Context-aware computing of actionable information using electronic health records data

A thesis submitted to the University of Manchester for the degree of

Doctor of Philosophy

In the Faculty of Biology, Medicine and Health

2016

Paolo Fraccaro

School of Health Sciences

Centre for Health Informatics

Table of Contents

List of Tables

List of Figures

List of Supplementary Tables

List of Supplementary Figures

List of Abbreviations

ACR: Albumin-creatinine ratio

AIC: Akaike Information Criterion

AIDS: Acquired immunodeficiency syndrome

AKI: acute kidney injury

AMD: Age-related macular degeneration

AoI: Area of Interest

AUC: area under receiving operating characteristic curve

BMI: body mass index

CCI: Charlson comorbidity index

CDS: Clinical Decision Support

CI: Confidence interval

CKD: Chronic Kidney Disease

CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration.

CSUQ: Computer System Usability Questionnaire

CVD: cardiovascular disease

eGFR: estimated Glomerular Filtration Rate

EHR: Electronic Health Record

GP: General Practitioner

HIV: Human immunodeficiency virus

IDI: Integrated Discrimination Improvement

ISO: International Standards Organization

MAPE: mean absolute prediction error

MDRD: Modification of Diet in Renal Disease

NICE: National Institute for Health and Care Excellence

NRI: Net Reclassification Improvement

NSAIDs: nonsteroidal anti-inflammatory drugs

QOF: Quality and Outcome Framework

QUIS: Questionnaire for User Interface Satisfaction

RI: Reference Interval

SD: Standard deviation

SIR: Salford Integrated Record

SUS: System Usability Scale

VIF: Variance Inflation Factors

ABSTRACT

The vast amount of routinely collected information in Electronic Health Records (EHRs) is increasingly used for other purposes than direct care – and in particular, for research. Although research would not aim to improve the care for any specific patient, it can produce generalisable knowledge that can be translated into action within specific healthcare contexts. The main aim of this PhD thesis was computing actionable information from EHR data for specified healthcare contexts, including health research itself, population health management, and health information technology (HIT) engineering. We focused on two main research areas: predictive modelling using EHR data and patient portals.

First, we explored how to use more effectively longitudinal EHR data to investigate multimorbidity. In a 10-year retrospective cohort study in the UK primary care, we tested different longitudinal comorbidity metrics and their value in predicting mortality. We found that explicitly accounting for longitudinal changes in comorbidities, as measured with the Charlson Comorbidity Index (CCI), better captures comorbidity burden on mortality, with more rapid changes in CCI posing a greater mortality risk. We suggest that the survival model proposed in the study should be considered by health researchers when investigating multimorbidity in EHR data.

Second, we followed international guidelines to externally validate available models for predicting onset of CKD in the UK primary care (i.e. seven in total). We tested their performance on a five-year time horizon. All models had good discrimination on a five-year time horizon, however the majority over-predicted the CKD risk. QKidney, the only model originally developed in the UK, outperformed the other models and was shown to support a high risk approach to CKD prevention. This finding is actionable at a population health management level: on the basis of our results, policy makers should consider to update clinical practice guidelines by including QKidney among the CKD screening criteria.

Finally, we focused on providing actionable information for HIT engineers. Particularly, we carried out a controlled study assessing whether patient interpretation and decision-making is influenced by the way the laboratory test results are presented to them in patient portals. We did not find any statistically significant differences between the three presentations that we tested, but we did find that misinterpretation of risk was high across all three presentations. Furthermore, we developed a method to calculate dynamic, patient-tailored alerts. Our method underwent proof-of-concept testing using one type of laboratory test value (i.e. potassium) and a group of GPs. Although representing a substantial methodological advancement and promising results were obtained, further evaluation of this method is required before HIT engineers can implement it in EHR systems.

