

- *consistency*: ensuring that only the experimental variables changed between profiles (to meet the requirements of the study's factorial design);
- *authentic portrayal of 'patients'*: we used actors with medical role-playing experience because of their ability to work from a brief and give responses appropriate to their character.

4.2.7.1 : How actors were selected

We required twelve actors to fulfil the 'patient template' of our factorial design (risk level and socio-demographic factors), six who could realistically portray a 58/59 year old, and six a 78/79 year old. Within each group of actors there needed to be every combination of our three ethnicities and male/female. We represented socio-economic circumstance through appearance, accent and lifestyle.

Recruiting the diversity of actors we required was challenging - the agencies struggled to find actors who could play the older age authentically, particularly the black Caribbean and South Asian roles. We therefore opted to make our older patients 78/79 years old rather than 85/86 years old as we had originally planned.

4.2.7.2 : Ensuring consistency

4.2.7.2.1 : Actors' briefs

JS, JMc and I produced an actor's brief for each of the thirty-six 'patients' (examples in Appendix 8). This contained 'profile' information (e.g. symptom presentation and features) plus details relating to the specific character (e.g. occupation).

4.2.7.2.2 : Checklists

I produced a checklist for each of the six 'patient' profiles (Appendix 9), with the questions we needed to film a response for. The content of each checklist varied depending on the symptoms and smoking status for that profile, although some questions were common to all. The checklists not only helped ensure consistency, but also that we filmed all the responses we required.

4.2.7.3 : The filming process

We filmed for four days. UCL Media Services performed the filming in a studio set up to resemble a GP's consulting room. JS, JMc and I asked questions, prompted actors (e.g. to cough if they were portraying a 'patient' profile with this symptom) and completed the checklists.

In the virtual patient application the video of the 'patient' is only displayed in a section of the screen; we therefore filmed the majority of clips as head and shoulders close-ups in order to enable the GPs to view facial expressions. We experimented with filming our actors walking to their seats to start the consultation, but rejected this as in general the actors' healthiness was too evident in their gait and posture.

Each actor was filmed giving responses for three 'patients'. In each case they started with an introduction to their presenting symptom - how one might answer a GP's initial question, "What seems to be the trouble?" We then asked a series of additional questions in order to film the 'patient's' responses to questions about specific features of the presenting symptom, additional symptoms and their features, and other relevant subjects (e.g. smoking status).

We filmed each actor individually, but scheduled their sessions to overlap slightly so that the majority could observe the previous actor before they started filming, enabling them to get a feel for what we required. Additional takes were filmed where necessary, generally to improve the actor's responses so that they were more accurate or appropriate to the brief.

4.2.7.4 : Selecting the video clips

JS, JMc and I watched the unedited video for each actor and selected the sections of film we wanted to use (about 30 for each 'patient'). UCL Media Services then provided us with about 1,000 short video clips, each of which we converted into a format the application could play.

.

4.2.8 : Populating the application with the content

Once the study content was decided and the software developed, we populated the application so that it could present the content to GPs. This involved:

- creating databases and entering the keywords required for the language recognition software to work effectively;
- building the virtual 'patients'.

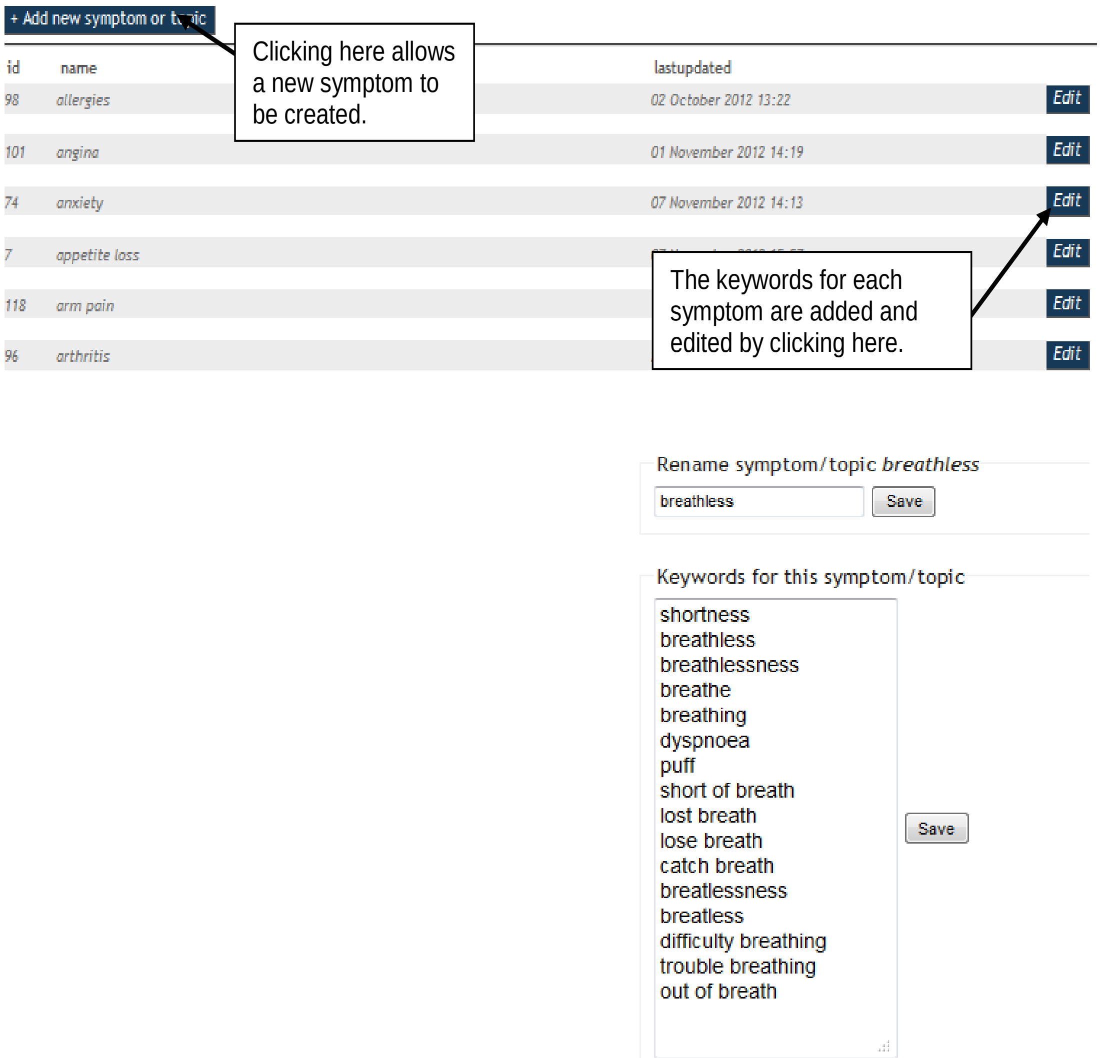
4.2.8.1 : Creating symptom and 'symptom topic' databases

The application used language recognition software to analyse a GP's question and play a video clip in response. We created two databases: one with symptoms GPs could ask about (the symptom bank), the other with features they could ask about these symptoms ('symptom topics'). We then generated keywords associated with each symptom/'symptom topic'. Developing these databases and keywords so that appropriate videos played in response to GPs' questions was very challenging.

4.2.8.1.1 : Developing the symptom bank

Based on our medical training, AB and I developed a list of symptoms and broader subjects GPs might ask patients about during a consultation. I used GPs' questions during piloting to extend this. The symptom bank comprised 66 symptoms (and subjects). I added these into the application before adding keywords for each (Figure 8).

Figure 8 : Creating the symptom bank: any symptom that a GP might ask the 'patient' about was entered into the application, then keywords were added for each symptom
(see Appendix 10 for full list of symptoms and keywords).



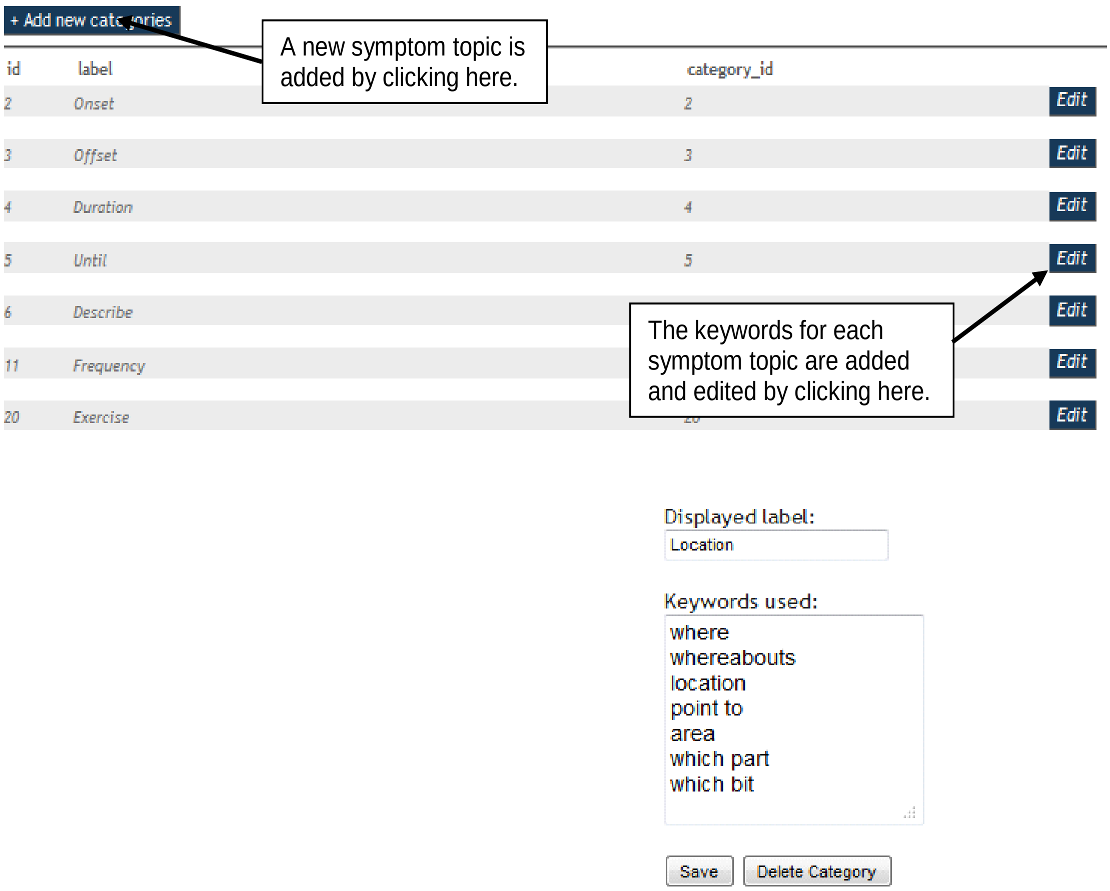
4.2.8.1.2 : Developing the 'symptom topics' database

This allowed GPs to ask questions (and receive responses) about the features of a symptom. It consisted of a list of 'symptom topics' such as exacerbating factors of the symptom, or how long it had been present. The developers introduced the capability for us to be able to define the content of this database during the second phase of development, in response to GPs' comments during the first pilot that the 'consultations' were not credible if they could not ask a wide range of questions.

JMc and I developed the database of 'symptom topics', informed by my medical training and the questions GPs asked during the pilots, and entered these into the application (Figure 9). We then developed a list of the keywords and phrases that GPs might use (or had used during piloting) to ask questions about each 'symptom topic'.

Figure 9 : Creating the symptom database: topics for GPs to ask about any of the symptoms were entered into the application, followed by keywords and phrases for each of these.

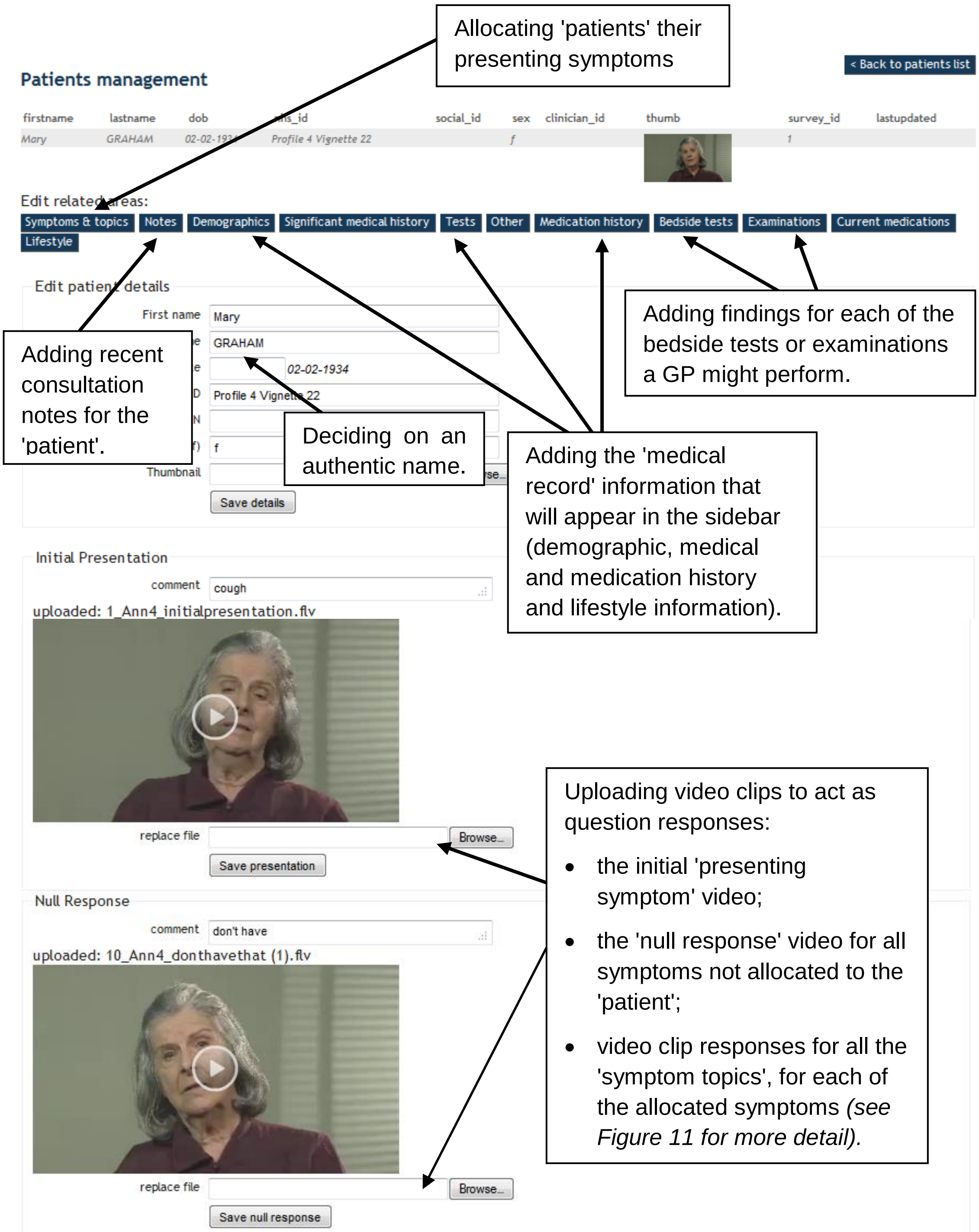
This process went through several stages of refinement; the final list of 'symptom topics' and keywords/phrases is in Appendix 11.



4.2.8.2 : Building the virtual 'patients'

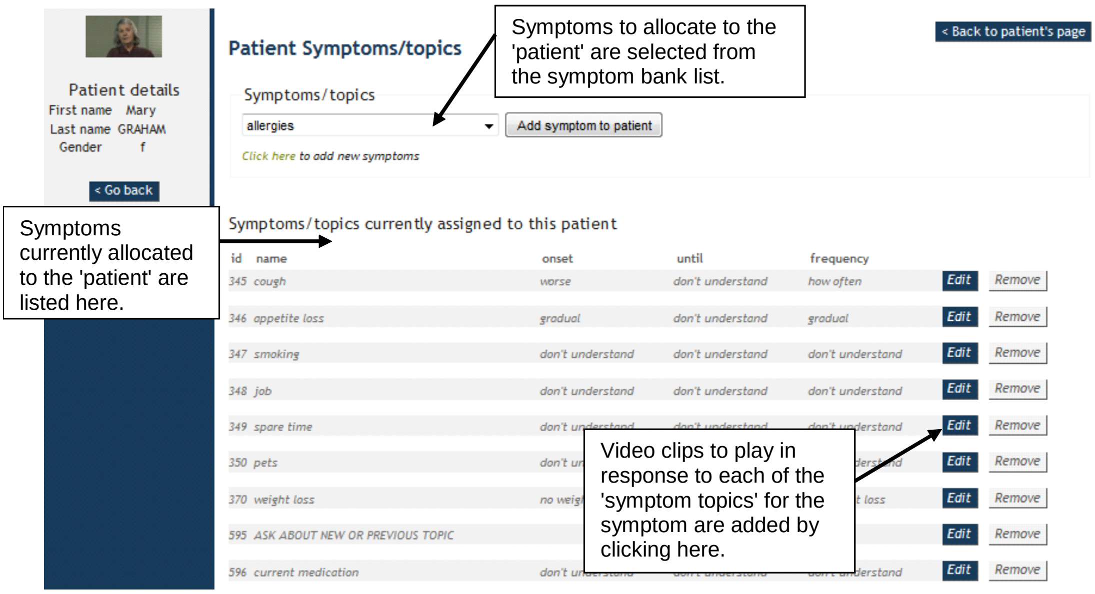
The final step in populating the application involved JS, JMc and I building our 36 'patients'. This involved a number of stages, as shown in Figure 10.

Figure 10 : Building a new 'patient' involved:



Each 'patient' was assigned (according to their profile) a number of symptoms from the symptom bank (Figure 11). For each symptom assigned we then uploaded a video clip for each of the 29 'symptom topics'.

Figure 11 : The symptoms allocated to a profile 4 'patient' (presenting symptoms cough/appetite loss)



The study's factorial design meant that we had to ensure that the clinical 'profile-related' information available to GPs was the same for all six 'patients' representing the same profile. I managed this by creating an upload document for each profile, detailing what information should be uploaded (Appendix 12). This also allowed me to note specifically the few situations where information was varied to reflect 'patient'-specific experimental factors and ensure authenticity (e.g. height varied with gender).

4.2.9 : Formative evaluation of the study tools through piloting

As shown in the timeline (Figure 7), we conducted three pilots of the virtual patient application and the post-consultation survey during the development process. These pilots had two purposes:

- to identify any changes needed to functionality, content and layout in order to ensure that both tools were as user-friendly and intuitive as possible, and check that it was feasible for GPs to complete the study in one hour;
- quantify the extent to which the virtual patient application could appropriately answer questions GPs asked.

4.2.9.1 : Pilot 1 (May 2012)
Three GPs including AB and MG

The initial pilot used 12 'patients' representing each of the symptom profiles and some combinations of gender, ethnicity and socio-economic circumstance.

GPs could use the application to ask questions about the presence of symptoms which might suggest lung cancer, but reported that the 'consultations' were not credible: they needed to be able to ask a wider range of questions if their decision making using the application was to reflect their real life behaviour. GPs also struggled to complete 'consultations' or the post-consultation survey without researcher advice.

4.2.9.2 : Pilot 2 (August-September 2012)
Seven GPs, including one of the initial piloters

This pilot used all 36 'patients' that formed part of our final study.

Its results were generally very encouraging. The majority of GPs were able to complete the 'consultations' and post-consultation survey successfully without requiring researcher input, although they still found some aspects of the virtual patient application non-intuitive. GPs fed back that the 'consultations' were credible and that they could use similar reasoning as in their day-to-day practice. Additional application functionality enabled us to broaden the range of questions GPs could ask and meant that they received appropriate video responses to many more of their questions.

In order to quantify the extent to which the virtual patient application could appropriately answer questions asked, JMc, JS and I reviewed all the questions GPs had asked during pilots 1 and 2 which led to an error message or an inappropriate video response. As a result we further refined the keywords the application used to interpret GPs' questions and determine which video to play in response. Figure 12 shows how these changes improved the proportion of questions that could be answered successfully, whilst Figure 13 gives a breakdown of the reasons for the unsuccessful questions at each of these stages.

Figure 12 : The number of questions GPs asked that the application answered with an appropriate video

'n' = number of questions asked

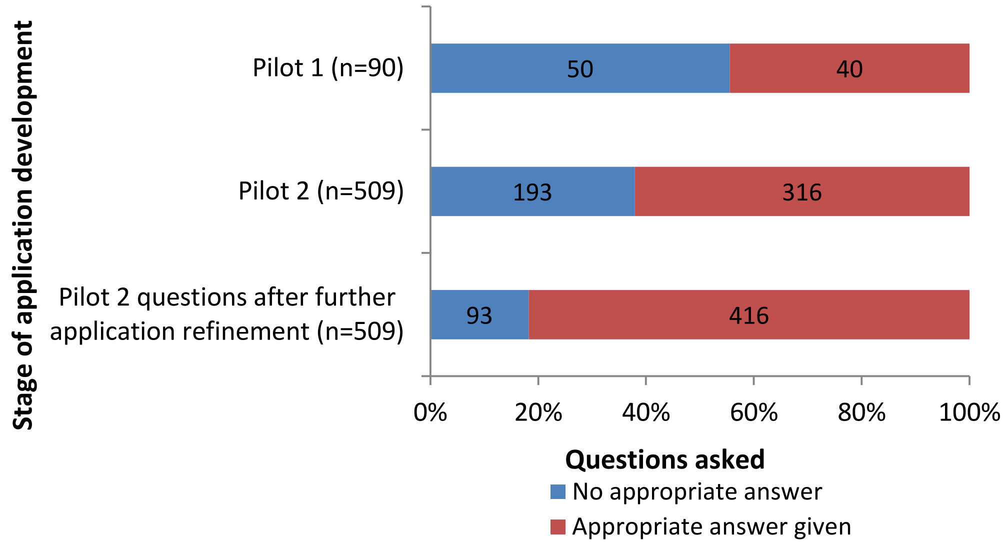
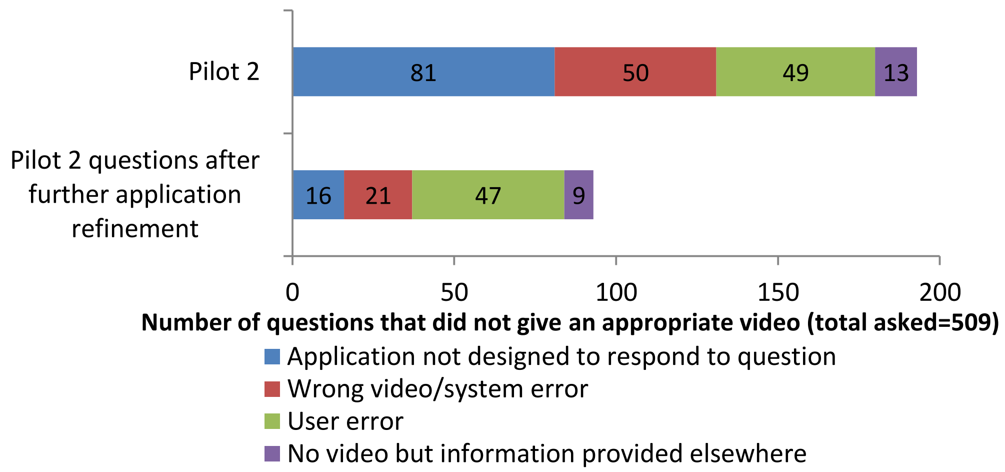


Figure 13 : Breakdown of why some questions did not lead to an appropriate video



4.2.9.3 : Pilot 3 (October 2012)
Members of our research department, including four medical professionals

Results indicated that the application now worked as required for the study and that people could successfully complete 'consultations' and make a management decision. Furthermore most questions asked now led to an appropriate video response.

4.2.10 : GP recruitment and participation

4.2.10.1 : Sample size

We aimed to recruit 216 GPs. This sample size was based on a power calculation performed by SD (see next paragraph). We planned that all GPs would view six vignettes. This meant that there would therefore be 1,296 vignette viewings in total: GPs were randomly allocated the six vignettes they viewed, one from each profile. Each of the three risk levels was to be viewed 432 times. It was not possible to ensure an exact balance of the other factors in the randomisation, but each of the two genders and two socio-economic circumstances were to be viewed approximately 648 times, and each of the three ethnicities approximately 432 times. The randomisation was constrained to ensure that no GP viewed the same actor twice.

The primary sample size calculation was based on the difference in referral intentions between variations of risk level, ethnicity, socio-economic circumstance and gender. However since not all variations of each factor in the randomisation were viewed the same number of times, this gave a range of statistical power for various main effect comparisons. For example between two risk levels (or two ethnicities), assuming a 20% variance inflation factor for clustering of GPs/'patients', 432 viewings of each risk level (or ethnicity) would give 95% power to detect a difference of 10% versus 20% referral. For a difference between socio-economic circumstance (or gender), 648 viewings of each variation would give 85% power to detect the smaller difference of 5% versus 10% referral.

4.2.10.2 : Recruitment procedure

We recruited GPs from five regions: the East of England, London, North West England, Surrey and Sussex, and the West Midlands. Recruitment was primarily through Primary Care Research Networks (PCRNs), supplemented by distribution of flyers to GPs at talks and educational sessions (this flyer is included in Appendix 13).

Once a GP expressed an interest in the study, JS or I contacted them to begin the registration process. This included sending them a participant information sheet (see Appendix 13) which contained further details about the study, what their participation would involve, and the benefits of taking part. The full recruitment and participation process is shown overleaf in Figure 14.

When recruiting GPs we presented this study as a study of GP decision making, with the aim of seeking to understand 'the ways in which GPs make decisions when faced with situations where there is a real, but low, likelihood of serious disease' (flyer) and 'the factors that influence these decisions' (participant information sheet). We acknowledged that GPs were often the first point of contact for patients feeling unwell, and thus that the decisions GPs make during consultations with these patients has an influence on patient outcomes.

We framed the study this way in order to avoid some of the methodological shortcomings of previous vignette studies, highlighted by my systematic review and reported in Section 3.3.3.2.3.

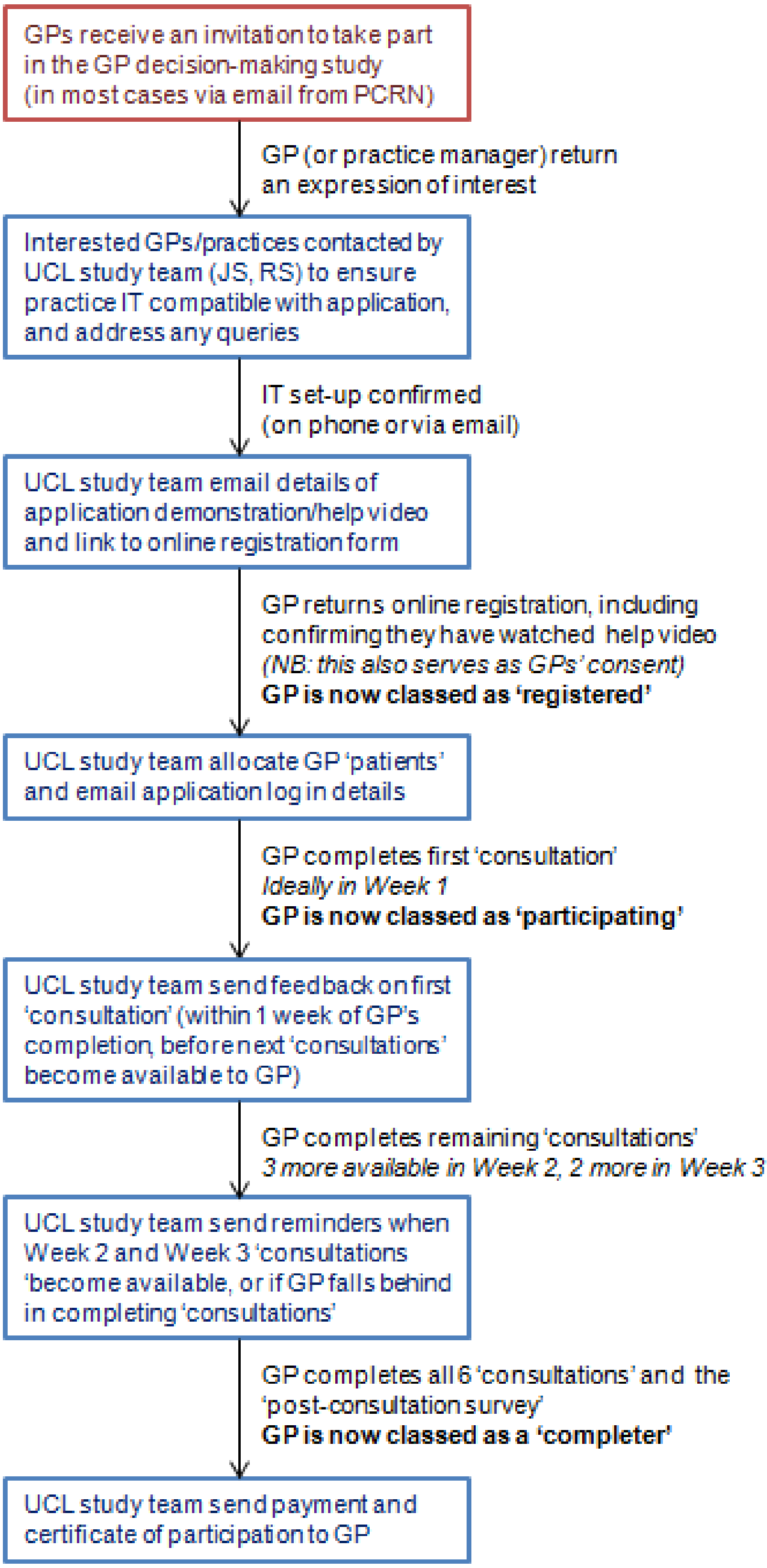
For example we chose not to share (before or during the study) that this study was focusing on symptoms that could indicate lung cancer, or that our primary outcome was referral for investigation or to secondary care. This was in part to avoid priming participants or influencing their management decisions, as well as to reduce the potential for creating participant bias (my systematic review found that in similar studies where the condition being studied was specified during recruitment, a greater proportion of participants than expected had a specialist interest in that condition).[110;122]

The GP decision making study did not require approval from an ethics committee because the study participants were healthcare professionals, recruited by virtue of their professional role.[158] However we did obtain both sponsorship and research and development approval through UCL, and for each CCG area in the regions we recruited GPs from (Appendix 14 contains examples of approvals obtained).

4.2.10.3 : Incentives

GPs were offered incentives to participate in the GP decision making study: we provided reimbursement of £80 for their time and a certificate as evidence of their participation which could be used as credit for their Continuing Professional Development (CPD). GPs were aware of these incentives at the time of recruitment

Figure 14 : GP recruitment and participation process



to the study, but only received them upon completion - which was defined as having completed all six vignettes and the post-consultation survey.

We applied for service support costs for GP practices to undertake the preparation required for the study, for example ensuring that participating GPs' computers could run the virtual patient application successfully and booking out time in GPs' diaries to complete each vignette. At the time we were recruiting for the GP decision making study (November 2012 to October 2013) the decision as to whether to provide service support costs was made by comprehensive local research networks (CLRNs). The following comprehensive local research networks were able to provide service support costs for this study: Noclor (covering North East London and North Central London, both in the London region), Greater Manchester CLRN and Cumbria and Lancashire CLRN (both in the North West England region), and Norfolk and Suffolk CLRN (in the East of England region). GP practices in these networks were able to claim service support costs once they had confirmed that the practice's browsers met the specifications required and the participating GPs had viewed the study's introductory help video. The amount provided was decided by each local network and ranged from £10 to just over £80 per GP.

4.2.11 : Data management

The virtual patient application recorded extensive information about GPs' behaviour during each of the six virtual 'consultations' they conducted. The application was developed such that all questions that the GPs typed (and the video response that played) and any additional information they sought (either by clicking on one of the patient information sidebars, or requesting an examination or bedside test) were logged, along with the exact time that this event occurred. The application also recorded GPs' free text typed differential diagnosis and management plan.

Once a GP had completed the GP decision making study a file with the logs from each of their six vignettes was downloaded from the virtual patient application by a member of the study team. Figure 15 shows the format that this data was initially presented in: it is possible to follow each stage of the 'consultation' that the GP completed.

Figure 15 : A portion of the log that the virtual patient application recorded for each vignette viewing


```
113757,"1864","expandpatientinfo","2013-11-28 19:14:36","90.244.87.202","3","","8","","13","clinician expands patient info","medication history"
113758,"1864","do you have a cough","2013-11-28 19:14:49","90.244.87.202","","","1","","","clinician input",""
113759,"1864","video 1_Olivette5_donthavethat.flv","2013-11-28 19:14:49","","3","","2","448","","patient input",""
113760,"1864","do you have chest pain","2013-11-28 19:15:07","90.244.87.202","","","1","","","clinician input",""
113761,"1864","video 1_Olivette5_donthavethat.flv","2013-11-28 19:15:07","","3","","2","448","","patient input",""
113762,"1864","tell me more about your shortness of breath","2013-11-28 19:15:31","90.244.87.202","","","1","","","clinician input",""
113764,"1864","video Olivette5_SOB_describe.flv","2013-11-28 19:15:31","","3","","2","448","","patient input",""
113765,"1864","when did your breathless worsen","2013-11-28 19:15:57","90.244.87.202","","","1","","","clinician input",""
113768,"1864","video Olivette5_SOB_howlong.flv","2013-11-28 19:15:57","","3","","2","448","","patient input",""
113769,"1864","what do you think worsens your breathlessness","2013-11-28 19:16:35","90.244.87.202","","","1","","","clinician input",""
113772,"1864","video Olivette5_illnessidea.flv","2013-11-28 19:16:35","","3","","2","448","","patient input",""
113773,"1864","have you noticed any blood","2013-11-28 19:17:05","90.244.87.202","","","1","","","clinician input",""
113774,"1864","video 1_Olivette5_donthavethat.flv","2013-11-28 19:17:05","","3","","2","448","","patient input",""
113775,"1864","ankle swelling","2013-11-28 19:17:23","90.244.87.202","","","1","","","clinician input",""
113776,"1864","video 1_Olivette5_donthavethat.flv","2013-11-28 19:17:23","","3","","2","448","","patient input",""
113777,"1864","how often are you doing your inhlares","2013-11-28 19:17:38","90.244.87.202","","","1","","","clinician input",""
113778,"1864","patient dialogue: The system has no response for that particular question. This could be because: <br/><br/>
              &bull; the system sometimes does not recognise a question without the symptom name within it: try rephrasing with the sym
              &bull;  the information is in another place: check patient information, examination & historical notes; or <br/><br/>
              &bull;  the answer to this question is not relevant to the patientâ€™s diagnosis.","2013-11-28 19:17:38","","3","","2","448","","
113779,"1864","how often are you doing your inhalers","2013-11-28 19:17:44","90.244.87.202","","","1","","","clinician input",""
113783,"1864","video 9_Olivette5_medication.flv","2013-11-28 19:17:44","","135","","2","448","","patient input",""
113784,"1864","have you noticed calf swelling","2013-11-28 19:18:18","90.244.87.202","","","1","","","clinician input",""
```

4.2.12 : Analysis

I conducted the following quantitative analyses.

4.2.12.1 : GP participant characteristics

I performed descriptive analysis of the characteristics of all GPs who participated (Appendix 6). Where possible I also compared the characteristics of our GP study population to the overall population of GPs practising in England, in order to examine selection bias.

4.2.12.2 : Primary outcome measure

The primary outcome was whether a GP referred the 'patient' for chest X-ray (CXR), or to a secondary care service where a chest X-ray would almost certainly be performed given the 'patient's' symptoms (e.g. referral to a respiratory specialist, or sending the 'patient' to an A&E department). This variable was constructed from the free text management plan responses that GP participants entered for each vignette completed, according to pre-defined criteria. The validity of each primary outcome was confirmed by a GP. I discuss the process and challenges of developing these criteria in Section 5.1.2.1.

We decided to use referral for chest X-ray as our primary outcome measure after consultation with academic GPs. If a GP suspects lung cancer a chest X-ray is the most appropriate first-line investigation. GPs might also refer for a chest X-ray if they suspect other chest/lung-related disease; however a radiologist should identify any visible pathology (including lung cancer) regardless of GPs' differential diagnoses.

4.2.12.2.1 : Descriptive analysis

This involved determining the proportion of 'patients' referred for chest X-ray: both overall and by the four 'patient' experimental factors and by GP characteristics. I performed these analyses using Stata.[159]

4.2.12.2.2 : Hierarchical modelling

These analyses were conducted by JM and SD, and further details of their methods are available in Appendix 15 (the primary results paper submitted for publication by the GP decision making team).

JM and SD analysed the data by fitting multilevel logistic regression models using Markov Chain Monte Carlo for estimation, allowing variation between GPs and between vignettes within GPs. This allowed for a correlation between outcomes within a given GP but independent outcomes for two vignettes viewed by different GPs. Estimation of odds ratios and 95% credible intervals was carried out using the RStan library in R version 3.0.2. Significance testing was carried out using Wald tests based on the means and posterior variances of the estimates.

Variations in outcome were examined by the four ‘patient’ experimental factors, an indicator variable for whether GPs had elicited the presence of the second symptom during the 'consultation' (as opposed to only having information about the presenting symptom to make their management plan), and by certain GP characteristics (their demographics, experience and region).

Two models were built in order to examine differences by clinical profile and by age. These were:

- Model 1 examined variations by clinical profile, controlled for all ‘patient’ and GP characteristics associated with investigation (with a p value of ≤0·1) and whether GPs elicited the second symptom;
- Model 2 examined variations by ‘patient’ age. Investigation in profiles of younger ‘patients’ (~aged 58-59 years) were compared with profiles of older ‘patients’ (~aged 78-79 years), controlled for all other ‘patient’ and GP characteristics associated with investigation (with a p value of ≤0·1), smoking status and whether GPs elicited the second symptom.

A supplementary analysis that replicated Model 1 was conducted to examine whether findings were explained by GPs’ responses to profile 1, the deflecting vignette.

4.2.12.3 : Other analysis

I also performed quantitative descriptive analyses of GPs' consideration of lung cancer as a possible diagnosis. This data was obtained from the differential diagnosis GPs entered for each 'patient'.

I discuss my methods of analysis for the post-consultation survey in Section 6.2.5.

5 : Results of and reflections on the GP decision making study's vignette study (Study 2a)

In Chapter 4 I discussed the development and the methods of the GP decision making study, in particular the vignette study (Study 2a). In this chapter I report and discuss the findings of the vignette study.

5.1 : Results

I present the results as follows:

- *Section 5.1.1*
  I report details of GPs' participation in the study, compare the characteristics of GPs who completed the study with those who did not, and discuss some of the challenges faced in recruitment;
- *Section 5.1.2*
  I report the key findings of the vignette study that relate to my PhD, including details of the construction of the primary variable used for analysis;
- *Section 5.1.3*
  I evaluate the use of the virtual patient application as a tool to investigate GPs' decision making behaviour.

5.1.1 : GP recruitment, participation and completion

5.1.1.1 : Recruitment and participation figures for the GP decision making study

We actively recruited GPs to the GP decision making study over a 12 month period, from November 2012 until October 2013, following up all expressions of interest we received from either individual GPs or via the PCRNs during this time. GPs who had expressed an interest in the study during this time period but had not registered by the end of October 2013 were still able to register for the study until the end of November 2013; however the study was closed to new expressions of interest.

GPs were classified using a series of different descriptions as they progressed through the study. These descriptions were defined as follows:

Expression of interest	At the point that a GP contacted either the PCRN or the study team directly about participating in the study, or asking for more information, they were considered to have 'expressed an interest'. Whilst we publicised the study widely in the regions from which we were recruiting, we only followed up and recruited GPs who expressed an interest (we did not cold call or selectively target GPs).
Registered	After a member of the study team had made contact with a GP and confirmed that their computer was compatible with the study, GPs were invited to register for the study using the online Opinio registration form. Once a GP's registration was received they were described as having 'registered'.
Participant	After registration the GP was able to start the study. Once a GP had completed their first virtual 'consultation', including entering a management plan, they were described as a 'participant'.
Completed	A GP was only described as having 'completed' the GP decision making study once they had completed (i.e. recorded a management plan for) all six virtual 'consultations', and completed the post-consultation survey.

Table 7 shows the number of GPs who reached each stage of the recruitment and participation process, both in total and broken down by each region we recruited in.

Table 7 : GP recruitment to, participation in, and completion of the GP decision making study

Region	Total count of expressions of interest (EOIs) received	Total count of GPs registered (% of EOIs)	Total count of participants (% of EOIs)	Total count of study completers (% of EOIs)
All regions	*556*	*300 (54.0%)*	*262 (47.1%)*	*227 (40.8%)*
East of England	152	102 (67.1%)	89 (58.5%)	76 (50.0%)
London	226	113 (50.0%)	101 (44.7%)	84 (37.2%)
North West England	60	36 (60.0%)	31 (51.7%)	29 (48.3%)
Surrey & Sussex	22	11 (50.0%)	9 (40.9%)	9 (40.9%)
West Midlands	80	31 (38.8%)	25 (31.3%)	22 (27.5%)
Locum GPs	16	7 (43.8%)	7 (43.8%)	7 (43.8%)

556 GPs expressed an interest in the GP decision making study; 227 (40.8%) of these GPs completed the study.

300 of the GPs (54.0% of those who expressed an interest in the study) confirmed their computer's IT set up, watched a video introducing the virtual patient application, and registered for the study. The most common reasons why GPs did not progress to registration were that they had only been seeking information about the study and/or it was not what they expected, that they were too busy to participate in the study, or that they required IT updates to complete the study which were not possible (this was usually due to practice limitations for security). There were a few GPs who had expressed an interest in the study but with who we were not able to make any further contact, despite a number of telephone and email attempts by the study team.

Once GPs had registered for the study we could be fairly confident of their interest and intention to participate. Therefore, when considering the completion rate of this study it seems reasonable to report how many of the GPs who registered for the study ended up completing all six virtual 'consultations' and the post-consultation survey. 75.7% of GPs who registered for the study completed it .

262 GPs (47.1% of those who expressed an interest, and 87.3% of those who registered for the study) completed the first 'consultation'. Of the 38 who did not complete the first consultation the majority of these (31) did not start the study. Just seven GPs started the study but failed to complete the first 'consultation'; these GPs stated that they were unable to complete the 'consultation' because they were too busy, or due to challenges or frustrations using the application.

86.7% of GPs who completed the first 'consultation' went on to complete the full study. 35 GPs 'participated' in the GP decision making study but did not go on to complete it; 24 of these completed just one consultation, 11 completed between two and five consultations. These GPs who did not go on to complete the study after the first 'consultation' again often reported that this was because they were too busy; some GPs also commented that the virtual patient application was either unrealistic, or too difficult to use.

5.1.1.2 : Comparison of the characteristics of GPs who completed the study versus GPs nationally

Table 8 compares characteristics of the 227 GPs who completed the GP decision making study with the population of all 41,877 GPs working in England.

Table 8 : Comparison of the characteristics of GPs who completed the GP decision making study with those of all GPs in England

* the data for GPs in England is sourced from the Health and Social Care Information Centre (2015)[160]

Characteristic		GPs who completed the study	Partners and salaried GPs who completed study	GPs in England *
Total		227		41,877
Region NB: not including locum GPs	London	39.5%		17.1%
	East of England	34.5%		11.4%
	North West	11.8%		13.5%
	Surrey & Sussex	4.1%		7.7%
	West Midlands	10.0%		10.4%
Gender NB: country wide % calculated using the 36,567 GPs where gender is known	Male	54.6%		45.8%
	Female	45.4%		54.2%
Age * NB: country wide data only includes partners and salaried GPs. It is calculated using the 30,502 GPs where age is known	Under 35 years	24.2%	20.9%	13.4%
	35-44 years	35.2%	36.8%	33.7%
	45-54 years	30.0%	32.3%	32.2%
	55-64 years	9.7%	9.0%	17.3%
	Over 64 years	0.9%	1.0%	3.4%
Role in practice	Partner/salaried	88.5%		85.0%
	Registrar (trainee)	2.6%		11.8%
	Locum	7.5%		3.2%

Since our recruitment approach was to recruit GPs from a few, contrasting regions of England (rather than country-wide), the percentage of GPs from each region who participated in the GP decision making study differs from the GP population in England as a whole. However the representation of GPs from both the North West of England and the West Midlands was very similar in the study population to England as a whole: GPs from the North West of England made up 11.8% of the study population compared to 13.5% of GPs in England, whilst 10.0% of the study population practised in the West Midlands compared to 10.4% of all GPs working in England. By contrast the representation of London (39.5% of GPs in the study versus 17.1% of all GPs working in England) and the East of England (34.5% versus 11.4%) was much greater in the study than for England as a whole.

54.6% of the GPs who completed the GP decision making study were male, compared to 45.8% of the GP population in England whose gender is known. It seems that our study population included more male GPs than might have been expected, although it is important to note that the gender of 12.7% of GPs practising in England is unknown. This higher representation of male GPs in the study population could reflect the fact that male GPs are more likely to work full- time than females,[161] and may therefore have more opportunity for participating in research.

The study population had an age distribution similar to that of GPs across England, with GPs most likely to be aged between either 35 and 44 years, or between 45 and 54 years: 65.2% of GPs in the study population were in one or the other of these age groups, corresponding to 65.9% of partners and salaried GPs working in England. However in general the GPs who completed the GP decision making study had a younger age profile than GPs across England as a whole, with 24.2% of the study population aged under 34 years, and just 10.6% 55 years or older. This may reflect the novel and technical nature of the study design. It should also be noted that the data from the Health and Social Care Information Centre only provides information about the age of GP partners and salaried GPs (not locums or trainees). However GP partners and salaried GPs did make up the majority of the study population and, as can be seen in Table 8, the age category percentages for the whole study population versus the study population restricted to GP partners and salaried GPs are very similar.

88.5% of GPs in the study population were partners or salaried GPs; this is very similar to the percentage of partners and salaried GPs across England (85.0%). The study population had a lower percentage of GP trainees/registrars than work in England, which is likely to reflect our decision that GP trainees were not eligible to participate in the study if they were at a stage in training where their consultations were not directly comparable with those of post-training GPs (e.g. undertaking consultations under supervision and/or having longer appointment slots).

Data on the ethnicity of GPs working in England is not readily available. However I have compared the ethnicity of the study population with data published by the General Medical Council detailing the ethnicity of registered doctors across the UK (see Table 9) and the distributions are similar. The majority of GPs in both the study population (57.7%) and registered nationally in the UK (52.2%) described their ethnicity as white, followed by South Asian (28.6 of the study population and 20.7% of UK doctors). Very few of the study population were black (3.1%), which reflects the ethnicity of doctors across the UK (just 3.3%). As would be expected, given we directly asked GPs about their ethnicity (albeit with the option 'I prefer not to say') ethnicity was unknown in a far lower percentage of the study population than nationally. The greater percentages of white and South Asian GPs in the study population may simply result from the fact that the study's ethnicity data is more complete than national statistics, although it is also possible the ethnic profile of doctors working as GPs differs from that of doctors working in other specialties.

Table 9 : Comparison of the ethnicity of GPs who completed the GP decision making study with those of all doctors registered to work in the UK

* the data for all doctors registered to work in the UK is sourced from General Medical Council's registration statistics for their list of registered medical practitioners [162]

Characteristic		GPs who completed the study	* All doctors registered in the UK
Ethnicity	White	57.7%	52.2%
	Black	3.1%	3.3%
	South Asian	28.6%	20.7%
	Other	5.3%	6.6%
	Unknown	5.3%	16.9%

The main paper which reports the findings of the vignette study (awaiting publication, see Appendix 15 for the full paper) also compares practices' age standardised cancer referral ratio and their proportion of patients aged over 65 years old for the practices of GPs in the study population compared to those nationally. This analysis found that the practices of GPs who completed the study had higher cancer referrals than non-participating practices, despite the fact that in order not to publicise the study's focus on cancer to GP participants, it was presented as a study of GP decision making.

5.1.1.3 : Comparison of the characteristics of GP participants versus completers

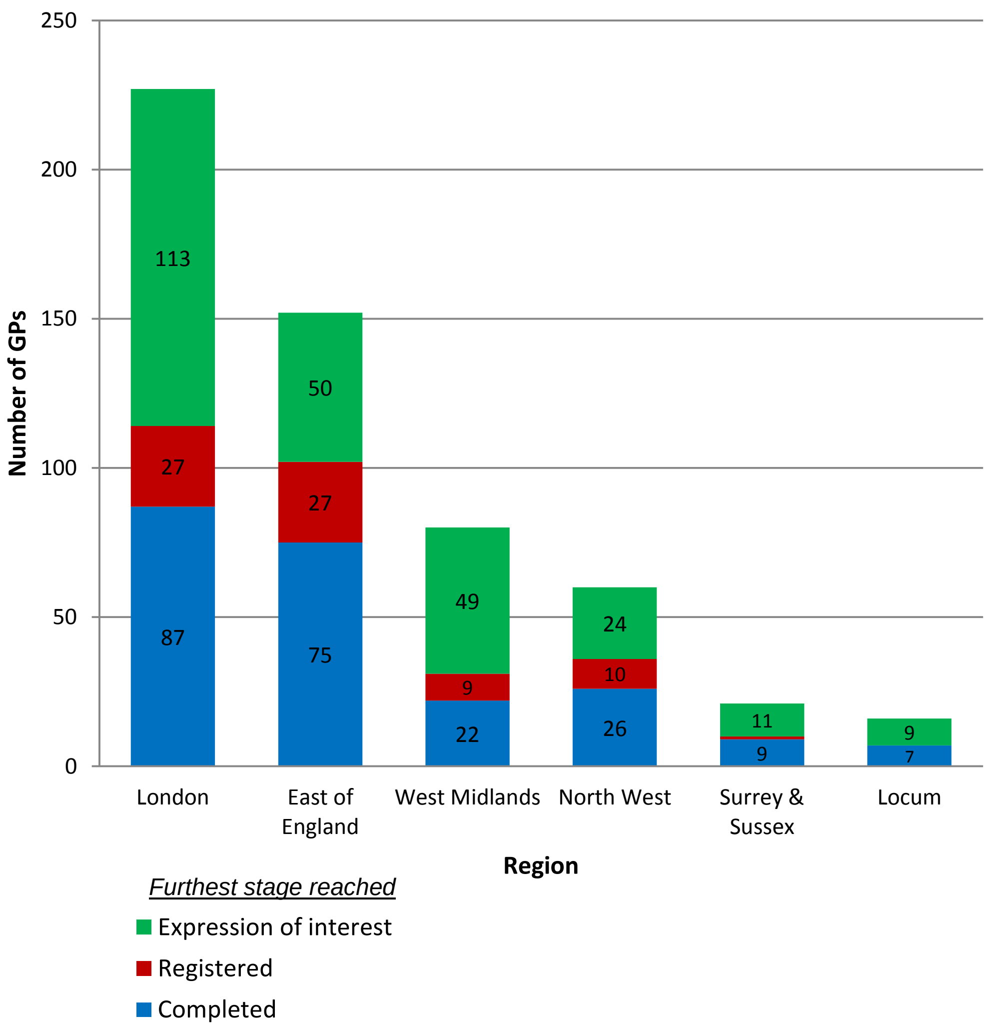
I conducted analyses to compare those GPs who registered for the study but did not complete it, with those who completed the study. This was important in order to evaluate the potential for bias. When GPs registered for the study we requested details of a number of personal characteristics, both socio-demographic and practice-related (these are listed in Appendix 6). I compared (both graphically and statistically, using the χ_2 test) whether GPs who completed the GP decision making study varied significantly from those GPs who registered for the study but did not complete it.

Ideally I would have also liked to compare these groups with those GPs who initially expressed an interest in the GP decision making study (but did not register for, participate in or complete it). However aside from gender (information which I collected from the General Medical Council register where it is publically available, and we can be confident is likely to be accurate) and region, we have very little information about the characteristics of GPs who expressed an initial interest in the GP decision making study. It has therefore not been possible to draw many conclusions about whether the GPs registering for, participating in or completing the GP decision making study were representative of those who expressed an interest in it.

5.1.1.3.1 : Region

As shown in both Table 7 and Figure 16, we received the most expressions of interest (226) from the London region, followed by the East of England (152). As would therefore be expected, these two regions also had the highest number of GPs completing the study (84 and 76 respectively). The regions with the highest rate of study completion for GPs who had expressed an interest in the study were the East of England (50.0%) and North West England (48.3%), whilst the West Midlands had the lowest rate of GPs who had expressed an interest completing the study (27.5%). GPs from the East of England were statistically more likely than those from London to both register for (p=0.001) and complete (p=0.002) the GP decision making study.

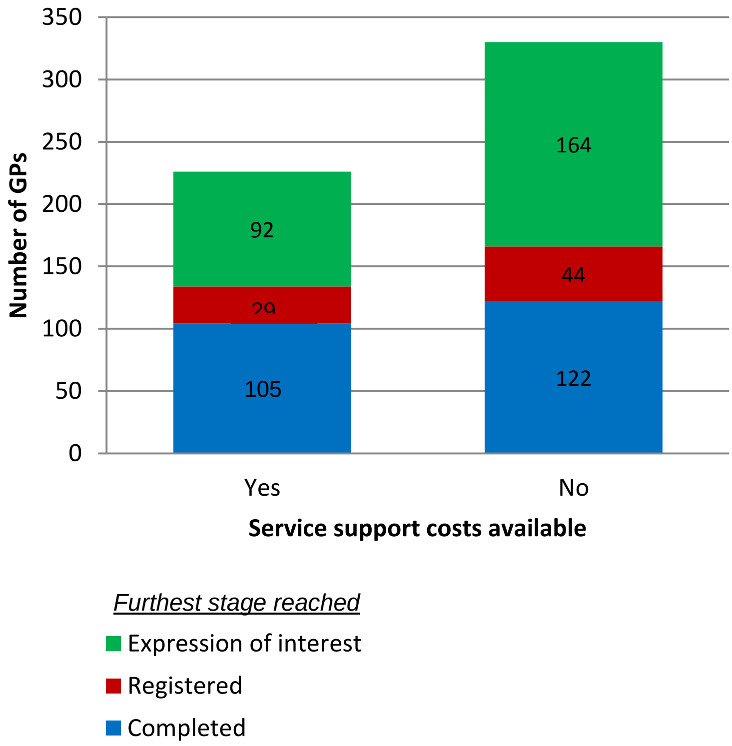
Figure 16 : Number of GPs reaching each stage of the study, by region



One potentially significant factor which varied between the study regions we recruited from was the availability for GPs' practices to claim service support costs for the time involved in setting up for the study (in particular the computer IT checks and updates that were often required). As discussed in Section 4.2.10.3, practices in certain areas of three of the regions we recruited in (London, East of England and North West England) were able to apply for service support costs. Service support costs were available to support 227 of all the 556 GPs who expressed an interest in the study (40.6%), 134 of the 300 GPs who completed IT set up and registered for the study (44.7%), and 105 of the GPs who completed the study (46.2%).