In this thesis, we used routinely collected EHR data to investigate decision-making in different contexts and involving different stakeholders. These included patients, clinicians and policy makers, as well as HIT engineers and researchers. Ultimately, we produced actionable information across health research and population health management, with methodological advances in predictive modelling using EHR data and findings from evaluation studies that are relevant to policy makers.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Acknowledgements

First of all I would like to thank the NIHR Greater Manchester Primary Care Patient Safety Translational Research Centre and the Health eResearch Centre, in the persons of Prof. Stephen Campbell and Prof. Iain Buchan, for believing in me and giving me the opportunity of this PhD project.

Second, I want to thank all my supervisors (Prof. Iain Buchan, Dr. Niels Peek, Dr. Sabine Van der Veer, Dr. Mattia Prosperi and Prof. John Ainsworth) for the support provided during my PhD studies. At different times and in different ways, each one of them has contributed in making me succeed in this PhD project and grow as a scientist.

Third, I thank my friends Lora, Jonny and Shoba, with whom I shared this experience at the University of Manchester.

Fourth, I want to thank my parents, Astor and Daniela, and my family, who, despite being thousands of kilometres far from Manchester, have always been there for me when I needed. On the same lines, I would like to thank my friends Stefano and Federica, also there for me since I have left from Italy.

For last, but not least, I want to thank my fiancé Stefania, with whom I have embarked on this new life in the UK. I could not have made it without her support and advice.

Thanks to you all!

Paolo

Chapter 1

INTRODUCTION

Clinical information systems are now widespread in many western countries [1–3], with electronic health records (EHRs) in primary care adopted by the majority of general practitioners (GPs) in high income countries [3]. Although initially introduced for billing reasons, the current use of clinical information systems reaches far beyond administrative purposes, with health reforms making EHRs an essential part of care delivery and service evaluation [4–6].

This widespread adoption of EHRs has made a vast amount of clinical information available in electronic format [7]. Depending on the purpose for which the information is utilised, the use of such electronic data can be seen as *primary* or *secondary*, the definitions of which vary across the literature [8–11]. In this thesis, we adopted the definitions by the American Medical Informatics Association (AMIA) [10], who defined primary use as: "the use of personal health information by the organization or entity that produced or acquired these data in the process of providing real-time, direct care of an individual."; and secondary use, which is the main focus of this thesis, as "non-direct care use of personal health information including but not limited to analysis, research, quality/safety measurement, public health, payment, provider certification or accreditation, and marketing and other business including strictly commercial activities.".

Primary use happens in the clinical practice context, when data from a single patient is used for clinical decision making. While this is usually done by clinicians, patients are more and more important in primary use since they can increasingly access portions of their EHR with systems called patient portals [12]. Although overall uptake of patient portals is still low [12], this is seen as a key way of having patients involved in their care [54–59]. Primary use involves deduction, for example knowing from Mr. Socrates' EHR data that he was born 80 years ago, we can use generic, prior information (i.e. knowledge) to deduce that he is old. Conversely, especially in research, secondary use of EHR data often concerns data from multiple patients (e.g. a cohort) utilised with the aim of producing new knowledge. This is obtained by using induction (i.e. generalisation) as the main reasoning method. For example, a research study on a large cohort of United Kingdom (UK) primary care patients might find that older patients who have been prescribed non-steroidal anti-inflammatory drugs (NSAIDs, often prescribed

to treat arthritis) without gastro-protective medication have a higher incidence of gastrointestinal bleeding.

Knowing that Mr. Socrates is old does not help us to take any action on its own. Conversely, the piece of generic information coming from the research project in the above example can be used in decision making when applied to a specific patient. For example, if we prescribe NSAIDs to Mr. Socrates because he has joint pain, we should also prescribe gastro-protective medication to avoid gastrointestinal bleeding. This is an example of actionable clinical information, which we will define as "information that can automatically prompt the best decisions about care at the point in time when clinical decisions need to be made" (following [13]). However, actionability of information is not limited to clinical practice, but can also apply to other contexts where generic information coming from secondary use of EHR data can be used to guide decision-making. These contexts include technology development and engineering, population health management and research itself (see Table 1.1). In general, each of the lower levels in Table 1.1 aims at providing actionable information to the levels above. The closer levels are to the clinical practice level, the more direct impact the actionable information from that level can have on patient care.