However, as seen in Figure 17, the availability of service support costs did not significantly affect either GP registration or study completion (p=0.08).

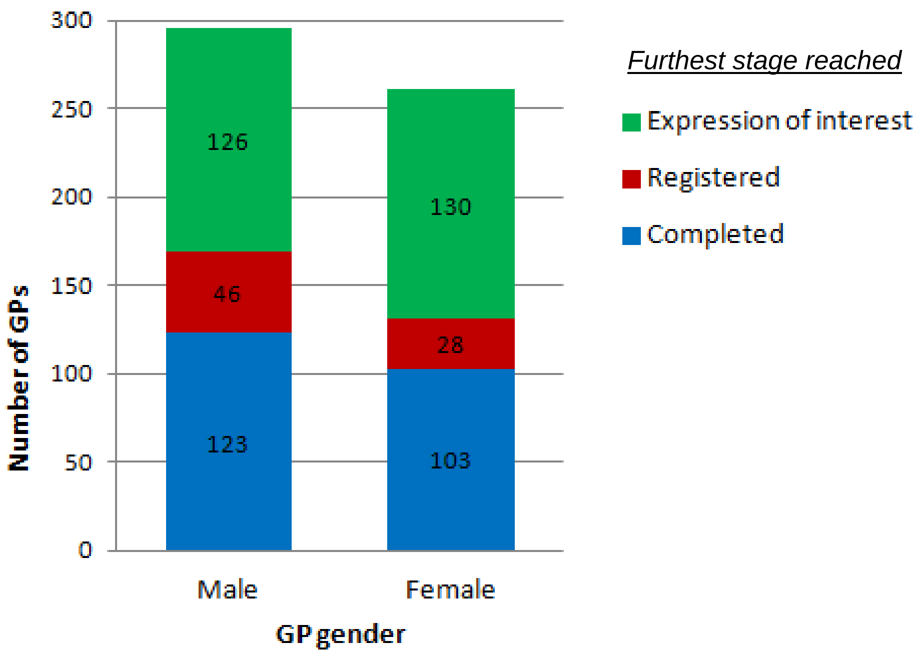
Figure 17 : The availability of service support costs and GPs' progress in the decision making study



5.1.1.3.2 : Gender

As shown in Figure 18, men were over-represented in terms of the number of expressions of interest we received, as well as in GPs registering for and completing the study. However there was no statistically significant difference between the number of men and women who registered for the study, or who completed it once they had expressed their interest.

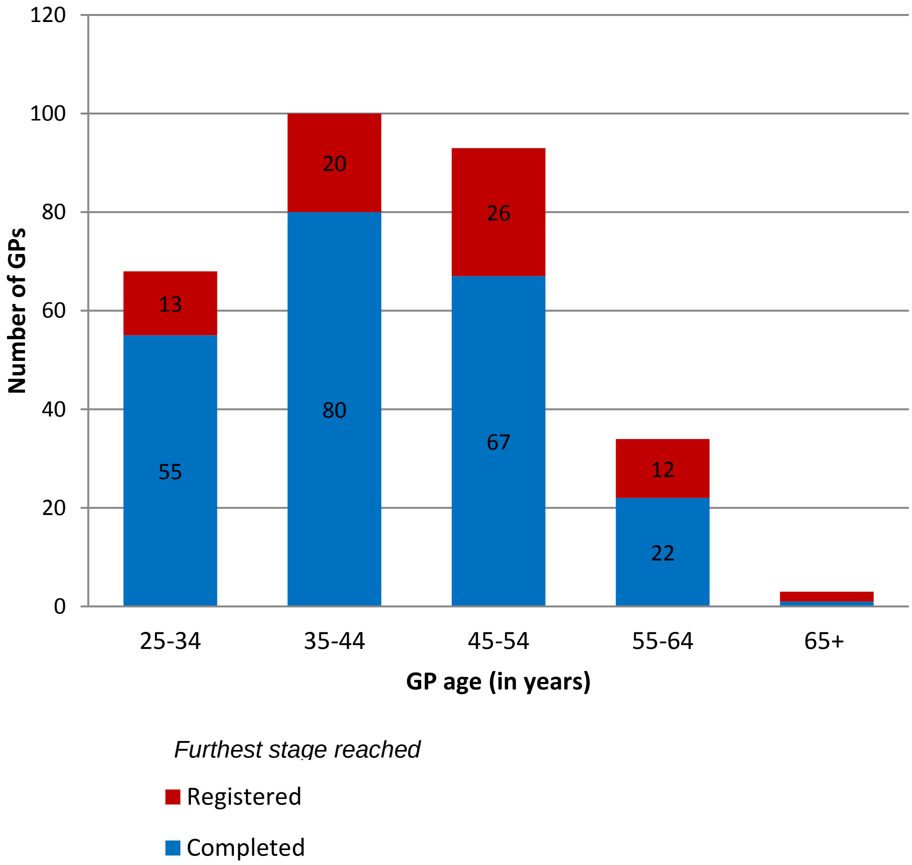
Figure 18 : GPs' study progress, by GP gender



5.1.1.3.3 : Age

Figure 19 shows the difference between GPs who completed the GP decision making study and those who registered but did not complete it, by age. GPs aged less than 45 years old were most likely to register for the study, and statistically more likely to complete the study once they had registered (p=0.02 for GPs aged <45 years vs. those aged ≥45 years).

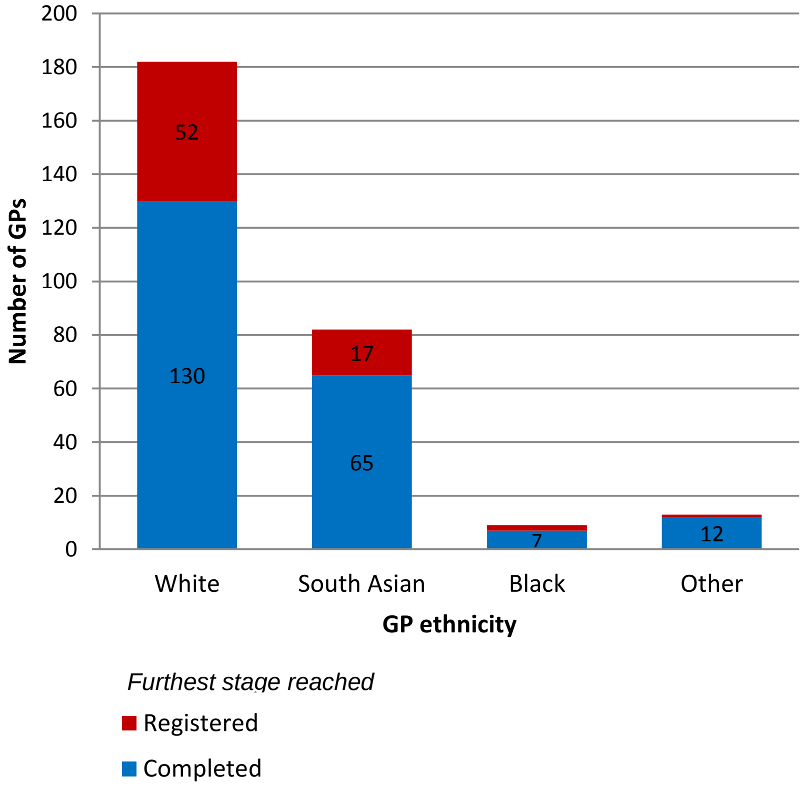
Figure 19 : GPs' study progress, by GP age



5.1.1.3.4 : Ethnicity

GPs of white and South Asian ethnicity were most likely to register for the study (Figure 20), however ethnicity did not significantly affect the likelihood that a GP completed the study.

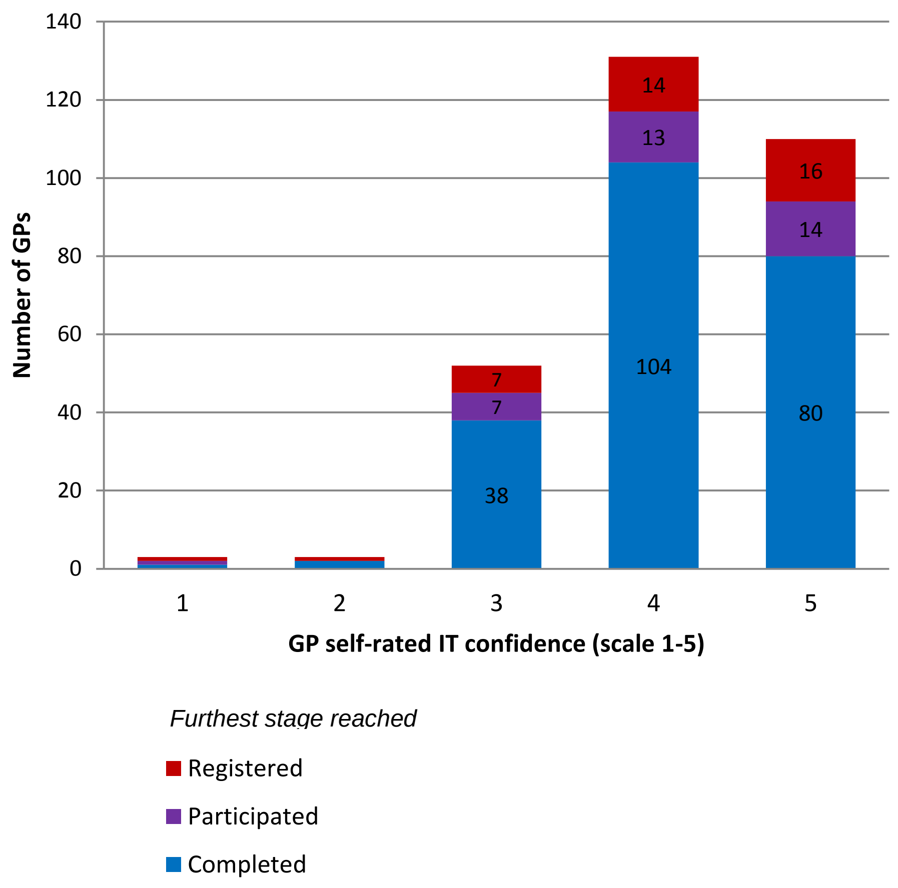
Figure 20: GPs' study progress, by GP ethnicity



5.1.1.3.5 : GPs' self-rated IT confidence

GPs were asked to rate their IT confidence on a scale of 1 to 5, 5 being the most confident. As Figure 21 shows, GPs who registered for the study were most likely to rate their IT confidence as average to high. However GPs' IT confidence did not significantly affect either their participation in or completion of the study.

Figure 21 : GPs' study progress, by GP's self-rated IT confidence



5.1.1.3.6 : GPs' role and their frequency of practice

The average number of sessions registering GPs worked per week did not significantly affect the likelihood that a GP completed the study (Figure 22). GP partners were most represented in study registration, but salaried GPs were statistically as likely to complete the study (Figure 23).

Figure 22 : GPs' study progress, by the number of sessions worked by a GP per week

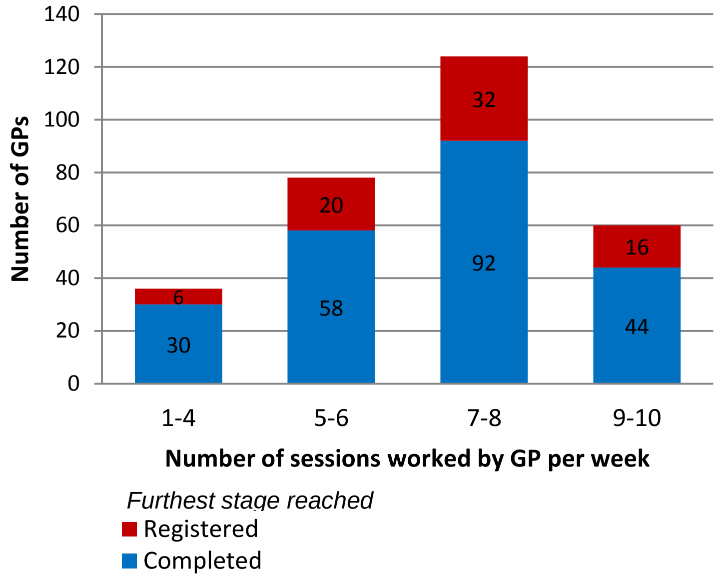
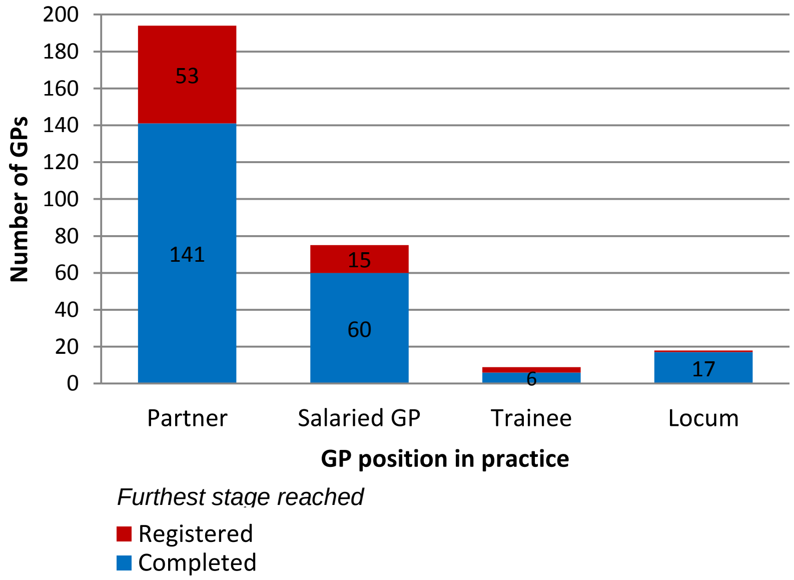


Figure 23 : GPs' study progress, by GPs' position in the practice



5.1.1.4 : Challenges of recruitment and steps we took to facilitate study completion

Recruitment and follow up of GPs was a very time-intensive process, more so than I had initially expected. We faced a number of challenges, in particular:

- Getting in touch with a GP who had expressed interest was often time-consuming due to their professional commitments.
- Some GPs were anxious about doing the IT compatibility check, despite its simplicity. Therefore, for many GPs, we completed this with them over the phone.
- Some GPs/practices (particularly in North West England) required IT updates that could not be downloaded on practice computers due to restrictions.
- GPs typically took about eight weeks from expressing interest to completing the study. We therefore kept in regular contact, reminding them about their next required action if they were delayed by more than a week. This often required us to contact them several times.

To make the recruitment process more manageable and to ensure that, where possible, GPs did not fail to complete the study due to a lack of follow up on our part, a number of temporary administrative staff helped make follow up phone calls to GPs, and entered some of the completed responses in the study database.

We succeeded in recruiting more GPs than the sample size calculation performed for the study suggested necessary: while based on the sample size calculation we were aiming to recruit 216 GPs, in the end 227 completed the study.

5.1.1.5 : Discussion of factors affecting recruitment and completion of the study

40.8% of GPs who expressed an interest in the GP decision making study completed it. This response rate was similar to the response rate in many other studies of GP decision making appraised during my systematic review.[76;80;101;121]

75.7% of GPs who registered for the study completed it. When compared to other studies this is a reasonable completion proportion,[92;99;120] especially given the innovative and potentially complex nature of the study tool, and the fact that GPs were required to complete the study on at least two, and ideally three, separate occasions.

The most common reason reported by GPs for failing to progress to the next stage of the study was that they were too busy. A small number of GPs experienced challenges relating to the online and interactive nature of the study which prevented them from registering for or completing the study; where these challenges occurred they were either due to GPs being unable to set up their computer to meet the study's requirements, or GPs experiencing difficulties using the virtual patient application. However very few GPs did not complete the study due to concerns about the study tool (the virtual patient application): just seven GPs started the study but were unable to complete the first 'consultation' due to time pressures or challenges and frustrations with the application, and whilst 35 GPs completed between one and five 'consultations' but did not complete the study, the majority of these still stated that this was because they were too busy, rather than solely due to issues with the study tool.

The regions with the highest rate of study completion (for GPs who had expressed an interest in the study) were the East of England and North West England. This may reflect the fact that we had PCRN support in these regions to follow up GPs who had expressed an interest; the East of England in particular had a number of research nurses who were highly proactive at following up GPs who had expressed an interest in the study, as well as assisting with ensuring that GPs' IT set up was suitable. Alternatively these regions might have had the highest rate of study recruitment because they were regions where we started recruitment early, and therefore GPs had longer to complete the study (for the average GP across all regions there was a period of a number of months between the GP expressing interest in the GP decision making study and completing it).

We have very little information about the personal characteristics of GPs who expressed an interest in the study. It does appear than men were over-represented in terms of expressions of interest, however there was no statistically significant difference in the gender of GPs registering for and completing the study.

The only GP characteristic for which there was a significant difference between GPs who completed the study and those who registered, but did not complete it, was age, with older GP's being less likely to either register or complete: this could reflect the novel and the technical nature of the study design. GPs' IT confidence did not significantly affect either participation in, or completion of, the vignette study. We cannot know for certain whether GPs' IT confidence influenced their behaviour and decisions within the vignette study itself, but it would seem unlikely that it had a significant effect. However it is important to note that almost all GPs who registered for the study rated their IT confidence as 3 or higher (on a scale of 1 to 5, where 5 was the most confident): we cannot be certain how many GPs with lower IT confidence may have been interested in participating in the study, but did not get to the registration stage.

5.1.2 : Results of the vignette study

Data from 1362 unique virtual 'consultations' conducted by 227 GPs were coded, and data from 1358 'consultations' analysed (see Section 5.1.2.1.1 for details of why four 'consultations' were excluded from the analysis). I conducted a series of descriptive analyses of both the primary and secondary outcomes, including considering the primary outcome (referral for chest X-ray) by profile, 'patient' characteristics and GP characteristics. I also summarised the results of the hierarchical modelling and interactions testing performed by SD and JM on the primary outcome of the vignette study (the full paper containing these findings can be viewed in Appendix 15).

5.1.2.1 : Constructing and coding variables for analysis

Once all 227 GPs had completed the study I produced a series of coding criteria in order to use the information from each vignette viewed (that is each virtual 'consultation' conducted - 1362 in total) to provide the data I required for the primary and secondary outcome measures that I defined in Section 4.2.12. I developed these coding criteria with the advice of all my supervisors, Professor Willie Hamilton (an academic GP specialising in the early diagnosis of cancer), and GP Dr Janakan Crofton.

The primary outcome measure of the vignette study for my PhD was whether a GP had decided to refer the virtual 'patient' in the vignette for chest X-ray (CXR), or to a secondary care service where a chest X-ray would almost certainly be performed given the 'patient's' symptoms (e.g. referral to a respiratory specialist, or sending the 'patient' to an A&E department). I constructed a variable to use in the analysis of this outcome measure using the information given in the free text management plan responses that GPs entered for each vignette completed. The full criteria for this variable are detailed in Table 10.

Table 10 : Criteria for coding a variable to use in the analysis of my primary outcome, referral for chest X-ray

Code as ‘chest X-ray’ - 1	**Code as ‘no chest X-ray’ - 0**
All requests for chest X-ray where there is no uncertainty and it is not a future/potential plan: - urgent, non-urgent or no urgency stated *e.g. CXR* *urgent CXR* *standard CXR* - hospital admission/A&E referral where chest X-ray specifically stated in management plan <u>or</u> lung disease is the most likely/likely diagnosis *e.g. refer to hospital for 12 lead ECG, CXR and arterial blood gases* Referral to chest clinic or to a respiratory or oncology specialist	Where chest X-ray is referred to using uncertain phrasing: *e.g. possible ECHO and/or CXR* *may arrange CXR* *may need a CXR* *may leave for now* *consider CXR* *if I was uneasy I would arrange CXR* Where chest X-ray is considered as a potential future management option: *e.g. CXR if persists* *review, if no better for CXR* *if still unwell for CXR* *give CXR form to go next week if no better* Referral to hospital medics (unless chest X-ray specified, or a lung disease considered most likely/likely diagnosis): *e.g. I suggest emergency hospital assessment* *refer on-call medics* Referral to non-respiratory specialist: *e.g. rapid access chest pain clinic* *cardiology* *gastroenterology* X-ray requested, but not chest or chest not specified Chest X-ray or referral not in management plan

NB: for all outcomes, where GPs did not state any management plan (this occurred for 3 of the 1362 total of virtual 'consultations') outcomes were coded 99 so that they could easily be identified and excluded.

The variable constructed from the coding criteria in Table 10 is the variable used for all further analysis of my primary outcome, referral for chest X-ray, (including in the hierarchical modelling) unless otherwise specified. However when developing this variable I became aware of the wide variation in GPs' management decisions, and the challenges of interpreting and coding free text responses. I therefore also coded two additional variables related to my primary variable: a less stringent variable that was coded positively when a GP made any suggestion of a chest X-ray in their free text, and a much stricter variable where I only coded positively where GPs clearly indicated that they were referring for an urgent chest X-ray. Again I developed the criteria for these variables with guidance from my supervisors and GPs. The full criteria for these variables can be seen in Appendix 16.

Once the coding criteria were finalised both Dr Crofton and I independently coded all 1362 vignettes completed (making a decision for each of these three variables) before comparing our responses for validity. Where we disagreed we discussed our reasons for this; most disagreements were down to human error, and the few cases where we had intentionally interpreted the coding criteria differently were all resolved with discussion.

I also created a variable for my secondary outcome measure, GPs' consideration of lung cancer as a potential diagnosis. I constructed this variable using the GPs' free text differential diagnoses responses that GPs entered for each vignette completed. I coded this variable as follows:

0 = lung cancer not stated anywhere
this included any vague mention of 'lung disease', a reference to cancer but no specification it was respiratory, interstitial lung disease

1 = lung cancer listed as main/most likely diagnosis

2 = lung cancer listed as an other, likely diagnosis

3 = lung cancer listed as an unlikely, but possible diagnosis
for codes 1-3 lung cancer was considered listed if there was reference to terms such as bronchial/bronchus/lung/respiratory in addition to a description of cancer such as cancer/carcinoma/neoplasm/malignancy/tumour

99 = consideration of lung cancer unknown or unclear

5.1.2.1.1 : Details of exclusions

The virtual patient application was a novel tool for studying GP decision making; it is therefore not unexpected that there were a few challenges associated with its use in the vignette study. I evaluate the application thoroughly in Section 5.1.3. However some of the challenges had implications for my analysis or the presentation of my findings so I will give details of these here.

Three GPs did not enter a management plan for one of the six vignettes they viewed because they did not feel they had been able to obtain the information required to make a management decision (more details in Section 5.1.3.2.2). We excluded these consultations (3 out of 1362 total, 0.2%) from any further analysis.

One GP was accidentally allocated a 'patient' from an earlier pilot of the application. This 'patient' did not have video responses for many of the questions that the updated application was able to support; we therefore excluded this 'consultation' from all further analysis and did not code variables from it.

The total number of 'consultations' analysed was therefore reduced from 1362 to 1358.

In addition, some errors in allocation of patients (more details in Section 5.1.3.2.2) meant that eight GPs did not view each of the six 'patient' profiles once: typically they viewed one profile twice and did not see another at all. We did not exclude the data from these 'consultations' from our analysis as these allocation errors did not affect the information presented to GPs, or their ability to make a management plan. However these allocation errors did mean that, despite a total of 227 GPs completing the GP decision making study, the number of GPs viewing each profile varied between 223 and 228.

5.1.2.2 : Descriptive analysis of the primary outcome - referral for chest X-ray

I conducted descriptive, univariate analysis to determine the proportion of 'patients' referred for chest X-ray: by profile, by the 'patient' experimental factors, and by GP characteristics.

5.1.2.2.1 : Referral differences by profile

1008 of the 'consultations' (74.2%) involved referral for chest X-ray. However as can be seen in Table 11, the percentage of referrals for chest X-ray differed widely between the six 'patient' profiles.

Table 11 : GPs' referral for chest X-ray (CXR): overall for the study, and by profile

	Number of GPs who stated a management plan	CXR (%)	No CXR (%)
All profiles combined	1358	1008 (74.2%)	353 (25.8%)
Profile 1 58/59yr non-smoker with breathlessness and fatigue for 10 days	228 *	152 (66.7%)	76 (33.3%)
Profile 2 58/59yr smoker with chest pain and cough for 10 days	226	188 (83.2%)	38 (16.8%)
Profile 3 78/79yr smoker with chest pain and cough for uncertain duration (~3 weeks)	228 *	196 (86.0%)	32 (14.0%)
Profile 4 78/79yr non-smoker with cough and appetite loss for uncertain duration (~3 weeks)	227	133 (58.6%)	94 (41.4%)
Profile 5 58/59 yr smoker with COPD with breathlessness and fatigue for >1 month	226	187 (82.7%)	39 (17.3%)
Profile 6 78/79 yr smoker with chest pain and weight loss for >1 month	223	152 (68.2%)	71 (31.8%)

* number is greater than 227 since some GPs viewed the same profile twice (see Section 5.1.2.1.1)

GPs' referral ranged from 58.6% for profile 4 (the 78/79 year old non-smoker with cough and appetite loss for about 3 weeks), up to 86.0% for profile 3 (the 78/79 year old smoker with chest pain and cough for about 3 weeks).

Profiles 2 and 3, where 'patients' had two chest symptoms (which could perhaps be considered 'typical' symptoms of lung cancer), had the highest rates of referral: 83.2% and 86.0% respectively. These profiles differed only in the age of the 'patient', and therefore it is not surprising that the rate of referral was very similar for both of these.

By contrast relatively few GPs referred 'patients' of profile 4 (the 78/79 year old non-smoker with cough and appetite loss for about 3 weeks) or profile 6 (the 78/79 year old smoker with chest pain and weight loss for more than a month), both of whom presented with a typical chest symptom, but who also had an additional atypical symptom. This is surprising given that both of these profiles meet the NICE guidelines for referral, especially profile 6 since weight loss is considered a red flag for lung cancer.

Profile 1 (58/59 year old with breathlessness and fatigue for 10 days) had a relatively low rate of referral (66.7%). However this was designed to be a 'deflecting' profile, with symptoms suggestive of heart failure, which is likely to explain the lower referral rate. Profile 5, the same symptom presentation in a smoker with COPD symptomatic for more than a month, had a much higher rate of referral for chest X-ray (82.7%).

Table 12 gives details of how GPs' referrals differ between the profiles when considering not simply the primary outcome variable (referral for chest X-ray, as described in Table 10), but also the two additional variables constructed: one looking at urgent chest X-ray referrals specifically, the other at any suggestion of chest X-ray (indicating that it was something the GP was considering, even if not suggesting it at the present time).

Interestingly, whilst urgent referral for chest X-ray was rarely part of GPs' management plan (8.4% across all profiles), it was most common for profile 6 (13.5%), despite referral for chest X-ray being less common for this profile when considering both the primary and less stringent (suggestion of chest X-ray) variables. Urgent referral for chest X-ray was also more likely than average for the the more 'classic' presentation of lung cancer in profiles 2 and 3.

Urgent referral for chest X-ray was extremely rare for profile 4 (3.5%), which is in line with a low percentage of referral for chest X-ray using the primary outcome variable. However 81.5% of GPs made a suggestion of chest X-ray in their management plan for profile 4, which is similar to the percentage of GPs who suggested chest X-ray overall for the six profiles (83.7%). This suggests that GPs appeared to be considering that significant lung pathology could be present, even if they do not feel it warranted urgent or current investigation.

Table 12 : GPs' referral for chest X-ray (CXR) using both the primary and additional outcome variables

	Number of GPs who stated a management plan	Referral for CXR	Urgent CXR	Any suggestion of CXR
All profiles combined	1358	1008 (74.2%)	114 (8.4%)	1136 (83.7%)
Profile 1 58/59yr non-smoker with breathlessness and fatigue for 10 days	228 *	152 (66.7%)	10 (4.4%)	163 (71.5%)
Profile 2 58/59yr smoker with chest pain and cough for 10 days	226	188 (83.2%)	23 (10.2%)	208 (92.0%)
Profile 3 78/79yr smoker with chest pain and cough for uncertain duration (~3 weeks)	228 *	196 (86.0%)	26 (11.4%)	210 (92.1%)
Profile 4 78/79yr non-smoker with cough and appetite loss for uncertain duration (~3 weeks)	227	133 (58.6%)	8 (3.5%)	185 (81.5%)
Profile 5 58/59 yr smoker with COPD with breathlessness and fatigue for >1 month	226	187 (82.7%)	17 (7.5%)	205 (90.3%)
Profile 6 78/79 yr smoker with chest pain and weight loss for >1 month	223	152 (68.2%)	30 (13.5%)	165 (74.0%)

* number is greater than 227 since some GPs viewed the same profile twice (see Section 5.1.2.1.1)

<u>*5.1.2.2.2 : Referral differences by symptom information obtained*</u>

We considered that some of the variation in rates of referral for chest X-ray seen between the profiles could be explained by what information GPs elicited during the 'consultation' - in particular whether they obtained information about the presence of a second symptom. As Table 13 shows, GPs only elicited the presence of the second symptom in 778 'consultations' (57.7%). This varied significantly between profiles: 95.1% elicited the second symptom (cough) in profiles 2 and 3, whilst only 21.1% discovered the breathlessness of the 'patient' in profile 1.

For some profiles, there was a significant interaction between GPs eliciting the second symptom and their referral for chest X-ray. Whether GPs elicited a second symptom explains much of the surprisingly low numbers of chest X-rays ordered for profile 6, despite the presence of the red flag symptom of weight loss: 90.8% of GPs referred for chest X-ray if they had elicited weight loss, compared to just 46.0% of those who did not. However, GPs eliciting the second symptom was not significant for profile 4, the other profile with an unexpectedly low rate of chest X-ray: 66.9% of GPs who elicited the second symptom appetite loss referred for chest X-ray, compared to 46.7% of GPs who did not. This may reflect the presentation of cough described in profile 4, although this was deliberately written to reflect the real life presentation of a lung cancer patient (DA's father).

Table 13 : Chest X-ray referral by profile according to whether GPs elicited the second symptom information

Profile number 2nd symptom	**Number of GPs who stated a management plan**	**Number of GPs who elicited the 2nd symptom**	**Referral for CXR**	
			2nd symptom NOT elicited n (%)	**2nd symptom elicited** n (%)
Profile 1 Fatigue	227	48 (21.2%)	120 (66.7%)	31 (66.0%)
Profile 2 Cough	225	214 (95.1%)	7 (63.6%)	181 (84.2%)
Profile 3 Cough	227	216 (95.2%)	7 (63.6%)	189 (87.1%)
Profile 4 Appetite loss	225	136 (60.4%)	42 (46.7%)	91 (66.9%)
Profile 5 Fatigue	224	56 (25.0%)	136 (80.5%)	50 (89.3%)
Profile 6 Weight loss	220	108 (49.1%)	52 (46.0%)	99 (90.8%)

5.1.2.2.3 : Referral differences by 'patient' characteristic

In Table 14 I report details of the frequency of chest X-ray referral in the vignette study by 'patient' characteristic.

GPs' referral of the high and low risk 'patients' was very similar (75.2% and 75.0%), however 'patient profiles' with a PPV indicating a medium risk of lung cancer were less likely to be referred for chest X-ray (72.4%).

GPs' referral of the female and male 'patients' was very similar (74.1% and 74.3% respectively), as was their referral of disadvantaged 'patients' compared to advantaged 'patients' (74.5% compared to 73.9%). White patients were most likely to be referred for chest X-ray compared to the other ethnicities studied: 76.6% of white 'patients' were referred, but only 74.2% of South Asian and 71.5% of black 'patients'.

Table 14 : Frequency of chest X-ray referral by 'patient' characteristic

		Number of 'consultations'	Number of CXR referrals
Risk level	Low	452	339 (75.0%)
	Medium	452	327 (72.4%)
	High	444	334 (75.2%)
Gender	Female	660	489 (74.1%)
	Male	688	511 (74.3%)
Socio-economic circumstance	Disadvantaged	682	508 (74.5%)
	Advantaged	666	492 (73.9%)
Ethnicity	White	482	369 (76.6%)
	Black	428	306 (71.5%)
	South Asian	438	325 (74.2%)

5.1.2.2.4 : Referral differences by GP characteristic

In Table 15 I report details of the frequency of chest X-ray referral in the vignette study by GP characteristic.

Male GPs referred more 'patients' than females GPs (77.1% compared to 70.5%).

The highest referring age group was those GPs aged 45 to 54 years old (78.7%); by contrast the youngest group of GPs, those aged 25 to 34 years old, referred 70.1%. The lowest referral was in the 65+/missing data category, however since this category only contained 12 'consultations' (i.e. data for two GPs), and combined missing data, it cannot reasonably be taken as representative of this age group's referral.

Black GPs were most likely to refer 'patients' for chest X-ray (81.0%), although again the number of GPs in this category (and therefore the number of 'consultations') was small. There was little difference in referral between white and South Asian ethnicity GPs (73.9% compared to 73.6%).

GPs who had qualified within the last 5 years referred less 'patients' for chest X-ray, in particular those who had been qualified for between 2 and 5 years (69.1%). Those GPs who had been qualified for 10 - 20 years referred the greatest percentage of 'patients' (77.6%).

By region there was fairly small variation in GPs' referral, although GPs from London and West Midlands referred a lower percentage of 'patients' for chest X-ray than those from the East of England, North West and Surrey and Sussex. Locum GPs had the lowest percentage referral (66.7%), although again the number of 'consultations' was relatively small.

GPs who rated their IT confidence more highly (4 or 5 on a five-point scale, where 5 indicated the most confidence) referred less patients than those who rated it as moderate (3 on the five-point scale): 74.0% and 73.7%, compared to 77.4%. Since there were only 17 'consultations' (i.e. data for three GPs) from GPs who rated their IT confidence as lower than 3, these results cannot reasonably be taken as representative. However when the 'consultations' of all GPs whose self-rated IT confidence was 3 or less were combined, these GPs still together referred more patients (76.9%) than those GPs with higher IT confidence.

Table 15 : Frequency of chest X-ray referral by GP characteristic

		Number of 'consultations'	Number of CXR referrals
GP gender	Female	603	425 (70.5%)
	Male	743	573 (77.1%)
GP age (in years)	25 - 34	324	227 (70.1%)
	35 - 44	461	336 (72.9%)
	45 - 54	413	325 (78.7%)
	55 - 64	136	102 (75.0%)
	65+ *or* missing	12	8 (66.7%)
GP ethnicity	White	789	583 (73.9%)
	Black	42	34 (81.0%)
	South Asian	402	296 (73.6%)
	Other *or* missing	119	90 (75.6%)
Years since qualification	0 - 2	168	120 (71.4%)
	2 - 5	269	186 (69.1%)
	5 - 10	240	177 (73.8%)
	10 - 20	330	256 (77.6%)
	20+	339	259 (76.4%)
Region	London	497	365 (73.4%)
	East of England	455	341 (75.0%)
	North West	172	131 (76.2%)
	West Midlands	132	96 (72.7%)
	Surrey & Sussex	54	41 (75.9%)
	Locum GP	36	24 (66.7%)
IT confidence (GPs self-rated this on a scale of 1 to 5, 5 being most confident)	1	6	5 (83.3%)
	2	11	7 (63.6%)
	3	221	171 (77.4%)
	4	635	470 (74.0%)
	5	479	353 (73.7%)

5.1.2.3 : Results of the hierarchical modelling analysis of the primary outcome

JM and SD conducted hierarchical modelling analysis in order to examine differences in GPs' referral of 'patients' for chest X-ray (the primary variable) by clinical profile and age. Table 16 shows the results of the models they constructed; full details are available in the primary results paper for the study (Appendix 15).

As seen in Table 16a, the hierarchical modelling analysis confirmed that a GP eliciting the second symptom of a 'patient' in the vignette study was associated with the 'patient' being more likely to be referred for chest X-ray: adjusted odds ratio 3.18 (95% CI 2.27-4.70) p<0.001). However this did not fully account for the lower referral of 'patients' with appetite loss (profile 4) and weight loss (profile 6) when compared with 'patients' with the more 'typical' (where 'typical' is the presence of two lung-related symptoms) lung cancer presentation of chest pain and cough in profile 3: adjusted odds ratios 0.25 (95% CI 0.14-0.42) p<0.001 and 0.50 (95% CI 0.29-0.91) p=0.02, respectively . These results were adjusted for all other factors that were found to be associated with chest X-ray referral in this study in a univariate analysis: that is 'patient' profile, 'patient' ethnicity, GP age, GP gender, and whether the second symptom was elicited.

The results of JM and SD's analysis also showed that there was significant non-clinical variation in referral by both 'patient' age and 'patient' ethnicity (Table 16b). GPs were less likely to investigate older 'patients' than younger 'patients': adjusted odds ratio 0.52 (95% confidence interval (CI) 0.39-0.70), p<0.001. GPs were also less likely to investigate black 'patients' compared to those of white ethnicity: adjusted odds ratio 0.68 (95% CI 0.48-0.95), p=0.03). Both these results were also adjusted for 'patient' profile and ethnicity, GP age and gender, and whether the second symptom was elicited.

Table 16 : Multilevel logistic regression of chest X-ray by 'patient' characteristic

^ = adjusted for all other factors associated ($p<0.1$) with chest X-ray in univariate analysis (i.e. 'patient' profile and ethnicity, GP gender and age) and whether the second symptom was elicited

^^ = adjusted for 'patient' profile, ethnicity, GP gender and age, and whether the second symptom was elicited

* = significant at $p \leq 0.05$

a) by 'patient' profile		**Adjusted^ odds ratio (95% confidence intervals)** ***reported to 2 decimal places***
Profile (second symptom)	1 (fatigue)	0.62 (0.35; 1.10)
	2 (cough)	0.65 (0.38; 1.15)
	3 (cough)	1
	4 (appetite loss)	0.25 (0.14; 0.42) *
	5 (fatigue)	1.64 (0.90; 3.11)
	6 (weight loss)	0.50 (0.29; 0.91) *
Ethnicity	White	1
	Black	0.67 (0.47; 0.96) *
	South Asian	0.86 (0.62; 1.20)
Second symptom elicited	No	1
	Yes	3.18 (2.27; 4.70) *

b) by age		**Adjusted^^ odds ratio (95% confidence intervals)** ***reported to 2 decimal places***
Age	Younger (58/59)	1
	Older (78/79)	0.52 (0.39; 0.70) *
Ethnicity	White	1
	Black	0.68 (0.48; 0.95) *
	South Asian	0.88 (0.63; 1.27)
Smoking status	Non-smoker	1
	Smoker	2.24 (1.64; 3.02) *
Second symptom elicited	No	1
	Yes	2.83 (2.09; 3.83) *

5.1.2.4 : Descriptive analysis of the secondary outcome - GPs' consideration of lung cancer as a potential diagnosis

Table 17 shows the extent to which GPs considered lung cancer as a potential diagnosis for the 'patients' in the vignette study, both overall and broken down by profile. There are data for 1361 'consultations' because I included the three 'consultations' where GPs could not provide a management decision in this analysis, as these GPs did provide a differential diagnosis for these 'patients'.

GPs considered lung cancer as a possible diagnosis in 50.8% of all 1361 'consultations', and as the most likely diagnosis in 165 of these (12.1%). Lung cancer was most frequently listed as a possible diagnosis for profiles 2 and 3 (by 65.4% and 65.8% of GPs respectively). Unsurprisingly the vast majority of GPs (88.2%) did not consider lung cancer in profile 1, the deflecting vignette, although 10.0% did state it was an unlikely, but potential diagnosis. A relatively large percentage of GPs, 28.1%, listed lung cancer as the most likely diagnosis for profile 6; perhaps a surprise when considering that this was accompanied by a surprisingly low rate of referral for chest X-ray. Most GPs (83.7%) did not consider lung cancer as a likely diagnosis for profile 4, although a majority of GPs (50.7%) did consider it as a potential diagnosis.

Table 17 : GPs' consideration of lung cancer as a diagnosis

	Number of GPs stated a differential diagnosis	**Most likely diagnosis (%)**	**A likely diagnosis (%)**	**Unlikely but a possible diagnosis (%)**	**Not considered (%)**
All profiles combined	1361	165 (12.1%)	218 (16.0%)	309 (22.7%)	670 (49.2%)
Profile 1 58/59yr non-smoker with breathlessness and fatigue for 10 days	228	1 (0.4%)	3 (1.3%)	23 (10.0%)	201 (88.2%)
Profile 2 58/59yr smoker with chest pain and cough for 10 days	228	32 (14.0%)	54 (23.7%)	63 (27.6%)	79 (34.6%)
Profile 3 78/79yr smoker with chest pain and cough for an uncertain duration (~3 weeks)	228	39 (17.1%)	55 (24.1%)	56 (24.6%)	78 (34.2%)
Profile 4 78/79yr non-smoker with cough and appetite loss for uncertain duration (~3 weeks)	227	17 (7.5%)	20 (8.8%)	78 (34.4%)	112 (49.3%)
Profile 5 58/59 yr smoker with COPD with breathlessness and fatigue for >1 month	226	13 (5.8%)	60 (26.5%)	58 (25.7%)	95 (42.0%)
Profile 6 78/79 yr smoker with chest pain and weight loss for >1 month	224	63 (28.1%)	26 (11.6%)	31 (13.8%)	104 (46.4%)

* number is greater than 227 since some GPs viewed the same profile twice (see Section 5.1.2.1.1)

5.1.3 : Evaluation of the virtual patient application as a tool to investigate GPs' decision making

The virtual patient application was developed as a novel study tool for examining GP decision making. We set out to use a combination of interactive multimedia technology and non pre-scripted vignettes to present information to GPs in such a way that we captured the experience of a real life consultation as closely as possible.

Here I consider the effectiveness and limitations of the virtual patient application, considering in turn: the data it provided, technical issues encountered, and GPs' experiences and views of using the study tool.

5.1.3.1 : Data

Overall, the virtual patient application appears to have been successful in its purpose. 227 GPs completed the GP decision making study, each completing six virtual 'consultations' using the virtual patient application. The application provided a wealth of data on GPs' behaviour and questions during the 'consultations', as well as their differential diagnosis and management plan for each vignette viewed. In just three of the 1362 'consultations' conducted (0.02% of the total) were GPs unable to reach a management decision as a result of the constraints of the system.

5.1.3.2 : Technical issues with the application, and how we addressed these

The virtual patient application was a novel tool for studying GP decision making; it was therefore not surprising that there were some challenges associated with its use in the vignette study. Here I report on these issues, and the steps we took to try to resolve them or limit their effect.

5.1.3.2.1 : GPs not seeking information as would be expected

Nine GPs (4.0%) experienced difficulties in using the application, in particular for their first virtual 'consultation'. For example five GPs conducted a consultation where they clearly attempted to interact with the virtual patient application (e.g. performing examinations and seeking information from the patient sidebars), but did not ask any questions. Where this occurred for a GPs' first 'consultation', four of these five cases, we provided appropriate email feedback using the standardised form (Appendix 4), and the issue did not arise in any of these GPs' later 'consultations'.

Similarly four GPs noted in their management plan that they were unable to conduct examinations during a 'consultation'. For three GPs this occurred for their first 'consultation'; we therefore followed this up with both email feedback and a call to confirm that their computer's IT setup was suitable - and indeed in each of these cases issues with IT configuration were found to be the cause of the problem.

Examination of the log files of GPs revealed that three additional GPs conducted very short virtual 'consultations' with very little content: two GPs had some extremely short consultations lasting less than 5 minutes, and asking very few questions, and one GP completed the study (that is completing all six vignettes and entering a management plan) without asking any questions, seeking any additional patient information, or conducting any examinations/bedside tests. It is not clear whether these three GPs experienced difficulties in using the application, or whether they were simply trying to complete the study as quickly as possible. However since all these GPs viewed the introductory video for each 'patient', (where the presenting symptom was stated) and submitted a plausible differential diagnosis and management plan in response to this, we decided to include these consultations in the analysis.

5.1.3.2.2 : Incorrect allocation of 'patients'

Once a GP had registered for the study, a member of the study team allocated each GP six randomly assigned 'patients', one for each of the six 'patient profiles', from an overall bank of 36 (Appendix 5). These 'patients' were set to become available to GPs on three separate occasions, each one week apart. For 96.5% of the 227 GPs who completed the study 'patients' were allocated successfully such that GPs viewed each profile once. However (as noted in Section 5.1.2.1.1) eight GPs were incorrectly allocated 'patients', such that these GPs did not see all six of the 'patient' profiles as we would have expected.

Three GPs were allocated to view one profile twice, whilst they did not see another profile at all (e.g. viewing profile 1 twice, but not viewing profile 6); for one of these cases the incorrect allocation led to the GP viewing the same actor twice, which the random allocation had been constructed to avoid. Interestingly, despite seeing two 'patients' with exactly the same history and symptoms (i.e. differing only by non-clinical characteristics), these GPs did not act identically in their 'consultations' of the same profile, and in two cases actually proposed different differential diagnoses and management plans.

One GP was accidentally allocated one of the 'patients' from an earlier pilot of the application, who did not have video responses for many of the questions that the updated application was able to support. We therefore excluded this 'consultation' from all further analysis and did not code variables from it.

The vignette study was designed to be completed over a minimum of three weeks, with only one 'consultation' (profile 1) available to GPs initially, after which we provided standardised feedback on how best to use the virtual patient application. However due to allocation errors four GPs actually viewed two 'consultations' prior to receiving feedback.

5.1.3.2.3 : GPs not entering a management plan

As mentioned in Section 5.1.2.1.1, three GPs did not enter a management plan for one of the six vignettes they viewed because they did not feel they had been able to obtain the information required to make a management decision. For one GP this was the first 'consultation' so we provided feedback in the standardised email (see Appendix 4) and their future 'consultations' all appeared more successful. Two GPs did not enter a management plan for their final virtual 'consultation'. We excluded these 'consultations' from analysis.

5.1.3.3 : GPs' experiences and views of the virtual patient application

In addition to evaluating whether the virtual patient application provided the data we required for the GP decision making study, we were also interested to receive GPs' views about using the application, and how they felt it compared to real life consultations. In the post-consultation survey (discussed in detail in Chapter 6), GPs were given the option to respond in free text to the question: *'If you have any further comments or reflections you wish to add on how you make decisions about sending patients for diagnostic tests, or referring them to secondary care, please type them in the box below'.* Overall, 24 GPs (10.6%) who completed the GP decision making study commented on the virtual patient application study method or design, with 20 of these (8.8%) using the free text question in the survey. I will discuss this feedback in some detail here.

5.1.3.3.1 : 5% of GPs reported challenges in gaining information they were seeking

The most frequent complaint, from 12 GPs (5.3%), was that it was difficult to use the study tool to extract the information GPs would have wanted to receive. For example one GP noted:

'Some difficulty and frustration using software.' [GP 28]

Several GPs were specific about the challenges they faced using the application. Many had difficulty working out how to phrase questions to the 'patient' in order to play videos answering the question they wanted:

'I really struggled with these videos to get some clinical answers out of them.' [GP 134]

'I did not find the online consultations easy to follow. I wanted to ask questions but did not know how to phrase them.' [GP 77]

'I found the study quite frustrating because I was often unable to ask the questions I would normally ask and so did not obtain as good a history as usual and so felt I was making decisions with only half the information I normally have available.' [GP 15]

'This was a difficult study as hard to question patient.' [GP 187]

'I found it difficult to use the tool as for all the breathlessness, cough vignettes, there were no answers to questions regarding heart failure.' [GP 107]

Four GPs (1.8%) commented specifically that they were not able to receive information about the patient's own ideas and concerns about their symptoms, or that they had to ask closed questions (in contrast to their usual open style of questioning) and received a lot of answers that were simply 'no':

'I found these cases very difficult as it was difficult to obtain a full history (which hopefully would lead to a better diagnosis of the symptom) including their ideas concerns and expectations.' [GP 81]

'Found the vignette in video search a bit difficult. As usual questions I ask tend to be open so I ended up asking a lot of closed questions at the same time a lot of red flag questions I ask did not turn up.' [GP 39]

'The vignettes are out of keeping with my style of open questions, so I found this difficult to explore symptoms.' [GP 65]

5.1.3.3.2 : 3% of GPs found the application frustrating

In conversation during a real life consultation it is likely that a patient would answer questions (even those about symptoms they did not have) in a wider variety of ways than the software used for the virtual patient application could replicate. These software constraints mean that the application was only able to be, at best, a simulation of real life rather than a full replication. The virtual patient application was designed so that a video giving a null response ("I don't have that" or "no") played in response to questions where the 'patient' did not have relevant information to give. However some GPs commented that they found this frustrating, or that it caused uncertainty whether the negative response was an error or a genuine negative response:

'The frustration surrounding the uncertainty of the answers definitely lowered my threshold to refer and review again.' [GP 170]

'Why did you have the video clips at all? Why It did not add much and when they said the same thing over and over it was irritating.' [GP 38]

5.1.3.3.3 : 4% of GPs reported the application did not their reflect real life practice

Four of the GPs who commented that they had difficulties extracting information from the 'patient' using the virtual patient application noted that they felt their decision making behaviour in this artificial situation was unlikely to be fully representative of their real life practice:

'Wasting time trying to get the relevant history when the computer could not respond de-motivated me to engage or care if I performed well.' [GP 112]

'I felt I may have over investigated as unable to obtain answers to [certain] questions.' [GP 107]

'[I] felt I was making decisions with only half the information I normally have available.' [GP 15]

'Getting lots of no's or I don't know mean I felt a bit frustrated and gave up on the consultation.' [GP 77]

Unsurprisingly, GPs also emphasised that (even if they were able to receive the information they would have sought from a real life patient) the virtual consultations were not like real life consultations. For example one GP commented:

'A simulated surgery such as this can never be as good as a real patient in a surgery.' [GP 136]

Some GPs felt that having a real life patient physically present in front of them was significant for their decision making:

'I think a lot of what we learn comes from visual cues or other things within the consultation - e.g. how breathless they are walking into the room.' [GP 77]

'It also makes it different when you actually see someone face to face.' [GP 187]

Other GPs noted that in real life they have contextual information about the specific patient in front of them, and that this is likely to significantly influence any conclusions or decisions that they make in the consultation:

'Each patient is an individual - your scenarios were difficult to put in a realistic context to make a valid assessment of what I personally would do in real life.' [GP 101]

'There is a lot of contextual material in the decision to refer for tests and further opinions. Much of that could not be captured in these vignettes.' [GP 67]

<u>5.1.3.3.4 : 2% of GPs did not feel that the application was realistic</u>

Two GPs (0.9%) noted that some of the features of the virtual patient application meant that it was not particularly realistic, be that due to the challenges they faced taking a history, or that they were required to suggest a suspected diagnosis after just one consultation:

'History taking in practice is easier than the vignettes and often an option would be seeing [the patient] again.' [GP 139]

'Most likely diagnosis' is a bit artificial - often I'm highly non-committal on this until the first round of basic tests is performed. 'Most concerning diagnosis that's reasonably likely' probably better explains my management decisions.' [GP 58]

When creating the vignettes for this study I (in consultation with one of our GP experts) decided that none of the 'patients' would have positive lung-related examination signs. In the study 'patients' presenting with breathlessness had a raised respiration rate and profile 1 patients (who complained of swollen ankles) had evidence of peripheral oedema; otherwise all patients had an otherwise normal respiratory and cardiovascular examination. This was to ensure that our study examined GPs' responses to the information 'patients' provided rather than testing how they responded to a positive examination. In addition, early stage lung cancer (when the disease is potentially curable) does not necessarily present with florid clinical signs. However we acknowledge that the negative test results could be potentially misleading, as one GP noted:

'It seems all examinations seem to be normal making it very confusing to diagnose.' [GP 39]

Two of the GPs who commented that the virtual patient application did not reflect their 'real life' consultations specifically noted that they felt their decision making behaviour in this artificial situation was unlikely to be fully representative of their 'real life' practice:

'I found the consultation interface not helpful and very much unlike a real consultation. I do not feel this exercise represents a fair representation of my diagnostic skills.' [GP 222]

'I found this whole process frustrating and not representative of daily practice and therefore I think will not enlighten you much.' [GP 38]

5.1.3.3.5 : Nearly 90% of GPs did not provide negative feedback

Whilst it is important to reflect on the limitations of the virtual patient application when evaluating its use as a tool to examine GP decision making, it is also important to note that 203 of the GPs who completed the GP decision making study (89.4%) did not provide any negative feedback about their experience of using the virtual patient application - neither during the post-consultation survey, nor by email or post after completion.

The majority of the 108 GPs (47.6%) who answered the free text question in the post-consultation survey used it to reflect on their real life decision making, rather than their experience of using the virtual patient application. Whilst we cannot conclude that these GPs did not have any views on the use of the application as a study tool, the fact that they took the time to type an answer to the question, but did not discuss the application, suggests that they are unlikely to have strong opinions about it.

Furthermore, a few GPs contacted us specifically with positive feedback about the study: for example describing it as 'interesting' and 'innovative', stating that they 'enjoyed it', and that they valued the support the researchers provided in setting up and completing the study.

This low proportion of negative comments suggests that the virtual patient application was an acceptable and effective tool to examine GPs' decision making.

5.2 : Discussion

5.2.1 : Main findings of the vignette study

In common with the results of my systematic review, the vignette study demonstrates non-clinical variation (in addition to clinical variation) in GPs' decisions to refer patients - in this case the referral for patients with symptoms indicative of lung cancer for a diagnostic chest X-ray.

Overall, GPs proposed a referral for chest X-ray in nearly 75% of 'consultations'. However there was significant clinical and non-clinical variation in referral.

'Patients' presenting with two chest symptoms were more likely to be referred for chest X-ray than those with one chest symptom and one 'atypical' symptom. The 'patients' presenting with appetite loss and weight loss were particularly unlikely to be referred, despite both 'patient profiles' meeting the NICE guidelines' recommendations for referral for chest X-ray. Once it was taken into account whether GPs had elicited the presence of weight loss or appetite loss, the difference in referral for chest X-ray compared to those 'patients' presenting with two chest symptoms was not so stark, although these 'patients' were still investigated less.

When considering the secondary variables coded, urgent referral for chest X-ray was rarely a part of GPs' management plan, but was most common for the 'patients' presenting with chest pain and weight loss. Whilst the 'patients' with chest pain and appetite loss were those least likely to be referred for chest X-ray (the primary variable of the vignette study) the majority of GPs did mention chest X-ray in their management plan for these 'patients', suggesting that GPs might be considering that significant lung pathology could be present, even if they did not feel it warranted investigation yet.

There was also significant non-clinical variation in GPs' referral for chest X-ray, with GPs less likely to investigate older 'patients' than younger, and less likely to investigate black 'patients' compared to white.

.

GPs' personal characteristics were not found to significantly influence their referral of 'patients' for chest X-ray.

.

5.2.2 : Strengths and limitations

5.2.2.1 : Strengths

The vignette study used a novel study tool, the virtual patient application, to examine GP decision making. Using this tool we were able to examine how GPs' referral behaviour varied with both clinical and non-clinical characteristics. Virtually all GPs were able to complete all six virtual 'consultations' using the application, and the majority did not report any challenges or problems in using it.

We designed the vignette study to simulate GPs' real life decision making as far as possible. The entire GP decision making study, including the vignette study, was completed online to enable GPs to complete the study in their own practices, thus replicating their routine consultation environment. The vignettes were presented in a multimedia format: the virtual patient application website provided GPs with information they would be able to access in real life (e.g. patient notes and examination findings), using videos for the 'consultations' (which provided them with both verbal and non-verbal cues), and the language recognition software simulated to some extent the back-and-forth dialogue of a real life consultation.

One of the main limitations of most text-based vignettes is that all participants receive the same information; however in real life the information that a GP receives will vary depending on the questions they ask, examinations they conduct, or additional sources of information that they consult. The interactive design of the virtual patient application enabled us to simulate this variation: GPs only obtained certain information about the 'patient' (e.g. the presence of a second symptom, or the duration of the symptoms) if they asked one or more relevant questions seeking it. The importance of this is apparent when considering the significance that GPs' obtaining information about the presence of a second symptom had for their likelihood to refer a 'patient' for chest X-ray.

5.2.2.2 : Limitations

5.2.2.2.1 : Some GPs found the virtual patient application difficult to use

Whilst most GPs did not comment on their experiences using the virtual patient application, 10.6% reported challenges using the software and/or commented that it was not entirely realistic. In particular, some GPs struggled with the constraints of the language recognition software, with the result that some GPs reported challenges in obtaining all the information they would normally seek in order to make a management decision in their day-to-day practice.

5.2.2.2.2 : The study does not examine real life

Whilst the factorial study design of the vignette study enabled the systematic manipulation of 'patient' characteristics in order to examine in their effects on GPs' referral in isolation, a disadvantage of this design is that it involves artificial scenarios. This raises the question of whether these vignettes were able to simulate real life GP consultations effectively enough for GPs' decisions to match the decisions they would make if the vignette 'patients' were real life patients in their GP practice.

Whilst it is not possible to know the extent to which the findings of this study reflect GPs' real life behaviour, it is possible to compare the study's findings with those of existing literature. Our finding that GPs proposed referring for a chest X-ray in 74.2% of the vignette viewings is in line with literature from 2013 (the year in which most GPs undertook the study),[31] although it is higher than might have been expected if GPs were following the 2005 NICE guidelines. This could reflect a limitation of the vignette study; an inability to fully simulate all the pressures (in particularly organisational) GPs face in real life primary care practice, such as resource constraints. However it is also possible that GPs' management decisions in the vignette study reflected an awareness of evidence supporting a lower threshold for cancer investigation,[1;153;154] and that our finding did reflect GPs' real life behaviour.

In the vignette study GPs were less likely to propose referring older 'patients' for chest X-ray. This is consistent with the findings of my systematic review, which identified several studies that had examined how referral of patients presenting to their GP with symptoms varied with patient age. By contrast Lyratzopoulos et al's

(2012)[44] study using cancer patient experience survey data found that patients aged 55 to 64 years had more referral delays than older patients (those over 75 years); this is perhaps more intuitive given that risk of cancer is known to increase with age, and so one would expect GPs to have a higher index of suspicion of cancer in older patients. However the fact that the findings of the vignette study differ does not necessarily mean that they do not represent GPs' real life referral decisions: Scott et al's (2013)[163] model of pathways to treatment proposes that as people age they become increasingly likely to attribute bodily changes to 'normal aging process' (rather than to disease), and it is possible that GPs might also have this approach and thus be less likely to investigate symptoms in older patients.

We also observed that GPs were less likely to propose a chest X-ray when viewing vignettes with black 'patients' that white 'patients', consistent with Lyratzopoulos et al's (2012)[44] findings that non-white cancer patients report more delays in referral than white patients. This consistency might reflect the fact that, by using videos to present vignettes to GPs, our study was able to simulate GPs' real life consultation experience effectively for ethnicity.

Contrary to what might be expected, in the vignette study we found an overall lack of gradient in the percentage of 'patients' GPs referred for chest X-ray across the three different levels of risk. This suggests that there could be limitations in the vignette study's validity to examine the influence of clinical risk on GPs' referral for chest-ray. When designing the vignette study we based the three risk levels we examined on PPVs from the CAPER symptom case-control dataset for lung cancer.[63] However since these PPVs have wide and overlapping confidence intervals they (on their own) are not necessarily sufficient to clearly delineate risk levels. That said, we also aligned the three risk levels and six 'patient' profiles with the NICE guidelines in place at the time. If there are problems expressing clinical risk effectively (i.e. such that it replicates real life) in vignettes then it is possible that, in future, the effect of clinical risk on GPs' decision making could be more effectively examined using other methods.

However it is important to consider that the vignette study's lack of gradient in referrals across the levels of risk could be due to GPs' consideration of other clinical factors (in addition to risk) when deciding whether to refer a patient for chest X-ray. It is possible that GPs had a lower threshold for referral than we expected when designing the study; this would be in line with the 2015 NICE guidelines, under which the vignette study's 'medium' and 'high' risk 'patient' profiles meet the criteria for chest X-ray).[35] It could also reflect a difference in GPs' approach to 'chest' and 'non-chest' symptoms : 'patients' with two chest symptoms were most likely to be referred for chest X-ray, despite the level of risk. It is also possible that it is a consequence of variation between the six 'patient' profiles in the likelihood that GPs elicited the second symptom a 'patient' was presenting with, since only eliciting one symptom would of reduce the perceived level of risk, and thus potentially influence GPs' decision making.

Whilst the artificial nature of the vignette study means that it is unclear to what extent the study's findings reflect those that GPs would make in real life, we were aware of this limitation when designing the GP decision making study and were therefore able to take steps to address it in the second part of the study (the post-consultation survey, Study 2b).

5.2.3 : Implications for future research, policy and practice

5.2.3.1 : Future research

The vignette study provided evidence (supporting that of my systematic review) that there are significant non-clinical variations in GPs' decisions to refer patients, in this case referral of patients with symptoms indicative of lung cancer for chest X-ray. Strategies are needed to identify the factors which underlie and influence GPs' decision making about referral, in particular for older and non-white patients.

The virtual patient application enabled us to study GP decision making in a novel manner so that we were able to gain an insight into the potentially significant aspects of GPs' behaviour within a consultation (e.g. the questions they ask and whether they receive their intended response from the patient, or the examinations they conduct and information they seek) and their effects on GPs' referral behaviour. Further research focusing on the content of the consultation, rather than simply the outcome, could help increase our understanding of the factors influencing the variations seen in GPs' referral behaviour.

5.2.3.2 : Policy and practice

The wide variation in GPs' decisions to refer 'patients' with symptoms indicative of lung cancer for chest X-ray (including relatively low referral in some of the higher risk profiles that meet the NICE guidelines' recommendations for conducting a chest X-ray), in particular where GPs did not elicit the presence of both the 'patient's' symptoms, demonstrates the importance of GPs having as much of the available information as possible in order to make an effective management plan. It suggests the benefit of developing strategies to prompt GPs to seek out key symptom information: for example if a patient reports a particular symptom, a prompt to GPs to ask about other symptoms that are frequently related (or symptoms that when they occur in combination significantly increase the likelihood of disease). Educating GPs as to the importance of following up with questions about additional symptoms (rather than relying on the patient to mention them, when the patient may not be aware of their significance or implications) could also be effective.

There is also a potential role for the virtual patient application as a teaching tool or within research: for example GPs could be prompted at various stages of a virtual 'consultation' to reflect on their thoughts and behaviour, and factors influencing these, as a means of reinforcing good practice.

6 : The GP decision making study's post-consultation survey (Study 2b)

6.1 : Introduction

The aim of the GP decision making study (Study 2) was to examine variations in GPs' decision making for patients presenting with symptoms that could indicate a diagnosis of lung cancer. The factorial design of the vignette study (Study 2a) allowed us to quantitatively examine the extent to which GPs' decisions varied both with different clinical presentations, and also by the socio-demographic characteristics of the 'patients' and the GPs. However the design of the vignette study did not allow us to explore what influences might be driving differences in GPs' decisions, nor the impact of organisational factors on GPs' referral behaviour. As I demonstrated in my systematic literature review (Study 1), there are few quantitative and well-conducted studies in the existing literature that consider the impact of these factors on GPs' referral behaviour. A better understanding of the underlying reasons for differences in GPs' referral decisions (especially for those patients presenting with symptoms that could indicate cancer) is therefore key to reducing the variation in GPs' management decisions that is likely to contribute to the variation in cancer survival rates within the UK.

I therefore developed the post-consultation survey (Study 2b) as a second part of the GP decision making study. This enabled me to identify factors that GPs believe affect their decision making and to consider how these may contribute to the non-clinical differences seen in GPs' referral decisions (both in the vignette study, and in many studies identified by my systematic literature review). When developing the GP decision making study we were aware that one limitation of the vignette study was that we were not examining GPs' real life referral behaviour. However in the post-consultation survey I was able to specifically ask GPs about their decision making processes and behaviours in their real life practice.

In this chapter I outline the aim, methods and development of the post-consultation survey, as well as discussing its findings and their implications.

6.1.1 : Aim

To examine the extent to which GPs believe certain factors influence their referral decisions for real patients who present in a similar manner to those in the vignette study (Study 2a) in order to increase our understanding of why GPs make the decisions that they do.

6.1.2 : Objectives

To conduct a questionnaire survey to:

- examine the extent to which GPs use sources of information in their decision making;
- identify factors that GPs believe influence their decisions to send patients for investigation and/or refer them to secondary care;
- provide GPs with an opportunity to provide any further comments about or reflections on their decision making process.

6.2 : Methods

6.2.1 : Delivery of the post-consultation survey

The post-consultation survey was a web-based survey, designed to be completed by all GPs participating in the GP decision making study immediately after each had finished the vignette study. The survey asked GPs how and why they make referral or investigation decisions in their day-to-day practice.

Initially we had planned for the survey to form part of the online application used for the vignette study. This proved to be beyond the scope of the software designers, so I instead developed the survey using UCL Opinio software (survey viewable in Appendix 17).[150] Each GP accessed the survey website directly from the virtual patient application via a link that was displayed on the virtual patient application website once they had completed all six 'consultations' in the vignette study. The survey website also stored GPs' responses.

6.2.2 : The survey structure

When GPs accessed the post-consultation survey they were first presented with an introductory page (Figure 24). GPs did not receive any additional information before completing the post-consultation survey - they moved straight from completing the final vignette to the survey's introductory page. Therefore, as with recruitment and the vignette study, I framed the survey to GPs as a study seeking to understand how GPs make decisions.

After the introductory page GPs then proceeded to pages containing questions about their behaviour during the vignette study, their day-to-day practice, and some of their personal characteristics. Each page included a reminder that the survey was not a test of 'correct' behaviour: we were keen to understand what GPs actually do in 'real life', not what they thought we wanted them to tell us.

My survey had the following three sections:

I - Decision making in these vignettes

GPs may use several **sources of information to assist their decision making** (e.g. guidelines, textbooks or seeking advice from colleagues). My questions sought to determine how this varied between the virtual 'consultations' and GPs' day-to-day practice.

II - Decision making in your everyday practice

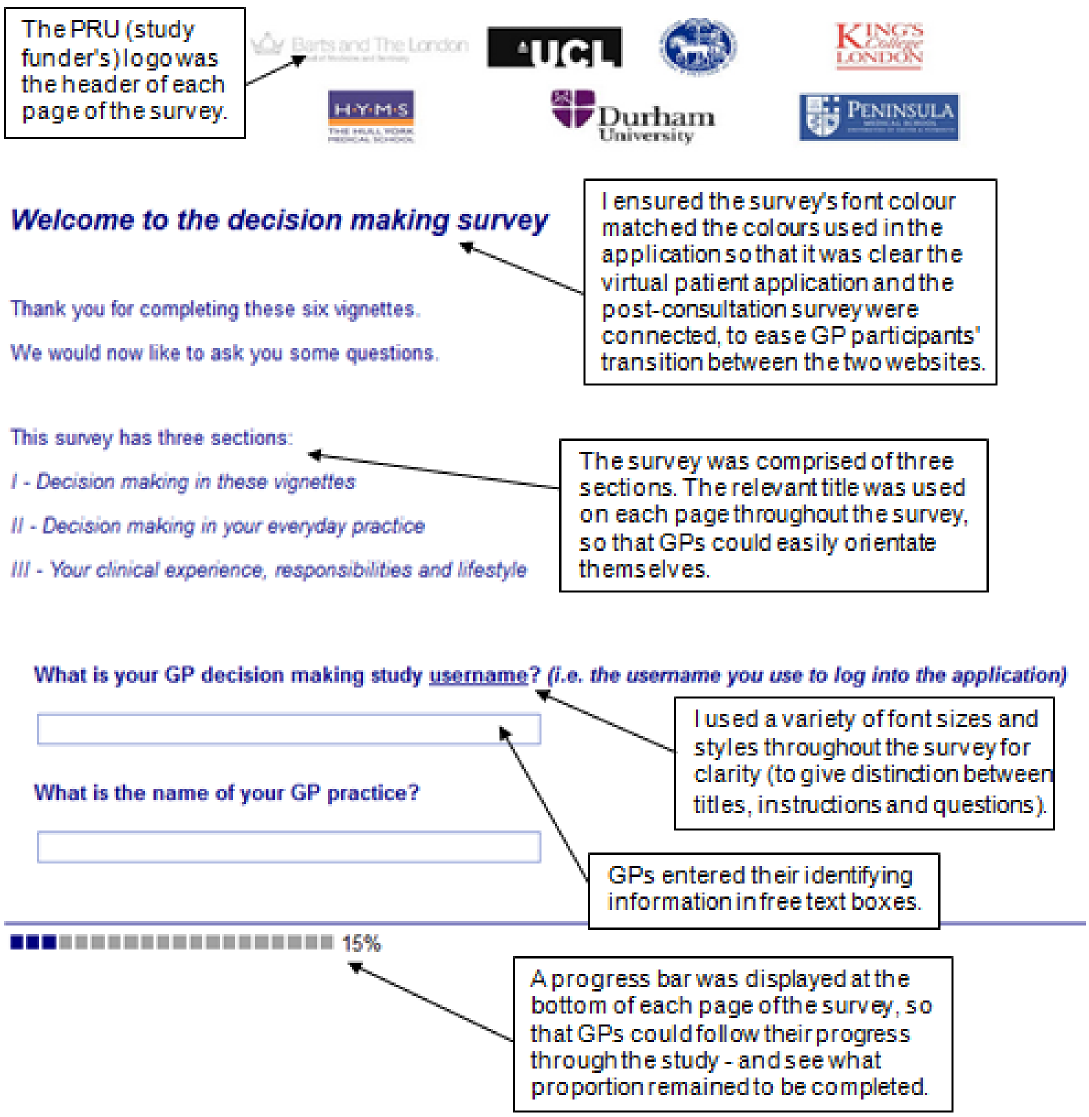
GPs were presented with a list of **factors that could influence the likelihood that they refer a patient** for investigation or to secondary care and asked to rate to what extent they are influenced by them.

III - Your clinical experience, responsibilities and lifestyle

In this section I included **questions about additional GPs' characteristics** which may influence decision making and which therefore we wanted to examine in the GP decision making study, but could not ask at registration due to the risk of priming GPs to our study aims:

- clinical experience;
- budgetary responsibilities;
- smoking status.

Figure 24 : The introductory page of the post-consultation survey, with particular features noted



6.2.3 : Development of the survey

I structured the survey as a questionnaire, using a variety of question formats. I used guidance from Bowling's work (2005)[164] on questionnaire design in order to structure the survey and phrase questions appropriately.

The survey version used in the first pilot of the GP decision making study contained more questions, but the three GPs took a long time to complete it. Since we intended for the full GP decision making study (the vignette study and the post-consultation survey) to take GPs one hour in total this was not practical, so I altered the survey content to ensure it could be completed in approximately five minutes. The majority of questions in the survey explored the extent to which factors influence GPs' investigation/referral behaviour. In earlier versions of the survey I used a multiple choice structure for these questions, with substantial conditional formatting so that GPs received particular follow up questions based on their initial responses. However GPs in the first pilot were often unsure how to answer these questions. I therefore simplified the structure and format of this section to ask all GPs about each factor I was exploring, using an adapted Likert scale for rating the likelihood of referral.

I developed the content of the post-consultation survey questions predominantly using existing literature (informed by Study 1, the systematic literature review), and also with advice and suggestions from three GP advisors, based on their experience. The process of generating and selecting which of the multiple factors that could influence GPs' investigation/referral behaviour to ask about in the survey involved several steps.

Whilst my development of the content of these questions was informed by Study 1, the timeline for completion of the GP decision making study required me to develop the survey before I had completed my systematic review. I therefore conducted a number of targeted searches of the records remaining after title screening (using terms related to gender or ethnicity, for example) then reviewed the full papers of the records retrieved by these searches. I identified 55 studies (UK and non-UK) that considered whether non-clinical patient and/or GP characteristics were associated with GPs' referral decisions. 26 of these studies hypothesised about the reasons for non-clinical differences in GPs' referral behaviour (Appendix 18);

however it is of note that very few tested these, which highlights the importance of including this post-consultation survey in the GP decision making study.

I used the hypotheses from these studies as a starting point for selecting topics to address in my survey. For each hypothesis, I noted which non-clinical variations in referral researchers were proposing it might contribute to, as well as any additional socio-demographic or organisational characteristics one might reasonably expect it to be relevant to. For example in the literature transport difficulties were proposed as a potential influence for variation in referral by both patients' level of deprivation (Sowden et al, 2008)[92] and their distance to travel for an appointment (Srinivasa et al, 2007);[165] and in this example I also hypothesised that they could contribute to variation by patient age. As well as the hypotheses from these studies I also used the suggestions from our GP advisors, considering what non-clinical variation in GPs' referral behaviour one might expect to see if the factors they had suggested do indeed influence GPs' referral behaviour.

When selecting which of these possible factors to explore in the survey (by asking GPs the extent to which factors relating to these influenced their decision making) I chose to focus in particular on factors which either more than one study had proposed as a potential reason for non-clinical differences in referral, or that clearly correlated to socio-demographic or organisational characteristics (since this reflected the aim of the GP decision making study, of which this post-consultation survey was a part, to examine patient and GP characteristics associated with GPs' management decisions). In addition I included an optional free text question to capture any other factors GPs considered an important influence.

Using this approach, I selected 33 factors that I hypothesised might influence GPs' decision making and contribute to non-clinical variation in their referral behaviour. These formed the basis of the majority of the content of the post-consultation survey. In the survey, when asking GPs about the extent to which these factors influenced their referral behaviour I listed six to eight of these factors per page (across five consecutive pages) in an attempt to avoid presenting GPs with too much information at once. I grouped similar factors together on a page, each page addressing one of the following five topics:

- patients' responsibilities and patients' use of/engagement with health services;
- barriers to access (language and travel related);

- patients' understanding and knowledge;
- patients' concerns;
- organisational issues (focused on investigations and secondary care).

In the initial version of the survey I asked GPs about the extent to which particular factors would 'influence [their] decisions to investigate or refer a patient'. However in the first pilot GPs commented that the context was too broad: for example some factors make them less likely to refer a patient for an involved procedure such as a colonoscopy, but would not influence their decision for a chest X-ray. In the final version of the post-consultation survey I therefore asked GPs to answer the questions thinking about patients they had seen within the last month who they had considered referring for simple investigations (such as ultrasound or X-ray) and/or referring to secondary care, actions relating to the vignette study's primary outcome measure.

6.2.4 : Data processing of the survey responses

All 227 GPs who completed the GP decision making study submitted a response to the post-consultation survey.

I exported the data on GPs' responses from the UCL Opinio website into both an Excel spreadsheet and a Stata worksheet. Nine GPs had completed the survey more than once; this was either due to an error with the application at the very start of recruitment (which displayed the link to the survey after the GP had completed each of the three batches of 'consultations', rather than simply after they had completed all six), or as a result of the GP not finishing the survey on first sitting and returning to complete it at a later point. In these cases I kept one survey entry per GP, selecting the first complete entry after completion of all six 'consultations'. The first question asked the GP to enter their study username so that it was possible to match their survey response to their vignette decisions. Only one GP had not entered their username, however they had entered their practice and therefore it was easy to trace the response (to double check I also cross-referenced the time they completed their last virtual 'consultation' with their survey completion time).

Once the data was exported into Stata I labelled each variable, and in some cases converted text information into numerical codes in order to enable quantitative analysis. For one question (information about GPs' clinical experience) I created new variables in order to make the information collected clearer and easier to interpret. In order to examine the five-point Likert scale quantitatively I coded GPs' responses to each statement as follows:

1 = Less likely to refer in most circumstances

2 = Less likely to refer in some circumstances

3 = No more or less likely to refer

4 = More likely to refer in some circumstances

5 = More likely to refer in most circumstances

0 = Don't know

6.2.5 : Analysis of the survey responses

As discussed in Section 6.2.2, the post-consultation survey had three sections, each seeking different information and each using a different question format. I therefore analysed the results for each of these sections differently.

6.2.5.1 : Sources of information GPs use in their decision making

GPs were asked questions requiring a 'Yes' or 'No' response about their use of a variety of different sources of information during both the vignette study, and in their real life practice. For each source of information I performed descriptive analysis to identify the percentage of GPs who reported using it.

6.2.5.2 : Factors influencing the likelihood that GPs refer a patient

6.2.5.2.1 : Descriptive analysis

GPs were asked to state the extent to which they felt a number of different factors influenced their real life referral behaviour using a five-point Likert scale. Where GPs did not provide a response to the question, or gave a 'don't know' response, I excluded their response from further analysis of that particular factor.

I constructed histograms for each factor to show the spread of responses across the full five-point Likert scale (excluding those GPs who stated 'don't know').

I also conducted quantitative descriptive analysis for each factor; for this analysis I combined the two 'less likely' response categories into one overall category, and likewise the two 'more likely' categories into another overall category. For each factor I therefore report the total number of GPs who gave a response, and break this total down into:

- the number of GPs who stated that they were more likely to refer;
- the number of GPs who stated they were no more or less likely to refer;
- the number of GPs who stated that they were less likely to refer;
- the number of GPs who reported that they did not know.

6.2.5.2.2 : Significance testing

In order to further analyse the extent to which each of these factors might influence GPs' referral behaviour, I used a variant of the McNemar test. This is usually used to assess whether discordances between matched binary outcomes are in one direction more than the other. The test simply assesses whether the proportion of discordances in either direction is significantly different from 0.5. I used the two combined 'overall more likely' and 'overall less likely' categories and compared the proportion more likely to refer with 0.5. Thus the test does not use the neutral observations, and essentially asks the question: where the factor does have an influence on GPs' decisions to refer for simple investigation and/or to secondary care, is it a significant influence in one direction or the other.

6.2.5.2.3 : Free text analysis

All GPs were given the opportunity to make further comments on factors affecting their decision making (or to reflect more widely on the GP decision making study); I analysed these comments qualitatively.

6.2.5.3 : Individual GPs' personal characteristics

We also used the post-consultation survey to collect GP-specific information which could be used to analyse GPs' decisions in the vignette study, but about which we could not ask prior to the study due to the risk of priming GPs to our study aims. I performed descriptive analysis of their responses.

6.3 : Results

The post-consultation survey asked GPs questions about:

- the sources of information they use in their decision making;
- factors influencing the likelihood that they refer a patient;
- their personal characteristics.

I will discuss each of these in turn.

6.3.1 : Sources of information GPs use in their decision making

Table 18 shows how the 227 GPs who completed the GP decision making study responded to the questions about their use of different sources of information.

As we might expect, given the artificial nature of the vignette 'consultations', not many GPs (25 in total, 11.0%) referred to additional sources of information while completing the vignette study. Where GPs did seek information during the study, the most common sources were NICE or other (including local) guidelines.

By contrast more than half of GPs (128 GPs, 56.4%) stated that they would use at least one source of information in a real life consultation with a patient with similar symptoms to those in the vignettes, with some noting that they would consult several. GPs most commonly selected that they would discuss with a colleague (32.2%). This was followed by 29.5% who used books or websites. Fewer GPs reported using NICE guidelines (29.1%) or other guidelines (27.8%).

Table 18 : GPs' reported use of information sources during both the vignette 'consultations' and their real life practice

Information source	Responded	Used in vignettes	Would use in real life	Would use in real life *including responses given in the free text section of the survey that indicated use of these sources
Colleague	227	2 (0.9%)	73 (32.2%)	76 (33.5%)
NICE guidelines	227	16 (7.0%)	66 (29.1%)	71 (31.3%) *
Other guidelines	227	10 (4.4%)	63 (27.8%)	65 (28.6%) *
Book/website	227	6 (2.6%)	67 (29.5%)	68 (30.0%)
None of the above	227	202 (89.0%)	99 (43.6%)	94 (41.4%)

* For the three GPs whose free text comments did not specify NICE vs. other guidelines I have included their response in both counts in order to give the maximum likely estimate of GPs' use of information sources from the data available in this study

The results reported above refer to GPs' responses to the specific questions in the post-consultation survey about their use of information in their real life practice. However 18 of the 227 GPs also commented about their use of different information sources (colleagues, guidelines and books/websites) in the free text section of the survey. 13 GPs' written responses correlated exactly with their response to the specific question about what information sources they would use in real life. However five GPs who had selected 'none of the above' for information sources they would use in real life did refer to consulting other sources in their free text response, and so I have included these additional data in the final column of Table 18. Of these five GPs, all wrote in the free text section that they consulted guidelines (three did not specify which, whilst two stated they refer to local guidelines), three stated that they consulted hospital or practice colleagues, and one stated that they also used websites. As seen in Table 18, even with this information included, GPs' use of information sources is still very similar overall.

6.3.2 : Factors influencing the likelihood that GPs refer a patient

I asked GPs about the extent to which they felt 33 different factors influenced their real life referral behaviour. Here I will discuss the overall patterns in GPs' responses, and factors that significantly affected GPs' referral decisions. In Appendix 19 I report descriptive analysis of all 33 factors.

The spread of GPs' responses to how each individual factor influenced their decision making generally followed one of three patterns, each of which I will discuss in more detail on the following pages:

a) A substantial majority (over 80%) of GPs reported that the factor did not make them any more or less likely to refer a patient for investigation or to secondary care;

b) There was a distinct skew, with a significant number of GPs either more or less likely to refer a patient for investigation or to secondary care;

c) Several GPs reported that the factor would influence their referral behaviour, but there was no consensus in which way it influenced them (i.e. it influenced different GPs differently).

6.3.2.1 Factors which a substantial majority of GPs reported did not influence their referral

For eight of the factors, the majority of GPs (more than 80%) did not report any effect on their referral behaviour, stating that they were 'no more or less likely to refer' a patient for simple investigation or to secondary care. In several of these cases there was also no significant directional influence for the GPs who did state that it would affect their decision to refer, although for four factors there was a clear direction of influence on referral behaviour for the GPs who had not responded neutrally.

Factors which do not influence most GPs' referral decisions (for our sample) are:

- the patient has not followed health promotion or disease prevention advice in the past (e.g. has not stopped smoking)
  87.1% stated a 'neutral' response. The 12.9% of GPs who stated that it would influence their referral behaviour were more likely to refer a patient (p=0.0093)
- the patient will require an interpreter for their appointment/diagnostic test
  90.1% stated a 'neutral' response. The 9.9% of GPs who stated that it would influence their referral behaviour were more likely to refer a patient (p=0.0190)
- the patient does not have a source of transport to or from the appointment/ diagnostic test
  83.6% stated a 'neutral' response. The 16.4% of GPs who stated that it would influence their referral behaviour were less likely to refer a patient (p<0.0001)
- the patient is concerned it is expensive to travel to the appointment/diagnostic test
  80.9% stated a 'neutral' response. The 19.1% of GPs who stated that it would influence their referral behaviour were less likely to refer a patient (p<0.0001)
- the patient does not ask about other management options available
  91.0% stated a 'neutral' response. For the 9.0% of GPs who stated that it would influence their referral behaviour, the direction of the effect was not significant (p=0.8231)

- the patient does not know what services are available to them
  92.4% stated a 'neutral' response. For the 7.6% of GPs who stated that it would influence their referral behaviour, the direction of the effect was not significant (p=1.0000)
- the patient appears concerned about the stigma associated with certain differential diagnoses
  83.0% stated a 'neutral' response. For the 17.0% of GPs who stated that it would influence their referral behaviour, the direction of the effect was not significant (p=0.3247)
- the patient is concerned about overusing the health service
  90.5% stated a 'neutral' response. For the 9.5% of GPs who stated that it would influence their referral behaviour, the direction of the effect was not significant (p=1.0000)

6.3.2.2 : Factors for which a significant number of GPs were either more or less likely to refer

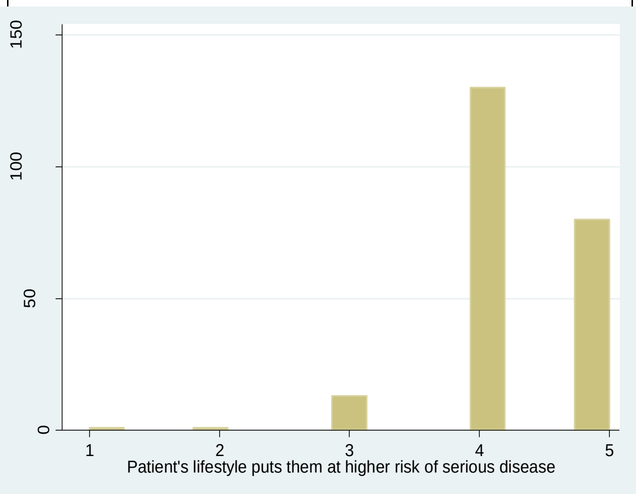
A number of factors have a significant directional influence on the majority of GPs' referral decisions, either to make them more likely to refer patients, or less likely.

6.3.2.2.1 : Factors that GPs report would increase the likelihood they refer a patient for simple investigation and/or to secondary care

Nine factors had a distinct positive skew, with a substantial number of GPs stating that they were more likely to refer patients in these situations.

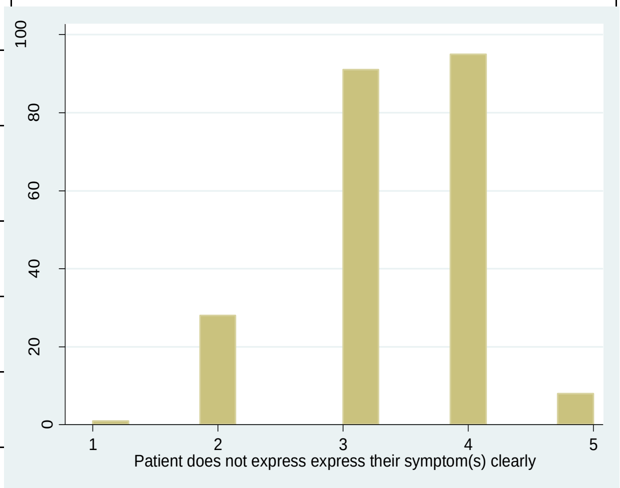
The most significant impact on GP referral behaviour was if a patient's lifestyle put them at increased risk of serious disease. 93.4% of GPs stated that this would make them more likely to refer a patient and McNemar's chi-squared test statistic was 203.12, showing a very significant influence in this direction (p<0.0001).

Statement	*The patient's lifestyle puts them at higher risk of serious disease*
Total responses	226
'Don't know' response	1
Less likely to refer (%) coded as 1 or 2	2 (0.9%)
No more or less likely to refer (%) coded as 3	13 (5.8%)
More likely to refer (%) coded as 4 or 5	211 (93.4%)
Mean of responses	4.28
Median of responses	4
Ratio of 'less likely to refer' to 'more likely to refer'	1 : 105.5
McNemar test result (p value)	203.12 (p<0.0001)

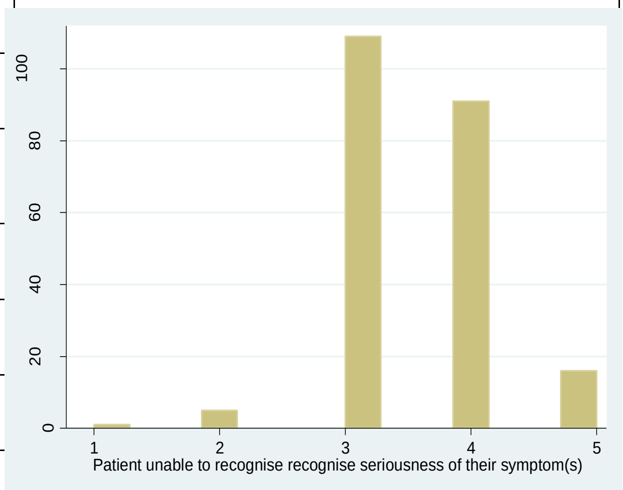


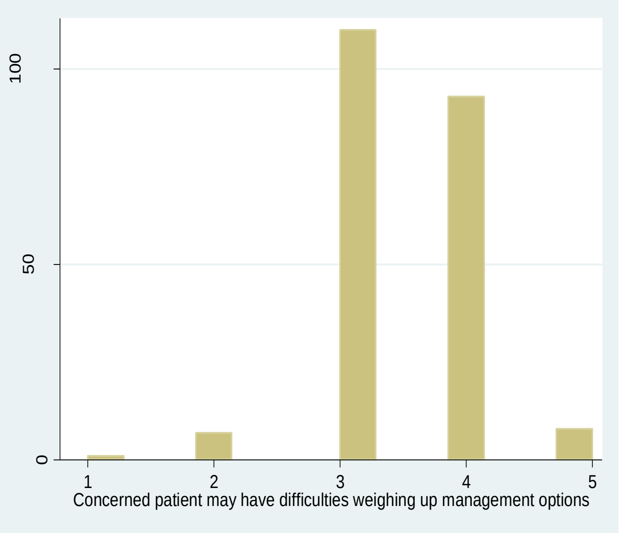
A significant proportion of GPs reported they were more likely to refer patients who have difficulty expressing their symptoms clearly (46.0%) or recognising their potential severity (48.0%), or who struggle to weigh up potential management options (45.9%) (p<0.0001 for all).

Statement	*The patient does not express their symptom(s) clearly*
Total responses	224
'Don't know' response	2
Less likely to refer (%) coded as 1 or 2	29 (12.9%)
No more or less likely to refer (%) coded as 3	92 (41.1%)
More likely to refer (%) coded as 4 or 5	103 (46.0%)
Mean of responses	3.36
Median of responses	3
Ratio of 'less likely to refer' to 'more likely to refer'	1 : 3.55
McNemar test result (p value)	40.37 (p<0.0001)



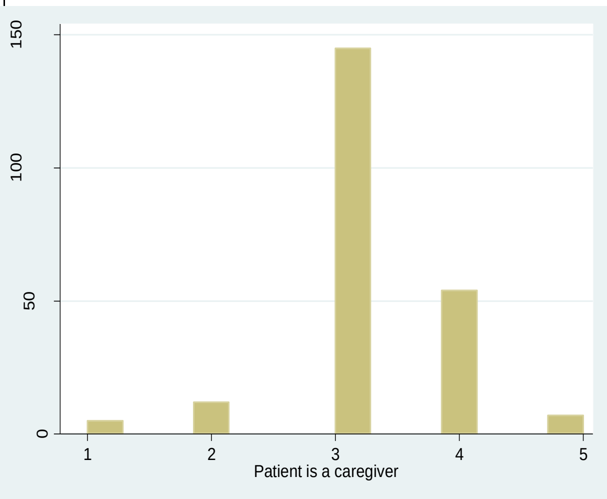
Statement	*The patient is unable to recognise the seriousness of their symptom(s)*
Total responses	223
'Don't know' response	3
Less likely to refer (%) coded as 1 or 2	6 (2.7%)
No more or less likely to refer (%) coded as 3	110 (49.3%)
More likely to refer (%) coded as 4 or 5	107 (48.0%)
Mean of responses	3.52
Median of responses	3
Ratio of 'less likely to refer' to 'more likely to refer'	1 : 17.83
McNemar test result (p value)	88.50 (p<0.0001)



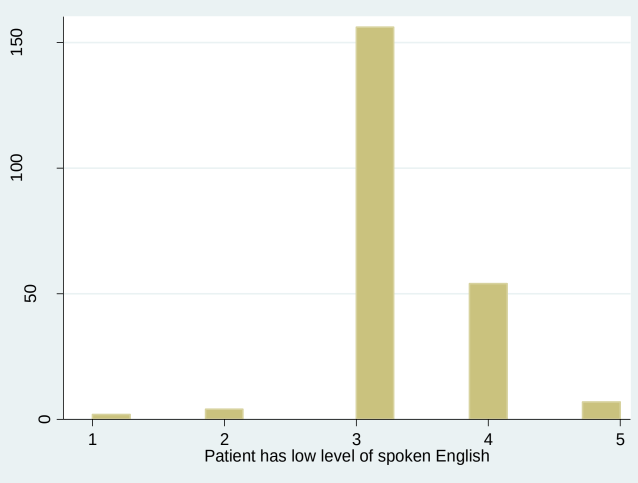
Statement	*You are concerned that the patient may have difficulties weighing up the consequences of different management options*	
Total responses	220	
'Don't know' response	6	
Less likely to refer (%) coded as 1 or 2	8 (3.6%)	
No more or less likely to refer (%) coded as 3	111 (50.5%)	
More likely to refer (%) coded as 4 or 5	101 (45.9%)	
Mean of responses	3.46	
Median of responses	3	
Ratio of 'less likely to refer' to 'more likely to refer'	1 : 12.63	
McNemar test result (p value)	77.65 (p<0.0001)	

A majority of GPs stated that their referral behaviour would not be affected by a patient being a caregiver (65.2%), having a low level of spoken English (70.5%) or their appointments running late (70.1%). However for each of these factors, where GPs were affected by the factor this was to significantly increase their likelihood of referring these patients (p<0.0001, p<0.0001 and p=0.0015 respectively).

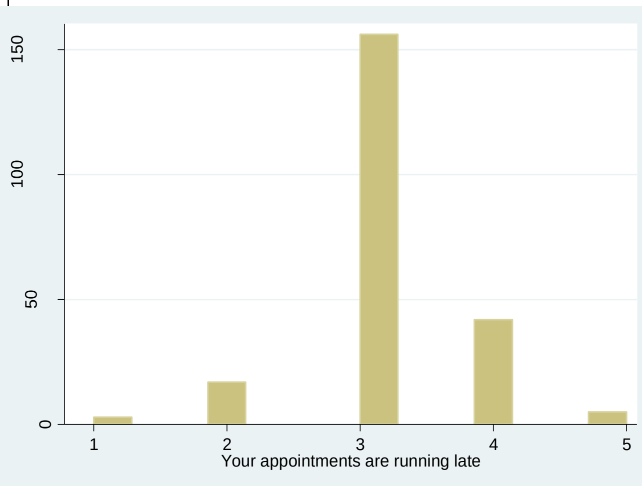
Statement	*The patient is a caregiver*
Total responses	224
'Don't know' response	2
Less likely to refer (%) coded as 1 or 2	17 (7.6%)
No more or less likely to refer (%) coded as 3	146 (65.2%)
More likely to refer (%) coded as 4 or 5	61 (27.2%)
Mean of responses	3.21
Median of responses	3
Ratio of 'less likely to refer' to 'more likely to refer'	1 : 3.59
McNemar test result (p value)	23.71 (p<0.0001)



Statement	*The patient has a low level of spoken English*
Total responses	224
'Don't know' response	3
Less likely to refer (%) coded as 1 or 2	6 (2.7%)
No more or less likely to refer (%) coded as 3	158 (70.5%)
More likely to refer (%) coded as 4 or 5	60 (26.8%)
Mean of responses	3.27
Median of responses	3
Ratio of 'less likely to refer' to 'more likely to refer'	1 : 10
McNemar test result (p value)	42.56 (p<0.0001)

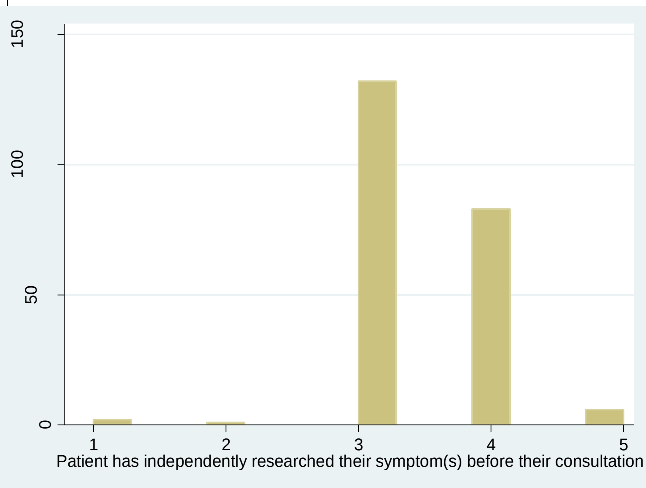


Statement	*Your appointments are running late*
Total responses	224
'Don't know' response	3
Less likely to refer (%) coded as 1 or 2	20 (8.9%)
No more or less likely to refer (%) coded as 3	157 (70.1%)
More likely to refer (%) coded as 4 or 5	47 (21.0%)
Mean of responses	3.13
Median of responses	3
Ratio of 'less likely to refer' to 'more likely to refer'	1 : 2.35
McNemar test result (p value)	10.09 (p=0.0015)



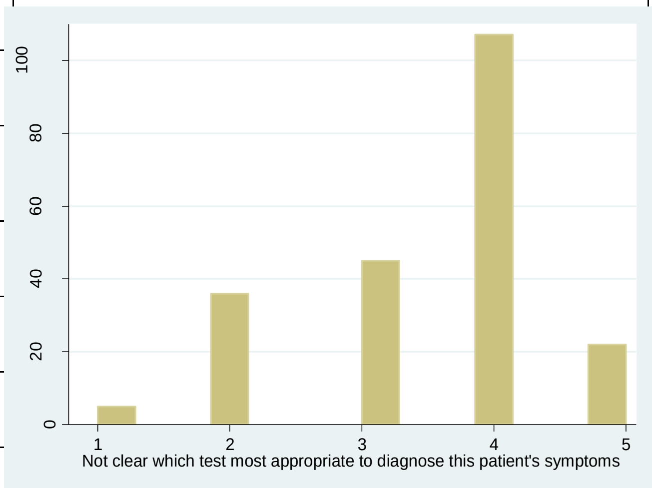
The GPs' free text comments in this survey, anecdotal evidence from practising GPs, and other studies of GP decision making suggest that GPs may be more likely to refer and/or investigate patients who specifically request this in their consultation.[166;167] In consultation with my supervisors, I opted not to ask GPs specifically whether patients' requests influenced their referral behaviour on the grounds that this is a leading question. However I did include whether patients had researched their symptoms as a factor, and it is possible these patients may also be more likely to request referral (be that as a result of their research, or due to an underlying 'proactive' interest in their healthcare). Whilst 59.1% of GPs reported that patient research would not influence their referral decision, 39.6% stated that they were more likely to refer these patients, with a McNemar's test result of 78.53, a very significant difference (p<0.0001).

Statement	*The patient has independently researched their symptom(s) before their consultation*
Total responses	225
'Don't know' response	1
Less likely to refer (%) coded as 1 or 2	3 (1.3%)
No more or less likely to refer (%) coded as 3	133 (59.1%)
More likely to refer (%) coded as 4 or 5	89 (39.6%)
Mean of responses	3.40
Median of responses	3
Ratio of 'less likely to refer' to 'more likely to refer'	1 : 29.67
McNemar test result (p value)	78.53 (p<0.0001)



GPs gave a wide range of responses as to whether lack of clarity on what test would be most appropriate to diagnose a patient would make them more or less likely to refer a patient for simple investigation and/or to secondary care (19.0% stated they would be less likely to refer, 21.3% no more or less likely, and 59.7% more likely to refer). However, despite this lack of consensus, McNemar's test showed that overall GPs were significantly more likely to refer patients when the most appropriate test was unclear ($p<0.0001$).

Statement	*It is not clear which test would be most appropriate to diagnose this patient's symptom(s)*
Total responses	216
'Don't know' response	11
Less likely to refer (%) coded as 1 or 2	41 (19.0%)
No more or less likely to refer (%) coded as 3	46 (21.3%)
More likely to refer (%) coded as 4 or 5	129 (59.7%)
Mean of responses	3.49
Median of responses	4
Ratio of 'less likely to refer' to 'more likely to refer'	1 : 3.15
McNemar test result (p value)	44.52 (p<0.0001)

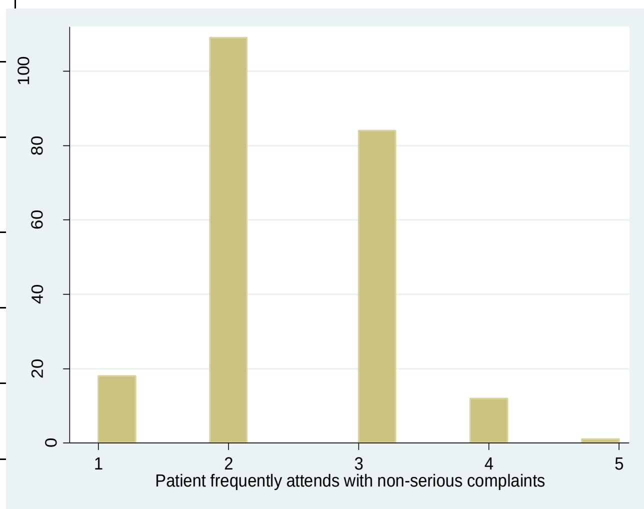


6.3.2.2.2 : Factors that GPs report would decrease the likelihood they refer a patient for simple investigation and/or to secondary care

13 of the 33 factors had a distinct negative skew, with a significant number of GPs stating that they were less likely to refer patients in these situations.

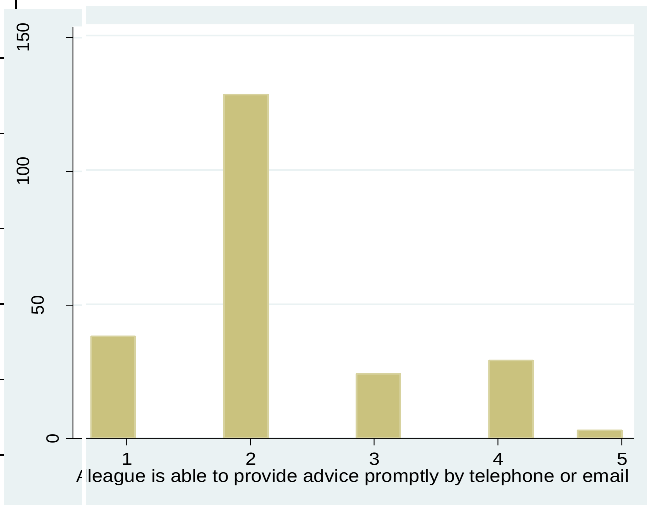
The majority of GPs (56.4%) reported that they were less likely to investigate or refer patients who frequently attended with non-serious complaints. This was the factor that most significantly decreased the likelihood that a patient would be referred for investigation and/or to secondary care: McNemar's chi-squared test statistic was 93.50, showing a very significant influence in this direction (p<0.0001).

Statement	*The patient frequently attends with non-serious complaints*
Total responses	225
'Don't know' response	2
Less likely to refer (%) coded as 1 or 2	127 (56.4%)
No more or less likely to refer (%) coded as 3	86 (38.2%)
More likely to refer (%) coded as 4 or 5	12 (5.3%)
Mean of responses	2.42
Median of responses	2
Ratio of 'less likely to refer' to 'more likely to refer'	10.58 : 1
McNemar test result (p value)	93.50 (p<0.0001)



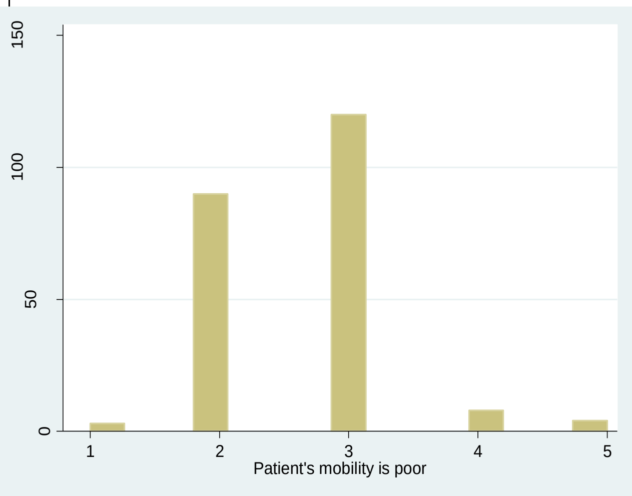
The majority of GPs also reported that they were less likely to refer patients when they could receive prompt advice from a hospital colleague (75.3%).

Statement	*A hospital colleague is able to provide advice promptly by telephone or email*
Total responses	223
'Don't know' response	4
Less likely to refer (%) coded as 1 or 2	168 (75.3%)
No more or less likely to refer (%) coded as 3	24 (10.8%)
More likely to refer (%) coded as 4 or 5	31 (13.9%)
Mean of responses	2.24
Median of responses	2
Ratio of 'less likely to refer' to 'more likely to refer'	5.42 : 1
McNemar test result (p value)	92.94 (p<0.0001)

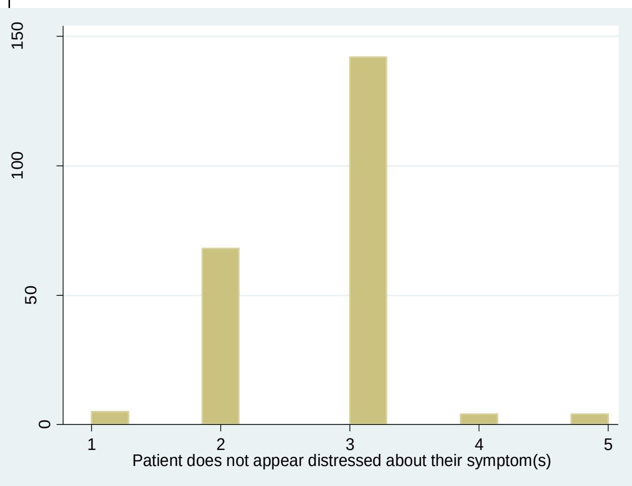


A significant proportion of GPs reported they were less likely to refer patients who have poor mobility (41.2%), or who do not appear distressed about their symptoms (32.6%), or where if the diagnostic test is positive there are limited effective treatment options available for the patient (42.0%) (p<0.0001 for all).

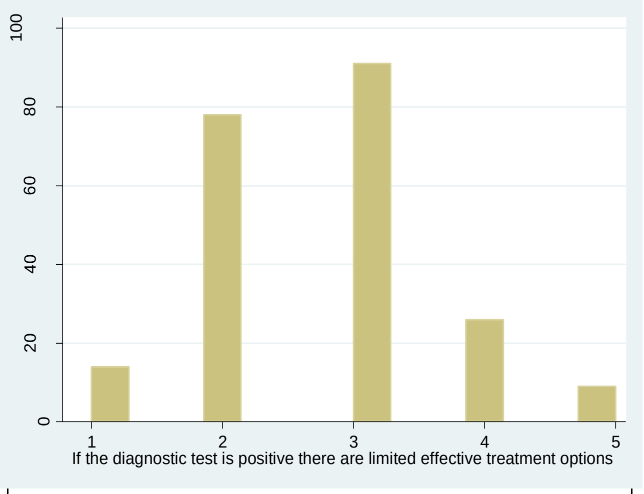
Statement	*The patient's mobility is poor*
Total responses	226
'Don't know' response	1
Less likely to refer (%) coded as 1 or 2	93 (41.2%)
No more or less likely to refer (%) coded as 3	122 (54.0%)
More likely to refer (%) coded as 4 or 5	11 (4.9%)
Mean of responses	2.64
Median of responses	3
Ratio of 'less likely to refer' to 'more likely to refer'	8.45 : 1
McNemar test result (p value)	63.09 (p<0.0001)



Statement	*The patient does not appear distressed about their symptom(s)*
Total responses	224
'Don't know' response	1
Less likely to refer (%) coded as 1 or 2	73 (32.6%)
No more or less likely to refer (%) coded as 3	144 (64.3%)
More likely to refer (%) coded as 4 or 5	7 (3.1%)
Mean of responses	2.70
Median of responses	3
Ratio of 'less likely to refer' to 'more likely to refer'	10.43 : 1
McNemar test result (p value)	52.81 (p<0.0001)

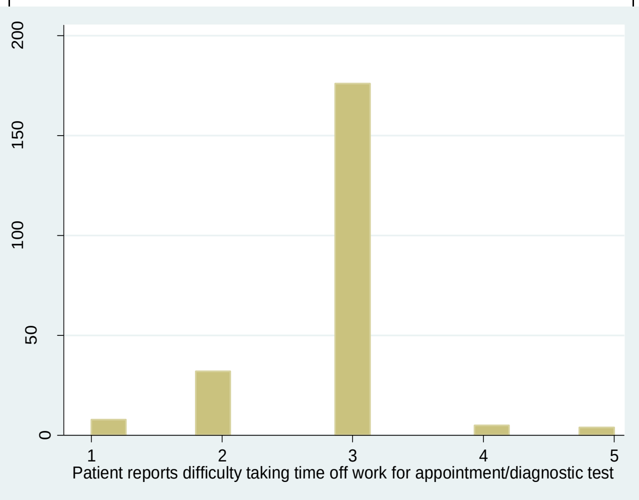


Statement	*If the diagnostic test is positive there are limited effective treatment options available for the patient*
Total responses	219
'Don't know' response	7
Less likely to refer (%) coded as 1 or 2	92 (42.0%)
No more or less likely to refer (%) coded as 3	92 (42.0%)
More likely to refer (%) coded as 4 or 5	35 (16.0%)
Mean of responses	2.72
Median of responses	3
Ratio of 'less likely to refer' to 'more likely to refer'	2.63 : 1
McNemar test result (p value)	24.69 (p<0.0001)

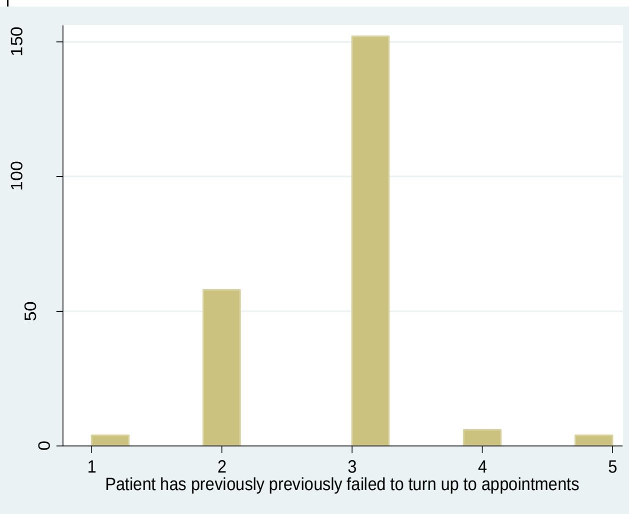


A substantial number of GPs stated that their referral behaviour would not be affected by the patient reporting difficulty taking time off work (78.8%), having previously failed to attend primary or secondary care appointments (68.4%), failing to follow medical advice in the past (74.8%), the consultation taking place via an interpreter (71.0%), the patient not expecting the diagnostic test to be accurate (79.2%), the GP being aware of the cost of the diagnostic test being considered (65.8%), or if the patient would have to wait a long time for the referral or diagnostic test (67.6%). However for all these factors where GPs did report that their referral behaviour was affected by the factor, it was that GPs were significantly less likely to refer patients for simple investigation and/or to secondary care.

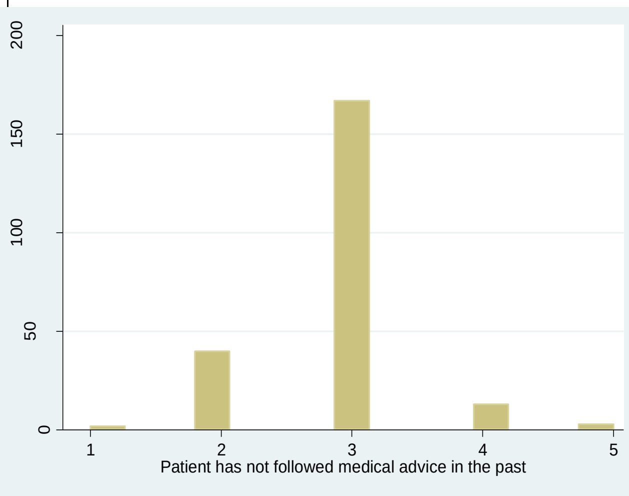
Statement	*The patient reports difficulty taking time off work for an appointment/ diagnostic test*
Total responses	226
'Don't know' response	1
Less likely to refer (%) coded as 1 or 2	40 (17.7%)
No more or less likely to refer (%) coded as 3	178 (78.8%)
More likely to refer (%) coded as 4 or 5	8 (3.5%)
Mean of responses	2.84
Median of responses	3
Ratio of 'less likely to refer' to more likely to refer'	5.00 : 1
McNemar test result (p value)	20.02 (p<0.0001)



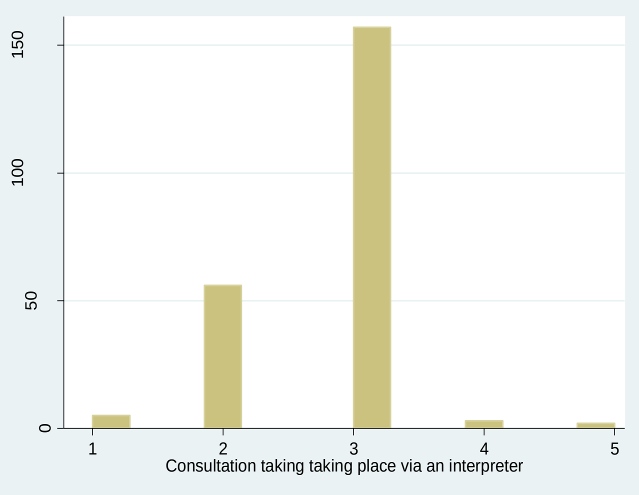
Statement	*The patient has previously failed to turn up to primary or secondary care appointments*
Total responses	225
'Don't know' response	1
Less likely to refer (%) coded as 1 or 2	62 (27.6%)
No more or less likely to refer (%) coded as 3	154 (68.4%)
More likely to refer (%) coded as 4 or 5	9 (4.0%)
Mean of responses	2.77
Median of responses	3
Ratio of 'less likely to refer' to 'more likely to refer'	6.89 : 1
McNemar test result (p value)	38.08 (p<0.0001)



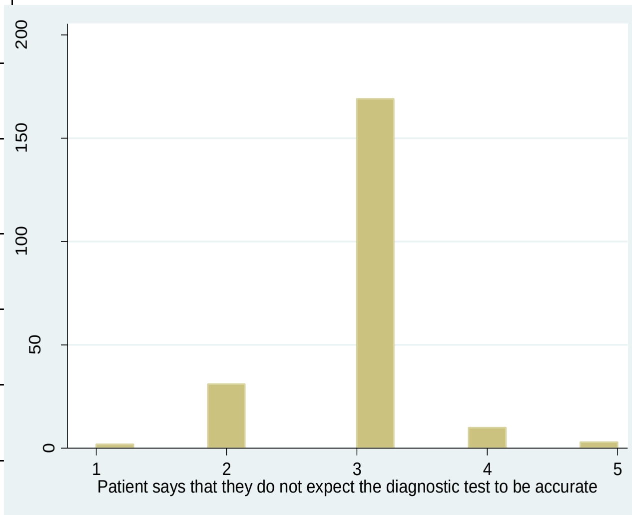
Statement	*The patient has not followed medical advice in the past (e.g. did not take medication as prescribed)*
Total responses	226
'Don't know' response	1
Less likely to refer (%) coded as 1 or 2	42 (18.6%)
No more or less likely to refer (%) coded as 3	169 (74.8%)
More likely to refer (%) coded as 4 or 5	15 (6.6%)
Mean of responses	2.89
Median of responses	3
Ratio of 'less likely to refer' to 'more likely to refer'	2.80 : 1
McNemar test result (p value)	11.86 (p=0.0006)



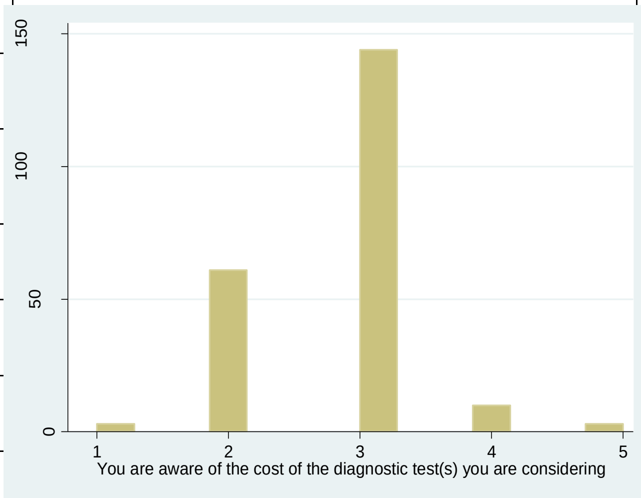
Statement	*The consultation is taking place via an interpreter*
Total responses	224
'Don't know' response	3
Less likely to refer (%) coded as 1 or 2	60 (26.8%)
No more or less likely to refer (%) coded as 3	159 (71.0%)
More likely to refer (%) coded as 4 or 5	5 (2.2%)
Mean of responses	2.74
Median of responses	3
Ratio of 'less likely to refer' to 'more likely to refer'	12.00 : 1
McNemar test result (p value)	44.86 (p<0.0001)



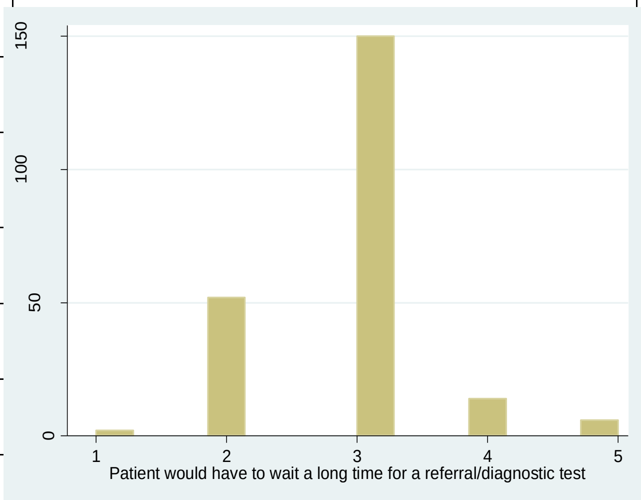
Statement	*The patient says that they do not expect the diagnostic test to be accurate*
Total responses	216
'Don't know' response	9
Less likely to refer (%) coded as 1 or 2	33 (15.3%)
No more or less likely to refer (%) coded as 3	171 (79.2%)
More likely to refer (%) coded as 4 or 5	12 (5.6%)
Mean of responses	2.91
Median of responses	3
Ratio of 'less likely to refer' to 'more likely to refer'	2.75 : 1
McNemar test result (p value)	8.89 (p=0.0029)



Statement	You are aware of the cost of the diagnostic test(s) you are considering
Total responses	222
'Don't know' response	4
Less likely to refer (%) coded as 1 or 2	64 (28.8%)
No more or less likely to refer (%) coded as 3	146 (65.8%)
More likely to refer (%) coded as 4 or 5	12 (5.4%)
Mean of responses	2.77
Median of responses	3
Ratio of 'less likely to refer' to 'more likely to refer'	5.33 : 1
McNemar test result (p value)	34.22 (p<0.0001)

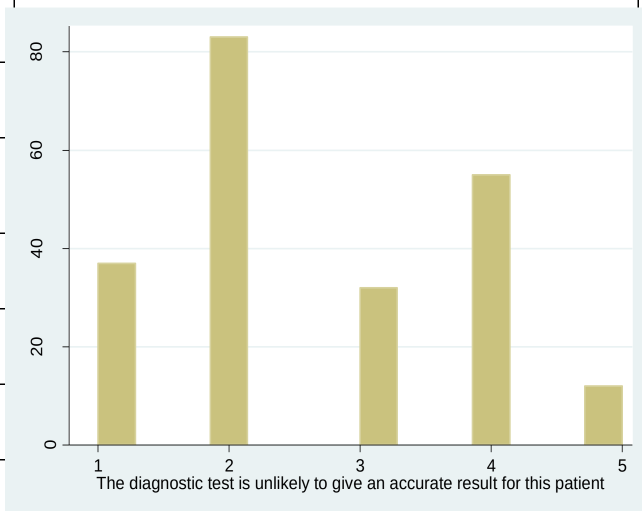


Statement	The patient would have to wait a long time for a referral/ diagnostic test
Total responses	225
'Don't know' response	2
Less likely to refer (%) coded as 1 or 2	54 (24.0%)
No more or less likely to refer (%) coded as 3	152 (67.6%)
More likely to refer (%) coded as 4 or 5	19 (8.4%)
Mean of responses	2.87
Median of responses	3
Ratio of 'less likely to refer' to 'more likely to refer'	2.84 : 1
McNemar test result (p value)	15.84 (p<0.0001)



As shown below, GPs gave a wide spread of responses to the question of whether a diagnostic test being unlikely to give an accurate result would make them more or less likely to refer a patient for simple investigation and/or to secondary care (55.0% stated they would be less likely to refer, 14.5% no more or less likely, and 30.5% more likely to refer). McNemar's test showed overall this factor had a significant negative influence on GPs' referral behaviour: overall GPs were less likely to refer patients in this case.

Statement	*The diagnostic test is unlikely to give an accurate result for this patient*
Total responses	220
'Don't know' response	6
Less likely to refer (%) coded as 1 or 2	121 (55.0%)
No more or less likely to refer (%) coded as 3	32 (14.5%)
More likely to refer (%) coded as 4 or 5	67 (30.5%)
Mean of responses	2.64
Median of responses	2
Ratio of 'less likely to refer' to 'more likely to refer'	1.81 : 1
McNemar test result **(p value)**	14.94 (p=0.0001)



6.3.2.3 : Factors where there was no consensus in the direction that GPs' referral was influenced

For three factors there was substantial variation between GPs as to how that factor would affect their decision to refer a patient for simple investigation or to secondary care. These factors had a wide spread of responses, however McNemar's test results showed that the factors had no significant influence in either the 'more likely to refer' or 'less likely to refer' direction.

Factors that made some GPs more likely to refer a patient, but others less likely to refer them, with no significant difference between the two directions of referral are:

- you know the patient well and are familiar with their past medical history
  26.8% of GPs reported that this factor would make them less likely to refer a patient, 27.3% that it would make them more likely to refer (p=1.0000)
- the patient appears anxious about the referral/diagnostic test
  17.1% of GPs reported that this factor would make them less likely to refer a patient, 18.5% that it would make them more likely to refer (p=0.8231)
- the patient is unwilling to discuss certain differential diagnoses
  9.5% of GPs reported that this factor would make them less likely to refer a patient, 14.5% that it would make them more likely to refer (p=0.1692)

6.3.2.4 : GPs' free text responses

All GPs were given the option to respond in free text to the question: *'If you have any further comments or reflections you wish to add on how you make decisions about sending patients for diagnostic tests, or referring them to secondary care, please type them in the box below'.*

GPs used the free text to comment on several aspects of the study, both reflecting on how they make decisions in real life, and also on their experience participating in the study. 108 of 227 GPs (47.6%) entered comments in response to this question; an additional six GPs gave free text responses indicating that they had no comments (e.g. 'none' or 'n/a').

The subjects of GPs' responses to this question can be divided into the following three categories, each of which I will discuss in turn:

- GPs' comments about the virtual patient application (I discussed these in Section 5.1.3.3, where I evaluated the GP decision making study);
- GPs' reflections on factors influencing their decisions to refer real life patients;
- GPs' thoughts about organisational set ups that might influence their referral behaviour.

6.3.2.4.1 : GPs' reflections on factors influencing their decisions to refer real life patients

GPs predominantly used the free text section to comment on factors they felt influenced their management decisions: 99 GPs (43.6%).

Some common themes of potential influences on their decision making (from GPs' free text) were:

- instinct/gut feeling that something was serious (their own concern or a patient's concern);
- experience: either recent or significant experiences with patients, or several years working as a GP;
- attitude to risk/the extent to which they felt the need to practise 'defensive medicine';
- patient anxiety or reassurance, particularly if this could be alleviated by 'simple tests' such as blood tests or chest X-ray.

GPs had differing attitudes as to how 'correct' it actually is for them to be influenced by some of these things. For example some GPs noted that if a patient was anxious/demanding they were keen to perform tests as it offered reassurance, whereas others referred to this as 'caving in' to the patient. Likewise some GPs saw conducting multiple diagnostic tests as a negative (e.g. commenting they ought to have less of a 'scattergun approach') whilst other saw it as an approach that allowed them to rule out several potential diagnoses.

GPs also noted:

- the importance of seeking advice from other sources: particularly colleagues, and ideally hospital colleagues (though it appears accessing hospital specialists was difficult and it seems this may be a barrier for some GPs);
- the importance of reviewing a patient (either to allow time for symptoms to improve/exacerbate, or to discuss next steps after performing basic tests).

6.3.2.4.2 : GPs' thoughts about organisational set ups that might influence their referral behaviour

Four GPs (1.8%) used the free text section to share organisational set ups that they believed might influence their referral behaviour. Many of these related to improving the connection between primary and secondary care. GPs commented:

- *'a daily indicator of waiting times for investigations would be useful'*
- *'it is useful to be up to date on tests available, so helpful to discuss individual case with hospital colleague - or have general teaching session with specialist colleague, particularly to know local pathways available'*
- *'GP hotlines for hospital specialities would be useful - direct numbers that do not involve phoning switchboards, being transferred or waiting for bleeps to be answered'*
- *' I suspect I would be heavily influenced by peer review or being able to view my diagnostic usage against peers'*

These suggestions give further, useful insight into some of the factors that may influence and underlie GPs' referral behaviour.

6.3.3 : Individual GPs' personal characteristics

We also used the post-consultation survey to collect further information about the personal characteristics of the 227 GPs who completed the GP decision making study.

6.3.3.1 : Clinical specialty experience

Speciality	Responded	No experience	F1/SHO	Specialist
Cardiology	215	74 (34.4%)	121 (56.3%)	20 (9.3%)
Emergency medicine	223	35 (15.7%)	131 (58.7%)	57 (25.6%)
Geriatrics	220	37 (16.8%)	122 (55.5%)	61 (27.8%)
Oncology	195	132 (67.7%)	55 (28.2%)	8 (4.1%)
Psychiatry	211	78 (37.0%)	85 (40.3%)	48 (22.7%)
Respiratory medicine	214	66 (30.8%)	119 (55.6%)	29 (13.6%)

The number of GPs who provided information about their speciality experience varied between specialties. GPs most commonly had experience of emergency medicine (84.3%) and geriatrics (83.2%). Many had also had some cardiology, psychiatry and respiratory medicine experience; however experience working in oncology was less common (just 32.3%).

6.3.3.2 : Financial responsibility

Budgetary responsibility in the:	Responded	Yes	No
Practice	226	104 (46.0%)	122 (54.0%)
As part of a clinical commissioning group (CCG)	225	73 (32.4%)	152 (67.6%)

Nearly half of the GPs who completed the GP decision making study had budgetary responsibilities within their practice (46.0%). Fewer GPs had budgetary responsibility as part of a clinical commissioning group (32.4%). Overall 122 of the 227 GPs (53.7%) reported having some financial responsibility.

6.3.3.3 : Smoking status

Smoking status (n=225, 2 no response)	**Number of GPs**
Never smoked	203 (90.2%)
Ex-smoker	22 (9.8%)
Current smoker	0 (0.0%)

The vast majority of GPs (90.2% of those who provided a response) reported that they had never smoked. The remaining 9.8% stated that they were ex-smokers; there were no current smokers reported.

6.4 : Discussion

6.4.1 : Main findings

The analysis of the post-consultation survey has identified a number of factors that appear to significantly impact on many GPs' real life referral behaviour.

The factors most commonly cited by GPs as increasing the likelihood of referral are when a patient's lifestyle puts them at increased risk of disease, or where there are challenges in communication with and/or understanding of a patient.

The factor most commonly cited as decreasing the likelihood of referral is when a patient frequently attends with non-serious complaints. In addition, based both on GPs' survey responses and their free text comments, it appears that the availability of input from secondary care, such as whether they could receive prompt advice from a hospital colleague, can also decrease the likelihood of GPs' referring a patient.

It is possible that some of the factors identified by the post-consultation survey as having an impact on some GPs real life referral behaviour may contribute to the non-clinical variations in GPs' referral for investigations or to secondary care that my systematic review identified; however we are not able to conclude that from this study.

Nearly half of the GPs reported that in real life situations (similar to those in the vignettes) they would not use external sources of information. Where GPs did report that they would seek additional information more of them stated that they would consult a colleague than refer to guidelines.

6.4.2 : Possible explanations for these findings and comparisons with other studies

6.4.2.1 : Sources of information GPs use in their decision making

Only a minority of GPs reported referring to additional sources of information whilst completing the vignette study. This is perhaps not unexpected given the artificial nature of the vignette 'consultations': whilst we asked GPs to imagine that their vignette 'consultations' were real life consultations, GPs knew that their management decisions in the study would not ultimately affect a real life patient.

Perhaps more of note however is that GPs' responses to the post-consultation survey suggested that less than a third of GPs would use guidelines (including NICE guidelines) in a real life consultation with a patient with similar symptoms to those in the vignettes. The percentage of GPs who reported that they would use guidelines is small, which is particularly interesting given the tendency in the vignette study for 'patients' with two chest symptoms of lung cancer to be more likely to be referred for chest X-ray than those 'patients' with one chest symptom and one non-specific symptom, despite both the 2005 and 2015 NICE guidelines for lung cancer not making any distinction between the importance of investigating either chest or non-specific symptoms.[35;152]

Variable use of, and poor compliance with, clinical guidelines is a long-standing and well-known issue.[168-170] A UK example considering GPs' use of NICE guidelines is a 2015 survey by the online GP website Pulse of 515 English and Welsh GPs, which found that whilst 76% of GPs stated that NICE guidelines were relevant to their practice, 39% reported going against their recommendations at least once a week.[171]

The results of the post-consultation survey alone do not enable me to conclude whether GPs' reported use of guidelines in their decision making is reflected in a difference in their referral behaviour. It was also not possible to determine from the survey why a large percentage of GPs would not use guidelines; it could reflect the content of guidelines being well known to many GPs so that they do not feel the need to consult them; alternatively it may indicate that GPs do not value guidelines as a key part of their decision making process, that they disagree with them or see them as restrictive, or that it does not occur to them to use them.

Where GPs reported consulting additional sources of information in their day-to-day practice, there was very little difference in the number who reported looking at books or websites, using guidelines, or seeking advice from a colleague. Over 40% of GPs reported that they did not use additional sources of information in their day-to-day practice, suggesting that these GPs are relying solely on internal factors (e.g. their knowledge, experience or gut instinct) and information supplied by, or about, the patient to make their management decisions.

6.4.2.2 : Influences on GPs' referral behaviour

I designed the post-consultation survey to identify factors that GPs believe influence their decisions to refer real life patients for simple investigations or to secondary care. As described in Section 6.2.3, I based the content of the survey on factors hypothesised (both in the literature and anecdotally) as potential influences on GPs' referral behaviour, and which may therefore also contribute to variations in referral. GPs' responses to the post-consultation survey suggest that a number of these hypothesised factors are likely to affect the referral behaviours of many GPs, while also suggesting that, by contrast, other factors have either little or no influence at all. In the rest of this section I will discuss the potential influences on GPs' referral behaviour that are suggested by GPs' responses in the post-consultation survey.

6.4.2.2.1 : Influences that increase the likelihood of referral

Patients' lifestyle risk

Most GPs stated that they would be more likely to refer a patient whose lifestyle puts them at increased risk of serious disease. This is particularly relevant in the context of lung cancer, the focus of the GP decision making study, since smoking is a lifestyle factor known to increase the risk of this disease.[172]

Challenges in communication or understanding

Many GPs stated that they would be more likely to refer a patient who has difficulty either expressing their symptoms clearly to the doctor, or recognising their potential severity. Communication difficulties due to language barriers would also make some GPs who completed the post-consultation survey more likely to refer a patient, although most of them stated that a patient having a low level of spoken English would not affect their referral decisions.

The findings support the hypotheses of a number of previous studies that these factors might influence GPs' referral behaviour,[123;124;173;174] and could reflect GPs erring on the side of caution by referring if comprehension or communication difficulties mean that they are not able to obtain the information they require to make a management decision. However the evidence is by no means conclusive; the National Audit of Cancer Diagnosis in Primary Care (2011)[47] found patients with communication difficulties were more likely to have a longer primary care interval (i.e. a delay in diagnosis).

Patients' assertiveness and health awareness

Many GPs reported that they would be significantly more likely to refer patients who had researched their symptoms; in addition GPs' free text comments suggested that some would also be more likely to refer a patient who had specifically requested a referral during the consultation. GPs also stated that they were more likely to refer patients who had difficulty weighing up potential management options, but that a patient's lack of awareness of management options or the services available would rarely influence their referral behaviour.

These findings support hypotheses in the literature that both patients' assertiveness,[174] and their research prior to attending an appointment,[124] might influence GPs' referral decisions; however they do not support the hypothesis that patients' knowledge of the services available influences GPs' decision to refer.[173;174]

The findings of the post-consultation survey suggest that whilst low health knowledge does not appear to make patients less likely to be referred, patients taking a high level of interest in their care, or who are particularly assertive, may be more likely to be referred for investigations and to secondary care. While this is in keeping with the current emphasis on moving towards shared decision making between GPs and their patients, in the case of those patients who do not have the health awareness to process all potentially significant information some GPs may still prefer to take a traditional paternalistic role and make referral decisions on behalf of these patients.

Patients' caregiving responsibilities

While most GPs stated that patients' having responsibilities as a caregiver did not influence their referral behaviour, where GPs were influenced they were more likely to refer a patient who has a responsibility to provide care for someone else. These findings are in line both with hypotheses in the literature,[78;124] and anecdotal comments from GPs, suggesting that patients who provide care for others often have a higher threshold for symptoms before attending the GP, and also that for these patients caution and early investigation is often important as they need to be healthy in order to care for others.

GPs' appointments running late

A number of GPs reported that they would be more likely to refer patients if their appointments were running late, providing some evidence for the hypothesis that GPs being overburdened may influence their referral behaviour.[175]

6.4.2.2.2 : Influences that decrease the likelihood of referral

Patients' previous poor or unnecessary engagement with health services

The majority of GPs stated that they would be less likely to refer patients who frequently attend with non-serious complaints. Some GPs were also less likely to refer a patient who has previously failed to attend primary or secondary care appointments, although most GPs stated that this would not affect their referral decisions. These results suggest that GPs are mindful both of thresholds for referral and not wasting resources.

Patients' poor mobility

Many GPs reported that a patient having poor mobility would decrease the likelihood that they would refer that patient for investigations or to secondary care. This is in line with anecdotal evidence from GPs, and may be a contributing factor as to why older patients are less likely to be referred (a finding in both my systematic review and the vignette study).

Organisational factors

The majority of GPs reported that they were less likely to refer patients when they could receive prompt advice from a hospital colleague, and this was also mentioned by GPs in the free text section as an important influence on referral behaviour. Very few of the papers I critically appraised in my systematic review considered the impact of organisational factors on GP decision making; in the few studies I have seen that considered access to secondary care, it was usually in the context of its distance from the GP practice.[112;115;121] However GPs' response to this statement and their free text comments appear to suggest that perhaps the key issue is not ease of access to sites, but rather to specific professionals and their expertise. This is supported by evidence of the benefits of advice and outreach from secondary care to primary care in the literature.[176;177]

Most GPs stated that their referral behaviour was not affected by their awareness of either the cost of an investigation or a long waiting list; however those who were influenced by these factors were significantly less likely to refer a patient. This suggests that both cost and waiting time could act as prompts to influence some GPs' referral behaviour, if GPs are mindful of them. Therefore as one GP commented in the free text response section, making GPs aware of both the cost and waiting list time for a number of commonly ordered investigations could influence their real life decision making.

Patients' work commitments

Most GPs stated that a patient's work commitments would not influence their referral behaviour. Where it did, GPs were less likely to refer a patient who was unable or unwilling to attend investigations or secondary care appointments due to difficulties in taking time off work.

It has been hypothesised that GPs' referral behaviour could be influenced by patients being unable or unwilling to attend investigations or secondary care appointments due to difficulties in taking time off work.[128] The findings of the post-consultation survey suggest that this could be an influence for some GPs.

Consultation requiring an interpreter

Whilst a patient's low level of spoken English made some GPs more likely to refer, most GPs reported that a language barrier requiring an interpreter in the consultation would not influence their referral decisions - and where it did they were generally less likely to refer the patient. This is despite the fact that it has been reported that the need for an interpreter in a consultation can lead to difficulty in the patient being able to express their symptoms clearly and accurately to the GP.[123;165;174]

6.4.2.2.3 : Issues that have a variable influence on the likelihood of referral

GPs' knowledge of patients

GPs varied widely in their responses regarding the extent to which prior knowledge of patients impacted on their referral behaviour, with no significant difference between the number of GPs more likely to refer them and those less likely to refer them.

It has been proposed, both in the literature and anecdotally by practising GPs,[123;130] that GPs' prior knowledge of patients' medical history and personality could influence their referral decisions. The findings of the post-consultation survey support this: the majority of GPs reported that it did influence their referral behaviour. The variable direction of influence could reflect that a GP's referral decision for a patient is likely to differ depending on what their prior knowledge of that patient is; for example, GPs may be aware if their patient has a tendency towards being 'worried well' and not necessarily need investigations or referral, or, alternatively, they may be aware that a patient's medical history puts them at increased risk of serious disease or makes the potential consequences of disease more significant. In some qualitative research studies GPs have commented on the importance to them of continuity of care within general practice,[178;179] and their belief that it 'allow[s] more effective and efficient diagnosis and management of problems presented'.[178]

Patients' concerns

The majority of GPs in this study reported that their referral behaviour was not influenced by a patient's concerns about investigation, treatment, particular diagnoses, or their utilisation of health services. However a number of GPs stated that a patient's anxiety about a referral or a patient's unwillingness to discuss certain differential diagnoses would impact their referral decisions, although with no clear direction of influence.

Where a patient's concerns do influence GPs' referral decisions, how they do so is likely to be situation and individual specific. It is apparent from GPs' free text responses in the post-consultation survey that attitudes to handling patient anxiety varied significantly between GPs: some saw it as important to minimise patient anxiety as far as possible and stated that they often conducted simple investigations to provide the reassurance of negative results, whilst others saw altering their intended behaviour in response to patient anxiety as a weakness.

GPs' opinions about the value of investigation

GPs' responses regarding the influence on their referral behaviour of being unclear what test would be most appropriate to diagnose a patient, or of a diagnostic test being unlikely to give an accurate response for a patient, varied greatly.

It has been hypothesised in the literature that GPs' referral for particular investigations may be influenced by their being unclear which diagnostic test is most appropriate, or awareness that a diagnostic test is unlikely to give an accurate response (e.g. older patients and women being less likely to have exercise testing).[113;130] However the variation in GPs' responses for these factors likely reflects a lack of clarity in my questions, or my phrasing not being specific enough, and therefore we cannot evaluate whether this hypothesis is likely to be true.

6.4.2.2.4 : Issues that do not influence most GPs referral decisions

In addition to the influences on GPs' referral behaviour discussed in Sections 6.4.2.2.1 to 6.4.2.2.3, there were also a number of factors in the post-consultation survey that a considerable majority of GPs (over 80% in each case) reported would not influence their referral behaviour. These include patients' concerns about stigma or overusing the health service, patients' lack of awareness of services available to them, or patients' not asking GPs about other management options. Most GPs reported that patients' transport difficulties or concerns about the costs of getting to appointments would not affect their referral decisions, despite hypotheses in the literature that these factors could be an influence.[92;165] The majority of GPs in the post-consultation survey were also unlikely to be influenced either by a patient requiring an interpreter for the investigation or appointment for which they were being referred, or by a patient has not following preventative advice in the past.

6.4.2.2.5 : Summary

The findings of the post-consultation survey have identified a number of factors that significantly influence GPs' referral behaviour, and have provided evidence to support some, and contradict others, of the hypotheses in the literature about influences on GPs' decision making.

These findings indicate that not all GPs are influenced equally by each of these factors. However the data from the post-consultation survey alone does not enable me to quantify the extent of their impact on GPs' referral behaviour, nor whether any of these influences on GPs' referral behaviour might contribute to non-clinical variations in GPs' decision making. In Chapter 7 I bring together data from both the vignette study and the post-consultation survey in order to start to address this.

6.4.3 : Strengths and limitations of the post-consultation survey

6.4.3.1 : Strengths

A number of studies have proposed factors which could influence GPs' decision making and thus contribute to variations by non-clinical characteristics in GPs' referral of patients for investigations or to secondary care. In this study, using the post-consultation survey, I set out to examine the extent to which GPs believed that these factors (and additional factors suggested anecdotally by GPs) influence their real life referral behaviour. The results of the post-consultation survey increase our understanding of some of the factors likely to underlie why GPs make the decisions that they do, and suggest directions for future research.

6.4.3.2 : Limitations

6.4.3.2.1 : GPs' responses could be subject to bias and/or their unawareness of their influences

The most significant limitation of the post-consultation survey is that whilst we are seeking to understand GPs' real life behaviour and factors that might influence their decisions, we cannot know whether their responses in the survey reflect their true behaviour. For example there was a potential for a desirability bias in GPs' responses, although we did repeatedly try to reinforce the point that the survey was not a test, and that we were interested in GPs' real life behaviour. There was also a potential for a form of memory bias: GPs were asked whether factors influenced their decision making or if they used additional sources of information in their day-to-day practice over the last month, which might have been challenging for them to assess outside of the situation, and to do so retrospectively. It is also possible that GPs were not conscious of some factors that influence their decision making, and therefore unable to accurately report all influences on their behaviour.

6.4.3.2.2 : GPs may be subject to different influences depending on the management decision in question

In addition, as both my systematic review and the vignette study have shown, there appears to be considerable heterogeneity in the variation of GPs' referral decisions by non-clinical characteristics, dependent on the characteristics, symptoms or outcome measure in question. Whilst we asked GPs to respond to the survey by considering how they would manage patients similar to those in the vignette study, and specified an interest in their referral for simple investigations or to secondary care, in hindsight this still covered a wide range of specific management decisions and it is possible that factors influencing GPs' behaviour vary further within these.

6.4.3.2.3 : Practical constraints of the survey

One of the significant constraints of the post-consultation survey was ensuring that it could be completed by GPs in about 5 minutes. This limited the content and complexity of the survey; a longer and more extensive survey asking GPs about factors influencing their real life decision making behaviour could provide more detailed information. The post-consultation survey focused on breadth, asking GPs for a quick reflection on whether any of a wide range of factors affected their referral decisions. Now that my research has identified some potentially significant influences on GPs' referral behaviour, an alternative strategy for future research could be to focus on a just a few of these factors in more detail, for example asking GPs the extent to which each factor would make them more or less likely to refer patients with particular symptoms or for specific investigations.

6.4.3.2.4 : Lack of clarity in certain questions

There were a few questions in the post-consultation survey where GPs' spread of responses suggests that the question was not completely clear to all GPs. If I were to repeat the post-consultation survey I would seek to make these questions clearer in order to enable me to better examine influences on GPs' referral behaviour; a particular example is the question asking whether GPs would be more or less likely to refer patients if there was a possibility that the diagnostic test might be inaccurate for certain patients.

6.4.3.2.5 : This study's sample might not be representative of all GPs

The post-consultation survey examined the extent to which GPs believe certain factors influence their real life referral decisions. The actions, beliefs and thought processes of GPs who choose to participate in research may not reflect those of all GPs. Since it is hard to eliminate this potential source of bias, it is important to be aware of it when considering how far the results of this study can be generalised.

6.4.4 : Implications for future research, policy and practice

6.4.4.1 : Future research

The analysis of the post-consultation survey identified several factors that impact on many GPs' real life referral behaviour. It is possible that these factors may contribute to some of the non-clinical variations seen in GPs' decisions to refer patients for investigations, including diagnostic tests such as chest X-ray, or to secondary care. Further research should aim to quantify the extent to which these factors influence both GPs' decision making and their referral behaviour. This will increase understanding of why GPs make the referral decisions that they do, and help to determine whether the non-clinical variation in referral that occurs is intentional, appropriate and in patients' best interests. Future research may need to focus on examining factors that influence GPs' referral behaviour for patients with particular symptoms for specific investigations; with those factors identified by this study as potentially significant influences on GPs' referral behaviour providing a place to start.

The post-consultation survey has also raised interesting questions about GPs' use of guidelines in their decision making, and the importance of these guidelines. This survey highlights the need for further research into what determines when and why GPs consult guidelines, the extent to which they follow the recommendations of guidelines, and whether GPs who refer to guidelines refer differently to those who do not.

6.4.4.2 : Policy and practice

This study has provided some interesting data on sources of information that GPs use and/or value having available to aid their decision making - in particular identifying that only a small percentage of GPs state that they would use NICE guidelines in their decisions.

It has also highlighted that accessing potentially valuable sources of information (in particular seeking advice from hospital colleagues) is often difficult for GPs. Many GPs stated that they would be less likely to refer patients if they were able to discuss the case with a hospital colleague, suggesting that strategies to improve communication between primary and secondary care could have significant implications for improving the efficiency of GPs' referral.

7 : How GPs' personal characteristics and attitudes to referral related to their behaviour in the vignette study - *combining data from both parts of the GP decision making study (Study 2a and Study 2b)*

7.1 : Introduction

From the descriptive results of the vignette study (Study 2a), reported in Section 5.1.2.2, it is clear that the referral behaviour of the GPs who participated in the GP decision making study (Study 2) varied widely between individual GPs. The second part of the GP decision making study, the post-consultation survey (Study 2b), identified several factors that GPs believe influence their referral decisions, but did not enable me to quantify the extent of their impact on GPs' referral behaviour.

These observations raised the question of whether the differing referral patterns seen in the vignette study (Study 2a) were matched by differences in GPs' responses in the post-consultation survey (Study 2b). I therefore set out to examine this.

In addition, my systematic review (Study 1) identified a specific gap in the literature: the need for examination of the association between individual GPs' personal characteristics and their referral of patients for investigations (including diagnostic tests) and/or to secondary care. I therefore also sought to examine the extent to which the differing referral patterns seen in the vignette study were associated with the personal characteristics of the GP participants.

In this chapter I outline the methods I used to examine the extent to which GPs' referral behaviour in the vignette study (Study 2a) related to both the responses they gave in the post-consultation survey (Study 2b) and their personal characteristics. I then go on to report and discuss my findings.

7.1.1 : Aim

To examine the extent to which GPs' referral behaviour in the vignette study related to both their personal characteristics and to the factors that they reported in the post-consultation survey would influence their real life referral decisions.

7.1.2 : Objectives

To use simple and multiple logistic regression to construct a series of multivariate models to evaluate whether:

- especially highly referring GPs in the vignette study differed from all other GPs in the study;
- GPs who referred very few 'patients' in the vignette study differed from all other GPs in the study;
- GPs whose referral of 'patients' in the vignette study was fully adherent to the NICE guidelines' recommendations differed from those who referral was not fully adherent.

7.2 : Methods

7.2.1 : Determining the categories to examine

In order to examine whether the differing referral patterns seen in the vignette study were matched by differences in either GPs' responses in the post-consultation survey, or their personal characteristics, I first had to determine how to classify the differing referral patterns seen. The following observations contributed to this decision.

220 GPs recorded a management decision for each of the six different 'patient profiles'. Whilst the mean number of 'patients' referred by GPs was 4.4 (to one decimal place), and the median and mode are both 5, individual GPs' referral behaviour varied enormously, ranging between just one and all six of the 'patients' seen being referred. I therefore decided to evaluate both whether especially high referring GPs (those who referred all six 'patients' in the vignette study) differed from the rest of the GPs in the study, and also whether GPs who were particularly low referrers in the vignette study (in comparison to the mean) differed from the rest of the GPs in the study.

GPs' decisions on whether to refer the two high risk 'patient profiles' (profiles 5 and 6) also varied more widely than one might have expected considering that both profiles were designed to clearly meet the NICE guidelines' criteria for referral for chest X-ray. Furthermore whilst only a minority of GPs reported in the post-consultation survey that they would use NICE guidelines in their referral decision making for real life patients presenting in a similar way to those in the vignette study, I could not conclude from the survey results alone whether GPs' reported use of guidelines was reflected in a difference in their referral behaviour. I therefore also decided to evaluate whether GPs who were adherent to the recommendations of the 2005 NICE guidelines for lung cancer differed from those who were not.

I therefore considered three different categories of GP identified by the vignette study. In consultation with my supervisors, and with consideration of the data and the relative sizes of the groups for each category, I chose to define and divide each category as follows:

a) whether **high referring** GPs, *who I classified as GPs referring all six 'patients' they saw for chest X-ray*, differed from all the other GPs (those who referred five or less 'patients' seen);

b) whether **low referring** GPs, *who I classified as GPs referring 50% or less of the six 'patients' they saw (i.e. three or less) for chest X-ray*, differed from all the other GPs;

c) whether GPs who were **adherent to NICE guidelines for the two 'patient profiles' categorised as high risk**, *who I classified as GPs referring both 'patient profile 5' and 'patient profile 6' for chest X-ray,* differed from GPs who were not adherent to the guidelines (and referred just one, or neither, of these 'patients').

7.2.2 : Data processing of GPs' referral patterns

For each of these three categories I assigned GPs a code based on their chest X-ray referral decisions for the six 'patients' they saw in the vignette study. GPs either met the criterion for a category, or they did not (and thus for that category were coded into one of two groups). I did this for all three categories, thus assigning each GP three codes: one code for whether they were a 'high referrer', a second code for whether they were a 'low referrer', and a third code for whether they were 'adherent to NICE guidelines'.

A small number of GPs could not be assigned into a group for one or more of the categories; where this occurred they were excluded from further analysis of that category and noted as missing data. Seven GPs were excluded from both the high referring and low referring categories: three of these GPs did not provide a management plan for all six 'patients' seen, whilst the other four had incorrect 'patient' allocations during the vignette study (as discussed in Section 5.1.3.2.2) and therefore did not view each of the six 'patient profiles'. Five GPs were excluded from the adherent to NICE guidelines category (which reflected their management of the two high risk 'patient profiles'): one GP did not provide a management plan for their profile 6 'patient', the other four GPs had incorrect 'patient' allocations and did not complete 'consultations' with both a profile 6 and a profile 5 'patient'.

7.2.3 : Analysis

Once GPs had been assigned to a group for each category (or excluded), I then proceeded to analyse the data: initially using simple logistic regression, and then by multiple regression. These analyses were conducted in Stata.[159] I considered each of the three categories in turn.

7.2.3.1 : Simple logistic regression

For each category, I performed a series of simple logistic regression calculations to compare the group meeting the criterion (e.g. high referring GPs) with the group containing the rest of the GPs. These calculations compared the information-seeking behaviour of GPs in the group, both in real life and in the vignette study (the variables I examined are listed in Table 19); GPs' personal characteristics (the variables I examined are listed in Table 20); and the extent to which they believed a series of different factors affected their real life referral behaviour (the variables I examined are listed in Table 21). All these logistic regression calculations were adjusted for GPs' age and gender.

Table 19 : A list of the variables related to GPs' information-seeking behaviour examined when conducting analysis to compare groups of GPs with different patterns of referral behaviour

Information-seeking behaviour (data source: GPs' responses to the post-consultation survey)
• Asked *colleague* for advice in vignette study • Would ask *colleague* for advice in real life
• Referred to *NICE guidelines* in vignette study • Would refer to *NICE guidelines* in real life
• Referred to *other (including local) guidelines* in vignette study • Would refer to *other (including local) guidelines* in real life
• Referred to *books or the internet* in vignette study • Would refer to *books or the internet* in real life
• *Did not use any of the above sources of information* in vignette study • *Would not use any of the above sources of information* in real life

Table 20 : A list of the variables related to GPs' personal characteristics examined when conducting analysis to compare groups of GPs with different patterns of referral behaviour, and where this information was collected from

GPs' personal characteristics examined	Source of this information
• Age • Gender • Ethnicity • Years since qualification • Type of GP (partner, locum etc) • Number of sessions worked per week • IT confidence • Month of registration for the study	Registration questionnaire
• Level of clinical specialty experience for: a) cardiology b) emergency medicine c) geriatrics d) oncology e) psychiatry f) respiratory • Budgetary responsibility: a) within their practice b) for the CCG • Smoking status	Post-consultation survey

Table 21 : A list of the variables related to GPs' personal characteristics examined when conducting analysis to compare groups of GPs with different patterns of referral behaviour

The extent to which [*factor listed below*] affects the GP's real life decision making (data source: GPs' responses to the post-consultation survey)
• The patient reports difficulty taking time off work for an appointment/diagnostic test
• The patient is a caregiver
• The patient's lifestyle puts them at higher risk of serious disease
• The GP knows the patient well and is familiar with their past medical history
• The patient frequently attends with non-serious complaints

The extent to which [*factor listed below*] affects the GP's real life decision making (data source: GPs' responses to the post-consultation survey)
• The patient has previously failed to turn up to primary or secondary care appointments
• The patient has not followed medical advice in the past (e.g. did not take medication as prescribed)
• The patient has not followed health promotion or disease prevention advice in the past (e.g. has not stopped smoking)
• The patient has a low level of spoken English
• The consultation is taking place via an interpreter
• The patient will require an interpreter for their appointment/diagnostic test
• The patient does not have a source of transport to or from the appointment/diagnostic test
• The patient's mobility is poor
• The patient is concerned it is expensive to travel to the appointment/diagnostic test
• The patient is unable to recognise the seriousness of their symptom(s)
• The patient does not express their symptom(s) clearly
• The GP is concerned the patient may have difficulties weighing up the consequences of different management options
• The patient does not ask about other management options available
• The patient has independently researched their symptom(s) before their consultation
• The patient does not know what services are available to them
• The patient does not appear distressed about their symptom(s)
• The patient appears anxious about the referral/diagnostic test
• The patient appears concerned about the stigma associated with certain differential diagnoses
• The patient is unwilling to discuss certain differential diagnoses
• The patient says that they do not expect the diagnostic test to be accurate
• The patient is concerned about overusing the health service
• It is not clear which test would be most appropriate to diagnose this patient's symptom(s)
• The diagnostic test is unlikely to give an accurate result for this patient
• If the diagnostic test is positive there are limited effective treatment options available for the patient
• The GP's appointments are running late
• The GP is aware of the cost of the diagnostic test(s) they are considering
• The patient would have to wait a long time for a referral/diagnostic test
• A hospital colleague is able to provide advice promptly by telephone or email

7.2.3.2 : Multiple regression

The next step was to perform a multiple regression for each category, again adjusting for GPs' age and gender.

In multiple regression with potentially large numbers of variables and a limited number of observations (in this case around 200), there is a need to keep the number of estimated parameters down to a manageable level. To help achieve this, for variables based on GPs' post-consultation survey responses about factors influencing their real life referral behaviour (e.g. whether a GP's appointments running late makes them less likely, no more or less likely, or more likely to refer a patient), I used the natural order of these outcomes to fit a single logistic regression trend across the three categories. Similarly, I fitted a single trend for the level of the GP’s IT confidence (which GPs rated using a five-point scale).

When selecting which variables to include in the multiple regression for each category I used a generous p value criterion: <0.1 instead of the usual <0.05. This was because there were several potentially correlated variables in the post-consultation survey, many pertaining to whether GPs felt certain factors influenced their real life referral behaviour, and therefore a likelihood of factors being confounded. Using the generous p value avoided overlooking any potentially important effects on GPs' referral behaviour in the vignette study which did not quite achieve significance due to confounding in this dataset. Thus inclusion of a variable in the multiple regression model was the starting point for an iterative process of reducing the number of variables until all those remaining had a p value of <0.05; where variables lost significance in multiple regression I carried out backward stepwise regression, eliminating the least significant variable first, and continued until all remaining variables had the desired p value of <0.05.

The result of this process was a multivariate model for each category in which all remaining variables had a p value of <0.05, adjusted for GP age and gender. Estimates of odds ratios and 95% confidence intervals are reported in Section 7.3.

7.3 : Results

I will report the results for each of the three categories I examined in turn:

- whether especially high referring GPs differed from all the other GPs;
- whether low referring GPs differed from all the other GPs;
- whether GPs who were fully adherent to the NICE guidelines in their management of the two high risk 'patient profiles' differed from those who were not.

Table 22 : Variables influencing a GP's probability of being a high referrer in the vignette study (adjusted for age and gender)

Variable		**Absolute number of GPs** *those included in the final model*		**Odds ratio (OR)** *reported to 2 decimal places*	**95% confidence intervals** *reported to 2 decimal places*	**P value** *reported to 3 decimal places*
		High referrers	**All others**			
Would ask a *colleague* for advice in real life	No	35	112	1.00	-	0.018
	Yes	7	63	0.31	0.12 - 0.82	
Would refer to *other (including local) guidelines* in real life	No	27	129	1.00	-	0.008
	Yes	15	46	3.15	1.35 - 7.35	
GP gender	Male	30	90	1.00	-	0.009
	Female	12	85	0.30	0.12 - 0.74	
GP's level of IT confidence *(reported on a scale of 1-5)*	1 (lowest)	0	1	0.48 *	0.28 - 0.83	0.008
	2	0	1			
	3	11	25			
	4	18	84			
	5 (highest)	13	63			
The extent to which GPs' real life decision to refer is influenced by whether 'the patient has a low level of spoken English'	Less likely to refer	2	4	0.45 *	0.23 - 0.87	0.019
	No more or less likely to refer	33	121			
	More likely to refer	7	50			

* Trend across categories

7.3.1 : Whether especially high referring GPs differed from all the other GPs

Data from 220 GPs were analysed. 42 GPs were classified as high referrers, and this group was compared to the other 178 GPs. Table 22 shows the variables which differed significantly ($p<0.05$) between high referring GPs and the other GPs in the GP decision making study.

High referring GPs were less likely to state that they would consult a colleague in real life situations similar to those in the vignette study (odds ratio, OR=0.31, p=0.018), but more likely to state that they would use non-NICE or local guidelines (OR=3.15, p=0.008). High referring GPs were also less likely to be female (OR=0.30, p=0.009) and, overall as a group, rated their IT confidence lower than the other GPs in the study (trend OR=0.48, p=0.008).

There was also a difference in the propensity to refer among those GPs who said that a patient having a low level of spoken English would affect their likelihood of referral. As discussed in Section 6.3.2.2.1, over two-thirds of GPs in the study (70.5%) stated that their decision to refer patients in real life would not be affected by the patient having a low level of spoken English. However, in those cases where GPs stated that it would affect their decision making, this was very significantly in the direction of GPs being more likely to refer. However, as can be seen in Table 22, high referring GPs were significantly less likely than the other GPs to state that a patient's low level of spoken English would increase their likelihood of making a referral in real life (trend OR=0.45, p=0.019).

Table 23 : Variables influencing a GP's probability of being a low referrer in the vignette study (adjusted for age and gender)

<table>
<tr><th rowspan="2" colspan="2">Variable</th><th colspan="2">Absolute number of GPs
those included in the final model</th><th rowspan="2">Odds ratio (OR)
reported to 2 decimal places</th><th rowspan="2">95% confidence intervals
reported to 2 decimal places</th><th rowspan="2">P value
reported to 3 decimal places</th></tr>
<tr><th>Low referrers</th><th>All others</th></tr>
<tr><td rowspan="5">Years since qualification as a doctor</td><td><2</td><td>7</td><td>20</td><td>10.66</td><td>1.75 - 64.89</td><td>0.010</td></tr>
<tr><td>2-5</td><td>14</td><td>28</td><td>13.98</td><td>2.86 - 68.30</td><td>0.001</td></tr>
<tr><td>5-10</td><td>10</td><td>27</td><td>5.32</td><td>1.20 - 23.51</td><td>0.028</td></tr>
<tr><td>10-20</td><td>7</td><td>48</td><td>1.00</td><td>-</td><td>-</td></tr>
<tr><td>>20</td><td>10</td><td>42</td><td>2.00</td><td>0.47 - 8.57</td><td>*0.351*</td></tr>
<tr><td rowspan="2">GP gender</td><td>Male</td><td>19</td><td>102</td><td>1.00</td><td>-</td><td rowspan="2"><0.001</td></tr>
<tr><td>Female</td><td>29</td><td>64</td><td>4.64</td><td>2.07 - 10.40</td></tr>
<tr><td rowspan="3">The extent to which GPs' real life decision to refer is influenced by whether 'the patient has previously failed to turn up to primary or secondary care appointments'</td><td>Less likely to refer</td><td>17</td><td>42</td><td rowspan="3">0.38 *</td><td rowspan="3">0.19 - 0.76</td><td rowspan="3">0.006</td></tr>
<tr><td>No more or less likely to refer</td><td>31</td><td>114</td></tr>
<tr><td>More likely to refer</td><td>0</td><td>9</td></tr>
<tr><td rowspan="3">The extent to which GPs' real life decision to refer is influenced by whether 'the diagnostic test is unlikely to give an accurate result for this patient'</td><td>Less likely to refer</td><td>32</td><td>85</td><td rowspan="3">0.54 *</td><td rowspan="3">0.37 - 0.78</td><td rowspan="3">0.001</td></tr>
<tr><td>No more or less likely to refer</td><td>8</td><td>23</td></tr>
<tr><td>More likely to refer</td><td>7</td><td>58</td></tr>
</table>

* Trend across categories

7.3.2 : Whether low referring GPs differed from all the other GPs

Data from 220 GPs were analysed. 49 GPs were classified as low referrers, and this group was compared to the other 171 GPs. Table 23 shows the variables which differed significantly between low referring GPs and the other GPs in the GP decision making study.

The length of time for which a GP had been qualified significantly affected their likelihood of being a low referrer in the vignette study (overall p=0.023). GPs were asked to state their time since qualification as less than 2 years, 2-5 years, 5-10 years, 10-20 years or more than 20 years. The likelihood of being a low referring GP was significantly higher for all three groups who had been qualified for less than 10 years compared to the baseline group of 10-20 years (qualified less than 2 years: OR=10.66, p=0.010; qualified 2-5 years: OR=13.98, p=0.001; qualified 5-10 years: OR=5.32, p=0.028). There was no significant difference in likelihood for the GPs who had been qualified for more than 20 years (p=0.351). Although variable, the general tendency is for those more recently qualified to be more likely to be low referrers.

Low referring GPs were also more likely to be female (OR=4.64, p<0.001).

A third area of difference between the low referring GPs and the population of GPs in the study as a whole was that about two-thirds of GPs in the study (68.4%) stated in the post-consultation survey that their real life decisions to refer patients would not be affected by whether that patient had previously failed to attend appointments. As discussed in Section 6.3.2.2.2, where it did affect GPs' real life referral behaviour it tended to decrease the likelihood of referral. By contrast GPs who were low referrers in the vignette study were significantly more likely to have also stated that failure to attend previous primary or secondary care appointments would make them less likely to refer a patient (trend OR=0.38, p=0.006).

A fourth area of difference between the low referring GPs and the rest of the GPs was the effect on referral of a diagnostic test being unlikely to give an accurate result for a patient (a question asked in the post-consultation survey). As I discussed in Section 6.3.2.2.2, very few GPs (14.5%) gave a neutral response to this question, and there was a significant trend that it would make GPs less likely to refer. Those who stated that they were more likely to refer despite the diagnostic test result being inaccurate were less likely to be low referrers (trend OR=0.58, p=0.001).

Table 24 : Variables influencing a GP's probability of fully adhering to NICE guidelines in the vignette study (adjusted for age and gender)

Variable		**Absolute number of GPs** *those included in the final model*		**Odds ratio (OR)** *reported to 2 decimal places*	**95% confidence intervals** *reported to 2 decimal places*	**P value** *reported to 3 decimal places*
		Fully adherent	**All others**			
GP's smoking status	Never smoked	108	85	1.00	-	0.039
	Ex-smoker	17	4	3.52	1.06 - 11.64	
	Current smoker	0	0	*Not included as no GPs stated they were current smokers*		
Respiratory experience	None	41	22	1.00	-	
	F1/SHO level (junior)	63	49	0.49	0.24 - 1.01	0.052
	Specialist	11	16	0.33	0.12 - 0.90	0.030
The extent to which GPs' real life decision to refer is influenced by whether '[the GP's] appointments are running late'	Less likely to refer	11	9	1.74 *	1.02 - 2.99	0.044
	No more or less likely to refer	84	67			
	More likely to refer	31	13			
The extent to which GPs' real life decision to refer is influenced by whether 'the diagnostic test is unlikely to give an accurate result for this patient'	Less likely to refer	60	57	1.58 *	1.18 - 2.11	0.002
	No more or less likely to refer	18	14			
	More likely to refer	48	17			

* Trend across categories

7.3.3 : Whether GPs who were fully adherent to the NICE guidelines differed from those who were not

Data from 222 GPs were analysed. 131 GPs were classified as fully adherent, whilst 91 GPs were not. Table 24 shows the variables which differed significantly between fully adherent and non fully adherent GPs in the study.

GPs who were adherent to NICE guidelines and referred both high risk profiles were more likely to be ex-smokers (no GPs reported that they were current smokers) than those who did not refer both high risk profiles (OR=3.52, p=0.039).

GPs who referred both high risk profiles were also less likely to have respiratory medicine experience (when GPs with no respiratory medicine experience were the comparison group). This likelihood was significant for GPs with specialist level respiratory experience (OR=0.33, p=0.030), and only just beyond the p<0.05 cut-off for significance for those whose highest level of experience was at a junior level (OR=0.49, p=0.052).

Another area of difference between those GPs who were fully adherent to NICE guidelines and those who were not was the impact of appointments running late on the likelihood of referral. As discussed in Section 6.3.2.2.1, whilst 70.1% of GPs in the study stated that their decision to refer patients in real life would not be affected by their appointments running late; in the cases where GPs stated this did affect their referral behaviour the trend was for it to increase the likelihood of referral. GPs who were fully adherent to NICE guidelines in the vignette study and referred both high risk profiles were significantly more likely than the other GPs to have stated that they would be more likely to refer a patient if their appointments were running late (trend OR=0.38, p=0.006).

GPs who were fully adherent and referred both high risk profiles were also more likely than the other GPs to state that they refer patients even when the diagnostic test is unlikely to give an accurate result (trend OR=1.74, p=0.044).

7.4 : Discussion

7.4.1 : Main findings

Combining data from the results of the two parts of the GP decision making study, the vignette study and the post-consultation survey, demonstrated that differences in GPs' referral behaviour in this study were associated with their personal characteristics and attitudes to referral.

GPs who were high referrers in the vignette study were more likely than other GPs to be male, and to rate their IT confidence more poorly. They were less likely to consult a colleague when making their referral decision, but more likely to consult other (i.e. non-NICE, and including local) guidelines. They were also less likely to report that their referral behaviour was influenced by a patient having a poor level of spoken English.

GPs who were low referrers in the vignette study were more likely than other GPs to be female, as well as to have qualified more recently. They were more likely to report that their referral behaviour was influenced by whether patients had previously failed to attend appointments, and less likely to report that they would refer a patient even if a diagnostic test was unlikely to give an accurate result.

GPs whose referral was fully adherent to NICE guidelines were more likely than other GPs to be ex-smokers, and less likely to have experience working in respiratory medicine. They were also more likely to report that their referral behaviour was influenced by their appointments running late, and more likely to report that they would refer a patient even if a diagnostic test was unlikely to give an accurate result.

It is unclear whether these are true effects that would be seen in real life - more research is needed. However it certainly seems likely that GP-related factors are associated with variation in referral behaviour.

7.4.2 : Possible explanations for these findings

7.4.2.1 : GPs' personal characteristics

7.4.2.1.1 : GP gender

In the GP decision making study GPs' gender had a significant effect on their referral behaviour, with male GPs more likely to have referred more of the 'patients' they saw for chest X-ray. It is possible that this reflects a true effect - for example female GPs were significantly more likely than male GPs to report that they would seek advice from colleagues in real life, a behaviour which was itself also associated with a lower likelihood of being a high referrer. However as discussed in Section 5.1.1.3.2, more men completed the GP decision making study than women (although this difference was not statistically significant); in addition five of the seven GPs excluded were female. Since the group size of both the high referrers and the low referrers is relatively small it does have to be considered that this apparent effect of gender might be exaggerated.

7.4.2.1.2 : GPs' IT confidence

High referring GPs were more likely to report a lower IT confidence. This was despite adjusting for age and gender (there was a strong correlation between gender and IT confidence, with women GPs' average reported IT confidence significantly lower than the average for men). It seems unlikely that IT confidence would have a strong effect on clinical judgment and decision making, but it is possible that this reflects differences in the way that GPs with different levels of IT confidence used the virtual patient application in this study.

The observed variation in GPs' referral behaviour by their IT confidence provides additional evidence (alongside GPs' completion of the vignette study, and their comments on using the virtual patient application) when evaluating the use of the application as a tool to examine GP decision making.

7.4.2.1.3 : Years since qualification

Low referrers were more likely to be GPs who had qualified within the last 10 years (and in particular those who had been qualified for less than 5 years). In recent years there has been growing recognition of the increasing demand facing the NHS and the need to control use of resources.[180;181] It is likely that more recently trained

GPs have been made more aware of this issue, and this may be reflected in their hesitancy to refer 'patients' in the vignette study.

7.4.2.1.4 : GPs' own smoking behaviour

There is evidence in the literature that GPs' own smoking behaviour may influence their attitudes towards, and decisions regarding, smoking cessation.[182] Only 22 GPs (9.8% of the 224 who supplied an answer) reported their smoking status as an ex-smoker, and no GPs reported being current smokers. However despite this small number there does appear to be a strong link between GPs' smoking status and their referral of the two high risk 'patient profiles', with ex-smokers being significantly more likely to refer both high risk 'patients' they saw compared to those GPs who had never smoked. This could reflect the fact that both the high risk profiles are smokers and that GPs who are ex-smokers take particular account of smokers' heightened risk of lung cancer, or alternatively it could be that GPs who are ex-smokers are more alert than non-smoking GPs to lung cancer and other lung disease as possible diagnoses.

7.4.2.1.5 : GPs' clinical experience

GPs who have had experience of respiratory medicine as a specialist were less likely to be adherent to NICE guidelines and refer the two high risk 'patient profiles'. This is surprising, as one might expect these GPs to be most familiar with the NICE guidelines, and the relative risk of lung cancer for certain symptoms and 'patient' presentations (the symptoms presented in the two high risk profiles were firstly chest pain combined with weight loss, and secondly increased breathlessness in a patient with COPD). It is possible that those with respiratory experience look for particular clinical signs or symptoms to determine how sick a patient is, and that things such as the normal chest examination in our 'virtual patients' affected their decision making. However it is a surprise that two 'patient profiles' both with PPVs of >3% (in one case 14%) were not both referred for chest X-ray by 41.0% of GPs and by an even higher percentage of those with respiratory medicine experience.

7.4.2.2 : GPs' information seeking behaviour

Whilst GPs' decision making for the 'virtual patients' in the vignette study cannot be assumed to exactly replicate their decision making patterns and referral behaviour in real life, it seems plausible to consider that GPs who were high referrers in the vignette study may also be high referrers in their day-to-day practice. The observation that GPs who were high referrers in the vignette study were less likely to report that they ask advice from colleagues in real life is therefore interesting. It suggests one of two possibilities. The first possibility is that some GPs may prefer to simply refer patients they are concerned or uncertain about, whilst others who have a higher threshold for referral may be more likely to discuss cases with their colleagues. The second possibility is that when GPs ask advice from colleagues about a patient they become less likely to refer; so by contrast those GPs who discuss less with colleagues will be relatively high referrers. It is also interesting that GPs who were high referrers in the vignette study were more likely to state that they use non-NICE/local guidelines in their real life decision making; this might indicate that many local guidelines have a relatively low threshold for referral.

7.4.2.3 : Factors that GPs believe influence their real life referral behaviour

As discussed in Section 6.2.2, in the post-consultation survey GPs were asked about the extent to which their decision to refer a patient for investigation or to secondary care in real life is influenced by a number of different factors. When I conducted simple logistic regression the significance of several of these factors appeared to differ within each of the three categories (of GPs' referral patterns in the vignette study) that I examined, but the majority were no longer significant when multiple regression was used. However, there were exceptions.

7.4.2.3.1 : The patient having a poor level of spoken English

In the post-consultation survey all GPs were asked whether they would be more or less likely to refer a real life patient who had a poor level of spoken English, a factor which previous studies have proposed could contribute to non-clinical variation in referral.[165;173;174] Whilst the majority of GPs stated it would have no impact on their referral behaviour, a significant proportion stated that they would be more likely to refer these patients. However GPs were less likely to state that they would be influenced in this way if they were also a high referrer in the vignette study.

Depending on the extent to which GPs' referral behaviour in the vignette study mirrors their real life referral behaviour, this may reflect the fact that these GPs have a lower threshold for referral overall, and so are less affected by situational factors. By contrast GPs who are less quick to refer patients overall, may be swayed more by specific factors. For example if a patient's level of spoken English is low the GP may be uncertain whether they have accurately elicited a complete and reliable history from the patient, and therefore less able to evaluate the level of risk and importance of referral, which in turn may lead some GPs to refer a patient 'just in case'.

7.4.2.3.2 : The patient having previously failed to attend appointments

Anecdotally, some GPs have suggested that patients' prior lack of attendance at primary or secondary care appointments might affect their decision making. In the post-consultation survey the majority of GPs stated that a patient's previous lack of attendance at primary or secondary care appointments would not affect their referral behaviour. However, as a group, those GPs who were classified as low referrers in the vignette study were significantly less likely to refer patients who had previously failed to attend appointments. This seems logical, as this group of GPs clearly has a much higher threshold for referral than the rest of the GPs in the study, and they may therefore be particularly keen to ensure any referral they make is likely to be taken seriously and the appointment kept.

7.4.2.3.3 : Poor accuracy of a diagnostic test

Low referring GPs also stated that they were less likely to refer patients if the result of the diagnostic test was unlikely to be accurate for that patient, which again fits the profile of a group of GPs with a high threshold for referral who may thoroughly weigh up the pros and cons of the investigation and referral decisions they make. However GPs who referred both high risk 'patient profiles' (i.e. adherent to NICE guidelines) were significantly more likely than those who did not refer both to have stated that they would refer patients even when 'the diagnostic test is unlikely to give an accurate result for this patient'. While it makes sense that GPs with this approach would refer more patients, it is not clear why there is such a significant difference between fully adherent GPs and other GPs (p=0.004). It is possible that GPs who have less concern about accuracy of diagnostic tests refer for more tests.

7.4.2.3.4 : Late-running appointments

It has been proposed in the literature that GPs who are overburdened might behave differently.[175] GPs who were fully adherent to NICE guidelines and referred both of the high risk profiles were also more likely to refer a patient if their appointments were running late. It is possible that this indicated a group of GPs who are keen not to miss potentially serious conditions, and therefore if faced with time pressures that may have an impact on the depth, quality or extent of their consultation with a patient, would prefer to refer them for investigation as a failsafe measure.

7.4.3 : Strengths and limitations of these analyses

7.4.3.1 : What this analysis adds to our understanding of this field

My systematic review showed that whilst there were several hypotheses in the existing literature about what might explain variations in GPs' referral behaviour, there was a lack of empirical evidence to support these. My analyses in this chapter have sought to address the gap by combining data collected from both parts of the GP decision making study. They demonstrate that there were some distinct differences between GPs who had particularly high or low rates of referral for 'patients' with symptoms of lung cancer in the vignette study, and other GPs. There were also differences between GPs who were fully adherent to NICE guidelines and referred both the high risk 'patients', compared to those who did not.

Whilst we cannot be certain to what extent GPs' referral behaviour and responses in the GP decision making study are representative of the behaviour of GPs more generally, this analysis does enable us to reasonably hypothesise that there might also be distinct differences between groups of GPs who refer differently in real life. It has also highlighted some of the characteristics and factors that are likely to influence differences in patterns of referral between GPs, thus providing a starting point for future research.

7.4.3.2 : Limitations

7.4.3.2.1 : This is not a study of real life, in situ, behaviour

The GP decision making study was not an observational study of real life - GPs' referral behaviour in the vignette study may reflect their behaviour in real life, but we cannot know to what extent this is the case. Since GPs who were high referrers in the vignette study were more likely to have lower IT confidence, it is also possible that GPs' facility to use the virtual patient application may have affected their referral decisions. Likewise in the post-consultation survey GPs were asked to comment on the extent to which they felt factors affected their likelihood of referral, but this may not reflect their actual behaviour (either consciously or subconsciously). As with any study that is not examining real life, we must therefore exercise caution when reflecting on the extent to which the results of this study can be generalisable to real life.

7.4.3.2.2 : Small sample size and the potential for statistical error

There are also limitations in the relatively small sample size; in particular for both the high referrer and the low referrer calculations one group was fairly small in each category (less than 50 GPs). In addition some of the variables I examined had very little variation in data (for example only two GPs were in the oldest age category, and very few GPs consulted information sources during the vignette study) as demonstrated by some of the very wide confidence intervals when performing logistic regression. Therefore, although the analysis has allowed me to identify and report on some differences that are statistically significant, real differences between groups in each category may have been missed, and minimal differences may have been exaggerated. There is also a potential for type 2 errors in my analysis since I conducted multiple statistical tests with a large number of explanatory variables, and three outcome measures.

7.4.3.2.3 : Strong correlation of variables

It is also important to note that, particularly for the factors where GPs were asked to rate the extent to which they influenced their likelihood of referral, several of the variables that I considered in this analysis are strongly correlated with each other. This limited my ability to distinguish their individual effects.

7.4.4 : Implications for future research, policy and practice

7.4.4.1 : Future research

Additional research is now needed to see if our findings from the GP decision making study are replicated in real life settings, and whether there are indeed distinct differences between groups of GPs who refer patients differently. If this is replicated in real life, it offers the potential to increase our understanding of why there is non-clinical variation in GPs' referral of patients - in particular the role of GP characteristics.

7.4.4.2 : Policy and practice

This analysis suggests that there are distinct differences between GPs who have different referral patterns: both in the frequency of referral (high and low referring GPs) and in the sensitivity of referral (GPs' adherence to NICE guidelines). If this is also reflected in real life practice then there is the potential to develop training or interventions targeted at particular groups of GPs: for example high referrers with low adherence who might be able to refer more specifically, or low referrers with low adherence who might be missing referring patients at high risk of disease. There is also the potential to learn from those GPs whose referral is the most efficient: low referrers whose referral is highly adherent.

8 : Thesis conclusions

8.1 : An overview of my thesis

The aim of my PhD (outlined in Chapter 2) was to examine the patient and GP characteristics associated with GPs' decisions to refer patients for investigations or to secondary care, with a particular focus on those patients presenting with symptoms indicative of lung cancer. I set out to address this aim through two studies.

Study 1 was a systematic literature review (Chapter 3) in which I sought to identify non-clinical characteristics associated with variation in GPs' referral of patients for investigation or to secondary care, as well as to identify areas of uncertainty and inconsistency requiring further research and different methodological approaches.

Study 2 was the GP decision making study, whose methods (Chapter 4) sought to address some of the methodological limitations highlighted in my systematic review. There were two parts to Study 2. For the first part, the vignette study, I worked in a team to develop a novel study tool which used an interactive, multimedia form to present GPs with vignettes of patients with symptoms that could indicate lung cancer, and enabled us to examine GPs' management decisions (Chapter 5). In the second part of Study 2 I developed an online post-consultation survey for GPs to complete, in order to identify factors that they believed influenced their real life referral decisions (Chapter 6). I then examined the extent to which GPs' referral decisions in the vignette study related both to their personal characteristics, and to factors that they reported in the post-consultation survey as influencing their referral behaviour (Chapter 7) in order to start to understand why differences in GPs' referral behaviour might exist.

8.2 : Key findings

8.2.1 : Systematic literature review (Study 1)

The systematic review found that there is strong evidence that both patient age (with the oldest patients less likely to be referred) and patient gender (direction of referral varying between conditions) are associated with variation in GPs' referral of patients for investigations or to secondary care.

It also enabled me to identify some key gaps in the literature, since due to a combination of methodological issues affecting a number of studies and the limited number of studies which examine the association of several characteristics, I was not able to conclude whether there is variation in GPs' referral behaviour for patient characteristics other than age and gender, or for either individual GP or practice characteristics. This systematic review, and my appraisal of a number of different study methods, identified that there are not currently enough studies of sufficient rigour and relevance to answer my question fully.

Furthermore the studies identified in my systematic review tend to simply draw conclusions about the extent to which there is non-clinical variation in GPs' referral behaviour, rather than also exploring what factors underlie these associations.

8.2.2 : GP decision making study - vignette study (Study 2a)

The analysis showed that, overall, GPs referred 74% of the 'patients' in the vignette study for chest X-ray. The referral percentages for each of the six different 'patient profiles' in isolation varied significantly. However the likelihood of referral did not increase as the clinical risk of lung cancer increased. This in large part reflected some GPs' failure to ask about, and therefore elicit the presence of, non-chest and non-specific symptoms such as weight loss.

In the vignette study there was also non-clinical variation in referral for chest X-ray. GPs were less likely to refer older 'patients' than younger ones, which is in line with the findings in my systematic review.

The factorial design of the vignette study meant that, in contrast to many studies I identified in my systematic review, we were able to examine the effect of patient ethnicity on GPs' referral decisions; in this study GPs were marginally less likely to refer 'patients' of black ethnicity compared to white.

Contrary to some of the literature we did not find a gender difference, even when 'patients' presented with chest symptoms (previous research has suggested that women with chest pain may be less likely to be referred for diagnostic tests).[75] However we did take specific care when designing the vignettes to ensure that they were symptomatically distinct enough from a typical cardiac presentation, so this difference might reflect that the majority of GPs were not considering a cardiac cause of pain.

In the vignette study we used a novel study tool, the virtual patient application, to examine GPs' decision making in a factorial design study. There were challenges with using the virtual patient application, in particular the lengthy computer set up process for GPs and, perhaps most importantly, the fact that we are not examining real life behaviours when using it. That said, when developing the application as a study tool we took great care to develop an application that simulated a real life GP consultation as closely as possible, as well as addressing the key methodological limitations of many previous vignette studies. This included:

- presenting information to GPs in a multimedia format that included using videos to deliver much of the vignette content, providing non-verbal cues;

- developing an interactive response system that reflected the length, content and interactive nature of a real life GP consultation without requiring researcher input to deliver each vignette;
- recruiting sufficient number of GPs and taking steps, both in the design and the delivery of the study, to avoid priming them.

Overall the tool appears to have been successful: in 99.98% of 'consultations' completed GPs were able to make a management decision, and the majority of the GPs did not report issues using the virtual patient application after receiving our standard guidance.

8.2.3 : GP decision making study - post-consultation survey (Study 2b)

The post-consultation survey identified a number of factors which GPs reported significantly impact on their real life referral decisions.

Many factors significantly impacted GPs' reported real life referral behaviour; I have identified the following as being particularly important. GPs were most influenced by a patient's lifestyle putting them at increased risk of disease, with a very significant majority more likely to refer patients in these cases. GPs also reported that they were less likely to refer patients who frequently attended with non-serious complaints, or if they could receive prompt advice from a hospital colleague. A significant proportion of GPs reported that they would be more likely to refer a patient if there were challenges in communication and/or understanding. This does not initially appear to correspond with the vignette study's findings that both older and non-white patients (both groups in which communication challenges could occur) were less likely to be referred for chest X-ray; however none of the 'patients' in the vignette study had communication difficulties.

Nearly half of GPs reported that in real life situations (similar to those in the vignettes) they would not refer to external sources of information. Where GPs did report that they would seek additional information, more stated that they would consult a colleague than refer to guidelines.

It is possible that some of the factors identified in the post-consultation survey as influences on GPs' referral behaviour may contribute to the non-clinical variations seen in GPs' decisions to refer patients for investigations or to secondary care. However we are not able to examine this possibility when considering the results of the post-consultation survey in isolation.

8.2.4 : GP decision making study - combining data

This analysis suggests that several GP-related factors (personal characteristics, behaviours and beliefs, and factors that influence them) are likely to be associated with variation in referral behaviour.

GPs' gender had an effect on their referral frequency in the vignette study, with men more likely to be high referrers and women more likely to be low referrers.

High referring GPs were less likely to report that they would ask a colleague for advice. They were also less likely to be influenced by patients having a low level of spoken English. Low referring GPs were more likely to have qualified recently. They were also significantly less likely to refer patients who had previously failed to attend appointments, or if a diagnostic test was inaccurate.

GPs whose referral of the high risk 'patients' was adherent to NICE guidelines were more likely to report that their referral behaviour would be affected by their appointments running late. GPs' personal experiences also seem to have an effect on their adherence: GPs were more likely to refer the two profiles with the highest risk of cancer if they (the GP) had a personal history of smoking, but less likely to refer both if they had worked as a specialist in respiratory medicine.

8.3 : My learning from this PhD

In this section I will reflect on my personal learning from this PhD, both some key changes I would make if I were repeating these two research studies, and notable things I have learnt from this experience.

8.3.1 : What I would do differently

8.3.1.1 : Determining the scope of my research question (Study 1)

My systematic review (Study 1) enabled me to extensively examine the literature on the associations between non-clinical characteristics and GPs' referral behaviour across studies from a five year period. As a result of this review I have a broad understanding of the research studies that address this subject.

However conducting a systematic review with such a wide scope did have limitations; most noticeably in the scale of the numbers of papers identified for screening. The initial question I set out to answer in my systematic review is now too big to be answered, given the quantity of literature that has been published since and the tools that I have available to conduct the review.

If I were to start this systematic review again I could perhaps deal with the breadth of the literature by more thoroughly exploring the effectiveness of using the 'data mining' software I discussed in Section 3.4.3.2.4. Alternatively, I would most likely seek to narrow the scope of my research and develop a more tailored research question focusing on a specific gap in the literature (e.g. the association between patient ethnicity and GP referral) and answer it fully.

8.3.1.2 : Development of the language recognition software (Study 2a)

The virtual patient application was an effective tool for capturing GPs' decisions, and was able to create a reasonable simulation of a GP consultation, providing that GPs understood how to use the software. The use of language recognition software enabled us to simulate the interactive nature of a real life conversation; however it also presented us with a number of challenges that prevented the application being as user-friendly and faithful to real life as we would have liked.

We were limited in our development of the virtual patient application by cost, and were not able to develop all the functionality we had originally planned. As a result, the language recognition software had some limitations that were not intuitive for GPs using the software: for example GPs had to repeat the name of the symptom they were asking about in all questions, and thus could not type questions in the same format as they would speak during a real life consultation. As discussed in Section 4.2.5 we took steps to minimise the impact of these limitations. To some extent these steps were effective, as GPs were unable to provide a management plan in only three out of 1362 total 'consultations'; however a small number of GPs did report that the software was unrealistic and frustrating to use.

If I were to conduct the vignette study again with the same budget then I would probably look to develop a very similar application, since the virtual patient application, as it was developed, enabled us to examine GPs' decision making effectively. However if the financial implications were less significant it would be worth investigating how to mimic the real life interactive conversation of a GP consultation more closely. Since the results of the vignette study indicate that the questions GPs ask during a consultation have a potential impact on their referral behaviour, studying the process and content of the consultation itself may provide useful information about influences on, and variations in, GPs' referral behaviour.

8.3.1.3 : Design and delivery of the post-consultation survey (Study 2b)

Developing the post-consultation survey enabled me to start to identify factors that could explain the variation in GPs' referral seen in both my systematic review and the vignette study. The survey asked GPs to reflect on the extent to which several factors influenced their referral behaviour: both in the vignette study and if they were to experience similar situations in real life. In order to capture GPs' consultation thought and decision making processes as reliably as possible, I designed the GP decision making study so that GPs completed the survey straight after their final virtual 'consultation'. The purpose of this was to increase the likelihood that GPs' behaviour and any influences during the 'consultation' would still be fresh in their mind, and to reduce the potential for recall bias. However whilst we tried to give GPs as authentic a 'consultation' experience as possible in the GP decision making study (by using the virtual patient application) it was not real life, and therefore we still cannot be sure to what extent GPs' responses actually reflect their real life behaviour.

If I were to repeat the post-consultation survey, it would be useful to deliver the survey to GPs immediately after real life consultations, as well as after the GP decision making study. This would allow us to compare these results with those from the GP decision making study in order to determine if GPs reported similar influences on their behaviour.

The post-consultation survey had some additional constraints. We designed the survey to be completed in about 5 minutes, as part of the plan to limit the time needed to complete the GP decision making study to one hour; this limited the content that could be included. In addition, I asked GPs to reflect on the extent to which a large number of different factors influenced their decisions to refer patients. Whilst this approach enabled me to create a broad picture of the types of factors influencing GPs' referral behaviour, it prevented me from investigating these in depth. It also meant that I conducted multiple statistical tests with a large number of explanatory variables, creating a possibility of Type 2 statistical error in my analysis..

If I were to conduct a follow up study using the same methods as the GP decision making study and using findings it has already provided, I would refine the post-consultation survey into a more focused survey that could still be completed by GPs in about 5 minutes, asking GPs in more depth about the extent to which specific factors influenced their referral decisions, and capturing how different GPs rank the relative importance of these factors as influences on their decision making. I would focus on asking GPs about factors shown in the post-consultation survey to significantly influence GPs' referral decisions; these could include a patient's lifestyle putting them at risk of disease or their previous poor or unnecessary engagement with health services, communication challenges (include a patient having low level of spoken English), GPs' appointments running late, and GPs' access to, and use of, sources of information such as guidelines or hospital colleagues. This would allow a deeper investigation into the impact of each factor, and reduce the risk of Type 2 errors.

8.3.2 : What I will take forward for my future research

8.3.2.1 : The importance of setting the scope of my research question

During the process of this PhD I have come to recognise the importance of defining the scope of one's research question clearly, early and thoughtfully. In both my systematic review and the GP decision making study I have had numerous questions that I have been keen to address, but which have been beyond the scope of my PhD. For example the virtual patient application captured a whole wealth of data relating to GPs' questions and behaviour during the virtual 'consultation'; it was a particular challenge not to explore these in depth - however it would not have enhanced my understanding of the non-clinical variation in GPs' decision making.

8.3.2.2 : The recognition that all methodological approaches have both strengths and failings

Undertaking the research studies in this PhD has also heightened my sense of the strengths and limitations of different research designs. I have developed an understanding that there is not one perfect method: all studies will have flaws, and it is impossible to answer all research questions in a field with one study.

For example I have gained an understanding of the benefits of the certainty of observational studies using routine, retrospective data which come from real life. However it is also the case that retrospective, observational studies are unable to effectively examine some aspects of GP decision making. By contrast, experimental studies (including those using vignettes) allow innovative methods, but they are not examining real life behaviour. Finally, systematic reviews provide a comprehensive and trustworthy summary of the literature; however they are time intensive and, even when substantial in size, still may not identify enough relevant studies to enable firm conclusions to be drawn.

8.4 : Implications for future research, policy and practice

8.4.1 : Future research

In addition to adding to and enhancing our existing knowledge of the non-clinical variation in GPs' referral of patients for investigations or to secondary care, the findings of my PhD provide a number of clear suggestions for future research to further increase our understanding, which in turn has implications for improving early diagnosis of diseases such as lung cancer.

These suggestions encompass not simply potential topics for that research, but also methodological considerations for future studies in order to ensure that they are of high quality.

8.4.1.1 : Gaps in the literature for future research

8.4.1.1.1 : Further clarification of which non-clinical characteristics are associated with variation in GPs' referral behaviour

Whilst my research in this thesis has identified that both patient age and gender are associated with variation in GPs' referral behaviour, I was not able to draw firm conclusions about associations with other non-clinical characteristics.

Outstanding areas of uncertainty that have been addressed in the literature, but for which it has not yet been possible to draw firm conclusions about their effect include:

- patient ethnicity;
- the individual GP's personal characteristics (e.g. gender, years since qualification, clinical experience);
- the individual GP's attitudes, beliefs and influences;
- practice characteristics: in particular GPs' relationship with, and access to, secondary care.

8.4.1.1.2 : Understanding the reasons underlying variation in GPs' referral behaviour

The GP decision making study has identified (both through the post-consultation survey, and the analysis combining its results with data from the vignette study) a number of factors that GPs report influence their referral behaviour in real life.

However I was not able to draw any firm conclusions about the reasons underlying variation in GPs' referral behaviour. Whilst combining data from the vignette study and the post-consultation survey enabled me to start examining this, the relatively small sample size of the GP decision making study resulted in small groups for much of the statistical analysis, meaning that real differences may have been missed, or minimal differences exaggerated. A larger sample size would enable future studies examining this to have more confidence in their findings.

Future studies should also continue with identifying and quantifying what influences GPs' *real life* referral behaviour, for example by examining whether any of the factors that GPs report influence their decision making are associated with variation in GPs' real life referral frequencies. This could be considered by referrals made, or by patients' presenting symptoms.

Further research is also needed to consider whether the variation in GPs' referral behaviour is intentional and/or whether it is in patients' best interests. Such information will be valuable when seeking to develop strategies to reduce the non-clinical variation in GPs' referral.

8.4.1.1.3 : Consider whether GPs' variable use of guidelines is a source of non-clinical variation in their referral behaviour

My research identified that a relatively small number of GPs reported that they would use guidelines when in a consultation with a patient with similar symptoms to those in the vignette study (i.e. symptoms that could indicate a diagnosis of lung cancer, for which there are national guidelines).

There is a suggestion in the literature that physicians' adherence to guidelines varies with patients' non-clinical characteristics.[183] In addition, my research found that GPs who were high referrers in the vignette study reported using guidelines differently to the rest of the GPs who participated in the GP decision making study. It is therefore possible that GPs' variable use of guidelines could be a source of non-clinical variation in their decision making.

Considering in depth the role of guidelines in GPs' decision making was beyond the scope of my PhD. However questions for future research to address are whether GPs' consultation of guidelines (or the lack of it) is reflected in their real life decision making, and whether the extent to which GPs use guidelines has an impact on the effectiveness of their referral behaviour.

8.4.1.2 : Methodological considerations for future research

In order for future research into non-clinical variation in GPs' referral decisions and the reasons for it to be as valuable as possible, it is vital that the methods used by future research studies seek to address some of the methodological shortcomings that my research (in particular the systematic review) has identified.

Particular areas of consideration should be:

- accounting or adjusting for potential relevant confounders for the characteristics being studied (where possible);
- using a sufficiently sized and diverse sample population so that the study's findings will be generalisable nationally;
- considering a specific outcome (e.g. chest X-ray), or considering how the influences on GPs' referral behaviour vary according to the referral being made;
- whether studies exploring the factors that influence GPs' referral behaviour, and the reasons for variations in this, are best answered using retrospective data or whether novel study designs and methods are needed.

8.4.2 : Policy and practice

8.4.2.1 : GP awareness and training

My PhD has identified that there is non-clinical variation in GPs' referral of patients for investigations and/or to secondary care. This variation (particularly in the case of patient age) is seen across a wide range of symptoms and diseases, suggesting that it is likely, at least to some extent, to reflect fundamental differences in GP attitudes towards referral for certain groups of patient.

Increasing GP awareness of this non-clinical variation in referral (and its potential role in contributing to the non-clinical variation in early cancer diagnosis seen within the UK) is therefore of high importance. A number of different strategies could be developed to address this, these include:

- educational software encouraging GPs to reflect on their decision making processes;
- using alerts to remind GPs to ask about specific, relevant additional symptoms (rather than relying on the patient to mention them, when the patient may not be aware of, or want to face, their significance or implications);
- education of GPs to increase their awareness of the patient characteristics associated with a lower likelihood of being referred, and of factors that might potentially influence their decision making.

8.4.2.2 : Tensions in decision making

My PhD has also highlighted some potential tensions which it would be valuable to address: both for individual GPs, and also more widely within the health service.

8.4.2.2.1 : Poor interface between primary and secondary care

Many GPs stated that they valued a close working relationship with a hospital colleague and that this not only aided their decision making, but also potentially reduced the likelihood that they would refer a patient to secondary care or for investigation. However GPs also reported that accessing advice from hospital colleagues is often difficult. Strategies to improve communication between primary and secondary care could therefore have significant implications for improving the efficiency of GPs' referral.

8.4.2.2.2 : An aging population who do not necessarily wish to be treated as old

Both the GP decision making study and my systematic review found that the oldest patients were the least likely to be referred for investigations or to secondary care. In some situations this may be intentional, and based on the patient's own wishes and best interests. However in many of these studies the 'oldest' patients were those aged over 70 or 75 years. Life expectancy in the UK has increased substantially over the last 20 years,[184] and people aged in their 70s may not think of themselves as especially 'old'. It must be considered whether lower referral of older patients is appropriate and in line with their preferences, or if it is based on GPs' perceptions of 'old age' and the management they believe these patients would want.

8.4.2.2.3 : The challenge of identifying serious disease in frequent attenders

GPs in the post-consultation survey reported that they were less likely to refer patients who frequently attended with non-serious complaints. This seems reasonable, particularly in a climate where GPs are encouraged to limit unnecessary use of referral; although the 2015 NICE guidelines on the recognition and referral of suspected cancer do encourage referral at a lower threshold, and more safety-netting.[35] Furthermore we have to be aware that new, serious disease can occur at any time in a patient, irrespective of their past medical history; it is therefore important for GPs to try to keep clinical and socio-behavioural factors separate in their minds when evaluating whether a patient is likely to require referral for investigation or to secondary care.

8.5 : Overall conclusions

Older patients are less likely to be referred for investigations, including chest X-ray, or to secondary care. Patient gender is also associated with variations in referral, though the direction of this variation differs with the symptom or disease; for patients presenting with symptoms of lung cancer we found no difference in referral by patient gender.

Black patients with symptoms of lung cancer were less likely to be referred for chest X-ray than white patients; however there were not enough high quality UK studies in the existing literature to draw any firm conclusions about the association between patient ethnicity and GPs' referral behaviour. Similarly, the association between other patient, GP or practice characteristics and GPs' referral for investigations or to secondary care is uncertain.

This study has shown that a number of different factors, such as a patient's lifestyle putting them at increased risk of disease or a patient attending frequently with non-serious complaints, along with GPs' personal characteristics (e.g. GP gender), are likely to influence GPs' referral decisions.

Whilst my research is a small piece in the much larger jigsaw of understanding and improving cancer outcomes, it is nonetheless important since it enhances our understanding of socio-demographic variations in cancer diagnosis within primary care. An increased understanding of the non-clinical characteristics associated with variation in GPs' referral decisions, and the factors that may underlie this, has the potential to enable us to develop targeted strategies to reduce non-clinical variation in referral. This in turn has the potential to reduce the variation in early diagnosis of cancer in the UK, and therefore perhaps to improve cancer survival.

Acknowledgments

I would particularly like to thank my supervisors, Professor Rosalind Raine, Professor Stephen Duffy and Dr Jessica Sheringham, for their guidance, patience and support throughout my PhD.

I would also like to thank:

- Joe McDonnell: for all his help with my systematic review and work on the GP decision making study;
- Joseph Forrest and his team at Atheneum Education Technologies: for their development of the virtual patient application study tool;
- Professors Willie Hamilton, Una MacLeod and Greg Rubin: for their clinical insight and input into both the systematic review and the GP decision making study;
- Dr Jonathan Myles: for his work (alongside Professor Stephen Duffy) conducting the hierarchical modelling analysis of the vignette study data;
- Dr Anjali Bajekal: for her clinical input during the development of the virtual patient application;
- Dr Janakan Crofton: for his clinical input coding the vignette study data;
- the team of temporary administrative staff who helped with GP recruitment;
- the GPs who participated in the GP decision making study;
- all those who piloted the study;
- the Department of Health and the PRU, for their generous funding;
- my dedicated personal support team of friends and family: for their encouragement, support, proof reading and patience over the last few years.

Appendix 1 : Search strategy for the systematic review

Medline search strategy *- Run 4th April 2012, retrieved 16,082 records*

No.	Type	Search term	Subheadings (MeSH only)	Notes
1	MeSH	Patients/	Include all	Do not explode MeSH term as subheadings not relevant
2	Title, abstract	(patient* or (service adj user*) or client* or consumer*).ti,ab.	n/a	Restrict to title and abstract search Additional words for patient do not add significantly to increase in retrieved results
3	OR	1 or 2	n/a	Create idea 1: ‘patient’
4	MeSH	exp Decision Making/	Include all	Explode MeSH term
5	MeSH	exp “Referral and Consultation”/	Include all	Explode MeSH term
6	MeSH	exp "Diagnostic Techniques and Procedures"/	Include all	Explode MeSH term
7	MeSH	exp "Outcome and Process Assessment (Health Care)"/	Include all	Explode MeSH term
8	MeSH	Physician’s Practice Patterns/	Include all	
9	MeSH	exp Professional Practice/	Include all	Explode MeSH term
10	Title, abstract	(decision* or refer* or investigat* or diagnostic* or outcome* or management*).ti,ab.	n/a	Exclude diagnosis as increases noise – reasonable?
11	OR	4 or 5 or 6 or 7 or 8 or 9 or 10	n/a	Create idea 2: ‘decision/outcome’
12	MeSH	exp General Practice/	Include all	Explode MeSH term
13	MeSH	exp Primary Health Care/	Include all	Explode MeSH term
14	MeSH	General Practitioners/	Include all	
15	MeSH	Physicians, Family/	Include all	
16	MeSH	Physicians, Primary Care/	Include all	
17	Keyword	GP* or (general adj practi*) or (family adj care) or (family adj (healthcare or (health adj care))) or ((family or (family adj care) or (family adj (healthcare or (health adj care)))) adj (doctor* or provi* or physician* or practi*)) or (primary adj care) or (primary adj (healthcare or (health adj care))) or ((primary or (primary adj care) or (primary adj (healthcare or (health adj care)))) adj (doctor* or provi* or physician* or practi*))	n/a	Do not restrict to title and abstract as being used to limit search to useful papers
18	OR	12 or 13 or 14 or 15 or 16 or 17	n/a	Create idea 3: ‘general practice/GP’
19	MeSH	exp Age Factors/	Include all	Explode MeSH term
20	MeSH	exp Adult/	Include all	Explode MeSH term
21	MeSH	Sex Factors/	Include all	
22	MeSH	Male/	Include all	

No.	Type	Search term	Subheadings (MeSH only)	Notes
23	MeSH	Female/	Include all	
24	MeSH	exp Social Class/	Include all	Explode MeSH term
25	MeSH	exp Ethnic Groups/	Include all	Explode MeSH term
26	Title, abstract	(age or sex or gender* or male* or female* or (social adj status) or socioeconomic* or socio-economic* or (social adj class) or depriv* or disadvantage* or poor or (less adj educated) or less-educated or underprivilege* or affluent or advantage* or rich or (more adj educated) or more-educated or ethnic* or race or racial or cultur* or socio-demographic* or (socio adj demographic*) or (patient adj factor*) or (patient adj characteristic*) or psychosocial or ((GP* or practi* or provi* or doctor* or physician*) adj factor*) or ((GP* or practi* or provi* or doctor* or physician*) adj characteristic*)).ti,ab.	n/a	Includes specific characteristics (socio-demographic and lifestyle) and broad terms for these ideas
27	OR	19 or 20 or 21 or 22 or 23 or 24 or 25 or 26	n/a	Create idea 4: ‘characteristics’
28	AND	3 and 11 and 18 and 27	n/a	Create search itself
29	Limit	limit 28 to (english language and yr="1980 -Current")	n/a	Set limits to English-only papers and those from 1980 onwards
30	MeSH	United States	Include all	Exclude paper which are clearly US studies – but do not use term US as ambiguous
31	Keyword	USA or America or (United adj States)	n/a	Exclude paper which are clearly US studies – but do not use term US as ambiguous
32	OR	30 or 31		Create United States exclusion
33	NOT	29 NOT 32		Final search

I also conducted searches in EMBASE, Web of Science, PsycInfo and Social Policy and Practice. For each of these databases I used identical free text search terms, combinations and limits and (where appropriate) comparable MeSH headings.

Appendix 2 : Selection criteria for the systematic review

Criterion	Include	Exclude	Comments	Implications for screening
Year of publication	2007 to 4th April 2012 (date the review searches were run)	2006 or previous	There have been many changes in GP practice in the UK health system in recent years. Since I am interested in current practice I will therefore restrict the search to studies from the last 5 years.	None - limits applied in search strategy and checked again when imported into Reference Manager.
Study area	United Kingdom (England, Wales, Northern Ireland, Scotland or any combination) form all/part of study population.	Entire study population is non-UK.	I am interested in GP decision making in the UK health system. Since this is quite unique (in terms of funding, priorities, burdens and structure), studies conducted solely in other countries will be excluded.	If title/abstract clearly references a different study area, not including the UK, it will be excluded. *Examples include:* *- country name* *- reference to "Medicaid"* *- physicians who are clearly non-UK e.g. internists* During full paper screening all author addresses will be reviewed – if none are UK and the paper does not specifically detail a UK study population it will also be excluded.
Population	Age: Adults (≥18 years old) form all/part of study population	Age: Entire study population is under 18 years old.	Age: Children will be excluded because a) their role in the consultation is different to adults and b) the adult population is more appropriate for our subsequent study.	If title/abstract refers to management of children or adolescents it will be excluded. This exclusion also covers the following scenarios: - parental consulting behaviour if consulting for child - referral to services known to be solely paediatric (e.g. CAMHS) - transition from paediatric to adult services for known medical conditions – patient is already within the secondary care system

Criterion	**Include**	**Exclude**	**Comments**	**Implications for screening**
Population (continued)				Studies referring to 'young people' will be included unless the title/abstract clearly indicates that participant age is <18.
	Field: Human medicine	Field: Studies in veterinary or dental practice, social work	Field: There are other forms of primary care, but these are not appropriate to this review or our study.	Exclude if title/abstract refers to veterinary, dental, optometry or social care or their management decisions.
	Specialty: Primary care – referrals for investigation, to secondary care or other services	Specialty: Purely secondary or tertiary care, management decisions made within secondary care	Specialty: We are interested in GPs as gatekeepers to secondary care, and their use of resources. Studies where a patient is already within secondary care are therefore not relevant.	If title/abstract clearly indicates study is solely based in secondary care it will be excluded. Examples include: - intermediate care placements - critical care - specialists' decision making e.g. gynaecologist
	Disease: All		Disease: There is some evidence that differences in referral behaviours may be specific to particular diagnoses.[140] Therefore, if appropriate, studies may be subdivided at analysis.	
Outcome	Management decision made by a physician in primary care to refer a patient for investigation, diagnostic tests or procedures (not including screening) or to secondary care or other services Studies looking at nurse practitioners' decisions to be considered on a study by study basis	Management decisions that solely relate to prescription of new drugs or alteration of medication regime. Decisions made by the patient in isolation (not during a consultation) e.g. to attend routinely offered breast cancer screening. Patient views of GP decision making	Need to consider how to approach studies about patients with a known condition, as opposed to initial presentation.	If title/abstract refers to drug prescription alone (as opposed to management of a condition in general, follow up or investigations whilst on a certain drug) it will be excluded. This includes the following subjects: - comparison of medication effectiveness (cost and/or symptomatic or trial) - adherence or compliance to medication regime - inequalities in decisions to prescribe

Criterion	Include	Exclude	Comments	Implications for screening
Outcome (continued)				If title/abstract suggests study topic does not have any connection with how GPs decide to manage patients, such as the following situations, it will also be excluded: - socio-demographic or clinical risk factors for disease (if disease occurrence only rather than disease recognition) -evaluation/comparison of treatment effectiveness - classification or validation of diagnostic scores, disease severity scores (but if when to use these include) - patient preferences (unless impact of patient requesting test or referral) - primary prevention if solely medication or lifestyle approaches *NB: if unclear from title/abstract whether paper might be relevant allow through to next stage.*
Study design	Observational studies – cohort, cross-sectional, case-control, case reports Randomised control trials Intervention studies Systematic reviews	Single case reports Narrative reviews Qualitative studies – but those containing information on why GPs make decisions will be noted	Different study designs will be assigned different quality scores.	
Language	English	Other languages	No resources for translation	Limits applied in search strategy so should be very few non-English papers, but exclude at title screening if so.
Publication type	Peer-reviewed journals	Conference proceedings Books Letters Comments	Exclude grey literature	

Appendix 3 : The critical appraisal and data extraction tool for the systematic review

The screenshots on the following pages show the questions and structure of the critical appraisal and data extraction tool (adapted from Heller et al's checklist (2008)[71] that I used for my systematic review.

Paper reference information

Reference | **Research question and study design** | **Characteristics and confounders** | **Methods, outcomes and bias** | **Results and conclusions** | **Summary**

(New) **Title**

Publication year

Reviewer initials

Date of form completion

NB: 2nd reviewer only needs to enter information above this point - i.e. title, publication year, their initials and date of form completion

Journal

Volume **Pages** **to**

Authors (surname, initials) **1st** **2nd** **3rd**

Last **Additional authors**

Research question and study design

Reference | Research question and study design | Characteristics and confounders | Methods, outcomes and bias | Results and conclusions | Summary

(New) **Title**

Study question

What is the research objective or hypothesis?

This objective/hypothesis is clearly stated ☐

Is this question relevant to my population? **Does this question clearly address my research question?**

Study design

What is the study type? **Further details of study type (if appropriate):**

The study type is appropriate for the research question ☐ **If not appropriate, why?**

Study population **Site country (select all that apply):** England ☐ Northern Ireland ☐ Scotland ☐ Wales ☐
Other country ☐ If other country, list all:

Study subjects (select all that apply): Patients ☐ GPs ☐ Other, please specify:

Sample size **Patients** **GPs** *Note: if sample size varies and/or is unclear state largest number used here and note in study limitations*

Power calculation performed ☐ **If performed, what is the power?**

Does the sample population reflect the study population? **If to some extent, give details:**

Inclusion/exclusion criteria

Are these criteria appropriate?

If 'to some extent' or 'no', give details:

Inclusion/exclusion criteria includes patients: presenting with particular symptoms ☐ List symptoms:
with existing disease/condition ☐ List conditions:
with a particular final diagnosis ☐ List diagnoses:

Is the study a factorial design? ☐ **If factorial design give details of patient characteristics:**

Recruitment **What recruitment procedures were used?**

These procedures were appropriate ☐ **If not appropriate, why?**

Is there potential for recruitment bias, and if so how is this addressed?

What is the response or completion rate?

Is this adequate? (considering proportion completed and potential for bias)

Characteristics and confounders addressed

Reference | Research question and study design | Characteristics and confounders | Methods, outcomes and bias | Results and conclusions

(New) **Title**

Patient and GP characteristics explored

Are these characteristics the exposure or the outcome?

Patient characteristics (select all that apply)
Age ☐ SEC ☐ *measures of SEC used:*
Gender ☐ Co-morbidity ☐ *specify which co-morbidities:*
Ethnicity ☐ Quality of life ☐
List any other patient characterisics:

GP characteristics (select all that apply)
Age (GP) ☐ Years working as GP ☐ GP has specialism interest or experience ☐
Gender (GP) ☐ Sessions worked/week ☐
Ethnicity (GP) ☐ GP's role (partner etc) ☐
List any other GP characteristics:

Practice characteristics (select all that apply)
Practice deprivation ☐ No. GPs in practice ☐ Distance from hospital ☐
Urban/rural practice site ☐ No. patients registered ☐
List any other practice characteristics:

Are these characteristics relevant for my research question?
If to some extent, give details:

Confounders *Note: Only fill in if appropriate*

What confounders are STATISTICALLY considered? (select all that apply)
Patient age ☐ Co-morbidity ☐ GP age ☐ Years worked as GP ☐ Practice deprivation ☐
Patient gender ☐ Quality of life ☐ GP gender ☐ Sessions worked/week ☐ Urban/rural practice ☐
Patient ethnicity ☐ Disease severity ☐ GP ethnicity ☐ GP interest/experience ☐ Number of GPs ☐
Patient SEC ☐ Presentation ☐ GP's job title ☐ GP personality/psychological factors ☐ Number of patients ☐
Patient preference ☐ Distance from hospital ☐
Patient personality/psychological factors ☐ Individual GP clustering ☐ Practice clustering ☐
Additional confounders considered:

How are confounders statistically addressed?
The statistical methods to address confounding are appropriate ☐ **If not appropriate, why?**
What other confounders were considered, and how?
Are any important confounders not addressed?
Is confounding dealt with adequately?

Study methods, outcomes,[v] and sources of bias

Reference	Research question and study design	Characteristics and confounders	Methods, outcomes and bias	Results and conclusions

(New) **Title**

Methods

What are the methods of data collection or extraction? (select all that apply)

GP research database ☐ Observation ☐ Vignettes ☐
Medical records ☐ Questionnaire ☐
Other method (give details):

These methods are appropriate ☐ **If not appropriate, why?**

Outcomes

What are the outcome measures? (select, and give details for, all that apply)

Bedside tests ☐ If selected give details:
Diagnostic tests ☐ If selected give details:
Procedures ☐ If selected give details:
Screening ☐ If selected give details:
Referral to hospital team ☐ If selected give details:
Referral to A+E ☐
Referral to other healthcare professional ☐ If selected give details:

Give details of any additional outcome measures:

These outcome measures are appropriate for the study ☐ **If not appropriate, why?**

Bias

Are there any potential sources of bias in how the exposures or outcomes are measured?

Selection bias ☐ Missclassification bias ☐ Recall bias ☐
Exclusion bias ☐ Surveillance bias ☐ Reporting bias ☐
Other sources of bias:

Give details of these sources of bias, and attempts made to manage them (if applicable):

Are there any issues of internal validity?

[v] 'Bedside tests' refers to tests that GPs perform during a consultation, such as measuring blood pressure or peak flow. 'Procedures' refers to non-diagnostic medical interventions, such as joint replacement or coronary angioplasty.

Results and conclusions

Reference	Research question and study design	Characteristics and confounders	Methods, outcomes and bias	Results and conclusions	Summary

(New) **Title**

Results

What are the main results of the study? Include numerical results where statistically significant.

Results are presented clearly ☐ **Results are presented appropriately** ☐

If the results are not presented clearly or appropriate, give details:

What statistical techniques are used?

Comment on the appropriateness of these techniques:

Do the statistical techniques adjust for confounding?

Are measures of both relative and absolute risk included? *Note: this is probably to go, to confirm after 6 test papers*

Is there any raw data? ☐

Conclusions

What conclusions do the authors make about the study questions/hypothesis?

Are these conclusions reasonable? **If not reasonable, why?**

The authors propose a reason/reasons to explain their findings ☐

If yes, what reasons do they propose?

The authors cite evidence for these reasons ☐ **If yes, what evidence do they cite?**

Other

The study was ethical ☐

There is a conflict of interest declaration ☐

Summary and overall impressions of the study [vi]

Reference | Research question and study design | Characteristics and confounders | Methods, outcomes and bias | Results and conclusions | **Summary**

(New) **Title**

Summary

What is the study's main strength?

What is the study's main weakness?

Would you use the results of this study?

Does this study provide information to help answer my research question?

How would you rate the quality of this study overall?

[vi] This section is for reviewers to make notes on their overall impressions of a study's relevance and quality. As discussed in Section 4.4 each study's final quality rating will be determined using a scoring system and based on discussion between both reviewers.

Appendix 4 : Template used to provide feedback to GPs after their first 'consultation' in the vignette study and an example of how this feedback is sent

Email template for GPs

After their first 'consultation' each GP was sent a feedback email providing advice about how to resolve any key errors they made when using the application. In order to ensure that GPs received standardised feedback we developed a template before recruitment commenced. Where applicable we gave examples of how to avoid these errors, using examples from their own behaviour/questions during the 'consultation', to avoid priming them.

NB: Yellow highlighting indicates where text varied between GPs.

Dear Dr *name here*

Congratulations on completing your first virtual consultation!

3 more consultations are uploaded and ready for you to undertake from *date here*

The GP decision making application interprets the questions you put in and selects an appropriate, pre-recorded, video to play in response. You may already have found out that it has some quirks.

This email gives *3 tips*, based on your first virtual consultation, to help you get the most out of the application:

Up to 3 tips were then listed here, in format shown below. If the GP encountered less than 3 types of issue then only the corresponding number of tips were included. Tips were only included if relevant to issues that arose during the GPs' first consultation.

1) ISSUE (e.g. Include a symptom name with your question)

You asked…	**Try…**
xxx	Xxx *(text from the suggested response here)*
xxx	xxx

Remember you can look at the help guide and troubleshooting questions at any point you are logged into the application by clicking on the link 'Help' in the top right hand corner of the screen. Or alternatively you can email us at gpstudy@ucl.ac.uk.

Best wishes,

The GP study team

List of issues GPs may encounter

The issues were ordered in priority - if a GP had made more than three errors during their 'consultation' only the first three on this list were noted in the feedback email (corresponding to the three most significant for their future use of the application). For the majority of issues examples of both the error and potential solutions were given. To avoid priming the example we gave solely based on GPs' own questions and/or behaviour during the first 'consultation'.

Issue	**Suggested response**
Include the symptom name with your question	Give examples from actual practice – italicising symptom name When you ask further questions about the current symptom (shown in the yellow bar) include the symptom name as part of, or after, your question. Without a symptom name the application may not recognise your question, so the answer it gives may not be appropriate. We appreciate it can be a matter of judgement to distinguish between when you are asking about a **new** symptom, and when you are asking further questions to probe a **current**. If in doubt, try both ways.
Returning to a previous symptom	Give examples from actual practice The yellow bar displays the current symptom or topic which the patient is talking about. Whenever you ask about a different symptom that the patient has (and the yellow bar changes to reflect this) a general video about this symptom will play initially, regardless of your question. Unfortunately, this is a quirk of the system we are not able to resolve. It will occur even if you are returning to a symptom you have previously asked about. Once the general video has played, or re-played, you will be able to ask further more specific questions about this symptom or topic. So, if you want to ask a question about a previous symptom, just type in the name of that symptom to play the general video, and then ask your follow up question. Note: You do not have to watch the whole general video playing again; you can stop it and ask your follow up question immediately
Make your questions specific to symptoms	Give examples from actual practice Patients consulted about this study indicated that they are not always clear what constitutes a symptom they should report to the doctor. Some patients therefore may find general questions hard to answer. Try asking questions about specific symptoms or topics.

Issue	Suggested response
Consultation lasts longer than 15 minutes	Each consultation is designed to take about 10 minutes. For the first time, it can take a little longer just to get familiar with the software but you may find the final consultations take much less than 10 minutes. It is perfectly acceptable to put your diagnoses thoughts and management plan as brief notes with abbreviations rather than full sentences. Note: the study is **not** a test of GPs' abilities. Rather than seeking the 'right answer', we are interested in what you would actually do faced with different scenarios. In some of the scenarios you will see, an optimal management plan may not be clear.
Dealing with a text error response	If you receive a text error response check: - you have included the current symptom name (if applicable) - for typos - if repeating the question gives a response - if rephrasing the question gives a response If none of these are successful the patient is unlikely to have any significant information to give. Note: each profile is different, so do seek this information again in subsequent consultations if you feel it is relevant – you might receive a different response.
Search elsewhere for your answers *given in conjunction with error response answer	Give examples from actual practice If you are unable to get an answer to a question despite trying error response steps, try looking in: - examinations or bedside tests - patient notes sidebar - historical notes from previous GP visits
Ask full questions	Give examples from actual practice Phrase and type questions how you would ask a patient in a true consultation – questions are more reliably interpreted by the system than single words
Avoid clinical jargon	Give examples from actual practice Phrase and type questions how you would ask a patient in a true consultation – the patient may not understand clinical terminology.
Check for typos	Give examples from actual practice – italicising typo Questions with typos may not be recognised (the application uses typing recognition software). If you get an answer which you do not expect, or that does not make sense, check your spelling.

Issue	Suggested response
Returning to the consultation whilst making your final note	It is possible to return to the consulting room and seek more information from the patient while making your final note. However if you do leave the 'Final Note' page before submitting it any information you have typed will be lost.

Feedback email example

An example of a finished feedback email is shown below (anonymised). This GP did not always include the current symptom name with their questions, asked the 'patient' non-specifically about additional symptoms, and took more than 15 minutes to complete the 'consultation'.

Dear Dr B,

Congratulations on completing your first virtual consultation!
Three more consultations are uploaded and ready for you to undertake from today, **x x 2013.**

The GP decision making application interprets the questions you put in and selects an appropriate, pre-recorded, video to play in response. You may already have found out that it has some quirks. This email gives 3 tips based on your first virtual consultation that might help you help you get the most out of the application in subsequent consultations.

1) Include the symptom name with your question

When you ask further questions about the current symptom (shown in the yellow bar) include the symptom name as part of, or after, your question. Without a symptom name the application may not recognise your question, so the answer it gives may not be appropriate.

You asked:	***Try:***
when did this start	when did the ankle swelling start
how long has this been the case	ankle swelling: how long has this been the case

We appreciate it can be a matter of judgement to distinguish between when you are asking about a **new** symptom, and when you are asking further questions to probe a **current** symptom. If in doubt, try both ways.

2) Make your questions specific to symptoms

Patients consulted about this study indicated that they are not always clear what constitutes a symptom they should report to the doctor. Some patients therefore may find general questions hard to answer. Try asking questions about specific symptoms or topics.

You asked:	***Try:***
do you have any other symptoms	*Questions such as:* do you have chest pain do you have any ankle swelling *Note: these were symptoms the GP themselves asked specifically about during the 'consultation', so as not to prime them for future consultations*

3) **Consultation length**

Each consultation is designed to take about 10 minutes. For the first time, it can take a little longer just to get familiar with the software but you may find the final consultations take much less than 10 minutes.

It is perfectly acceptable to put your diagnoses thoughts and management plan as brief notes with abbreviations rather than full sentences.

Note: the study is **not** a test of GPs' abilities. Rather than seeking the 'right answer', we are interested in what you would actually do faced with different scenarios. In some of the scenarios you will see, an optimal management plan may not be clear

Remember you can look at the help guide and troubleshooting questions at any point you are logged into the application by clicking on the link 'Help' in the top right hand corner of the screen. Or alternatively you can email us at gpstudy@ucl.ac.uk.

Best wishes,
Rachel Sequeira (on behalf of the GP study team)

Appendix 5 : 'Patient' template of all combinations of our four experimental factors which formed the basis of the vignette study's factorial design

Patient number	*Name*	*Profile*	*Description*	*Gender*	*Ethnicity*	*Socio-economic circumstance*
1	Jack Jones	1	58/59 years old Non-smoker Breathlessness and fatigue for 10 days	Male	White	Disadvantaged
2	Mercy Whyte	1		Female	Black Caribbean	Disadvantaged
3	Sachin Bhatia	1		Male	South Asian	Disadvantaged
4	Joanna Hampton	1		Female	White	Affluent
5	Winston Benjamin	1		Male	Black Caribbean	Affluent
6	Shalina Metha	1		Female	South Asian	Affluent
7	Jonathan Turner	2	58/59 years old Smoker Chest pain and cough for 10 days	Male	White	Affluent
8	Jeanette Wilson	2		Female	Black Caribbean	Affluent
9	Manish Prasad	2		Male	South Asian	Affluent
10	Jayne Peters	2		Female	White	Disadvantaged
11	Marcus Blake	2		Male	Black Caribbean	Disadvantaged
12	Meena Patel	2		Female	South Asian	Disadvantaged

Low risk:
'watch and wait' appropriate

Patient number	*Name*	*Profile*	*Description*	*Gender*	*Ethnicity*	*Socio-economic circumstance*
13	William Talbot	3	78/79 years old Smoker Chest pain and cough duration uncertain (~3 weeks)	Male	White	Affluent
14	Elizabeth Cleveland	3		Female	Black Caribbean	Affluent
15	Rohan Dhoni	3		Male	South Asian	Affluent
16	Lucy Norton	3		Female	White	Disadvantaged
17	Clive Marshall	3		Male	Black Caribbean	Disadvantaged
18	Arundati Sharma	3		Female	South Asian	Disadvantaged
19	Bill Davidson	4	78/79 years old Non-smoker Cough and appetite loss, duration uncertain (~3 weeks)	Male	White	Disadvantaged
20	Dorsey Gardner	4		Female	Black Caribbean	Disadvantaged
21	Ranjeev Chaudhury	4		Male	South Asian	Disadvantaged
22	Mary Graham	4		Female	White	Affluent
23	Dwight Smith	4		Male	Black Caribbean	Affluent
24	Gita Banerjee	4		Female	South Asian	Affluent

Medium risk:
<u>either</u> 'watch and wait' (with safety-netting)
<u>or</u> refer for chest X-ray appropriate

Patient number	*Name*	*Profile*	*Description*	*Gender*	*Ethnicity*	*Socio-economic circumstance*
25	Nicholas Mortimer	5	58/59 years old Smoker with COPD Breathlessness and fatigue for >1 month	Male	White	Affluent
26	Rosemary Campbell	5		Female	Black Caribbean	Affluent
27	Manjit Laxman	5		Male	South Asian	Affluent
28	Margaret Johnson	5		Female	White	Disadvantaged
29	Jerome Bishop	5		Male	Black Caribbean	Disadvantaged
30	Rupal Shah	5		Female	South Asian	Disadvantaged
31	Leslie Johns	6	78/79 years old Smoker Chest pain and weight loss for >1 month	Male	White	Disadvantaged
32	Ruth Lashley	6		Female	Black Caribbean	Disadvantaged
33	Sunil Bopanna	6		Male	South Asian	Disadvantaged
34	Eileen Evans	6		Female	White	Affluent
35	Maxwell Jacobs	6		Male	Black Caribbean	Affluent
36	Preeti Joshi	6		Female	South Asian	Affluent

High risk:
immediate referral for chest X-ray appropriate

Appendix 6 : List of the GP characteristics we examined in the GP decision making study

GP practice characteristics examined	Source of this information
Region List size Number of GPs Training practice Area socio-economic profile Cancer referral rate Cancer detection rate *(proportion of patients diagnosed via the two-week wait pathway)* Cancer conversion rate *(proportion of two-week wait referrals found to have cancer)*	Routine data (the National General Practice Profiles and the National Cancer Information Network's general practice profiles)

GP individual characteristics examined	Source of this information
Age Gender Ethnicity Years since qualification Role/position in practice Sessions worked per week Confidence with computers	Registration questionnaire
Specialty experience Budgetary responsibility Smoking status	Post-consultation survey

Appendix 7 : List of key questions GPs are likely to ask patients presenting with the symptoms we investigated in the vignette study

Appetite loss

Are you on a diet?

Have you changed your eating habits?

Do you get full easily?

What are you eating?

Have you lost weight?

What makes it worse?

Do you feel sick?

Do you have abdominal pain?

Have you had a change in bowel habit?

Have you been ill recently?

How long has this been going on?

Breathlessness

What makes it better?

What makes it worse?

How long have you been breathless?

How far can you walk?

Is it worse on exercise?

Is it worse when you lie down?

Does it stop your normal activities?

Can you carry things?

Have you ever had this before?

Do you have chest pain?

Do you have swollen ankles?

Have you had calf swelling?

Do you have asthma?

Do you have COPD?

Are you a smoker?

Do you have heavy periods?

Chest pain

Describe the pain?

Is it continuous?

How long have you had this?

What brings the pain on?

What makes it worse?

What makes it better?

Is it worse on movement?

Is it worse on exercise?

Is it worse on eating?

Is it worse when you take a breath?

Where is the pain?

Does it hurt to touch it?

Does it radiate anywhere?

Have you had any palpitations?

Do you feel sick?

Do you have a family history of heart disease?

Cough

How long have you had this?

Are you coughing anything up?

What colour is your phlegm?

Are you coughing up blood?

Have you been ill recently?

Do you have a fever?

Do you have chest pain?

Are you short of breath?

Can you describe your cough?

Fatigue

How long have you been feeling like this?

Does anything help?

How are you sleeping?

Do you snore?

How is work?

Are you working too hard?

How are things at home?

Has anyone changed?

Are you able to do your normal activities?

Are you breathless?

Do you feel cold?

How is your mood?

Weight loss

How much weight have you lost?

How long has this been happening?

Are your clothes looser?

Are you on a diet?

Is your weight loss intentional?

Have your eating habits changed?

Have you lost your appetite?

Do you feel sick?

Do you have abdominal pain?

Have your bowel habits changed?

Have you been ill recently?

Other questions

Do you smoke?

Does anyone in the house smoke?

Have you every smoked?

How much do you smoke?

How often do you smoke?

What is your job?

Do you have any pets?

What do you do in your spare time?

Appendix 8 : Examples of 'patient' briefs sent to the actors for the vignette study

Profile 1 vignette id 1: 58/59 year old non-smoker, experiencing breathlessness and fatigue

NB: this was the 'distracting' profile designed to suggest heart failure

About you

LOW SOCIAL CLASS: Your character is a white British man, aged ~60 years old. You work as security staff in a block of offices, but will be retiring in the next couple of years. You are married with children though they left home a while ago and now have children of their own, who you enjoy seeing when you are able to. In your spare time you like to box (these days you coach more than competing) although you have not been able to since becoming unwell. You enjoy watching sports on the TV and will often spend an evening down at the bookmakers with your friends.

Why have you come to see your GP?

You're feeling breathless. You've never felt like this before and are not sure what's going on. It's interfering with your life (e.g. you now have to get the bus into work rather than walking) and so your wife suggested you come and check it out.

When questioned further about your breathlessness

You notice it particularly when you're being active (e.g. you're unable to box at the moment, and struggle playing with the grandchildren). However even minor activities like walking down the street or to the doctors' surgery seem to bring it on. You have to stop to catch your breath every 200m or so. You also notice it when you lie down in bed, and have had to start using one of your wife's pillows as well as your own to help. It happens several times a day: whenever you do anything to exert yourself. It only seems to ease when you stop and rest at home. It's been happening for 10 days (e.g. you haven't been able to make boxing training for the last week because of it).

Do you have any other symptoms? (we will ask about these separately)

1) You are also feeling extremely tired. You had the 'flu a couple of months ago but thought you were recovered from that – you've been back at work for the last month. You're not sure why you're feeling so tired: your workload is the same as usual, nothing has changed at home. You aren't sleeping very well, but you've put this down to the breathlessness and the difficulty lying flat. You've been feeling this tired for the last 10 days or so.
2) You have noticed your ankles swelling a bit, particularly at the end of the day and when you've been on your feet a lot. They improve a bit if you put your feet up. They are not painful. You've never had anything like this happen to them before. You first noticed it a couple of weeks ago.

What else do you need to know/will we ask you about?

Smoking*:* You do not smoke, and never have. Nobody in your house smokes either.

We will also ask you to give some generic responses.

"No"*:* We will ask you to reply in the negative for a series of questions (e.g. "No", "No I don't have that" or a similar phrase). These might include: You don't have allergies, you don't have a cough, you haven't lost weight. These clips will be played if the GP asks about a range of symptoms that you don't have.

I don't understand*:* We will also ask you to query questions in 1-2 different ways (along the lines of "I don't understand", "Can you rephrase that?"). These will be played if the GP asks something that is not recognised by the system.

Profile 2 vignette id 10: 58/59 year old smoker, experiencing chest pain and cough

About you

HIGH SOCIAL CLASS: Your character is a South Asian man, aged ~60 years. You are still working as a teacher in a secondary school, though thinking of retiring. You are married, your children have left home but visit often with grandchildren. You like spending time with the family which you do most weekends.

Why have you come to see your GP?

You have a <u>pain in your chest</u>. You're trying to get on with your job and normal activities as much as you can, but it niggles a bit. It's unusual for you – you've not had anything like this before – that's why you've come to the doctor today.

When questioned further about your chest pain

It's a kind of dull aching pain (not a sharp or stabbing pain). You feel it pretty much all the time *[we'll show you whereabouts in your chest and will ask you to point there in response to a question about where you feel it*]. Sometimes when you breathe in deeply you feel it more. A painkiller helps. You have had the pain for the last 10 days or so and it affects your life. (e.g. you first noticed it after dinner for your daughter's birthday not last Saturday but the one before).

Do you have any other symptoms?

You always have a bit of a <u>cough</u> in the mornings (where you cough up a bit of white stuff) but your cough has got worse recently. You are not coughing up any more phlegm or any blood, but the cough has become more constant and it's getting on your nerves. You cough several times a day now, and it can seem to come on at any time or doing any activity. You haven't found anything that eases it. It's been going on about 1-2 weeks.

****NOTE: YOU HAVE A COUGH, SO WE WILL REMIND YOU TO COUGH THROUGHOUT!****

What else do you need to know/will we ask you about?

Smoking*:* You are a smoker. You usually smoke 20 cigarettes/ 10 cigarettes a day *[we'll film both*]. You have smoked for many years.

Family history*:* You're not aware of a family history of any diseases.

We will also ask you to give some generic responses.

"No"*:* We will ask you to reply in the negative for a series of questions (e.g. "No", "No I don't have that" or a similar phrase). These might include: You don't have allergies, you don't have a cough, you haven't lost weight. These clips will be played if the GP asks about a range of symptoms that you don't have.

I don't understand*:* We will also ask you to query questions in 1-2 different ways (along the lines of "I don't understand", "Can you rephrase that?"). These will be played if the GP asks something that is not recognised by the system.

Profile 4 vignette id 20: 78/79 year old non-smoker, experiencing cough and loss of appetite

About you

LOW SOCIAL CLASS: Your character is a black Caribbean woman, aged ~85 years. You are widowed, and live with one your daughters and her family. You were a housewife, and your husband used to work in a local hardware store. You enjoy going to bingo a couple afternoons a week with friends, and watch a lot of TV.

Why have you come to see your GP?

You have a <u>cough</u>. You've had it for a while and it is not going away. Your daughter has noticed and suggested you came to see the doctor. You did have a bad cough a couple years ago when you had flu, but you've had the jab every year since and been fine.

When questioned further about your cough

You're coughing regularly – short single coughs but every few minutes *[some of this may be more effectively demonstrated by your cough in the film rather than words]*. You are not sure how long you've had it, but you were fine at your great-granddaughter's birthday a month ago. You are not coughing up any phlegm or any blood, but the cough has become more constant and it's getting on your (and your family's) nerves. You cough whatever you are doing – nothing specific seems to make it worse. You have tried taking cough mixture but it hasn't made any difference.

Do you have any other symptoms?

You've been a bit <u>off your food</u> over the last few weeks as well, though you haven't changed your diet at all. You find yourself leaving half of what is on your plate, as you just can't manage any more, which has led to a couple of arguments with your daughter. You have had no nausea, vomiting or change in bowel habit.

****NOTE: YOU HAVE A COUGH, SO WE WILL REMIND YOU TO COUGH THROUGHOUT!****

What else do you need to know/will we ask you about?

Smoking: You have never smoked, and nor do any of your family.

Weight: Your weight is stable.

We will also ask you to give some generic responses.

"No": We will ask you to reply in the negative for a series of questions (e.g. "No", "No I don't have that" or a similar phrase). These might include: You don't have allergies, you don't have a cough, you haven't lost weight. These clips will be played if the GP asks about a range of symptoms that you don't have.

I don't understand: We will also ask you to query questions in 1-2 different ways (along the lines of "I don't understand", "Can you rephrase that?"). These will be played if the GP asks something that is not recognised by the system.

Appendix 9 : Filming checklists for the vignette study

PROFILE 1 – dyspnoea, fatigue, 60yr, non-smoker

Actor: ______________________ Checklist by: ________________

DYSPNOEA

- ☐ Presentation
  Is the clip 30-45s long?
- ☐ What makes it worse?
  Including lying down
- ☐ What makes it better?
- ☐ Duration
  10days / <2weeks
- ☐ How often?
- ☐ Worse on exercise
- ☐ Change from normal ability
  Less distance possible

SWOLLEN ANKLES

- ☐ Presentation
- ☐ Duration
  10days / <2weeks

RECENT ILLNESS

- ☐ Flu (recovered)

OTHER

- ☐ JOB
- ☐ PETS Yes / No (circle answer)
- ☐ SPARE TIME
- ☐ TAKING THEIR MEDICATION

FATIGUE

- ☐ Presentation
- ☐ Tired all the time
- ☐ Duration
  10days / <2weeks
- ☐ No life changes
- ☐ Change from normal ability
- ☐ Sleeping with extra pillow

SMOKING

- ☐ Non smoker and no exposure

ILLNESS IDEAS

- ☐ No idea what is wrong

GENERIC

- ☐ Don't understand
- ☐ Rephrase question
- ☐ Don't have that
- ☐ Not noticed that
- ☐ Don't think so
- ☐ No

***PROFILE 2** – chest pain, cough, 60yr, smoker*

Actor: ____________________ Checklist by: ______________

CHEST PAIN

- ☐ Presentation
  Is the clip 30-45s long?
- ☐ What makes it worse?
  NB: sharp breath not sharp pain
- ☐ What makes it better?
- ☐ Duration
  10days / <2weeks
- ☐ Location
  Left nipple, 2 fingers
- ☐ Can feel it at any time/activity
- ☐ Describe the pain
- ☐ Not worse on exercise
- ☐ No radiating

FAMILY HISTORY

- ☐ None

ILLNESS IDEAS

- ☐ No idea what is wrong

OTHER

- ☐ JOB
- ☐ PETS Yes / No (circle answer)
- ☐ SPARE TIME
- ☐ TAKING THEIR MEDICATION

COUGH

- ☐ Presentation
  Change from normal cough
- ☐ What makes it worse?
- ☐ What makes it better?
- ☐ Duration
  10days / <2weeks
- ☐ Hard to stop once start
- ☐ Disturbs partner's sleep
- ☐ Exercise can trigger coughing
- ☐ Productive (same for years)
- ☐ No blood

SMOKING

- ☐ Smoker
- ☐ How much?
- ☐ Since when?

GENERIC

- ☐ Don't understand
- ☐ Rephrase question
- ☐ Don't have that
- ☐ Not noticed that
- ☐ Don't think so
- ☐ No

PROFILE 3 – chest pain, cough, 85yr, smoker

Actor: ______________________ Checklist by: ________________

CHEST PAIN

- [ ] Presentation
  Is the clip 30-45s long?
- [ ] What makes it worse?
  NB: sharp breath not sharp pain
- [ ] What makes it better?
- [ ] Duration
  Uncertain (> 2 weeks, < 4 weeks)
- [ ] Location
  Left nipple, 2 fingers
- [ ] Can feel it at any time/activity
- [ ] Describe the pain
- [ ] Not worse on exercise
- [ ] No radiating

FAMILY HISTORY

- [ ] None

ILLNESS IDEAS

- [ ] No idea what is wrong

OTHER

- [ ] JOB (retired and past)
- [ ] PETS Yes / No (circle answer)
- [ ] SPARE TIME
- [ ] TAKING THEIR MEDICATION

COUGH

- [ ] Presentation
  Change from normal cough
- [ ] What makes it worse?
- [ ] What makes it better?
- [ ] Duration
  Uncertain (> 2 weeks, < 4 weeks)
- [ ] Hard to stop once start
- [ ] Exercise can trigger coughing
- [ ] Productive (same for years)
- [ ] No blood

SMOKING

- [ ] Smoker
- [ ] How much?
- [ ] Since when?

GENERIC

- [ ] Don't understand
- [ ] Rephrase question
- [ ] Don't have that
- [ ] Not noticed that
- [ ] Don't think so
- [ ] No

PROFILE 4 – *cough, loss of appetite, 85yr, non-smoker*

Actor: ______________________ Checklist by: ________________

COUGH

- ☐ Presentation
  Is the clip 30-45s long?
- ☐ What makes it worse?
- ☐ What makes it better?
- ☐ Duration
  Uncertain (> 2 weeks, < 4 weeks)
- ☐ Continuous
- ☐ Not worse on exercise
- ☐ Not productive
- ☐ No blood

SMOKING

- ☐ Non smoker and no exposure

ILLNESS IDEAS

- ☐ No idea what is wrong

OTHER

- ☐ JOB (retired and past)
- ☐ PETS Yes / No (circle answer
- ☐ SPARE TIME
- ☐ TAKES THEIR MEDICATION

LOSS OF APPETITE

- ☐ Presentation
- ☐ Gradually increasing/getting worse
- ☐ Duration
  Uncertain (> 2 weeks, < 4 weeks)
- ☐ Every meal
- ☐ How much can you eat?
  Eats about half a plate
- ☐ No change in type of food
- ☐ No change in bowel habit

WEIGHT LOSS

- ☐ None, but off food

GENERIC

- ☐ Don't understand
- ☐ Rephrase question
- ☐ Don't have that
- ☐ Not noticed that
- ☐ Don't think so
- ☐ No

PROFILE 5 – dyspnoea, fatigue, 60yr, smoker, COPD

Actor: ____________________

Checklist by: ______________

DYSPNOEA

- ☐ Presentation v.1
  Antibiotics 2 weeks ago no help
  Is the clip 30-45s long?
- ☐ Presentation v.2
  Sputum darker than normal
  Is the clip 30-45s long?
- ☐ What makes it <u>worse</u>?
- ☐ What makes it <u>better</u>?
- ☐ Duration
  Six weeks
- ☐ How often?
- ☐ Worse on exercise
- ☐ Change from normal ability
  Less distance possible

COPD

- ☐ Had for a few years
- ☐ Sputum

OTHER

- ☐ JOB
- ☐ PETS Yes / No (circle answer)
- ☐ SPARE TIME
- ☐ TAKING THEIR MEDICATION

FATIGUE

- ☐ Presentation
- ☐ Tired all the time
- ☐ Duration
  One month
- ☐ No life changes
- ☐ Change from normal ability
- ☐ Sleeping well

SMOKING

- ☐ Smoker
- ☐ How much?
- ☐ Since when?

ILLNESS IDEAS

- ☐ No idea what is wrong

GENERIC

- ☐ Don't understand
- ☐ Rephrase question
- ☐ Don't have that
- ☐ Not noticed that
- ☐ Don't think so
- ☐ No

***PROFILE 6** – chest pain, weight loss, 85yr, smoker*

Actor: ____________________

Checklist by: ______________

CHEST PAIN

- [ ] Presentation
  Is the clip 30-45s long?
- [ ] What makes it worse?
  NB: sharp breath not sharp pain
- [ ] What makes it better?
- [ ] Duration
  Six weeks
- [ ] Location
  Right nipple, 2 fingers
- [ ] Can feel it at any time/activity
- [ ] Describe the pain
- [ ] Not worse on exercise
- [ ] No radiating

FAMILY HISTORY

- [ ] None

ILLNESS IDEAS

- [ ] Thought indigestion but is not

OTHER

- [ ] JOB (retired and past)
- [ ] PETS Yes / No (circle answer)
- [ ] SPARE TIME
- [ ] NO REGULAR MEDICATION

WEIGHT LOSS

- [ ] Presentation
  Despite normal appetite
- [ ] No life changes
- [ ] Duration
  One month
- [ ] How much? (number)
- [ ] How much? (clothes)
- [ ] No change in diet
- [ ] No change in bowel habit

SMOKING

- [ ] Smoker
- [ ] How much?
- [ ] Since when

GENERIC

- [ ] Don't understand
- [ ] Rephrase question
- [ ] Don't have that
- [ ] Not noticed that
- [ ] Don't think so
- [ ] No

Appendix 10 : Symptom bank and symptom keywords for the vignette study

Symptom	Keywords
Allergies	allergy allergies allergic atopic atopy hay fever hay-fever hayfever
Angina	angina heart attack heart attacks heart-attack myocardial infarction myocardial infarctions heart attach heart attachs
Anxiety	anxiety anxious concerned concern alarmed afraid nervous alarm fear phobia phobic

Symptom	Keywords
Anxiety *(continued)*	panic anxeity worried stress stressed stressful
Appetite loss	anorexia hunger appetite not hungry not eating feel hungry feeling hungry off your food eating less eating normally
Arm pain	arm pain pain in your arm pain in the arm pain in right arm pain in your right arm pain in the right arm pain in your left arm pain in the left arm pain in left arm arm hurt arms hurt
Arthritis	stiff joints joint osteo-arthritis arthritis osteoarthritis

Symptom	Keywords
Arthritis *(continued)*	arthralgia arthralgias myalgia
ASK ABOUT NEW OR PREVIOUS TOPIC	otherwise well well otherwise other symptoms additional symptoms other symptom another symptom additional symptom other problems additional problems other problem additional problem another problem is there anything else do you have anything else anything else how do you feel in general how do you feel generally how do you generally feel how do you feel otherwise
Asthma	asthma asthmatic
Back pain	Back backache spine lumbar

Symptom	Keywords
Back pain *(continued)*	thoracic back-ache vertebrae vertebra backaches back-aches
Bloating	bloat bloated bloating distend distended gas gaseous burp burping flatulence fart farting pass wind passing wind
Bowel habits	bowel motion bowel motions bowel habit bowel habits toilet you regular things regular it regular clockwork motions stool stools down there

Symptom	**Keywords**
Bowel habits *(continued)*	bowl motion bowl motions bowl habit bowl habits bowel movements bowl movements pooh poo defecate defaecate defecation defecating faeces constipated constipation constipate
Breast problems	breast
Breathlessness	shortness breathless breathlessness breathe dyspnoea puff short of breath lost breath lose breath catch breath breatlessness breatless breathing difficulty breathing trouble breathing out of breath

Symptom	**Keywords**
Cardiac history	cardiac history heart history history of heart problem with heart problems with heart problem with the heart problems with the heart problem with your heart heart trouble heart problem heart problems cardiac trouble cardiac problem cardiac problems
Chest pain	chest lung pain lungs hurt sore lung sore lungs lung ache lungs ache lungs aching cherst
Common cold	got a cold had a cold getting a cold have a cold common cold coldy
COPD	copd COPD

Symptom	Keywords
COPD *(continued)*	chronic obstructive pulmonary disease lung disease pulmonary disease pulmanary bad lungs bad chest lung problem
Cough	chesty coughing cough coughy hacking hack
Current medication	medication medications medicine medicines treatment treatments tablet tablets pill pills drug drugs prescribed prescription prescriptions
Depression	depression depressed depress depressive miserable

Symptom	Keywords
Depression *(continued)*	upset sad mood bipolar tearful happy unhappy
Diabetes	diabetes blood sugar blood sugars mellitus DM sugars sugar levels sugar level
Diarrhoea	diarrhoea diarrhea loose stool loose stools runny stool runny stools loose poo runny poo
Faint	faint faints fainted fainting collapse collapse collapsed collapsing fits fall

Symptom	Keywords
Faint *(continued)*	falls funny turn funny turns seizure seizures blackout blackouts black out black outs blacked out blacking out dizzy dizziness woozy wooziness woosiness
Fatigue	tiredness tired energy lethargic lethargy drained exhaustion exhausted fatigue fatigued sluggish knackered pooped
Fever	fever temperature feverish pyrexial pyrexia

Symptom	Keywords
Fever *(continued)*	hot feel warm feeling warm feel too warm feeling too warm
Foreign travel	abroad travelled travel foreign exotic flight flights flying flown aeroplane aeroplanes plane
Haemoptysis	blood specks haemoptysis hemoptysis rusty rust-coloured rust coloured
Hand problem	wrist hand wrists finger fingers hands thumb thumbs nail

Symptom	Keywords
Hand problem *(continued)*	Nails fingernail fingernails
Headache	Headache migraine head migranous headaches migraines head-ache head-aches
Hip problem	Hips hip thigh thighs femur femurs
Hoarseness	Hoarse hoarseness voice croak croaky
Hospital	Hospital hopsital consultant A + E A+E a + e a+e A&E a&e A and E a and e

Symptom	Keywords
Hospital *(continued)*	accident and emergency casualty
Indigestion	heartburn heart burn heart-burn reflux acid indigestion oesophagus oesophageal esophagus esophageal
Injuries	injury injuries injured injuring accident accidents have you hurt broken rib broken a rib broken ribs ribs broken rib broken
Irritable bowel syndrome	irritable bowel IBS
Jaundice	jaundice jaundiced yellow skin yellow eyes yellow eye yellowish skin

Symptom	Keywords
Jaw pain	Mouth jaw jaws cheek cheeks
Job	job jobs labour profession professional occupation retired for a living retiring you work your work line of work do you do you working to retire for work as work
Kidney problems	kidney kidneys renal ureter
Knee pain	knee knees
Leg pain	leg pain leg pains pain in the leg pain in the legs pains in the leg

Symptom	Keywords
Leg pain *(continued)*	pains in the legs pain in your leg pains in your leg pain in your legs pains in your legs leg hurt legs hurt legs ache leg ache leg aches
Liver problems	liver
Nasal problems	nose nostrils nasal congested post-nasal drip postnasal drip
Neck pain	neck-ache vertebrae vertebra neck ache neck pain neck problem neck problems pain in your neck pain in the neck pain in neck
Night sweats	sweat sweaty sweating sweats hot flush hot flushes

Symptom	Keywords
Night sweats *(continued)*	hot and cold shivery shivers shivering swating swats
Nosebleed	nose-bleed nose-bleeds nose bleed nose bleeds nosebleed nosebleeds nose bleeding bleeding from the nose
Numbness	numb numbness tingle tingles tingling pins and needles pins-and-needles parasthesia parasthesiae
Palpitations	irregular beat fast beat quick beat quickly beating fast beating quick beating quickly heart-beat heart beat heartbeat

Symptom	Keywords
Palpitations *(continued)*	heartbeats heart-beats heart beats heart rate heartrate heart-rate skip a beat skips a beat skipping a beat miss a beat misses a beat missing a beat palpitation palpitations
Past antibiotics	ciprofloxacin flucloxacillin metronidazole penicillin trimethoprim
Pets	pets birds animals cats dogs rabbits parrots pigeons horses cows sheep pigs puppies

Symptom	**Keywords**
Pets *(continued)*	kittens a pet a bird an animal a cat a dog a rabbit a parrot a pigeon a horse a cow a pig a puppy a kitten a unicorn pet
Rash	rash rashes rashs itch itchy itchiness itching hives weals cellulitis
Shoulder pain	shoulder pain shoulder ache shoulder pains
Sickness	nausea nauseous sickness vomit

Symptom	**Keywords**
Sickness *(continued)*	vomiting vomited feel sick feeling sick puke puked throw up throwing up thrown up
Sinusitis	sinus sinuses sinusitis
Smoking	smoking smoker smoked smoky smokers you smoke still smoke home smoke family smoke around smoke partner smoke husband smoke wife smoke else smoke anyone smoke she smoke he smoke they smoke smokinh
Sore throat	throat tonsil

Symptom	**Keywords**
Sore throat *(continued)*	tonsils mouth
Spare time	home life spare time hobbies pastimes pass time spend time pass your time spend your time pass the time spend the time like to do like to get interests occupy
Stomach ache	stomach abdomen abdominal tummy gut belly
Swallowing problems	swallowing swallow swallowed swallows
Swollen ankles	swelling swollen swells ankles feet foot swelled

Symptom	**Keywords**
Swollen ankles *(continued)*	oedema edema edaema odaema
Taking inhalers	puffer puffers salbutamol beclometasone beclomethasone inhaler inhalers
Thirst	thirst thirsty drinking more drinking lots drinking a lot drink a lot drink lots drink more
Tuberculosis	tuberculosis TB infectious contact infectious contacts infected contact infected contacts infectious people infectious person infected people infection person anyone infected anyone infectious someone infected someone infectious

Symptom	Keywords
Tuberculosis *(continued)*	similar symptoms similar symptom similar problems similar problem same problem same problems family member family members anyone else anyone at home anyone at work
Urinary symptoms	urine burn burning hesitance urgency burns stinging stings sting wee bladder urethra penis penile erectile urinary
Weight loss	weight size skinny slimmed slimmer bony
Weight loss *(continued)*	lighter you weigh you weigh weighed wieght wieghed you wiegh
Wheeze	wheeze wheezy wheezing wheezes wheezey wheezed

Appendix 11 : Vignette study 'symptom topics' and their keywords and key phrases

Symptom topic	Keywords/ key phrases
Onset	what brings exacerbates what triggers makes it happen start to happen causes exacerbate aggravate aggravates agrivate aggrivate agrivates aggrivates especially bad aggrevate aggrevates makes it worse exacerbation pleuritic plueritic deep breath taking a breath take a breath breathing in breathe in breath in
Offset	what stops when does it stop

Symptom topic	Keywords/ key phrases
Offset *(continued)*	what makes it stop better helps help eases ease relieve relieves reduces alleviate aleviate aleviates alleviating alleviated alleviates does it stop when does it stop if lessens subside relief makes it go away you stop it anything for what is different how is it different painkiller painkillers aspirin paracetemol ibuprofen nurofen neurofen improves

Symptom topic	**Keywords/ key phrases**
Offset *(continued)*	improve analgesia analgesics analgesic analgesias
Duration	happened before had this before had it before had before first notice duration start recently started recently weeks months how many days have how many years have how long when did since when from when over what time
Until	until giving up give up stop does it last still there ongoing on going did it end did they end did this end

Symptom topic	**Keywords/ key phrases**
Until *(continued)*	stopping
Describe	how bad how badly intense intensity how severe severity describe bearable severe mild feel like what is it like what type tell me about tell me a little about tell me a bit about tell me more tell more tell me a little more tell a little more tell me a bit more tell a bit more explain more explain that more explain a bit more explain a little more explain what portion portions can you finish can you eat

Symptom topic	**Keywords/ key phrases**
Describe *(continued)*	do you eat are you eating can you manage do you consume had enough how painful
Frequency	how much what amount lost a lot how often frequently frequent many times how many often happen a lot a regular regularly come and go constant constantly continuous continuously continually continual when do you always there all the time what meals any meals some meals certain meals particular meals

Symptom topic	**Keywords/ key phrases**
Frequency *(continued)*	when does it happen all day starts every day how oftern how iften
Exercise	exercise exercising exercises exerting exert exertion on activity on activities exervise exercuise exercive execise exersise exersises exersising with activity with activities doing activity doing activities walking a walk walked running a run jogging jogged a jog

Symptom topic	Keywords/ key phrases
Exercise *(continued)*	you walk
Lying down	lie lay lying laying
Change in activities	affect your lifestyle effect on your lifestyle affect you affect your life effect on your life what you can do capabilities capability able to do ability abilities how far far can distance effect your lifestyle affect on your lifestyle effect you effect your life affect on your life stairs at work affecting work effecting work affecting your work effecting your work
Life changes	anything new stress

Symptom topic	Keywords/ key phrases
Life changes *(continued)*	stressed stressful any worry been worried been worrying been a change any change any changes major change major changes anything changed anything changing is there a change are there changes anything significant something significant life event life events how is work how is the job how is your job how is your work how is home how are things at home how is the family how are the family how is family life how is your family life how is home life how is your home life
Location	where whereabouts location

Symptom topic	**Keywords/ key phrases**
Location *(continued)*	point to area which part which bit
Diet	diet what you eat what you are eating what you're eating what your eating cutting back cut back restricted restricting restrict foods cooking intentional intention deliberate deliberately are you trying been trying are you glad are you pleased are you happy eating habits eating normal eating normally eating the same eating as usual eating the usual eating what you what are you eating

Symptom topic	**Keywords/ key phrases**
Diet *(continued)*	what you eat what do you eat
Bowel habits	bowel motion bowel motions bowel habit bowel habits toilet you regular things regular it regular clockwork motions stool stools down there bowl motion bowl motions bowl habit bowl habits bowel movements bowl movements pooh poo defecate defaecate defecation defecating faeces bowels bowls
Movement	move movement

Symptom topic	**Keywords/ key phrases**
Movement *(continued)*	motion moving
Radiating	radiate radiating spread spreading spreads pain move go anywhere anywhere else down your arm up your neck in your neck up the neck in the neck to your neck to the neck your jaw the jaw go naywhere another part go anyhere anyhere else
Sleep apnoea	apnoea apnoeas apnea apneas snore snorer snores snoring
Sleeping	sleep slept

Symptom topic	**Keywords/ key phrases**
Sleeping *(continued)*	sleeping enough rest insomnia kept awake keep awake keeps awake keep you awake keeps you awake kept you awake night nighttime night-time pillow pillows wake waking
Family history	family have heart family history parents parent mother father relatives relative anyone have heart hereditary anyone in your family anyone in the family anyone in family
Recent illness	been unwell unwell recently been ill ill recently

Symptom topic	Keywords/ key phrases
Recent illness *(continued)*	recent illness illness recently flu influenza 'flu flue
Illness ideas	idea ideas concern concerns expectation expectations thoughts thought guesses what is wrong what the problem is what the matter is what the trouble is what do you think why do you think could be
Medication	medication medications medicine medicines treatment treatments tablet tablets inhaler inhalers pill

Symptom topic	Keywords/ key phrases
Medication *(continued)*	pills drug drugs prescribed prescription prescriptions
Additional symptoms	other symptoms other problems other complaints other issues other difficulties another symptom another problem another complaint another issue another difficulty additional symptoms additional problems additional issues additional difficulties additional complaints additional symptom additional problem additional issue additional difficulty additional complaint anything else is that all is that everything is that the only thing is that the only problem is this the only thing

Symptom topic	**Keywords/ key phrases**
Additional symptoms *(continued)*	is this the only problem otherwise well well otherwise other symptom other problem how do you feel in general how do you feel generally how do you generally feel how do you feel otherwise
Haemoptysis	haemoptysis hemoptysis blood rusty rust-coloured rust coloured specks
Sputum	productive phlegm sputum mucus mucous anything up bringing up bring up brought up coughing up cough up coughed up spitting up

Symptom topic	**Keywords/ key phrases**
Sputum *(continued)*	spit up producing anything coughing anything spitting anything anythng up anthing up phlem flem flegm up anything dry
Worse with food	you eat you have eaten you ate meals food after eating during eating when eating if eating because of eating eating make
Catch all	Worse happen start make you more trigger triggers exacerbating exacerbated exacerbate

Appendix 12 : Sample profile upload instructions for the vignette study

The following is an example of the upload instruction documents I produced for each of the six 'patient' profiles.

Profile 1: 58/59 year old non-smoker with breathlessness, fatigue and swollen ankles

DU = Don't understand video

	Breathless (B)	*Fatigue (F)*	*Swollen ankle*	*Smoking*	*Occupation*	*Spare time*	*Pets*	*ASK SPECIFIC*	*Diabetes*	*Medication*
Presentation	Initial presentation	Presentation (F)	Ankles video	Smoker	Occupation	Spare time	Pets	Rephrase	Medication	Medication
Onset	Worse (B)	All the time (F)	Ankles video	DU	DU	DU	DU		DU	DU
Offset	Better (B)	All the time (F)	Feet up (ankles video if none)	DU	DU	DU	DU		DU	DU
Duration	How long (B)	How long (F)	How long (SA)	DU	DU	DU	DU		DU	DU
Until	DU	All the time (F)	DU	DU	DU	DU	DU		DU	DU
Describe	How often (B)	Long all the time or activity change	Ankles video	DU	DU	DU	DU		DU	DU
Frequency	How often (B)	All the time (F)	DU	DU	DU	DU	DU		DU	DU
General	No	No	No	No	DU	DU	DU		No	DU
Exercise	Exercise (B)	Not noticed	Not noticed	DU	DU	DU	DU		DU	DU
Lying down	Lying down (B) Worse if none	Not noticed	DU	DU	DU	DU	DU		DU	DU
Change in activities	Activity (B)	Activity (F) if none activity (B) or DU	DU	DU	DU	DU	DU		Activity (B)	DU
Life changes	Life change (F)	Life change (F)	Life change (F)	Life change (F)	Life change (F)	Life change (F)	Life change (F)		Life change (F)	Life change (F)
Location	DU	DU	DU	DU	DU	DU	DU		DU	DU
Sputum	DU	DU	DU	DU	DU	DU	DU		DU	DU
Haemoptysis	Don't have	Don't have	Don't have	Don't have	Don't have	Don't have	Don't have		Don't have	Don't have
Diet	DU	DU	DU	DU	DU	DU	DU		DU	DU
Bowel habits	DU	DU	DU	DU	DU	DU	DU		DU	DU
Movement	Exercise (B)	Not noticed	Not noticed	DU	DU	DU	DU		DU	DU
Radiating	DU	DU	DU	DU	DU	DU	DU		DU	DU
Sleep apnoea	Don't think so	Don't think so	Don't think so	Don't think so	Don't think so	Don't think so	Don't think so		Don't think so	Don't think so
Sleeping	Sleeping	Sleeping	Sleeping	DU	DU	DU	DU		Sleeping	Sleeping
Family history	Don't think so	Don't think so	Don't think so	Don't think so	Don't think so	Don't think so	Don't think so		Don't think so	Don't think so
Recent illness	Flu	Flu	Flu	Flu	Flu	Flu	Flu		Flu	Flu
Illness ideas	Illness idea	Illness idea	Illness idea	Illness idea	Illness idea	Illness idea	Illness idea		Illness idea	Illness idea
Medication	Medication	Medication	Medication	Medication	Medication	Medication	Medication		Medication	Medication
Other symp.	Rephrase	Rephrase	Rephrase	Rephrase	Rephrase	Rephrase	Rephrase		Rephrase	Rephrase
Catch all	Worse (B)	All the time (F)	Ankles video	DU	DU	DU	DU		DU	DU
Haemoptysis2	Don't have	Don't have	Don't have	Don't have	Don't have	Don't have	Don't have		Don't have	Don't have
Sputum2	DU	DU	DU	DU	DU	DU	DU		DU	DU

Patient sidebar entry

Demographics	**Information to enter**
Date of birth	19.05.54
Address	Determined by patient socio-economic circumstance: 219a Homestead Way – poor, 5 Tulip Way – rich
Occupation	Determined by patient socio-economic circumstance and age (but must not have risk of asbestos exposure)
Ethnicity	Varies between patients: white British, black Caribbean, Indian)

Lifestyle	**Information to enter**
Alcohol	Determined by patient's age/ethnicity/gender
Smoking	Never smoked
BMI	Determined by patient's height and actors build (what is realistic)
Family history	None recorded

Significant medical history	**Date of diagnosis**
Diabetes mellitus	24.11.09
Depression	05.01.09
Allergies	None recorded

Medication history	**When last prescribed or if current**
Flucloxacillin 250mg qds 7day	Last prescribed December 2011
Penicillin (V) 250mg qds 7day	Last prescribed December 2011
Fluoxetine 20mg od	Last prescribed November 2010
Metformin 500mg bd	Current prescription

Patient examinations/bedside tests entry

Bedside test	**Result to enter**
Blood glucose	6.7 mmol/L
Blood pressure	140/80 mmHg
Cultures	No sputum
Height	180 cm (male) 163 cm (female)
Peak flow	555 L/min (male) 375 L/min (female)
PHQ-9 score	3/27
Swabs	Swabs taken and sent to laboratory
Temperature	36.5°C
Urinalysis	Urinalysis normal
Weight	Vary by actor

Examination	**Result to enter**
Abdomen (including rectal)	Soft and non tender. No abnormalities detected.
Breast	Not applicable. (male) Examination normal. (female)
Cardiovascular system	Heart rate 80 beats/minute. Regular rhythm. Bilateral pitting oedema in both feet/ankles.
ENT examination	No abnormality detected.
Eye examination (including fundoscopy)	No abnormalities seen. No exophthalmos. No conjunctival pallor or redness. Schlera, iris and cornea normal in colour and appearance
Foot examination	Pulses palpable. Sensation normal.
Genitalia examination	No abnormality detected.
Heart rate	Heart rate 80 beats/minute.
Nail examination	All nails appear normal.
Neurological examination, central (including cranial nerves)	No abnormality detected.
Neurological examination, peripheral	No abnormality detected.
Peripheral pulses	All pulses palpable. No abnormality detected.
Respiratory rate	Respiratory rate 18 breaths/minute.
Respiratory system	Respiratory rate 18 breaths/minute. No peripheral or central cyanosis. Good chest movement. Chest clear.
Joint examination, cervical spine	Good range of pain-free movement.
Joint examination, shoulder	Both joints normal in appearance and movement.
Joint examination, elbow	Both joints normal in appearance and movement.
Joint examination, wrist	Both joints normal in appearance and movement.
Joint examination, hand	Joints normal in appearance and movement.
Joint examination, thoraco-lumbar spine	Normal gait. Good range of pain-free movement.
Joint examination, hip	Both joints normal in appearance and movement.
Joint examination, knee	Both joints normal in appearance and movement.
Joint examination, ankle	Both joints normal in appearance and movement.
Joint examination, foot	Joints normal in appearance and movement.

Appendix 13 : Materials provided to GPs during the vignette study recruitment process

GP recruitment flyer

GP flyer (generic) May 2013

How do GPs make decisions?

DEPARTMENT OF HEALTH FUNDED STUDY RECRUITING GPs NOW

GPs are often the first port of call for a health problem. Sometimes the diagnosis is obvious but often GPs face situations where there is real, but low, likelihood of disease.

The ways in which GPs make decisions when faced with these situations are not well understood. However, they can have big implications for patient outcomes and health service costs.

How do you make these decisions?

- Take part in an online study of GP decision making processes using interactive vignettes
- Participation takes 1 hour in total (over 3 weeks)
- You can complete it at your desk
- On completion you will be reimbursed for your time (£80) and receive a certificate for CPD

To register your interest in taking part or to receive more information, email gpstudy@ucl.ac.uk or call Dr Jessica Sheringham (020 7679 8286).

The study does not require ethics approval but has UCL sponsorship (ref: 12/0310) and has been approved by R&D in your area ref:101553

Participant information sheet

Study of GP decision making processes

Participant Information Sheet

Thank you for considering taking part in this web-based research study. Please read this leaflet, which tells you about the study and what it involves, and do not hesitate to email us at gpstudy@ucl.ac.uk if you are unclear about anything or would like further information. This study is being carried out by researchers at University College London, with funding from the Department of Health.

1. Why are we doing the study?

When patients feel unwell or experience a painful or unusual symptom, the GP is often the first contact, so the decisions that GPs make during these consultations is a major influence on patients' outcomes. However, the factors that influence these decisions are poorly understood. In this study we are seeking to understand how GPs make decisions when faced with a set of patient characteristics. Ultimately, the learning from this study should inform interventions (for example educational initiatives or decision aids) to help GPs in making decisions.

2. What is involved?

The study will use a web-based application to provide 6 simple, simulated consultations using patient actors. Participation involves:

- **Registration:** you (or a practice representative) will need to complete a short form with basic information about your practice and yourself. You will then receive login detailed by email and instructions on how to use the web-based application.
- **Simulated consultations:** when you log into the application, you will see 'patients' in a virtual 'waiting room' (Note: not all 6 patients will be visible initially). By clicking on a patient, you enter a 'consultation', which starts with a video presentation by the 'patient'. You can find out more about this 'patient' by asking questions (typing in text to which responses appear as pre-recorded video links) or clicking on links to examinations, demographic and lifestyle information or medical history. At the end of each 'consultation' you need to enter your management decision for this 'patient'.
- **Short survey:** after you have completed all 6 consultations, you enter a short survey about decision-making in your real, every-day practice.

Each 'consultation' should take 7-10 minutes with 5 minutes to complete the survey. It is anticipated, therefore, that your entire involvement should take no more than 60 minutes.

3. Does my practice need special computers or software to access the application?

No. You **will** need broadband internet access, a reasonably up to date browser (eg Internet Explorer 9, Mozilla Firefox 3.5 or above) and MS Windows XP or more recent. You will also need to make sure you can hear sound through your computer (through headphones or speakers). If your practice computer system does not meet these requirements or you are not sure, we can help - email gpstudy@ucl.ac.uk.

4. Is it a test?

No: the study is **not** a test of GPs' abilities. Rather than seeking the 'right answer', we are interested in what you would actually do faced with different scenarios. In some of the scenarios you will see, an optimal management plan may not be clear.

5. What are the benefits of taking part?

By participating in the study, you are helping to inform an important area of health service delivery. All GPs will be reimbursed £80 for their time on completion of the 6 vignettes and survey. Furthermore, according to RCGP guidelines participation in a research study is eligible for continuing professional development (CPD) – we will send a certificate upon completion as evidence of participation.

6. Do I have to take part?

No: if you decide at any point during the study that you do not wish to take part, just email gpstudy@ucl.ac.uk. If you have not completed the study within 3 weeks, you will receive reminders by email.

7. What will happen with my information?

All the information you give for this research and your contact details will be kept strictly confidential. The handling, processing, storage and destruction of data collected will be conducted in accordance with the Data Protection Act (1998).

8. What will happen to the results of the study?

We will send a summary of the whole study's aggregated findings to your practice. We will also send GPs a summary of the decisions all participants made in response to the profiles you saw.

9. What do I do if I wish to make a complaint about the research?

If you wish to complain about any aspect of the research, contact the Chief Investigator, Rosalind Raine, email: r.raine@ucl.ac.uk, tel: 020 76791713. If you feel you do not receive a satisfactory response and you wish to take the matter further you should contact the **UCLH Complaints Manager** giving the project title and the Chief Investigator's contact details at: Complaints Department, 2nd Floor West, 250 Euston Road, London NW1 2PQ Tel: 0845 1555 000 ext. 3413 Fax: 020 7380 9595

10. Contacts for further information

If you have any questions about the study, please contact the researchers, Dr Jessica Sheringham or Ms Rachel Sequeira: Dept Applied Health Research, 1-19 Torrington Place, London WC1E 7HB ☏ 020 7679 8286 ✉ gpstudy@ucl.ac.uk

Thank you for taking the time to read about this study

Study R&D approval reference: 101553

Appendix 14 : Evidence of UCL research and development approval for the GP decision making study

Research and development approval was obtained for each CCG area we planned to recruit GPs from before recruitment in that area commenced. Examples of the approval obtained are shown here.

Approval for the CCG areas in North Central London (part of the London region)

NHS
North Central London
Research Consortium

North Central London Research Consortium
3rd Floor, Bedford House
125 - 133 Camden High Street
London, NW1 7JR

16th October 2012

Professor Rosalind Raine,
Dept Applied Health Research
119, Torrington Place
London WC1E 7HB

Dear Prof Raine,

I am pleased to confirm that the following study has now received R&D approval, and you may now start your research in **the trust(s) identified below:**

Study Title: A factorial study of GP decision making delivered through a web based multimedia application

R&D Reference: CSP 101553
REC Reference: Not Required

	- -
NHS Barnet	- -
NHS Islington	- -
NHS Haringey	- -
NHS Enfield	- -
NHS Camden	- -
Central North West London NHS Foundation Trust	- -
- -	- -
- -	- -

If any information on this document is altered after the date of issue, this document will be deemed INVALID

Please ensure that all members of the research team are aware of their responsibilities as researchers which are stated in page 2. For more details on these responsibilities, please check the R&D handbook or NoCLoR website: http://www.noclor.nhs.uk

We would like to wish you every success with your project

Yours sincerely,

Mabel Saili
Senior Research Governance Officer

R&D Approval: 16/09/2012 CSP101553 *REC Reference: Not Required-* *R&D Reference:*
Page 1 of 2

North Central London Research Consortium

North Central London Research Consortium
3rd Floor, Bedford House
125 - 133 Camden High Street
London, NW1 7JR

May I take this opportunity to remind you that during the course of your research you will be expected to ensure the following:

- **Patient contact:** only trained or supervised researchers who hold the appropriate Trust/NHS contract (honorary or full) with each Trust are allowed contact with that Trust's patients. If any researcher on the study does not hold a contract please contact the R&D office as soon as possible.
- **Informed consent:** original signed consent forms must be kept on file. A copy of the consent form must also be placed in the patient's notes. Research projects are subject to random audit by a member of the R&D office who will ask to see all original signed consent forms.
- **Data protection:** measures must be taken to ensure that patient data is kept confidential in accordance with the Data Protection Act 1998
- **Health & safety:** all local health & safety regulations where the research is being conducted must be adhered to.
- **Serious Adverse events:** adverse events or suspected misconduct should be reported to the R&D office and the Ethics Committee.
- **Project update:** you will be sent a project update form at regular intervals. Please complete the form and return it to the R&D office.
- **Publications:** it is essential that you inform the R&D office about any publications which result from your research.
- **Ethics:** R&D approval is based on the conditions set out in the favourable opinion letter from the Ethics Committee. If during the lifetime of your research project, you wish to make a revision or amendment to your original submission, please contact both the Ethics Committee and R&D Office as soon as possible.
- **Monthly / Annually Progress report:** you are required to provide us and the Research Ethics Committee with a progress report and end of project report as part of the research governance guidance.
- **Recruitment data:** if your study is a portfolio study, you are required to upload the recruitment data on a monthly basis in the website: http://www.crncc.nihr.ac.uk/about_us/processes/portfolio/p_recruitment/
- **Amendments:** if your study requires an amendment, you will need to contact the Research Ethics Committee. Once they have responded, and confirmed what kind of amendment it will be defined as, please contact the R&D office and we will arrange R&D approval for the amendment.
- **Audits:** each year, NoCLoR select 10% of the studies from each service we have approved to be audited. You will be contacted by the R&D office if your study is selected for audit. A member of the governance team will request you complete an audit monitoring form before arranging a meeting to discuss your study.

R&D Approval: 16/09/2012 *REC Reference: Not Required-* *R&D Reference:*
CSP101553 Page 2 of 2

Approval for the CCG areas in Sussex (part of the Surrey & Sussex region)

NHS

Sussex NHS Research Consortium

Research Consortium Office
Worthing Hospital
Lyndhurst Road
Worthing
West Sussex
BN11 2DH

Tel: 01903 285027
Fax: 01903 209884
www.ssrc.nhs.uk

Professor Rosalind Raine
Head of Department of Applied Health Research
University College London
1-19 Torrington Place
London
WC1E 7HB

04/07/2013

Dear Professor Raine,

Our ID: CSP 101553
TITLE: A factorial study of GP decision-making delivered through a web-based multimedia application.

Thank you for your application to the Sussex NHS Research Consortium for research governance assurance of the above named study.

I am pleased to inform you that the study has been assessed, and so may proceed. This assurance is valid in the following Organisations:

- **Brighton & Hove CCG (formerly NHS Brighton and Hove)**
- **Eastbourne, Hailsham & Seaford CCG (formerly NHS East Sussex Downs & Weald)**
- **High Weald Lewes Havens CCG (formerly NHS East Sussex Downs & Weald)**
- **Hastings & Rother CCG (formerly NHS Hastings & Rother)**
- **East Surrey CCG (formerly NHS Surrey)**
- **Guildford & Waverley CCG (formerly NHS Surrey)**
- **North East Hampshire & Farnham CCG (formerly NHS Surrey)**
- **North West Surrey CCG (formerly NHS Surrey)**
- **Surrey Downs CCG (formerly NHS Surrey)**
- **Surrey Heath CCG (formerly NHS Surrey)**
- **Crawley CCG (formerly NHS West Sussex)**
- **Coastal West Sussex CCG (formerly NHS West Sussex)**
- **Horsham & Mid Sussex CCG (formerly NHS West Sussex)**

The final list of documents reviewed and approved is as follows:

- NHS R&D Form (submission code: 101553/390974/14/919)
- NHS SSI Form (submission code: 101553/463685/6/281/210352/274293; electronically authorised and dated 12/06/2013)
- Invitation Letter and Expression of Interest Form
- One Page Study Summary for GPs (version 5, dated 28/01/2013)
- Professor Rosalind Raine's CV (signed and dated 30/06/2012)
- Post-consultation Survey (version 1, dated 10/09/2012)
- GP Flyer (version 1, dated 10/09/2012)
- Protocol (version 1, dated 10/09/2012)

ZONE

Your research governance assurance is valid providing you comply with the conditions set out below:
1. You commence your research within one year of the date of this letter. If you do not begin your work within this time, you will be required to resubmit your application.
2. You notify the Consortium Office should you deviate or make changes to the approved documents.
3. You alert the Consortium Office by contacting me, if significant developments occur as the study progresses, whether in relation to the safety of individuals or to scientific direction.
4. You complete and return the standard annual self-report study monitoring form when requested to do so at the end of each financial year. Failure to do this will result in the suspension of research governance approval.
5. You comply fully with the Department of Health Research Governance Framework, and in particular that you ensure that you are aware of and fully discharge your responsibilities in respect to Data Protection, Health and Safety, financial probity, ethics and scientific quality. You should refer in particular to Sections 3.5 and 3.6 of the Research Governance Framework.
6. You ensure that all information regarding patients or staff remains secure and strictly confidential at all times. You ensure that you understand and comply with the requirements of the NHS Confidentiality Code of Practice, Data Protection Act and Human Rights Act. Unauthorised disclosure of information is an offence and such disclosures may lead to prosecution.

Good luck with your work.

Yours sincerely,

[redacted]

Miss Hannah Haines
Senior Research Governance Officer & Lead Monitor

Email: Hannah.haines@wsht.nhs.uk
Tel: 01903 285222 Ext 84394
Fax: 01903 209884

cc. Ms Rachel Sequeira, PhD Student, University College London.
Mrs Helen Woronowski, PCRN South East. University of Brighton.

Appendix 15 : Paper presenting the vignette study results (currently submitted for publication)

Sheringham J, Sequeira R, Myles J, Hamilton W, McDonnell J, Offman J, Duffy S, Raine R. Variations in GPs' Decisions to Investigate Suspected Lung Cancer: A Factorial Experiment Using Multimedia Vignettes. Submitted to *BMJ Quality & Safety* May 2016. Revisions requested. Resubmitted June 2016.

ABSTRACT

INTRODUCTION: Lung cancer survival is low and comparatively poor in the UK. Patients with symptoms suggestive of lung cancer commonly present to primary care but it is unclear how general practitioners (GPs), distinguish which patients require further investigation. This study examined how patients' clinical and socio-demographic characteristics influence GPs' decisions to initiate lung cancer investigations.

METHODS: A factorial experiment was conducted amongst a national sample of 227 English GPs using vignettes presented as simulated consultations. A multimedia interactive website simulated key features of consultations using actors ('patients'). GP participants made management decisions online for six 'patients', whose socio-demographic characteristics systematically varied across three levels of cancer risk. In low-risk vignettes, investigation (i.e. chest X-ray ordered, computerised tomography scan or respiratory consultant referral) was not indicated; in medium-risk, investigation could be appropriate and in high-risk vignettes investigation was definitely indicated. Each 'patient' had two lung cancer-related symptoms; one volunteered and another elicited if GPs requested it. Variations in investigation likelihood were examined using multilevel logistic regression.

RESULTS: GPs decided to investigate lung cancer in 74% (1000/1348) of vignettes. Investigation likelihood did not increase with cancer risk. Investigations were more likely when GPs requested information on relevant symptoms that 'patients' had but did not volunteer (adjusted odds ratio = 3.18; 95%CI 2.27-4.70). However GPs omitted to seek this information in 42% (570/1348) of cases. GPs were less likely to investigate older than younger 'patients' (adjusted odds ratio = 0.52 95%CI 0.39-0.7]) and Black 'patients' than White (adjusted odds ratio = 0.68; 95%CI 0.48-0.95).

CONCLUSIONS: GPs did not investigate everyone with the same symptoms equally. Insufficient data gathering could be responsible for missed opportunities in diagnosis.

INTRODUCTION

Lung cancer, the most common cancer worldwide, has comparatively poor survival in the UK.[1] Most lung cancer patients first present to primary care but diagnostic delays are well documented: lung cancer patients have more consultations in primary care before investigation than many other cancers.[2] In addition, whilst intervals from presentation to diagnosis have reduced for other common cancers over time, they remain unchanged for lung cancer.[3] It has been suggested that missed opportunities for lung cancer diagnosis in primary care may contribute to poor lung cancer survival.[4]

Primary care physicians, referred to throughout this paper as general practitioners (GPs), have direct access to lung cancer diagnostic tools including chest X-ray. GPs may not consider lung cancer as a differential diagnosis because patients with lung cancer commonly present in primary care with non-specific symptoms that are more often due to benign causes.[5] Non-specific symptoms and rare disease occurrence therefore present diagnostic difficulty for GPs.[6] Reducing diagnostic delays requires an understanding of how GPs decide which patients with common, non-specific symptoms to investigate for lung cancer. Not only is it unclear how GPs decide who requires further investigation by chest X-ray or by specialist referral, but inequalities by patient age, gender and socioeconomic circumstances have been identified in retrospective analyses of routine data.[1,2,7,8] Most previous research has examined the diagnostic process using retrospective data in cancer patients only,[5] thus missing a key dimension, i.e. how GPs decide which patients with symptoms do not require investigation.

Examining decision making in a standardised way in clinical practice presents substantial methodological challenges.[9,10] Direct observation of real physician-patient encounters offers no opportunity to control patients' clinical and socio-demographic characteristics, and so requires observation of very large numbers of consultations to obtain the necessary numbers in specific risk or demographic categories. The use of fictional patient profiles (vignettes) can provide a valid and efficient approach to examining clinician behaviour,[11] and studies have already produced useful insights into sources of error in clinicians' decision making processes, due to both patient factors (e.g. symptom characteristics)[12] and physician factors (e.g. cognitive biases).[12,13] As Blumenthal-Barby and others recognise, however, there are limits to the applicability of written vignettes and other vignette designs that do not simulate key features of real consultations.[14] In particular, when vignettes offer little or no opportunity for physicians to seek information from or about the vignette patient, they can inappropriately frame the decision for the physician by cueing what they should notice about the patient or by offering participants only a limited selection of response options. This risks priming participating physicians to consider certain actions, and biasing their responses.

In this vignette study, we therefore sought to simulate key features of consultations. We designed a website using interactive multimedia vignettes with videos of actor 'patients', which enabled participating GPs to ask questions in their own words and receive real-time responses. We used this intervention in a factorial randomised experimental study to examine GPs' decisions to initiate lung cancer investigation across different combinations of patient clinical and socio-demographic characteristics.

METHODS

Design

We constructed 36 simulated consultations comprising video vignettes of actor 'patients' and comprehensive clinical information, including previous medical history, comorbidities and examination findings, and sociodemographic characteristics,. The symptomatic information provided adhered to material in the latest available National Institute for Health and Care Excellence (NICE) referral guidelines for suspected cancer (published in 2005),[15] with cancer risk based on data from the CAPER case-control study.[16] Each consultation was designed to take participating GPs approximately 10 minutes to complete so that it mirrored the length of a 'real' clinical encounter in primary care in the UK National Health Service.

At the start of each 'consultation', a video was shown where the actor 'patient' volunteered a description of their presenting symptom. Participants could then elicit further information in real-time on the presenting symptom, other symptoms, and risk factors by typing in questions, to which they received the 'patient's' video response. They could also, if they wished, click on a drop-down menu to obtain information on behavioural and familial risk factors, previous medical history, family history, socio-demographic information and examination findings.(Figure 1) A demonstration is available at: www.ucl.ac.uk/stream/media/swatch?v=c22f1a2b58b8

<<FIGURE 1>>

We applied a factorial experimental design, where GPs undertook one consultation from each of six clinical profiles across three lung cancer risk levels (Table 1); no GP saw the same actor twice. Within these constraints, allocation of GPs to vignettes was random. This achieved approximate balance of patient characteristics by clinical profile, gender, ethnicity, and socioeconomic circumstances. The study protocol is available at: http://www.ucl.ac.uk/dahr/research-pages/gp_study

Recruitment and participation

Qualified GPs and registrars nearing the end of their specialist GP training were invited through nine Primary Care Research Networks across England in 2012 and 2013 to participate in a study of decision making (without explicit reference to lung cancer). GP participants were first trained to use the on-line simulated consultations. This was done using a web based video in advance of the study with access to support from the research team during or between study consultations. Each participating GP used the study website to 'consult' with six 'patients' and at the end of the 'consultation', entered their management plan. GPs also completed a brief questionnaire about their practice characteristics and years since qualifying.

Application development

The application's development followed the steps recommended by Adler et al[17] for developing simulations:

1. Case concept: developing the vignette design and content
2. Review and Revision by Content Experts
3. Outline and Flow Development: A typical online consultation in the study
4. Translation of content into simulation platform: vignette interactive website
5. Pilot testing and revisions

A detailed description of each step is given in supplementary file S1. In brief, the structure of the factorial experiment required 36 unique vignette combinations to cover the four experimental factors: known to be associated with variation in lung cancer survival, but whose effect on inequalities in GPs' rates of referral for investigation or to secondary care is uncertain[8]:

- **Ethnicity**: three variations (White, Black Caribbean, South Asian)
- **Gender**: two variations (male, female)
- **Socioeconomic circumstances**: two variations (advantaged or disadvantaged)
- **Clinical risk of lung cancer**: three variations (low, medium and high risk), with two profiles for each level of risk. Age was not included as a separate experimental factor but was instead incorporated into profiles because older age increases the risk of cancer associated with most symptom combinations.[16] We constructed six clinical profiles, two for each risk level using different combinations of symptoms, age, and smoking status.(Table 1) The positive predictive values (PPV) of lung cancer were drawn from PPVs generated by analysis of symptom combinations in the CAPER case-control dataset and interpretation of these symptoms and their characteristics informed by the latest available NICE guidance on investigation of suspected cancer.[15, 16] (described further in supplementary data)

To maximise the clinical authenticity of the cases, GPs specializing in cancer diagnosis and non-academic GPs reviewed the proposed vignettes. The website content and functionality were also informed by patient representatives' comments. For example, these influenced the types of responses 'patients' provided, because patient representatives corroborated previous research that patients may well not disclose certain symptoms with their doctors without being directly asked about them.[18]

The translation of content into the online study application website (virtual patient application) required filming actors portraying patients, creating and populating the website with that content. The website architecture and application software was produced by Athenaeum Educational Technologies. It involved the development of a bespoke system using natural language processing principles to recognise GPs' free-text questions and play a video clip in response (see Doan et al 2014 for an explanation of the principles).[19] This system was underpinned by databases on symptoms or risk factors and the features those symptoms (e.g. what exacerbates or relieves the symptom or how long it has been present).

Table 1. Components of the six different clinical profiles by risk level

Clinical Profile	Information volunteered by 'patient' or available onscreen			Information only available if participant GPs asked		Positive Predictive Value (PPV) of lung cancer	Other relevant information
	Age range	Smoking status	Symptom 1	Symptom 2	Duration		
Low risk: Expected action = no active investigation (safety netting appropriate)							
1	Younger (Late fifties)	Non smoker	Breathlessness	Fatigue	1-2 weeks	0.40%	Patient has swollen ankles, possibly due to heart failure
2	Younger (Late fifties)	Smoker	Chest pain	Cough	1-2 weeks	1.10%	
Medium risk: Expected action = either investigation (e.g. order chest x-ray) or safety netting							
3	Older (Late seventies)	Smoker	Chest pain	Cough	Uncertain (approx 3 weeks)	1.70%	
4	Older (Late seventies)	Non-smoker	Cough	Appetite loss	Uncertain (approx 3 weeks)	2.50%	
High risk: Expected action = lung cancer investigation							
5	Younger (Late fifties)	Smoker	Breathlessness	Fatigue	>5 weeks	3-4%	Chronic obstructive pulmonary disease (COPD) present
6	Older (Late seventies)	Smoker	Chest pain	Weight loss	>5 weeks	14%	

Analysis

Every action performed by GPs on the website (i.e. all the questions asked of 'patients', drop-down menus accessed, free-text entered in management plans) was captured by the study website. This information was used to measure the duration of each consultation and to generate three indicators about GPs' information requests in each consultation and the capacity of the research application to respond to these requests:

- *data sought*: average number of data items sought (questions asked or drop-down menu items accessed), by GP and by individual vignette
- *errors*: error messages displayed as a proportion of all data items sought, calculated for all consultations, consultation 1 and consultations 2-6 only, assuming that in the first consultation GPs were familiarising themselves with the application
- *key information elicited*: proportion of GPs that elicited information on the vignettes second, but unvolunteered, lung cancer symptom.

GPs also had the opportunity to provide free-text comments on any aspect of the application in an online survey after all the consultations were completed. These comments were not treated as a representative survey of all participants' experiences but were examined to provide insights into GPs' experiences of the application and their perceptions of its utility as a research tool for eliciting the decision making process.

The primary outcome was the proportion of 'patients' for whom lung cancer investigation was included in the management plan. This included ordering appropriate imaging, or referral for a specialist opinion e.g. from a respiratory consultant whether participants' management plan stated this investigation was for lung cancer or not. This outcome variable was constructed from free-text responses entered by participants in their management plan, according to pre-defined criteria. A clinician confirmed the validity of every constructed primary outcome.

Data were analysed by fitting multilevel logistic regression models using Markov Chain Monte Carlo for estimation,[20] allowing variation between participants and between vignettes within participants. This allowed for a correlation between outcomes within a given GP but independent outcomes for two vignettes viewed by different GPs. Estimation of odds ratios and 95% credible intervals was carried out using the RStan library in R version 3.0.2.[21] Significance testing was carried out using Wald tests based on the means and posterior variances of the estimates.

Variations in outcome were examined by 'patient' gender, ethnicity, socioeconomic circumstances and risk profile, an indicator variable for whether participants sought the second symptom, and GP characteristics (demographics, experience, and region). Two models were built in order to examine differences by a) clinical profile and b) by age. A supplementary analysis was conducted to examine whether findings were difficulties in obtaining information sought from the application, by including the indicator on errors as another covariate in each model. To examine selection bias, the gender and age of participating GPs and their practices' cancer referral characteristics were compared with national data.[22,23]

The required sample size was calculated on the basis that a minimum difference in investigations of 10% was considered of clinical importance and realistic given variations in cancer investigations in other studies.[24] A response from 216 participants was sought to give 1296 vignettes (i.e. each of the 36 vignettes viewed 36 times). Each risk and ethnic group would therefore be viewed 432 times, each gender and socioeconomic group 648 times. Assuming a 20% variance inflation factor for clustering of GPs/'patients', 432 in each risk and ethnic group would give 95% power to detect a difference of 10%. For differences between gender and socioeconomic groups, 648 in each group would give 85% power for a difference of 5%.

RESULTS

Sample characteristics

227 GPs completed the study, 76% of the 300 GPs who registered.(See: supplementary file S2A) There were no demographic differences between registered GPs who did and did not complete the study but GP participants were younger than the national GP population and practices had higher cancer referrals than non-participating practices.(See: supplementary file S2B)

Out of 1362 vignettes, 14 (1%) were excluded due to missing participant demographic data (n=6, 0.4%), when participants asked about second symptoms but did not receive a response (n=4, 0.3%) or did not enter a management plan (n=4, 0.3%).

Consultation process

GPs spent on average 16 minutes on the first consultation and 11 minutes on consultations 2-6 and sought 47 items of information per consultation (by asking text questions of the patient, looking up patient history or personal information, conducting 'examinations' or 'bedside tests'). GPs received error messages in response to an average of 4.6% of data sought for consultations 2-6 (range 4-22%).(See supplementary file, S2C)

Lung cancer investigations

Participants initiated investigations in 1000 (74%) vignettes. There was little difference in investigation between low, medium and high-risk levels (72-75%) but large variation between clinical profiles (59-86%). There were no variations by 'patient' gender or socioeconomic circumstances but there was a gradient in investigation by ethnicity, with 'patients' of Black ethnicities least and White ethnicities most likely to be investigated (71% vs 77%). (Table 2)

GPs asked for additional, relevant information about second symptoms in 778 (58%) of cases overall with marked variation by clinical profile, ranging from 48 (21%) in Profile 1 to 214 and 216 (95%) in Profiles 2 and 3. There was a significant interaction between seeking a relevant second symptom and clinical profile ($p<0.001$). 91% of GPs who discovered the presence of weight loss initiating investigation compared with just 46% who did not seek this information. In contrast, knowing 'patients' experienced fatigue did not significantly change the likelihood of investigation.(Table 3)

While obtaining second symptom information was associated with more investigation (adjusted odds ratio (AOR): 3.18 [2.27;4.70], $p<0.001$), there was still under-investigation in 'patients' with appetite or weight loss (Profiles 4 and 6) compared with 'patients' with chest pain and cough (Profile 3) (AORs: 0.25 [0.14;0.42], $p<0.001$; and 0.5 [0.29;0.91], $p=0.02$ respectively).(Table 4a) GPs were less likely to investigate older than younger 'patients' (AOR: 0.52 [0.39;0.70], $p<0.001$), and less likely to investigate 'patients' of Black compared with White ethnicities (AOR: 0.68 [0.48;0.95], $p=0.03$).(Table 4b)

Associations were similar when the variable for errors received was included. (See: supplementary file S2D)

Comments volunteered by GP participants on their experiences of the application and their perceptions of its utility as a research tool for eliciting the decision making process are summarised in S3.

Table 2. Frequency of lung cancer investigation

		Investigation n	Investigation %	N (vignettes)
Total		**1000**	**74.18**	**1348**
a. By 'patient' characteristic				
Risk level	Low	339	75.00	452
	Medium	327	72.35	452
	High	334	75.23	444
Clinical profile[1]	Clinical Profile 1 *PPV=0.4% (younger; ns; 1-2w breathless [& fatigue])*	152	66.96	227
	Clinical Profile 2 *PPV=1.1% (younger, s; 1-2w chest pain [& cough])*	187	83.11	225
	Clinical Profile 3 *PPV=1.7% (older, s; ~3w chest pain [& cough])*	195	85.90	227
	Clinical Profile 4 *PPV=2.5% (older, ns; ~3w cough [& appetite loss])*	132	58.67	225
	Clinical Profile 5 *PPV=3-4% (younger, s; >5w breathless [& fatigue])*	185	82.59	224
	Clinical Profile 6 *PPV~14% (older, s; >5w chest pain [& weight loss])*	149	67.73	220
Gender	Female	489	74.09	660
	Male	511	74.27	688
Socioeconomic circumstances	Disadvantaged	508	74.49	682
	Advantaged	492	73.87	666
Ethnicity	White	369	76.56	482
	Black	306	71.50	428
	South Asian	325	74.20	438

2nd symptom elicited	No	361	63.33	570
	Yes	639	82.13	778
b. By GP participant characteristic				
GP gender	Female	425	70.48	603
	Male	573	77.12	743
GP age range	25-34 years	227	70.06	324
	35-44 years	336	72.89	461
	45-54 years	325	78.69	413
	55-64 years	102	75.00	136
	65 years or over/missing	8	66.67	12
Years since qualifying	0 to 2 years ago	120	71.43	168
	2 to 5 years ago	186	69.14	269
	5 to 10 years ago	177	73.75	240
	10 to 20 years ago	256	77.58	330
	20+ years ago	259	76.40	339
Ethnicity	White	583	73.89	789
	Black	34	80.95	42
	South Asian	296	73.63	402
	Other/missing	90	75.63	119
Region	London	365	73.44	497
	East of England	341	74.95	455
	North West	131	76.16	172
	West Midlands	96	72.73	132
	Surrey and Sussex	41	75.93	54
	Locum GP	24	66.67	36

[1] younger = late fifties; older = late seventies; s = smoker ns = non-smoker; w = weeks; [symptom] = not volunteered by patient

Table 3. Lung cancer investigation by profile according to whether GPs did or did not elicit symptom information

Clinical profile (Second symptom)[1]	Second symptom		Lung cancer investigation				
	Not elicited	Elicited	Symptom not elicited		Symptom elicited		Total
	n (%)	n (%)	n (%)	Odds Ratio [95% CI]	n (%)	Odds Ratio [95% CI]	n (%)
1 (Fatigue)	179 (78.85)	48 (21.15)	120 (66.67)	1.00 [-]	31 (65.96)	0.94 [0.43;2.09]	152 (66.96)
2 (Cough)	11 (4.89)	214 (95.11)	7 (63.64)	0.73 [0.16;3.18]	181 (84.19)	2.83 [1.82;4.40]	187 (83.11)
3 (Cough)	11 (4.85)	216 (95.15)	7 (63.64)	0.93 [0.19;4.39]	189 (87.1)	3.67 [2.13;6.30]	195 (85.90)
4 (Appetite loss)	89 (39.56)	136 (60.44)	42 (46.67)	0.38 [0.21;0.69]	91 (66.91)	0.98 [0.59;1.62]	132 (58.67)
5 (Fatigue)	168 (75.00)	56 (25.00)	136 (80.47)	2.21 [1.31;3.72]	50 (89.29)	4.59 [2.86;7.37]	185 (82.59)
6 (Weight loss)	112 (50.91)	108 (49.09)	52 (46.02)	0.36 [0.20;0.62]	99 (90.83)	5.69 [2.07;15.63]	149 (67.73)
Total	**570 (42.28)**	**778 (57.72)**	**364 (63.41)**		**641 (82.18)**		**1000 (74.18)**

[1] Clinical profile is formed from symptoms, smoking status and patient age

Table 4. Multilevel logistic regression of cancer investigation by 'patient' characteristic

a) By clinical profile		Adjusted[1] odds ratio [95% CI]
Clinical profile (2nd symptom)	1 (Fatigue)	0.62 [0.35; 1.10]
	2 (Cough)	0.65 [0.38; 1.15]
	3 (Cough)	1
	4 (Weight loss)	0.25 [0.14; 0.42]*
	5 (Fatigue)	1.64 [0.90; 3.11]
	6 (Appetite loss)	0.50 [0.29; 0.91]*
Ethnicity	White	1
	South Asian	0.86 [0.62; 1.20]
	Black	0.67 [0.47; 0.96]*
Second symptom elicited	No	1
	Yes	3.18 [2.27; 4.70]*

b) By age		
Age	Younger (Late fifties)	1
	Older (Late seventies)	0.52 [0.39; 0.70]*
Ethnicity	White	1
	South Asian	0.88 [0.63; 1.27]
	Black	0.68 [0.48; 0.95]*
Smoking status	Non smoker	1
	Smoker	2.24 [1.64; 3.02]*
Second symptom elicited	No	1
	Yes	2.83 [2.09; 3.83]*

[1] adjusted for all other factors associated (p<0.1) with investigation in univariate analysis (i.e 'patient' profile and ethnicity, GP gender and age), and whether second symptom was elicited

[2] adjusted for 'patient' profile, ethnicity, GP gender and age and whether second symptom was elicited

* significant at p≤0.05

DISCUSSION

Summary

In this factorial experiment using vignettes in simulated consultations, GPs' decisions to investigate lung cancer was influenced by whether they sought out additional, relevant clinical information about the presence of common symptoms. Even when participating GPs elicited sufficient information about symptoms, inequalities by age and ethnicity in investigation decisions remained.

Strengths and limitations

Our novel approach, using vignettes in an interactive website that delivered real-time responses, obtained comprehensive information on decision making in over 99% of consultations and in a timeframe comparable to a typical consultation. The method simulated more components of the decision-making process in real time than has been achieved in previous studies.[25-27]

Of equal importance is the fact that we applied a randomised, factorial, experimental design, with exact balance on profile and risk, and approximate, balance with random allocation to GP's, on socio-demographic factors. This allowed us to examine the effects of patients' socio-demographic and clinical characteristics on GPs' decision-making.

Despite the advances we achieved in simulating real consultations, the on-line vignettes were limited mainly due to the constraints of the natural language system. These constraints meant the website was unable to provide responses to all GPs' information requests. In the post-consultation survey 12 GP participants (5%) reported difficulty in obtaining information, which caused some of them frustration, and a small number (n=4, 1.8%) observed it may have altered their decision-making behaviour. The process itself of typing in questions may also have prompted GP participants to consider their clinical reasoning more than they would in their routine clinical practice. Conversely, the opportunity to select from the extensive drop-down selections of examinations without facing any of the logistical constraints faced in a real consultation (e.g. time required to measure weight) may have led them to seek more information with less consideration than they would do in routine clinical practice. However, it is important to note that all approaches to simulating consultations have some drawbacks. For example, while other vignette studies have enabled physicians to 'ask' questions of the patient, this has required a researcher to type responses online as 'the patient', sometimes resulting in longer 'consultations' than real consultations. [25 27] Moreover, there are several reasons why these simulations still provide valuable insights into GPs' decision making. Firstly, our sensitivity analysis indicates that results were very close to the main analysis even after taking into account GPs' difficulties in obtaining responses from the application. Secondly, shortcomings in doctor-patient communication during the clinical encounter are well recognised, such that patients in real consultations do not volunteer all the information clinicians would need to make informed decisions.[18] Thirdly, it is the divergence from reality that makes simulated consultations useful for studying phenomena or circumstances not possible to observe or investigate in real life.[28] In this study, this divergence enabled the systematic manipulation of patient characteristics to examine their effects on GPs' decisions in isolation of the complex range of patient expectations and co-morbidities that might explain variations in decision making in real life. The divergence also meant GPs were not faced with the logistical and system/organisational constraints that affect referral decisions in practice. As a result, the findings provide insight into the cognitive processes underlying GPs' decision making when the variation in system and patient factors present in real life are removed.

We were not able to achieve total orthogonality in design of all patient characteristics, but the randomisation and approximate balance give some confidence in the general applicability of our results.

There was some bias in the GP sample registering for the study in that GP participants' practices had higher cancer referrals than non-participating practices, so they may be more ready than GPs nationally to investigate symptoms suggestive of cancer. However, there was no evidence to suggest participating GPs would have greater or smaller variation in decision making than non-participants.

Another possible limitation is that the risk levels were based on positive predictive values from the CAPER symptom case-control dataset, which had wide and overlapping confidence intervals (as shown in supplementary data, S1). Therefore, the PPVs alone are not sufficient to conclude that clinical risk and therefore decision making should have varied by profile. However, even where the PPV point estimates are most disparate and confidence intervals overlap minimally, GPs investigated similar proportions of patients. In addition, the risk profiles had additional information other than PPV which should have guided decision making if GPs were acting in line with the latest available clinical guidance (e.g. symptom duration). Furthermore, our three broad categories align well with the 2015 NICE guidance. These equate to: risk below 1%, safety-netting; 1-3%, test in primary care if possible; over 3% refer for specialist testing.[33]

Comparisons with existing literature

Our data were collected during 2012-2013 and our finding that GPs investigated a high proportion (72-75%) of cases is in line with literature from 2013.[24] However it is higher than might have been expected if GPs were following the latest national guidance for suspected cancer investigation available during the study period.[15] Participants may have proposed more tests for vignette 'patients' than they would in reality because they were not subject to the resource constraints of clinical practice or may have ordered X-rays primarily to investigate diagnoses other than cancer. Alternatively, they may have been aware of and responding to epidemiological evidence, presumed patient preferences, and policy published since the 2005 NICE guidance, all of which support a lower threshold for cancer investigation.[29-32] Indeed, updated NICE guidance on referral of suspected cancer, published in 2015 (after our data were collected), include a substantially lower investigation threshold than that recommended in their earlier guideline,[33] such that all our vignettes would now suggest investigation.

We found that in 42% of cases, GPs did not seek additional information that would help to make an informed decision regarding referral and that was available on request. This accords to some extent with international studies of missed opportunities in cancer diagnosis.[34-35] In the UK, the updated NICE guidance explicitly recognises that patients with combinations of common symptoms may be more likely to have lung cancer than patients with any one of these symptoms alone,[33,36] but patients may not volunteer all the symptoms they experience in consultations, perhaps due to real or perceived time constraints in the consultation.[36] The importance of data gathering for reaching a timely diagnosis was highlighted in the recent Institute of Medicine Report into improving diagnosis in health care.[37] Zwaan et al's study of breathlessness using expert review of medical records found evidence of inappropriately selective information gathering in a third of cases with some evidence that diagnostic error and patient harm occurred in a proportion of these cases.[38] Our study extends the field by providing objective evidence of non-clinical variations in data gathering by physicians in a large vignette study and demonstrates associations between gathering sufficient data and appropriate decision making.

We also found that the effect of eliciting this second symptom on decision making varied by symptom. It made little difference whether participants knew that patients had a cough or fatigue, but made significant difference to decision making if participants knew of appetite and weight loss. For weight loss in particular (a key question when clinicians are considering whether cancer is a possible diagnosis), in 91% of cases where GP participants had elicited information about weight loss, they initiated investigation, compared with just 46% where GPs were unaware the patient had lost weight. It is important to acknowledge that neither in real life nor in the vignettes are the factors (symptom, age and smoking) that constituted each profile independent of one another. Therefore whilst we contend the results are interpretable and reliable, they are not as definitive as a randomised controlled trial results so this finding has to be treated with some caution. However, the finding accords with Kostopoulou et al's recent 'think aloud' study which suggests that when physicians have an idea of cancer early in the consultation, they ask pertinent questions and initiate appropriate investigations to ensure a cancer diagnosis is reached.[39] Therefore, it still seems likely that routinely questioning patients with ongoing respiratory symptoms about weight loss would expedite the diagnosis of some lung cancers.

Our finding that GPs were less likely to investigate older 'patients' is consistent with several observational studies of primary care cancer referral and investigation.[40-41] Scott et al's Model of Pathways to Treatment proposes that as patients grow older, they are increasingly likely to attribute bodily changes to normal ageing processes than to disease.[42] If clinicians also apply this 'normal ageing' heuristic, it may explain why GPs in this study were less likely to investigate older patients, despite knowing their symptoms. In contrast, patient experience survey data indicate more referral delays in younger (aged 55-64 years) than older patients (over 75 years). However survey data may be biased if older patients (with lower overall survival) were underrepresented because they had died or were too ill to participate in the survey (which was undertaken 6-12 months after diagnosis).[2]

We also found smaller ethnic variations in GPs' investigation behaviour, with fewer investigations initiated in Black (and to some extent) South Asian 'patients' than White. This is consistent with survey data where non-White cancer patients report more referral delays than White patients.[2] One possible explanation is that GPs were less ready to consider a lung cancer diagnosis in individual non-White 'patients' who presented with high-risk clinical profiles because they placed weight on knowledge that lung cancer risk factors and prevalence are lower in Black and South Asian than White populations.[43] However, the presence of symptoms should overrule consideration of risk factors of modest effect. Another possible explanation is that investigation likelihood is influenced by GPs' ethnicity. In this study there were only seven GPs identified as Black, so it was not possible to examine this, but the mechanism by which observed ethnic variations in decision making occur remains an important question to address.

Conclusions and implications for research and practice

This study demonstrates that GPs do not investigate everyone with the same symptoms equally. It also indicates that insufficient data gathering could be responsible for diagnostic errors. It is not that GPs are doing a bad job: the average GP sees one patient with new lung cancer a year.[16] Distinguishing symptoms indicating possible cancer from self-limiting illness that GPs see daily, therefore is challenging. However, non-clinical variations in investigation could contribute to the sociodemographic inequalities in the timeliness of diagnosis and survival of lung cancer seen in the UK. It also marks a departure from the National Health Service commitment to promote equality through its services.[44] The findings also have wider implications for quality and safety in healthcare internationally. According to the Institute of Medicine, diagnostic errors contribute to approximately 10 percent of patient deaths, and sufficient data gathering is an essential part of reaching a timely diagnosis.[37]

It is therefore incumbent on health systems to consider strategies that can be implemented in practice such as clinician education,[37,45] decision support tools[25] and the assessment of equity in clinical practice.

REFERENCE LIST

1. Coleman MP, Forman D, Bryant H, et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet* 2011;377:127-38.

2. Lyratzopoulos G, Neal RD, Barbiere JM, et al. Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England. *Lancet Oncol* 2012;13:353-65.

3. Neal RD, Din NU, Hamilton W, et al. Comparison of cancer diagnostic intervals before and after implementation of NICE guidelines: analysis of data from the UK General Practice Research Database. *Br J Cancer* 2014;110:584-92.

4. O'Dowd EL, McKeever TM, Baldwin DR, et al. What characteristics of primary care and patients are associated with early death in patients with lung cancer in the UK? *Thorax* 2015;70:161-68.

5. Walter FM, Rubin G, Bankhead C, et al. Symptoms and other factors associated with time to diagnosis and stage of lung cancer: a prospective cohort study. *Br J Cancer* 2015;112(s1):S6-S13.

6. Kostopoulou O, Delaney BC, Munro CW. Diagnostic difficulty and error in primary care—a systematic review. *Fam Pract* 2008;25:400-13.

7. Raine R, Wong W, Scholes S, et al. Social variations in access to hospital care for patients with colorectal, breast, and lung cancer between 1999 and 2006: retrospective analysis of hospital episode statistics. *BMJ* 2010;340:b5479.

8. Macleod U, Mitchell ED, Burgess C, et al. Risk factors for delayed presentation and referral of symptomatic cancer: evidence for common cancers. *Br J Cancer* 2009;101 Suppl 2:S92-s101.

9. Singh H, Sittig DF. Advancing the science of measurement of diagnostic errors in healthcare: the Safer Dx framework. *BMJ Qual Saf* 2015;24:103-10.

10. Hrisos S, Eccles MP, Francis JJ, et al. Are there valid proxy measures of clinical behaviour? A systematic review. *Implement Sci* 2009;4:37. doi: 10.1186/1748-5908-4-37.

11. Peabody JW, Luck J, Glassman P, et al. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA* 2000; 283: 1715–22.
12. Kostopoulou O, Oudhoff J, Nath R, et al. Predictors of diagnostic accuracy and safe management in difficult diagnostic problems in family medicine. *Med Decis Making*. 2008;28:668-80. doi: 10.1177/0272989x08319958.
13. Zwaan L, Monteiro S, Sherbino J, et al. Is bias in the eye of the beholder? A vignette study to assess recognition of cognitive biases in clinical case workups. *BMJ Qual Saf* 2016. 10.1136/bmjqs-2015-005014
14. Blumenthal-Barby JS, Krieger H. Cognitive Biases and Heuristics in Medical Decision Making: A Critical Review Using a Systematic Search Strategy. *Med Decis Making*. 2015;35:539-57.
15. National Institute for Health and Care Excellence. Referral guidelines for suspected cancer: NICE guidelines CG27, 2005. Available: https://www.nice.org.uk/Guidance/CG27. Accessed 10 May 2016.
16. Hamilton W. The CAPER studies: five case-control studies aimed at identifying and quantifying the risk of cancer in symptomatic primary care patients. *Br J Cancer* 2009;101 Suppl 2:S80-6.
17. Adler MD, Trainor JL, Siddall VJ, et al. Development and evaluation of high-fidelity simulation case scenarios for pediatric resident education. *Ambul Pediatr* 2007;7:182-6.
18. Shim J, Brindle L, Simon M, et al. A systematic review of symptomatic diagnosis of lung cancer. *Fam Pract* 2014;31:137-48.
19. Doan S, Conway M, Phuong TM, et al. Natural Language Processing in Biomedicine: A Unified System Architecture Overview. In: Trent R, ed. Clinical Bioinformatics. New York, NY: Springer New York, 2014:275-94.
20. Spiegelhalter D, Abrams K, Myles J. *Bayesian Approaches to Clinical Trials and Health Care Evaluation*. Chichester, Wiley; 2004.
21. Stan Development Team. RStan: the R interface to Stan. Version 2.5; 2014. Available: http://mc-stan.org/rstan.html
22. Health and Social Care Information Centre. Workforce. 2014 Available: http://www.hscic.gov.uk/workforce. Accessed 10 May 2016.
23. National Cancer Intelligence Network. Practice Profiles. 2013 Available: http://www.ncin.org.uk/cancer_information_tools/profiles/gp_profiles Accessed 10 May 2016.
24. Hamilton W, Green T, Martins T, et al. Evaluation of risk assessment tools for suspected cancer in general practice: a cohort study. *Br J Gen Pract* 2013;63:e30-6.
25. Kostopoulou O, Rosen A, Round T, et al. Early diagnostic suggestions improve accuracy of GPs: a randomised controlled trial using computer-simulated patients. *Br J Gen Pract* 2015;65:e49-54.
26. Hooper LM, Weinfurt KP, Cooper LA, et al. Virtual standardized patients: an interactive method to examine variation in depression care among primary care physicians. *Prim Health Care Res Dev*. 2008; 9: 257-268.
27. Harries C, Forrest D, Harvey N, et al. Which doctors are influenced by a patient's age? A multi-method study of angina treatment in general practice, cardiology and gerontology. *Qual Saf Health Care*. 2007; 16: 23-27.
28. Hughes R. Considering the Vignette Technique and its Application to a Study of Drug Injecting and HIV Risk and Safer Behaviour. *Sociol Health Ill* 1998;20:381-400.
29. Hamilton W, Peters TJ, Round A, et al. What are the clinical features of lung cancer before the diagnosis is made? A population based case-control study. *Thorax* 2005;60:1059-65.
30. Banks J, Hollinghurst S, Bigwood L, et al. Preferences for cancer investigation: a vignette-based study of primary-care attendees. *Lancet Oncol* 2014;15:232-40.

31. Department of Health. Improving Outcomes: A Strategy for Cancer. London: Department of Health, 2011. Available: www.gov.uk. Accessed 10 May 2016.
32. Hippisley-Cox J, Coupland C. Identifying patients with suspected lung cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* 2011; 61: e715–23.
33. National Institute for Health and Care Excellence. Suspected cancer: recognition and referral. Secondary Suspected cancer: recognition and referral. 2015. Available: nice.org.uk/guidance/ng12. Accessed 10 May 2016.
34. Jensen H, Nissen A, Vedsted P. Quality deviations in cancer diagnosis: prevalence and time to diagnosis in general practice. *Br J Gen Pract* 2014;64:e92-e98.
35. Singh H, Daci K, Petersen LA, et al. Missed Opportunities to Initiate Endoscopic Evaluation for Colorectal Cancer Diagnosis. *Am J Gastroenterol* 2009;104:2543-54.
36. Lyratzopoulos G, Vedsted P, Singh H. Understanding missed opportunities for more timely diagnosis of cancer in symptomatic patients after presentation. *Br J Cancer* 2015;112 Suppl:S84-91.
37. Balogh E, Miller B, Ball JE. (eds) Improving diagnosis in health care / Committee on Diagnostic Error in Health Care. Washington, DC : The National Academies Press, 2015. DOI: 10.17226/2179442.
38. Zwaan L, Thijs A, Wagner C, Timmermans DR. Does inappropriate selectivity in information use relate to diagnostic errors and patient harm? The diagnosis of patients with dyspnea. *Soc Sci Med* 2013;91:32-8.
39. Kostopoulou O, Sirota M, Round T, et al. The Role of Physicians' First Impressions in the Diagnosis of Possible Cancers without Alarm Symptoms. *Med Decis Making*. 2016 doi:10.1177/0272989X16644563
40. McBride D, Hardoon S, Walters K, et al. Explaining variation in referral from primary to secondary care: cohort study. *BMJ* 2010;341:c6267. doi: 10.1136/bmj.c6267.
41. Tate AR, Nicholson A, Cassell JA. Are GPs under-investigating older patients presenting with symptoms of ovarian cancer? Observational study using General Practice Research Database. *Br J Cancer* 2010;102:947-51.
42. Scott SE, Walter FM, Webster A, et al. The model of pathways to treatment: conceptualization and integration with existing theory. *Br J Health Psychol* 2013;18:45-65.
43. Jack RH, Davies EA, Moller H. Lung cancer incidence and survival in different ethnic groups in South East England. *Br J Cancer* 2011;105(7):1049-53.
44. Department of Health. Equity and Excellence: Liberating the NHS. 2010. Available: www.gov.uk. Accessed 10 June 2016.
45. Reilly JB, Ogdie AR, Von Feldt JM, et al. Teaching about how doctors think: a longitudinal curriculum in cognitive bias and diagnostic error for residents. *BMJ Qual Saf* 2013;22:1044-50.

Appendix 16 : Coding criteria for the additional variables constructed when considering GPs' decisions in the vignette study

As discussed in Section 5.1.2.1, in addition to coding my primary variable 'referral for chest X-ray' Dr Crofton and I also coded two additional related variables: a less stringent variable ' any suggestion chest X-ray', and a much stricter variable ' urgent chest X-ray'. The coding criteria for both these variables are below.

Urgent/two week wait referral for chest X-ray (more stringent)

Code as 'urgent chest X-ray' - 1	**Code as 'non-urgent chest X-ray' - 0**
Chest X-ray requests listed as 'urgent', 'stat', 'immediate' or 'same day': *e.g. urgent CXR* *CXR stat* *CXR immediately* *send for CXR the same day* Chest x-ray requests where GP notes that they would ask for an urgent/same day review: *e.g. CXR…with request for urgent report* Hospital/A&E referrals where chest x-ray was specifically stated in management plan <u>or</u> a lung disease is most likely/likely diagnosis Referrals via the two week wait pathway Referral to chest clinic, or for respiratory or oncology specialist, <u>if</u> specified as urgent	All other requests for chest X-ray where no urgency or low urgency is stated. *e.g. CXR* *standard CXR* *non-urgent CXR* Where chest X-ray is referred to using uncertain phrasing: *e.g. possible ECHO and/or CXR* *may arrange CXR* *may need a CXR* *may leave for now* *consider CXR* *if I was uneasy I would arrange CXR* Where chest X-ray is considered as a potential future management option: *e.g. CXR if persists* *review, if no better for CXR* *if still unwell for CXR* *give CXR form to go next week if no better* Referral to hospital medics (unless chest x-ray specified, or a lung disease considered most likely/likely diagnosis) Referral to non-respiratory specialist: *e.g. rapid access chest pain clinic* *cardiology* *gastroenterology* X-ray requested, but not chest or chest not specified Chest x-ray or referral not in management plan

Any suggestion of chest X-ray (less stringent)

<table>
<tr><th>Code as
‘suggestion of chest X-ray’ - 1</th><th>Code as
‘no suggestion of chest x-ray’ - 0</th></tr>
<tr><td>All management plans that mention chest X-ray, including when this is:
- urgent, non-urgent or no urgency stated
e.g. CXR
urgent CXR
standard CXR
- hospital admission/A&E referral where chest X-ray specifically stated in management plan <u>or</u> lung disease is the most likely/likely diagnosis
e.g. refer to hospital for 12 lead ECG, CXR and arterial blood gases
- referred to using uncertain phrasing:
e.g. possible ECHO and/or CXR
may arrange CXR
may need a CXR
may leave for now
consider CXR
if I was uneasy I would arrange a CXR
- considered as potential future management:
e.g. CXR if persists
review, if no better for CXR
if still unwell for CXR
give CXR form to go next week if no better
Referral to chest clinic or for respiratory or oncology specialist</td><td>Referral to hospital medics where <u>neither</u> chest X-ray specified, <u>nor</u> a lung disease considered most likely/likely diagnosis:
Referral to non-respiratory specialist:
e.g. rapid access chest pain clinic
cardiology
gastroenterology
X-ray requested, but not chest or chest not specified
Chest X-ray or referral not in management plan</td></tr>
</table>

NB: for both additional outcomes, where GPs did not state any management plan (n=3) this was coded 99 so it can be easily identified and excluded.

Appendix 17 : The post-consultation survey

The full content of the post-consultation survey that all GPs completed after the vignette study is as follows.

Welcome to the decision making survey

Thank you for completing these six vignettes.

We would now like to ask you some questions.

This survey has three sections:

I - Decision making in these vignettes

II - Decision making in your everyday practice

III - Your clinical experience, responsibilities and lifestyle

What is your GP decision making study username? *(i.e. the username you use to log into the application)*

What is the name of your GP practice?

15% Start

There are no right or wrong answers - we are keen to understand what you actually do.

If you have any queries or problems completing this survey, please email gpstudy@ucl.ac.uk

---

I - Decision making in these vignettes

We would like to know how you made decisions in these vignettes

1 **Did you refer to any other source(s) of information when completing the vignettes? Tick as many boxes as apply.**

- [ ] GP or hospital colleague
- [ ] NICE guidelines
- [ ] Other, local, guidelines
- [ ] Textbook or website
- [ ] None of the above

2 **If you saw similar patients (to those in the vignettes) in your day-to-day practice, would you have referred to any of the following sources of information? Tick as many as apply.**

- [ ] GP or hospital colleague
- [ ] NICE guidelines
- [ ] Other, local, guidelines
- [ ] Textbook or website
- [ ] None of the above

30% Back Next

There are no right or wrong answers - we are keen to understand what you actually do.

If you have any queries or problems completing this survey, please email gpstudy@ucl.ac.uk

II - Decision making in your everyday practice

We would now like to know how you made decisions in your day-to-day practice

Please think about the patients you saw in the last month who you considered sending for simple investigations (such as ultrasound or X-ray) and/or referring to secondary care.

3 **To what extent do the factors listed below influence the likelihood that you will refer a patient? Select one answer for each statement.**

	More likely to refer in most cases	More likely to refer in some cases	No more or less likely to refer	Less likely to refer in some cases	Less likely to refer in most cases	Don't know
The patient reports difficulties taking time off work for an appointment/diagnostic test	○	○	○	○	○	○
The patient is a caregiver	○	○	○	○	○	○
The patient's lifestyle puts them at higher risk of serious disease	○	○	○	○	○	○
You know the patient well and are familiar with their past medical history	○	○	○	○	○	○
The patient frequently attends with non-serious complaints	○	○	○	○	○	○
The patient has previously failed to turn up to primary or secondary care appointments	○	○	○	○	○	○
The patient has not followed medical advice in the past (e.g. did not take medication as prescribed)	○	○	○	○	○	○
The patient has not followed health promotion or disease prevention advice in the past (e.g. has not stopped smoking)	○	○	○	○	○	○

38% Back Next

There are no right or wrong answers - we are keen to understand what you actually do.

If you have any queries or problems completing this survey, please email gpstudy@ucl.ac.uk

---

II - Decision making in your everyday practice

Please think about the patients you saw in the last month who you considered sending for simple investigations (such as ultrasound or X-ray) and/or referring to secondary care.

4 **To what extent do the factors listed below influence the likelihood that you will refer a patient? Select one answer for each statement.**

	More likely to refer in most cases	More likely to refer in some cases	No more or less likely to refer	Less likely to refer in some cases	Less likely to refer in most cases	Don't know
The patient has a low level of spoken English	○	○	○	○	○	○
The consultation is taking place via an interpreter	○	○	○	○	○	○
The patient will require an interpreter for their appointment/diagnostic test	○	○	○	○	○	○
The patient does not have a source of transport to or from the appointment/diagnostic test	○	○	○	○	○	○
The patient's mobility is poor	○	○	○	○	○	○
The patient is concerned that it is expensive to travel to the appointment/diagnostic test	○	○	○	○	○	○

46% Back Next

There are no right or wrong answers - we are keen to understand what you actually do.

If you have any queries or problems completing this survey, please email gpstudy@ucl.ac.uk

II - Decision making in your everyday practice

Please think about the patients you saw in the last month who you considered sending for simple investigations (such as ultrasound or X-ray) and/or referring to secondary care.

5 **To what extent do these factors influence the likelihood that you will refer a patient? Select one answer for each statement.**

	More likely to refer in most cases	More likely to refer in some cases	No more or less likely to refer	Less likely to refer in some cases	Less likely to refer in most cases	Don't know
The patient is unable to recognise the seriousness of their symptom(s)						
The patient does not express their symptom(s) clearly						
You are concerned that the patient may have difficulties weighing up the consequences of different management options						
The patient does not ask about other management options available						
The patient has independently researched their symptom(s) before their consultation						
The patient does not know what services are available to them						

53% Back Next

There are no right or wrong answers - we are keen to understand what you actually do.

If you have any queries or problems completing this survey, please email gpstudy@ucl.ac.uk

---

II - Decision making in your everyday practice

Please think about the patients you saw in the last month who you considered sending for simple investigations (such as ultrasound or X-ray) and/or referring to secondary care.

6 **To what extent do these factors influence the likelihood that you will refer a patient? Select one answer for each statement.**

	More likely to refer in most cases	More likely to refer in some cases	No more or less likely to refer	Less likely to refer in some cases	Less likely to refer in most cases	Don't know
The patient does not appear distressed about their symptom(s)						
The patient appears anxious about the referral/diagnostic test						
The patient appears concerned about the stigma associated with certain differential diagnoses						
The patient is unwilling to discuss certain differential diagnoses						
The patient says that they do not expect the diagnostic test to be accurate						
The patient is concerned about overusing the health service						

61% Back Next

There are no right or wrong answers - we are keen to understand what you actually do.

If you have any queries or problems completing this survey, please email gpstudy@ucl.ac.uk

II - Decision making in your everyday practice

Please think about the patients you saw in the last month who you considered sending for simple investigations (such as ultrasound or X-ray) and/or referring to secondary care.

7 To what extent do these factors influence the likelihood that you will refer a patient? Select one answer for each statement.

	More likely to refer in most cases	More likely to refer in some cases	No more or less likely to refer	Less likely to refer in some cases	Less likely to refer in most cases	Don't know
It is not clear which test would be most appropriate to diagnose this patient's symptoms	◯	◯	◯	◯	◯	◯
The diagnostic test is unlikely to give an accurate result for this patient	◯	◯	◯	◯	◯	◯
If the diagnostic test is positive there are limited effective treatment options available for the patient	◯	◯	◯	◯	◯	◯
Your appointments are running late	◯	◯	◯	◯	◯	◯
You are aware of the cost of the diagnostic test(s) you are considering	◯	◯	◯	◯	◯	◯
The patient would have to wait a long time for a referral/diagnostic test	◯	◯	◯	◯	◯	◯
A hospital colleague is able to provide advice promptly by telephone or email	◯	◯	◯	◯	◯	◯

69% Back Next

There are no right or wrong answers - we are keen to understand what you actually do.

If you have any queries or problems completing this survey, please email gpstudy@ucl.ac.uk

---

II - Decision making in your everyday practice

8 **If you have any further comments or reflections you wish to add on how you make decisions about sending patients for diagnostic tests, or referring them to secondary care, please type them in the box below.**

76% Back Next

There are no right or wrong answers - we are keen to understand what you actually do.

If you have any queries or problems completing this survey, please email gpstudy@ucl.ac.uk

III - Your clinical experience, responsibilities and lifestyle

Finally, we would like to ask you a few questions about your clinical experience, responsibilities and lifestyle

9 **What clinical experience have you had in the following specialties? Select as many as appropriate.**

	None	Foundation/house officer training *	Specialist training *
Cardiology	☐	☐	☐
Emergency medicine	☐	☐	☐
Geriatrics	☐	☐	☐
Oncology	☐	☐	☐
Psychiatry	☐	☐	☐
Respiratory medicine	☐	☐	☐

* GPs who entered specialist training before 2005, or who qualified outside of the UK, should consider clinical experience within 2 years of graduation to correspond to foundation/house officer training, and clinical experience 2 years or more after graduation as specialist training.

10 **Do you have any budgetary responsibility?**

	Yes	No
In your practice?	○	○
As part of a clinical commissioning group?	○	○

11 **Do you smoke? Select one answer.**

○ Current smoker
○ Ex smoker
○ Never smoked

100% Back Finish

There are no right or wrong answers - we are keen to understand what you actually do.

If you have any queries or problems completing this survey, please email gpstudy@ucl.ac.uk

---

Thank you for completing this survey. You have now completed the GP decision making study.

There are no right or wrong answers - we are keen to understand what you actually do.

If you have any queries or problems completing this survey, please email gpstudy@ucl.ac.uk

Appendix 18 : Potential reasons for non-clinical differences in GPs' referral behaviour proposed by studies in my interim systematic review

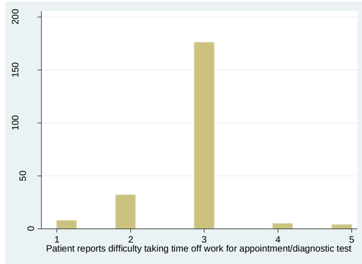
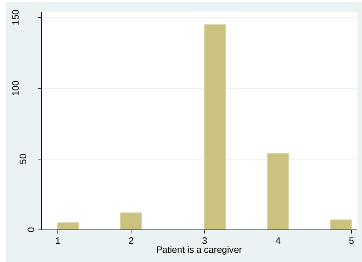
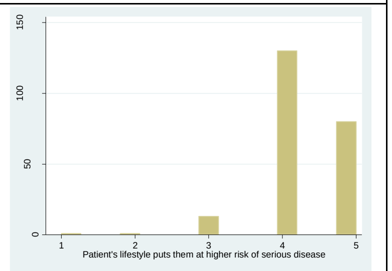
As discussed in Section 6.2.3, when designing the post-consultation survey I reviewed a number of studies which considered potential reasons for non-clinical differences in GPs' referral behaviour and used these to select factors to address in my survey. Factors these studies proposed could contribute to non-clinical variation in GP decision making are listed below.

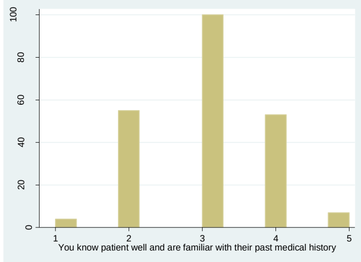
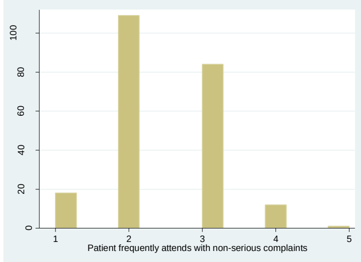
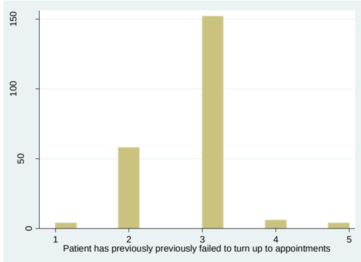
Factor contributing to variation	**Study**
Male bias in medical knowledge (not enough known about how females present, especially if is atypically)	Adams et al,[127] Crilly et al,[113] Ruiz-Cantero et al.[128]
Unawareness of patient's risk of disease	Srinivasa et al.[165]
Dilution effect (e.g. women present more frequently with symptoms but have same/less risk of disease)	Bosner et al,[130] Ruiz-Cantero et al.[128]
Diagnostic test not appropriate (e.g. exercise test in women, elderly)	Bosner et al,[130] Crilly et al.[113]
Medication not beneficial in certain patient populations	Hamilton et al,[83] Martin et al.[93]
GPs prior knowledge of patient – medical history and personality	Bosner et al,[130] Patel et al.[123]
Differences in threshold of symptoms reached before patient consults GP (therefore some do not meet referral threshold)	Bosner et al,[130] Ruiz-Cantero et al,[128] Patel et al.[123]
Over-investigation of some patients rather than under-investigation of others	Bosner et al.[130]
Differences in treatment may relate to differences in initial investigation of disease	Calvert et al.[185]
Concordance – GPs treat patients similar to themselves differently	Coyle et al,[186] Tabenkin et al.[187]
GP perceptions of likely disease severity and prognosis	Crilly et al,[113] Currin et al,[121] Patel et al.[123]
Procedure referring for more risky in certain populations (so less willing to take risk)	Crilly et al,[113] Judge et al.[124]
GP perception age contra-indication to treat	Harries et al,[125] Judge et al.[124]
Clinically appropriate – the disparities in referral are reasonable given likelihood of disease	Adams et al,[127] Crilly et al,[113] Maserejian et al,[188] Tabenkin et al,[187] de Lusignan et al,[77] Schofield et al.[189]

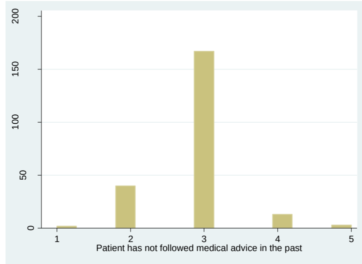
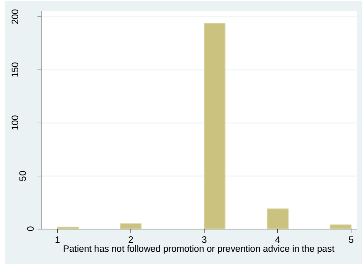
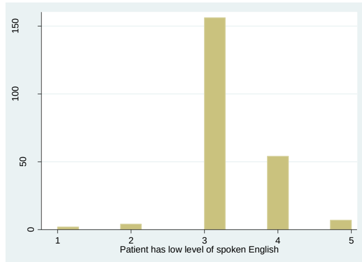
Factor contributing to variation	Study
GP personal attitude to behaviour (+/- in a certain population group)	Geirsson et al.[190]
Males desire to maintain a stoical or "strong appearance" may lead GPs to underestimate significance of a problem or symptom	Geirsson et al,[190] Judge et al.[124]
Symptom or behaviour considered normal for that patient population	Geirsson et al,[190] Patel et al,[123] Judge et al.[124]
Patient incapacitation (e.g. after operation) has significant implications	Judge et al.[124]
Patient places opinions of friends/relatives above opinions of GP	Judge et al.[124]
Concerns about side-effects of procedure	Juni et al,[78] Judge et al.[124]
Patient concerns about being dependent	Juni et al.[78]
Concerns about not being able to care for others during rehabilitation	Juni et al.[78]
Patient does not ask about other available options for treatment	Juni et al.[78]
Combinations of socio-demographic variables (e.g. gender when combined with age, SEC when combined with ethnicity)	Maserejian et al,[188] Mathur et al.[191]
Unstable occupational positions and unwillingness/unable to take time off work	Ruiz-Cantero et al.[128]
Socio-demographic variable may make you more prone to certain diseases	Aleimda et al.[192]
Concerns about overusing health service	Judge et al.[124]
Patient unable to recognise importance of symptoms	Judge et al.[124]
Patient difficulty articulating symptoms/complaint to doctor	Judge et al,[124] Norredam et al,[173] Patel et al,[123] Worth et al.[174]
Patient has done research	Judge et al.[124]
Patient has lower expectation of health care	Judge et al,[124] Norredam et al.[173]
Transport difficulties to attend secondary care	Sowden et al,[92] Srinivasa et al.[165]
Economic costs of attending referral appointment	Sowden et al.[92]
Shorter consultation times in some patients	Videau et al,[175] Norredam et al.[173]
Patient reticence in giving information in consultation	Videau et al.[175]
GPs overburdened	Videau et al.[175]
Adherence/compliance of patients	Millett et al,[87] Schofield et al.[189]
Knowledge of services available	Norredam et al,[173] Worth et al.[174]
Language barriers	Norredam et al,[173] Srinivasa et al,[165] Worth et al.[174]

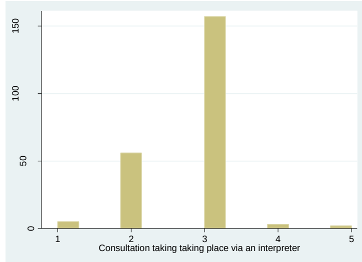
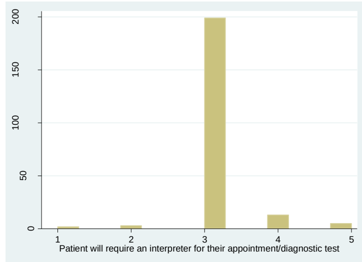
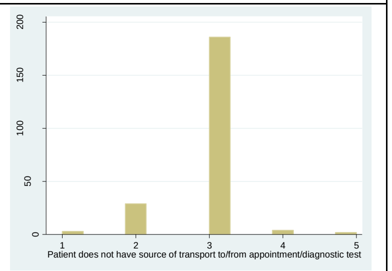
Factor contributing to variation	Study
Use of/need for interpreters	Patel et al,[123] Srinivasa et al,[165] Worth et al.[174]
Poor access to relevant and important family history if family members abroad etc	Srinivasa et al.[165]
Unwillingness of patient to discuss certain topics (e.g. if culturally inappropriate)	Norredam et al,[173] Srinivasa et al,[165] Worth et al.[174]
Patient assertiveness	Worth et al.[174]
GP's concerns about lack of cultural awareness	Worth et al.[174]

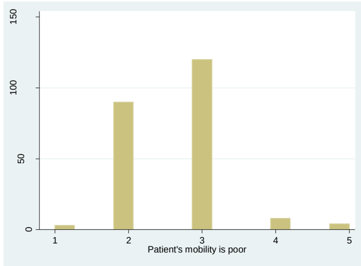
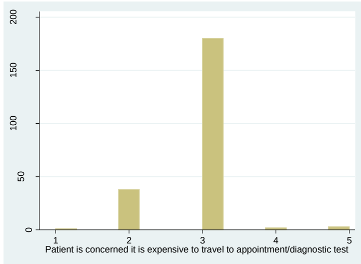
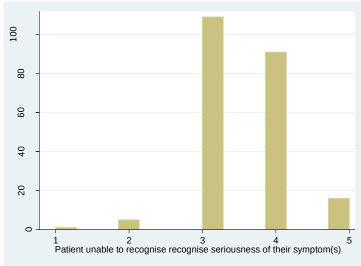
Appendix 19 : Descriptive analysis of all 33 factors that GPs were asked to evaluate the influence of on their decision making

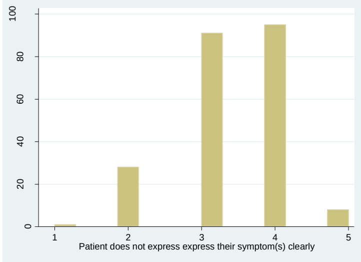
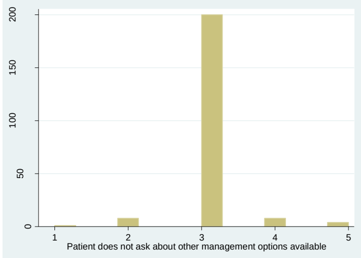
Statement	Total	'Don't know' response	Less likely to refer (%)	No more or less likely to refer (%)	More likely to refer (%)	McNemar test result (p value)	Histogram
The patient reports difficulty taking time off work for an appointment/ diagnostic test	226	1	40 (17.7%)	178 (78.8%)	8 (3.5%)	20.02 (p<0.0001)	
The patient is a caregiver	224	2	17 (7.6%)	146 (65.2%)	61 (27.2%)	23.71 (p<0.0001)	
The patient's lifestyle puts them at higher risk of serious disease	226	1	2 (0.9%)	13 (5.8%)	211 (93.4%)	203.12 (p<0.0001)	

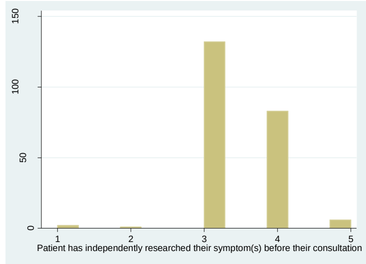
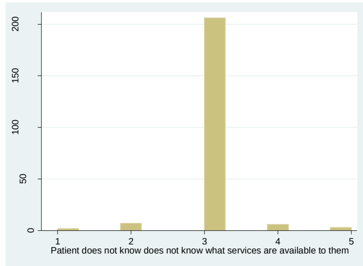
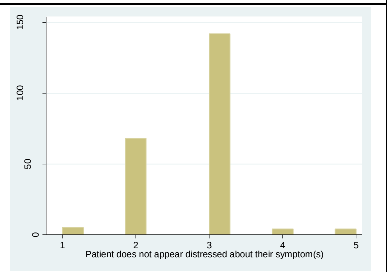
Statement	Total	'Don't know' response	Less likely to refer (%)	No more or less likely to refer (%)	More likely to refer (%)	McNemar test result (p value)	Histogram
You know the patient well and are familiar with their past medical history	220	5	59 (26.8%)	101 (45.9%)	60 (27.3%)	0.00 (p=1.0000)	
The patient frequently attends with non-serious complaints	225	2	127 (56.4%)	86 (38.2%)	12 (5.3%)	93.50 (p<0.0001)	
The patient has previously failed to turn up to primary or secondary care appointments	225	1	62 (27.6%)	154 (68.4%)	9 (4.0%)	38.08 (p<0.0001)	

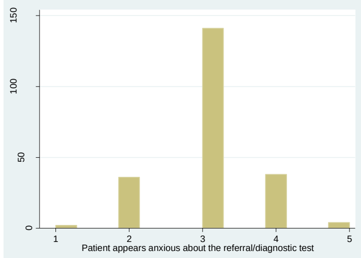
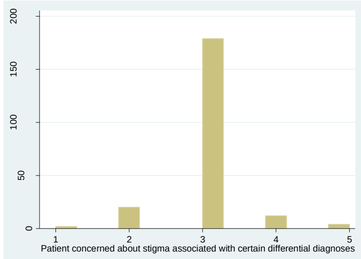
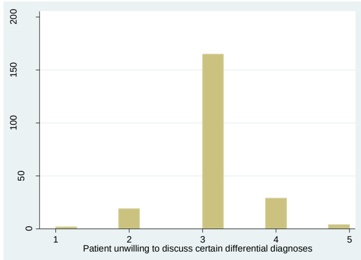
Statement	Total	'Don't know' response	Less likely to refer (%)	No more or less likely to refer (%)	More likely to refer (%)	McNemar test result (p value)	Histogram
The patient has not followed medical advice in the past (e.g. did not take medication as prescribed)	226	1	42 (18.6%)	169 (74.8%)	15 (6.6%)	11.86 (p=0.0006)	
The patient has not followed health promotion or disease prevention advice in the past (e.g. has not stopped smoking)	225	2	7 (3.1%)	196 (87.1%)	22 (9.8%)	6.76 (p=0.0093)	
The patient has a low level of spoken English	224	3	6 (2.7%)	158 (70.5%)	60 (26.8%)	42.56 (p<0.0001)	

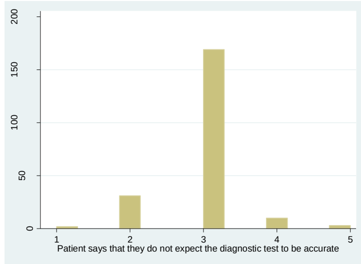
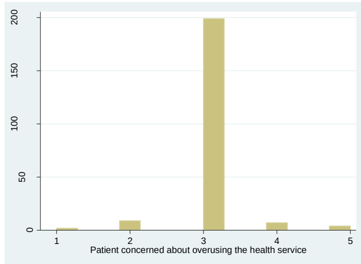
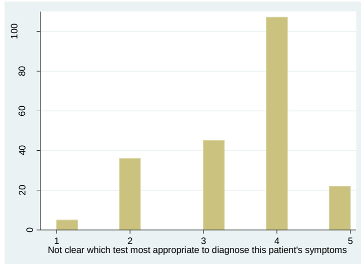
Statement	Total	'Don't know' response	Less likely to refer (%)	No more or less likely to refer (%)	More likely to refer (%)	McNemar test result (p value)	Histogram
The consultation is taking place via an interpreter	224	3	60 (26.8%)	159 (71.0%)	5 (2.2%)	44.86 (p<0.0001)	
The patient will require an interpreter for their appointment/ diagnostic test	223	3	5 (2.2%)	201 (90.1%)	17 (7.6%)	5.50 (p=0.0190)	
The patient does not have a source of transport to or from the appointment/ diagnostic test	225	2	32 (14.2%)	188 (83.6%)	5 (2.2%)	18.27 (p<0.0001)	

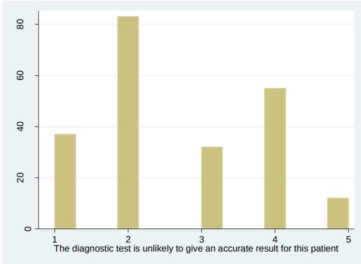
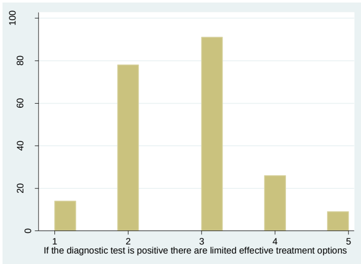
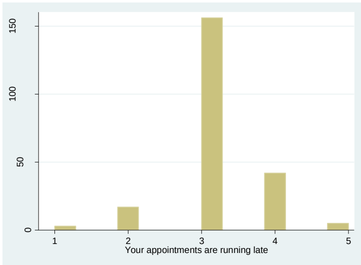
Statement	Total	'Don't know' response	Less likely to refer (%)	No more or less likely to refer (%)	More likely to refer (%)	McNemar test result (p value)	Histogram
The patient's mobility is poor	226	1	93 (41.2%)	122 (54.0%)	11 (4.9%)	63.09 ($p<0.0001$)	
The patient is concerned it is expensive to travel to the appointment/diagnostic test	225	2	39 (17.3%)	182 (80.9%)	4 (1.8%)	26.88 ($p<0.0001$)	
The patient is unable to recognise the seriousness of their symptom(s)	223	3	6 (2.7%)	110 (49.3%)	107 (48.0%)	88.50 ($p<0.0001$)	

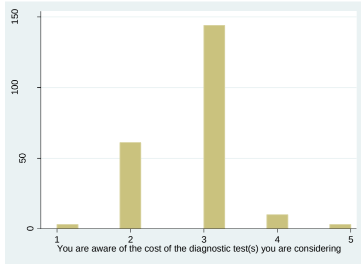
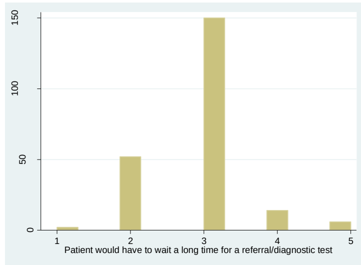
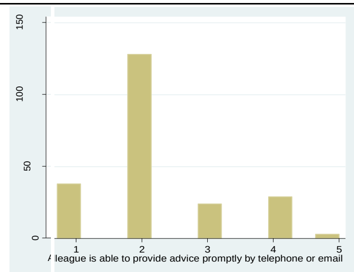
Statement	Total	'Don't know' response	Less likely to refer (%)	No more or less likely to refer (%)	More likely to refer (%)	McNemar test result (p value)	Histogram
The patient does not express their symptom(s) clearly	224	2	29 (12.9%)	92 (41.1%)	103 (46.0%)	40.37 (p<0.0001)	
You are concerned that the patient may have difficulties weighing up the consequences of different management options	220	6	8 (3.6%)	111 (50.5%)	101 (45.9%)	77.65 (p<0.0001)	
The patient does not ask about other management options available	222	3	9 (4.1%)	202 (91.0%)	11 (5.0%)	0.05 (p=0.8231)	

Statement	Total	'Don't know' response	Less likely to refer (%)	No more or less likely to refer (%)	More likely to refer (%)	McNemar test result (p value)	Histogram
The patient has independently researched their symptom(s) before their consultation	225	1	3 (1.3%)	133 (59.1%)	89 (39.6%)	78.53 (p<0.0001)	
The patient does not know what services are available to them	225	1	9 (4.0%)	208 (92.4%)	8 (3.6%)	0.00 (p=1.0000)	
The patient does not appear distressed about their symptom(s)	224	1	73 (32.6%)	144 (64.3%)	7 (3.1%)	52.81 (p<0.0001)	

Statement	Total	'Don't know' response	Less likely to refer (%)	No more or less likely to refer (%)	More likely to refer (%)	McNemar test result (p value)	Histogram
The patient appears anxious about the referral/diagnostic test	222	2	38 (17.1%)	143 (64.4%)	41 (18.5%)	0.05 (p=0.8231)	
The patient appears concerned about the stigma associated with certain differential diagnoses	218	8	22 (10.1%)	181 (83.0%)	15 (6.9%)	0.97 (p=0.3247)	
The patient is unwilling to discuss certain differential diagnoses	220	6	21 (9.5%)	167 (75.9%)	32 (14.5%)	1.89 (p=0.1692)	

Statement	Total	'Don't know' response	Less likely to refer (%)	No more or less likely to refer (%)	More likely to refer (%)	McNemar test result (p value)	Histogram
The patient says that they do not expect the diagnostic test to be accurate	216	9	33 (15.3%)	171 (79.2%)	12 (5.6%)	8.89 (p=0.0029)	
The patient is concerned about overusing the health service	222	4	11 (5.0%)	201 (90.5%)	10 (4.5%)	0.00 (p=1.0000)	
It is not clear which test would be most appropriate to diagnose this patient's symptom(s)	216	11	41 (19.0%)	46 (21.3%)	129 (59.7%)	44.52 (p<0.0001)	

Statement	Total	'Don't know' response	Less likely to refer (%)	No more or less likely to refer (%)	More likely to refer (%)	McNemar test result (p value)	Histogram
The diagnostic test is unlikely to give an accurate result for this patient	220	6	121 (55.0%)	32 (14.5%)	67 (30.5%)	14.94 (p=0.0001)	
If the diagnostic test is positive there are limited effective treatment options available for the patient	219	7	92 (42.0%)	92 (42.0%)	35 (16.0%)	24.69 (p<0.0001)	
Your appointments are running late	224	3	20 (8.9%)	157 (70.1%)	47 (21.0%)	10.09 (p=0.0015)	

Statement	Total	'Don't know' response	Less likely to refer (%)	No more or less likely to refer (%)	More likely to refer (%)	McNemar test result (p value)	Histogram
You are aware of the cost of the diagnostic test(s) you are considering	222	4	64 (28.8%)	146 (65.8%)	12 (5.4%)	34.22 (p<0.0001)	
The patient would have to wait a long time for a referral/diagnostic test	225	2	54 (24.0%)	152 (67.6%)	19 (8.4%)	15.84 (p<0.0001)	
A hospital colleague is able to provide advice promptly by telephone or email	223	4	168 (75.3%)	24 (10.8%)	31 (13.9%)	92.94 (p<0.0001)	

References

(1) Department of Health. Improving Outcomes: A Strategy for Cancer. 12-1-2011.
Ref Type: Report

(2) Office for National Statistics. Mortality in the United Kingdom, 1983-2013. 11-12-2014. 14-10-2015.
Ref Type: Report

(3) Office for National Statistics. Deaths Registered in England and Wales: 2014. 15-7-2015.
Ref Type: Report

(4) Ahmad AS, Ormiston-Smith N, Sasieni PD. Trends in the lifetime risk of developing cancer in Great Britain: comparison of risk for those born from 1930 to 1960. *Br J Cancer* 2015; 112(5):943-947.

(5) National Cancer Research Institute. NCRI Cancer Research Database. 15-6-2016. 19-6-2016.
Ref Type: Online Source

(6) Cancer Research UK website. The 20 Most Common Cancers, UK, 2013. 17-2-2016. 9-5-2016.
Ref Type: Online Source

(7) Office for National Statistics. Death Registration Summary Tables - England and Wales. 15-7-2016.
Ref Type: Report

(8) Cancer Research UK website. Lung cancer survival trends over time. 4-12-2014. 9-5-2016.
Ref Type: Online Source

(9) National Cancer Intelligence Network. Stage breakdown by CCG 2013. 2015.
Ref Type: Report

(10) Coleman MP, Forman D, Bryant H, Butler J, Rachet B, Maringe C et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *The Lancet* 2011; 377(9760):127-138.

(11) Abdel-Rahman M, Stockton D, Rachet B, Hakulinen T, Coleman MP. What if cancer survival in Britain were the same as in Europe: how many deaths are avoidable? *Br J Cancer* 2009; 101(S2):S115-S124.

(12) Walters S, Maringe C, Butler J, Brierley JD, Rachet B, Coleman MP. Comparability of stage data in cancer registries in six countries: lessons from the International Cancer Benchmarking Partnership. *Int J Cancer* 2013; 132(3):676-685.

(13) Walters S, Maringe C, Coleman MP, Peake MD, Butler J, Young N et al. Lung cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: a population-based study, 2004–2007. *Thorax* 2013.

(14) Department of Health. The Handbook to the NHS Constitution. 26-3-2013.
Ref Type: Report

(15) Cancer Research UK, Men's Health Forum, National Cancer Intelligence Network. Cancer Statistics Report – The Excess Burden of cancer in Men in the UK. 2013.
Ref Type: Report

(16) Office for National Statistics. Geographic Patterns of Cancer Survival in England: Patients followed up to 2011. 18-4-2013.
Ref Type: Report

(17) Office for National Statistics. Geographic Patterns of Cancer Survival in England: Patients followed up to 2011 - Table D: lung. 18-4-2013.
Ref Type: Report

(18) Cancer Research UK website. Cancer survival statistics. 2015. 20-10-2015.
Ref Type: Online Source

(19) Cancer Research UK website. Lung cancer survival by socio-economic variation. 26-6-2012. 20-10-2015.
Ref Type: Online Source

(20) Cancer Research UK website. Lung cancer survival by age. 4-12-2014. 20-10-2015.
Ref Type: Online Source

(21) National Cancer Intelligence Network, Cancer Research UK. Cancer Incidence and Survival by Major Ethnic Group, England 2002-2006. 2009.
Ref Type: Report

(22) Office for National Statistics. Cancer survival for Clinical Commissioning Groups in England: adults diagnosed with breast, colorectal or lung cancer during 2008-2012 and followed up to 31 December 2013. 2015.
Ref Type: Report

(23) Richards MA. The size of the prize for earlier diagnosis of cancer in England. *Br J Cancer* 2009; 101(S2):S125-S129.

(24) Richards MA, Westcombe AM, Love SB, Littlejohns P, Ramirez AJ. Influence of delay on survival in patients with breast cancer: a systematic review. *The Lancet* 1999; 353(9159):1119-1126.

(25) Neal R, Tharmanathan P, France B, Din N, Cotton S, Fallon-Ferguson J et al. Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *Br J Cancer* 2015; 112(s1):S92-S107.

(26) Torring ML, Frydenberg M, Hansen RP, Olesen F, Hamilton W, Vedsted P. Time to diagnosis and mortality in colorectal cancer: a cohort study in primary care. *Br J Cancer* 2011; 104(6):934-940.

(27) Moss SM, Wale C, Smith R, Evans A, Cuckle H, Duffy SW. Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial. *The Lancet Oncology* 16(9):1123-1132.

(28) Ait Ouakrim D, Pizot C, Boniol M, Malvezzi M, Boniol M, Negri E et al. Trends in colorectal cancer mortality in Europe: retrospective analysis of the WHO mortality database. *BMJ* 2015; 351.

(29) Neal RD. Do diagnostic delays in cancer matter? *Br J Cancer* 2009; 101(S2):S9-S12.

(30) Torring ML, Frydenberg M, Hansen RP, Olesen F, Vedsted P. Evidence of increasing mortality with longer diagnostic intervals for five common cancers: a cohort study in primary care. *European Journal of Cancer* 2013; 49(9):2187-2198.

(31) Hamilton W, Green T, Martins T, Elliott K, Rubin G, Macleod U. Evaluation of risk assessment tools for suspected cancer in general practice: a cohort study. *Br J Gen Pract* 2013; 63(606):e30-e36.

(32) The National Lung Screening Trial Research Team. Reduced lung cancer mortality with low-dose Computed Tomographic screening. *N Engl J Med* 2011; 365(5):395-409.

(33) Lyratzopoulos G, Abel GA, Barbiere JM, Brown CH, Rous BA, Greenberg DC. Variation in advanced stage at diagnosis of lung and female breast cancer in an English region 2006-2009. *Br J Cancer* 2012; 106(6):1068-1075.

(34) Hamilton W. Cancer diagnosis in primary care. *Br J Gen Pract* 2010; 60:121-128.

(35) Suspected cancer: recognition and referral. 2015.
Ref Type: Report

(36) Stapley S, Sharp D, Hamilton W. Negative chest X-rays in primary care patients with lung cancer. *Br J Gen Pract* 2006; 56(529):570-573.

(37) Speets AM, van der Graaf Y, Hoes AW, Kalmijn S, Sachs AP, Rutten MJ et al. Chest radiography in general practice: indications, diagnostic yield and consequences for patient management. *Br J Gen Pract* 2006; 56(529):574-578.

(38) Walter F, Webster A, Scott S, Emery J. The Andersen Model of Total Patient Delay: a systematic review of its application in cancer diagnosis. *Journal of Health Services Research & Policy* 2012; 17(2):110-118.

(39) Olesen F, Hansen RP, Vedsted P. Delay in diagnosis: the experience in Denmark. *Br J Cancer* 2009; 101(S2):S5-S8.

(40) Weller D, Vedsted P, Rubin G, Walter FM, Emery J, Scott S et al. The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. *Br J Cancer* 2012; 106(7):1262-1267.

(41) Baughan P, O'Neill B, Fletcher E. Auditing the diagnosis of cancer in primary care: the experience in Scotland. *Br J Cancer* 2009; 101(S2):S87-S91.

(42) Lyratzopoulos G, Abel GA, McPhail S, Neal RD, Rubin GP. Measures of promptness of cancer diagnosis in primary care: secondary analysis of national audit data on patients with 18 common and rarer cancers. *Br J Cancer* 2013; 108(3):686-690.

(43) Lyratzopoulos G, Saunders CL, Abel GA, McPhail S, Neal RD, Wardle J et al. The relative length of the patient and the primary care interval in patients with 28 common and rarer cancers. *Br J Cancer* 2015; 112(s1):S35-S40.

(44) Lyratzopoulos G, Neal RD, Barbiere JM, Rubin G, Abel GA. Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England. *The Lancet Oncology* 2012; 13(4):353-365.

(45) Raine R, Wong W, Scholes S, Ashton C, Obichere A, Ambler G. Social variations in access to hospital care for patients with colorectal, breast, and lung cancer between 1999 and 2006: retrospective analysis of hospital episode statistics. *BMJ* 2010; 340.

(46) Forrest LF, Adams J, Wareham H, Rubin G, White M. Socioeconomic inequalities in lung cancer treatment: systematic review and meta-analysis. *PLoS Med* 2013; 10(2):e1001376.

(47) Rubin G, McPhail S, Elliott K. National Audit of Cancer Diagnosis in Primary Care. 2011. Royal College of General Practioners.
Ref Type: Report

(48) Macleod U, Mitchell ED, Burgess C, Macdonald S, Ramirez AJ. Risk factors for delayed presentation and referral of symptomatic cancer: evidence for common cancers. *Br J Cancer* 2009; 101(S2):S92-S101.

(49) Meechan D, Gildea C, Hollingworth L, Richards MA, Riley D, Rubin G. Variation in use of the 2-week referral pathway for suspected cancer: a cross-sectional analysis. *Br J Gen Pract* 2012; 62(602):e590-e597.

(50) Barrett J, Hamilton W. Pathways to the diagnosis of lung cancer in the UK: a cohort study. *BMC Family Practice* 2008; 9:31.

(51) Elliss-Brookes L, McPhail S, Ives A, Greenslade M, Shelton J, Hiom S et al. Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. *Br J Cancer* 2012; 107(8):1220-1226.

(52) Oehler C. Presentation at 3rd NAEDI conference. 2015.
Ref Type: Unpublished Work

(53) Mitchell ED, Pickwell-Smith B, Macleod U. Risk factors for emergency presentation with lung and colorectal cancers: a systematic review. *BMJ Open* 2015; 5(4).

(54) Abel GA, Shelton J, Johnson S, Elliss-Brookes L, Lyratzopoulos G. Cancer-specific variation in emergency presentation by sex, age and deprivation across 27 common and rarer cancers. *Br J Cancer* 2015; 112(s1):S129-S136.

(55) Sheringham JR, Georghiou T, Chitnis XA, Bardsley M. Comparing primary and secondary health-care use between diagnostic routes before a colorectal cancer diagnosis: cohort study using linked data. *Br J Cancer* 2014; 111(8):1490-1499.

(56) Macleod U. Presentation at 3rd NAEDI conference. 2015.
Ref Type: Unpublished Work

(57) Walabyeki J. Presentation at 3rd NAEDI conderence. 2015.
Ref Type: Unpublished Work

(58) Lyratzopoulos G, Vedsted P, Singh H. Understanding missed opportunities for more timely diagnosis of cancer in symptomatic patients after presentation. *Br J Cancer* 2015; 112(s1):S84-S91.

(59) McCartney M. One problem. *BMJ : British Medical Journal* 2014; 348.

(60) Hobbs FDR, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007-2014. *The Lancet* 2016; 387(10035):2323-2330.

(61) Andersen R, Torring M, Vedsted P. Global health care–seeking discourses facing local clinical realities: exploring the case of cancer. Medical Anthropology Quarterly 29, 237-255. 2015.

(62) Forbes LJL, Simon AE, Warburton F, Boniface D, Brain KE, Dessaix A et al. Differences in cancer awareness and beliefs between Australia, Canada, Denmark, Norway, Sweden and the UK (the International Cancer Benchmarking Partnership): do they contribute to differences in cancer survival[quest]. *Br J Cancer* 2013; 108(2):292-300.

(63) Hamilton W. The CAPER studies: five case-control studies aimed at identifying and quantifying the risk of cancer in symptomatic primary care patients. *Br J Cancer* 2009; 101:Suppl-6.

(64) Weller D, Campbell C. Early lung cancer detection: the role of primary care. *Primary Care Respiratory Journal* 2006; 15(6):323-325.

(65) McBride D, Hardoon S, Walters K, Gilmour S, Raine R. Explaining variation in referral from primary to secondary care: cohort study. *BMJ (Clinical research ed)* 2010; 341(pp c6267):2010.

(66) Banks J, Hollinghurst S, Bigwood L, Peters TJ, Walter FM, Hamilton W. Preferences for cancer investigation: a vignette-based study of primary-care attendees. *The Lancet Oncology* 15(2):232-240.

(67) Hajjaj FM, Salek MS, Basra MKA, Finlay AY. Non-clinical influences on clinical decision-making: a major challenge to evidence-based practice. *Journal of the Royal Society of Medicine* 2010; 103(5):178-187.

(68) Orwin RG. Evaluating coding decisions. In: Cooper H, Hedges L, editors. The handbook of research synthesis. Russel Sage Foundation, New York; 1994. 139-162.

(69) Critical Appraisal Skills Programme. CASP checklists. 2012. 9-1-2013.
Ref Type: Online Source

(70) Healthcare Improvement Scotland. Critical Appraisal: Notes and Checklists.
26-4-2013. 2-5-2013.
Ref Type: Online Source

(71) Heller R, Verma A, Gemmell I, Harrison R, Hart J, Edwards R. Critical appraisal for public health: a new checklist. *Public Health* 2008; 122(1):92-98.

(72) Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *International Journal of Epidemiology* 2007; 36(3):666-676.

(73) Higgins JPT, Green S, The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions (version 5.1.0). 2011.
Ref Type: Online Source

(74) Smith R. The Cochrane Collaboration at 20. *BMJ : British Medical Journal* 2013; 347.

(75) Bonte M, von dem Knesebeck O, Siegrist J, Marceau L, Link C, Arber S et al. Women and men with coronary heart disease in three countries: are they treated differently? *Women's Health Issues* 2008; 18(3):May.

(76) Macfarlane GJ, Beasley M, Jones EA, Prescott GJ, Docking R, Keeley P et al. The prevalence and management of low back pain across adulthood: results from a population-based cross-sectional study (the MUSICIAN study). *Pain* 2012; 153(1):January.

(77) de Lusignan S, Nitsch D, Belsey J, Kumarapeli P, Vamos EP, Majeed A et al. Disparities in testing for renal function in UK primary care: cross-sectional study. *Family Practice* 2011;(6):638-646.

(78) Juni P, Low N, Reichenbach S, Villiger PM, Williams S, Dieppe PA. Gender inequity in the provision of care for hip disease: population-based cross-sectional study. *Osteoarthritis and Cartilage* 2010; 18(5):May.

(79) Tate A, Nicholson A, Cassell J. Are GPs under-investigating older patients presenting with symptoms of ovarian cancer? Observational study using General Practice Research Database. *Br J Cancer* 2010; 102(6):947-951.

(80) Kendrick T, Dowrick C, McBride A, Howe A, Clarke P, Maisey S et al. Management of depression in UK general practice in relation to scores on depression severity questionnaires: analysis of medical record data. *BMJ* 2009;(7697):1-11.

(81) Vamos EP, Pape UJ, Bottle A, Hamilton FL, Curcin V, Ng A et al. Association of practice size and pay-for-performance incentives with the quality of diabetes management in primary care. *Canadian Medical Association Journal* 2011; 183(12):E809-E816.

(82) Coleman J, McDowell S, Evans S, Gill P, Ferner R. Oversight: a retrospective study of biochemical monitoring in patients beginning antihypertensive drug treatment in primary care. *British Journal of Clinical Pharmacology* 2010; 70(1):July.

(83) Hamilton FL, Bottle A, Vamos EP, Curcin V, Anthea, Molokhia M et al. Impact of a pay-for-performance incentive scheme on age, sex, and socioeconomic disparities in diabetes management in UK primary care. *Journal of Ambulatory Care Management* 2010; 33(4):October-December.

(84) Verma A, Birger R, Bhatt H, Murray J, Millett C, Saxena S et al. Ethnic disparities in diabetes management: a 10-year population-based repeated cross-sectional study in UK primary care. *Journal of Public Health* 2010; 32(2):June.

(85) McGovern MP, Boroujerdi MA, Taylor MW, Williams DJ, Hannaford PC, Lefevre KE et al. The effect of the UK incentive-based contract on the management of patients with coronary heart disease in primary care. *Family Practice* 2008;(1):33-39.

(86) McGovern MP, Williams DJ, Hannaford PC, Taylor MW, Lefevre KE, Boroujerdi MA et al. Introduction of a new incentive and target-based contract for family physicians in the UK: good for older patients with diabetes but less good for women? *Diabetic Medicine* 2008; 25(9):1083-1089.

(87) Millett C, Gray J, Bottle A, Majeed A. Ethnic disparities in blood pressure management in patients with hypertension after the introduction of pay for performance. *Annals of family medicine* 2008;(6):490-496.

(88) Phatak H, Wentworth C, Burke TA. Lipid testing among patients beginning statin therapy in general practice in the United Kingdom. *Value in Health* 2008; 11(5):933-938.

(89) Millett C, Car J, Eldred D, Khunti K, Mainous AG, III, Majeed A. Diabetes prevalence, process of care and outcomes in relation to practice size, caseload and deprivation: national cross-sectional study in primary care. *Journal of the Royal Society of Medicine* 2007; 100(6):275-283.

(90) Saxena S, Car J, Eldred D, Soljak M, Majeed A. Practice size, caseload, deprivation and quality of care of patients with coronary heart disease, hypertension and stroke in primary care: national cross-sectional study. *BMC Health Services Research* 2007; 7:96.

(91) Simpson CR, Hippisley-Cox J, Sheikh A. Trends in the epidemiology of smoking recorded in UK general practice. *Br J Gen Pract* 2010; 60(572):Mar.

(92) Sowden SL, Breeze E, Barber J, Raine R. Do general practices provide equitable access to physical activity interventions? *Br J Gen Pract* 2008; 58(555):Oct.

(93) Martin A, Badrick E, Mathur R, Hull S. Effect of ethnicity on the prevalence, severity, and management of COPD in general practice. *Br J Gen Pract* 2012; 62(595):February.

(94) Baughan P, Keatings J, O'Neill B. Urgent suspected cancer referrals from general practice: audit of compliance with guidelines and referral outcomes. *Br J Gen Pract* 2011; 61(592):2011.

(95) de Lusignan S, Navarro R, Chan T, Parry G, Dent-Brown K, Kendrick T. Detecting referral and selection bias by the anonymous linkage of practice, hospital and clinic data using Secure and Private Record Linkage (SAPREL): case study from the evaluation of the Improved Access to Psychological Therapy (IAPT) service. *BMC Med Inform Decis Mak* 2011; 11.

(96) Grimshaw JM, Eccles MP, Steen N, Johnston M, Pitts NB, Glidewell L et al. Applying psychological theories to evidence-based clinical practice: identifying factors predictive of lumbar spine x-ray for low back pain in UK primary care practice. *Implementation Science* 2011; 6.

(97) Hammouche S, Holland R, Steel N. Does quality of care for hypertension in primary care vary with postcode area deprivation? An observational study. *BMC Health Services Research* 2011; 11.

(98) Hu MTM, Butterworth R, Kumar V, Cooper J, Jones E, Catterall L et al. How common and what are the determinants of sub-optimal care for Parkinson's disease patients: the Milton Keynes community study. *Parkinsonism & Related Disorders* 2011;(3):177-181.

(99) Jinks C, Vohora K, Young J, Handy J, Porcheret M, Jordan KP. Inequalities in primary care management of knee pain and disability in older adults: an observational cohort study. *Rheumatology* 2011; 50(10):ker179.

(100) Raymond M, Iliffe S, Kharicha K, Harari D, Swift C, Gillmann G et al. Health risk appraisal for older people 5: self-efficacy in patient-doctor interactions. *Primary Health Care Research and Development* 2011;(4):348-356.

(101) Wagg A, Duckett J, McClurg D, Harari D, Lowe D. To what extent are national guidelines for the management of urinary incontinence in women adhered? Data from a national audit. *BJOG: An International Journal of Obstetrics & Gynaecology* 2011; 118(13):1592-1600.

(102) McGorm K, Burton C, Weller D, Murray G, Sharpe M. Patients repeatedly referred to secondary care with symptoms unexplained by organic disease: prevalence, characteristics and referral pattern. *Family Practice* 2010;(5):479-486.

(103) Murray J, Saxena S, Millett C, Curcin V, de Lusignan S, Majeed A. Reductions in risk factors for secondary prevention of coronary heart disease by ethnic group in south-west London: 10-year longitudinal study (1998-2007). *Family Practice* 2010;(4):430-438.

(104) Kumar K, Daley E, Khattak F, Buckley CD, Raza K. The influence of ethnicity on the extent of, and reasons underlying, delay in general practitioner consultation in patients with RA. *Rheumatology* 2010; 49(5):May.

(105) Nicholson A, Rait G, Murray-Thomas T, Hughes G, Mercer CH, Cassell J. Management of epididymo-orchitis in primary care: results from a large UK primary care database. *Br J Gen Pract* 2010; 60(579):October.

(106) Sadler KE, Low N, Mercer CH, Sutcliffe LJ, Islam MA, Shafi S et al. Testing for sexually transmitted infections in general practice: cross-sectional study. *BMC Public Health* 2010; 10.

(107) Fischbacher CM, Bhopal R, Steiner M, Morris AD, Chalmers J. Is there equity of service delivery and intermediate outcomes in South Asians with type 2 diabetes? Analysis of DARTS database and summary of UK publications. *Journal of Public Health* 2009; 31(2):239-249.

(108) Ingram JC, Calnan MW, Greenwood RJ, Kemple T, Payne S, Rossdale M. Risk taking in general practice: GP out-of-hours referrals to hospital. *Br J Gen Pract* 2009; 59(558):e16-e24.

(109) Loo B, Parnell C, Brook G, Southall E, Mahy I. Atrial fibrillation in a primary care population: how close to NICE guidelines are we? *Clinical Medicine* 2009; 9(3):219-223.

(110) Green H, Johnston O, Cabrini S, Fornai G, Kendrick T. General practitioner attitudes towards referral of eating-disordered patients: a vignette study based on the theory of planned behaviour. *Mental Health in Family Medicine* 2008; 5(4):December.

(111) Tahrani AA, McCarthy M, Godson J, Taylor S, Slater H, Capps N et al. Impact of practice size on delivery of diabetes care before and after the Quality and Outcomes Framework implementation. *Br J Gen Pract* 2008; 58(553):576-579.

(112) Ashworth M, Seed P, Armstrong D, Durbaba S, Jones R. The relationship between social deprivation and the quality of primary care: a national survey using indicators from the UK Quality and Outcomes Framework. *Br J Gen Pract* 2007; 57(539):441-448.

(113) Crilly M, Bundred P, Hu X, Leckey L, Johnstone F. Gender differences in the clinical management of patients with angina pectoris: a cross-sectional survey in primary care. *BMC Health Services Research* 2007; 7:142.

(114) Gray J, Millett C, Saxena S, Netuveli G, Khunti K, Majeed A. Ethnicity and quality of diabetes care in a health system with universal coverage: population-based cross-sectional survey in primary care. *Journal of general internal medicine* 2007; 22(9):1317-1320.

(115) McLean G, Guthrie B, Sutton M. Differences in the quality of primary medical care services by remoteness from urban settlements. *Quality & Safety in Health Care* 2007; 16(6):446-449.

(116) Parker C, Hippisley-Cox J, Coupland C, Vinogradova Y. Rectal and postmenopausal bleeding: consultation and referral of patients with and without severe mental health problems. *Br J Gen Pract* 2007; 57(538):371-376.

(117) Ridsdale L, Clark LV, Dowson AJ, Goldstein LH, Jenkins L, McCrone P et al. How do patients referred to neurologists for headache differ from those managed in primary care? *Br J Gen Pract* 2007; 57(538):388-395.

(118) Roberts L, Roalfe A, Wilson S, Lester H. Physical health care of patients with schizophrenia in primary care: a comparative study. *Family Practice* 2007; 24(1):34-40.

(119) Rossdale M, Kemple T, Payne S, Calnan M, Greenwood R. An observational study of variation in GPs' out-of-hours emergency referrals. *Br J Gen Pract* 2007; 57(535):152-154.

(120) Weich S, Nazareth I, Morgan L, King M. Treatment of depression in primary care. Socio-economic status, clinical need and receipt of treatment. *British Journal of Psychiatry* 2007; 191:164-169.

(121) Currin L, Schmidt U, Waller G. Variables that influence diagnosis and treatment of the eating disorders within primary care settings: a vignette study. *Int J Eat Disord* 2007; 40(3):257-262.

(122) Yohannes AM, Hann M, Sibbald B. The management of depressive symptoms in patients with COPD: a postal survey of general practitioners. *Primary Health Care Research and Development* 2011;(3):237-244.

(123) Patel S, Peacock S, McKinley R, Carter DC, Watson PJ. GPs' experience of managing chronic pain in a South Asian community - a qualitative study of the consultation process. *Family Practice* 2008; 25(2):71-77.

(124) Judge A, Welton NJ, Sandhu J, Ben-Shlomo Y. Equity in access to total joint replacement of the hip and knee in England: cross sectional study. *BMJ* 2010; 341(7770):26.

(125) Harries C, Forrest D, Harvey N, McClelland A, Bowling A. Which doctors are influenced by a patient's age? A multi-method study of angina treatment in general practice, cardiology and gerontology. *Quality & Safety in Health Care* 2007; 16(1):23-27.

(126) Office for National Statistics. Older People's Day 2011. 29-9-2012.
Ref Type: Report

(127) Adams A, Buckingham CD, Lindenmeyer A, McKinlay JB, Link C, Marceau L et al. The influence of patient and doctor gender on diagnosing coronary heart disease. *Sociology of Health & Illness* 2008;(1):1-18.

(128) Ruiz-Cantero MT, Ronda E, Alvarez-Dardet C. The importance of study design strategies in gender bias research: the case of respiratory disease management in primary care. *Journal of Epidemiology & Community Health* 2007; 61:Suppl-16.

(129) Dunstan S. General Lifestyle Survey Overview : A report on the 2009 General Lifestyle Survey. 2011.
Ref Type: Report

(130) Bosner S, Haasenritter J, Hani MA, Keller H, Sonnichsen AC, Karatolios K et al. Gender bias revisited: new insights on the differential management of chest pain. *BMC Family Practice* 2011; 12(pp 45):2011.

(131) Wang Y, Hunt K, Nazareth I, Freemantle N, Petersen I. Do men consult less than women? An analysis of routinely collected UK general practice data. *BMJ Open* 2013; 3(8).

(132) Burton J, Nandi A, Platt L. Measuring ethnicity: challenges and opportunities for survey research. *Ethnic and Racial Studies* 2010; 33(8):1332-1349.

(133) Connelly R, Gayle V, Lambert PS. Ethnicity and ethnic group measures in social survey research. *Methodological Innovations* 2016; 9.

(134) Sekhri N, Timmis A, Hemingway H, Walsh N, Eldridge S, Junghans C et al. Is access to specialist assessment of chest pain equitable by age, gender, ethnicity and socioeconomic status? An enhanced ecological analysis. *BMJ Open* 2012; 2(3).

(135) Mathur R, Hull SA, Boomla K, Robson J. Ethnic differences in primary care management of diabetes and cardiovascular disease in people with serious mental illness. *Br J Gen Pract* 2012; 62(601):e582-e588.

(136) Ghali S, Fisher HL, Joyce J, Major B, Hobbs L, Soni S et al. Ethnic variations in pathways into early intervention services for psychosis. *The British Journal of Psychiatry* 2013; 202(4):277-283.

(137) Brown JSL, Ferner H, Wingrove J, Aschan L, Hatch SL, Hotopf M. How equitable are psychological therapy services in South East London now? A comparison of referrals to a new psychological therapy service with participants in a psychiatric morbidity survey in the same London borough. *Social Psychiatry and Psychiatric Epidemiology* 2014; 49(12):1893-1902.

(138) Chauhan M, Bankart MJ, Labeit A, Baker R. Characteristics of general practices associated with numbers of elective admissions. *Journal of Public Health* 2012; 34(4):584-590.

(139) Thomas J, McNaught J, Ananiadou S. Applications of text mining within systematic reviews. *Res Syn Meth* 2011; 2(1):1-14.

(140) McKinlay JB. The unexpected influence of physician attributes on clinical decisions: results of an experiment. *Journal of Health and Social Behavior* 2002; 43(1):92-106.

(141) Epstein SA, Hooper LM, Weinfurt KP, DePuy V, Cooper LA, Harless WG et al. Primary care physicians' evaluation and treatment of depression: results of an experimental study using video vignettes. *Medical Care Research and Review* 2008;(6):674-695.

(142) Lutfey KE, Campbell SM, Renfrew MR, Marceau LD, Roland M, McKinlay JB. How are patient characteristics relevant for physicians' clinical decision making in diabetes? An analysis of qualitative results from a cross-national factorial experiment. *Social Science & Medicine* 2008; 67(9):1391-1399.

(143) Cancer Research UK website. Module 3: Beliefs, behaviours and systems in primary care. 1-12-2010.
Ref Type: Online Source

(144) Cottrell E, Roddy E, Foster NE. The attitudes, beliefs and behaviours of GPs regarding exercise for chronic knee pain: a systematic review. *BMC Family Practice* 2010; 11:4.

(145) Raine R. Bias measuring bias. *Journal of Health Services Research & Policy* 2002; 7(1):65-67.

(146) Ijas-Kallio T, Ruusuvuori J, Perakyla A. Patient involvement in problem presentation and diagnosis delivery in primary care. *Communication & Medicine* 2010; 7(2):131-141.

(147) Blumenthal-Barby JS, Krieger H. Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. *Medical decision making* 2015; 35(4):539-557.

(148) Kostopoulou O, Oudhoff J, Nath R, Delaney BC, Munro CW, Harries C et al. Predictors of diagnostic accuracy and safe management in difficult diagnostic problems in family medicine. *Medical decision making* 2008; 28(5):668-680.

(149) Kostopoulou O, Rosen A, Round T, Wright E, Douiri A, Delaney B. Early diagnostic suggestions improve accuracy of GPs: a randomised controlled trial using computer-simulated patients. *Br J Gen Pract* 2014; 65(630):e49-e54.

(150) Opinio 6.6 [2012.

(151) Cancer Research UK website. Cancer survival and deprivation statistics. 25-9-2009. 24-4-2013.
Ref Type: Online Source

(152) The diagnosis and treatment of lung cancer (update): full guidance. 2011. National Institute for Clinical Excellence.
Ref Type: Report

(153) Hippisley-Cox J, Coupland C. Identifying patients with suspected lung cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* 2011; 61(592):2011.

(154) Hamilton W, Peters TJ, Round A, Sharp D. What are the clinical features of lung cancer before the diagnosis is made? A population based case-control study. *Thorax* 2005; 60(12):1059-1065.

(155) Hamilton W. 2012.
Ref Type: Personal Communication

(156) Biesalski HK. European Consensus Statement on Lung Cancer: risk factors and prevention. Lung Cancer Panel. *Ca* 1998; 48(3):167-176.

(157) Shim J, Brindle L, Simon M, George S. A systematic review of symptomatic diagnosis of lung cancer. *Family Practice* 2014; 31(2):137-148.

(158) UK Health Departments Research Ethics Service. Standard Operating Procedures for Research Ethics Committees (version 7.1). 2016.
Ref Type: Online Source

(159) Stata Statistical Software : Release 12. College Station, TX: StataCorp LP [2011.

(160) Ageism and age discrimination in secondary health care in the United Kingdom. 2009.

(161) Health and Social Care Information Centre. General and Personal Medical Services, England, 2005-2015. 27-4-2016.
Ref Type: Report

(162) General Medical Council. List of registered medical practitioners - statistics. 6-10-2016.
Ref Type: Online Source

(163) Scott SE, Walter FM, Webster A, Sutton S, Emery J. The model of pathways to treatment: conceptualization and integration with existing theory. British Journal of Health Psychology 18, 45-65. 2013.

(164) Bowling A. Techniques of questionnaire design. In: Bowling A, Ebrahim S, editors. Handbook of Research Methods in Health: Investigation, Measurement and Analysis. Open University Press; 2005.

(165) Srinivasa J, Rowett E, Dharni N, Bhatt H, Day M, Chu CE. Improving access to cancer genetics services in primary care: socio-economic data from North Kirklees. *Familial Cancer* 2007; 6(2):197-203.

(166) Rosen R, Florin D, Hutt R. An anatomy of GP referral decisions : A qualitative study of GPs' views on their role in supporting patient choice. 2007.
Ref Type: Report

(167) Sun Z, Ng C, Halkett G, Meng R, Jiwa M. An investigation of factors that influence general practitioners' referral of computed tomography scans in patients with headache. *Int J Clin Pract* 2013; 67(7):682-690.

(168) Grimshaw J, Eccles M, Thomas R, MacLennan G, Ramsay C, Fraser C et al. Toward evidence-based quality improvement. *Journal of general internal medicine* 2006; 21(S2):S14-S20.

(169) Gagliardi AR, Brouwers MC, Palda VA, Lemieux-Charles L, Grimshaw JM. How can we improve guideline use? A conceptual framework of implementability. *Implementation Science* 2011; 6(1):1-11.

(170) Cabana M, Rand C, Powe N, Wu A, Wilson M, Abboud P et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 282[15], 1458-1465. 20-10-1999.

(171) Kaffash J. Seven in ten GPs ignore NICE guidance at least once a month. 4-3-2015.
Ref Type: Online Source

(172) Doll R. Smoking and carcinoma of the lung; preliminary report. *BMJ* 1950; 2(4682):739-748.

(173) Norredam M, Nielsen SS, Krasnik A. Migrants' utilization of somatic healthcare services in Europe - a systematic review. *European Journal of Public Health* 2010; 20(5):Oct.

(174) Worth A, Irshad T, Bhopal R, Brown D, Lawton J, Grant E et al. Vulnerability and access to care for South Asian Sikh and Muslim patients with life limiting illness in Scotland: prospective longitudinal qualitative study. *BMJ* 2009; 338.

(175) Videau Y, Saliba-Serre B, Paraponaris A, Ventelou B. Why patients of low socioeconomic status with mental health problems have shorter consultations with general practitioners. *Journal of Health Services & Research Policy* 2010; 15(2):76-81.

(176) Akbari A, Mayhew A, Al-Alawi MA, Grimshaw J, Winkens R, Glidewell E et al. Interventions to improve outpatient referrals from primary care to secondary care. *Cochrane Database Syst Rev* 2008;(4):CD005471.

(177) Blank L, Baxter S, Woods HB, Goyder E, Lee A, Payne N et al. Referral interventions from primary to specialist care: a systematic review of international evidence. *Br J Gen Pract* 2014; 64(629):e765-e774.

(178) Guthrie B, Wyke S. Personal continuity and access in UK general practice: a qualitative study of general practitioners' and patients' perceptions of when and how they matter. *BMC Family Practice* 2006; 7(1):1-9.

(179) Freeman G, Hughes J. Continuity of care and the patient experience. 2010.
Ref Type: Report

(180) NHS England. The NHS belongs to people : a call to action. 2013.
Ref Type: Report

(181) Robertson R, The King's Fund. Six ways in which NHS financial pressures can affect patient care. 31-3-2016. 20-6-2016.
Ref Type: Online Source

(182) Stead M, Angus K, Holme I, Cohen D, Tait G, The PESCE European Research Team. Factors influencing European GPs' engagement in smoking cessation: a multi-country literature review. *Br J Gen Pract* 2009; 59(566):682-690.

(183) McKinlay JB, Link CL, Freund KM, Marceau LD, O'Donnell AB, Lutfey KL. Sources of variation in physician adherence with clinical guidelines: results from a factorial experiment. *Journal of general internal medicine* 2007; 22(3):289-296.

(184) Public Health England. Recent Trends in Life Expectancy at Older Ages. 2015.
Ref Type: Report

(185) Calvert MJ, Shankar A, McManus RJ, Ryan R, Freemantle N. Evaluation of the management of heart failure in primary care. *Family Practice* 2009;(2):145-153.

(186) Coyle E, Hanley K, Sheerin J. Who goes where? A prospective study of referral patterns within a newly established primary care team. *Irish Journal of Medical Science* 2011; 180(4):December.

(187) Tabenkin H, Eaton CB, Roberts MB, Parker DR, McMurray JH, Borkan J. Differences in cardiovascular disease risk factor management in primary care by sex of physician and patient. *Annals of family medicine* 2010;(1):25-32.

(188) Maserejian NN, Link CL, Lutfey KL, Marceau LD, McKinlay JB. Disparities in physicians' interpretations of heart disease symptoms by patient gender: results of a video vignette factorial experiment. *Journal of Women's Health* 2009; 18(10):1661-1667.

(189) Schofield P, Saka O, Ashworth M. Ethnic differences in blood pressure monitoring and control in south east London. *Br J Gen Pract* 2011; 61(585):Apr.

(190) Geirsson M, Hensing G, Spak F. Does gender matter? A vignette study of general practitioners' management skills in handling patients with alcohol-related problems. *Alcohol & Alcoholism* 2009; 44(6):620-625.

(191) Mathur R, Hull SA, Badrick E, Robson J. Cardiovascular multimorbidity: the effect of ethnicity on prevalence and risk factor management. *Br J Gen Pract* 2011; 61(586):May.

(192) Almeida OP, Pirkis J, Kerse N, Sim M, Flicker L, Snowdon J et al. Socioeconomic disadvantage increases risk of prevalent and persistent depression in later life. *Journal of Affective Disorders* 2012; 138(3):May.