Table 1.1: Hierarchical structure representing the different contexts in which information produced through secondary use of EHR data can be actionable.

Level	Context	Stakeholders	Decision/action types
1	Clinical practice	Healthcare professionals and patients	Clinical diagnosis and treatment; life style choices
2	Technology development and engineering	EHR developers	EHR design decisions; development of computerised decision support; user interaction design
3	Population health management	Policy makers, commissioners, guideline developers	Commissioning services; developing guidelines
4	Research	Health informaticians, health data scientists, epidemiologists	Choice of appropriate study methods; choice of topic to investigate

In the above example of Mr Socrates, we have seen how generic information produced as part of academic research was actionable in clinical practice. However, the same piece of information is also actionable in the other contexts. For instance, the findings from the research study could prompt EHR providers to include a specific alert suggesting clinicians to

prescribe gastro-protective medications for older patients on NSAIDs in EHR systems. Although the alert was the result of a research study (i.e. secondary use), once implemented in EHR systems it will start playing a role in the primary use of EHR data since it uses information to make clinical decisions for specific patients. This is also true for other forms of computerised decision support (e.g. prediction models and computerised clinical practice guidelines). Furthermore, in the engineering context the secondary use of EHR data is not limited to the analysis of the information, but it also comprises exploring optimal visualisation methods to improve user interaction with EHR systems and patient portals. Going back to our example, the results from the study are also actionable in the context of population health management, with policy makers making a decision of updating clinical practice guidelines accordingly. Finally, other researchers could consider the methodology adopted in the abovementioned research study when carrying out future research.

This thesis explores how to extract actionable information from routinely collected data in EHRs for the stakeholders in Table 1.1. This chapter first introduces the two research areas we focused on, such as predictive modelling using EHR data and patient portals. We concentrate on the primary care in the UK, which is the main domain in which this project has been carried out. In addition, we present an overview of multimorbidity and chronic kidney disease (CKD), which are the main clinical conditions considered in the research. Finally we describe the research questions and present an outline of the thesis.

1.1 USING EHR DATA FOR RESEARCH

1.1.1 EHRs in UK Primary Care

The UK is one of the early adopters of EHRs [14–16], and, despite major failures in some of its health IT programs [17,18], it is leading worldwide on the development of repositories for secondary use of anonymised, routinely collected EHR data for research purposes [19]. These include established databases like the Clinical Research Practice Datalink [20], which was established in 1987 and contains 10% of the UK population [20], and emerging ones like the Salford Integrated Record (SIR), which contains all data from primary care and a selection of secondary care data for all patients in the city of Salford (Greater Manchester, UK) [21]. SIR is the main data source that was used in this thesis.

The UK's world-leading position in this area was favoured by two main conditions [16]: the presence of the National Institute for Health and Care Excellence (NICE) [22], which standardises healthcare procedures in the context of a single National Health Service (NHS); and the incentive scheme that within the NHS brought almost all clinical information systems

in primary care to be compliant to specific standard specifications [23,24]. In addition, the introduction of the Quality and Outcomes Framework (QOF) incentive scheme [6] contributed to an even more central role of data in EHRs in delivering care. QOF is a scheme, set up in 2004, which remunerates GPs based on whether they obtain a set of quality indicators for 19 chronic conditions, and this is established from the data that they record into EHRs [6].

1.1.2 Differences between traditional studies and EHR studies

To fully appreciate the challenges and opportunities that routinely collected EHR data present when analysed for research purposes, we need first to consider the differences with the data from traditional studies (i.e. studies that implement primary data collection) [25]. Table 1.2 (modified from Casey et al. [25]) reports the main strengths and weaknesses of the two approaches.

Table 1.2: Strengths and weaknesses of traditional versus EHR studies (Modified from Casey et al. [25]).

Study feature	Traditional study	EHR study
Original purpose of data collection	Research; requires bespoke data collection.	Clinical care; research relies on secondary use of data that were collected for care purposes.
Cost	Expensive; requires nursing staff to collect data.	Less expensive, as data are collected during routine care.
Common study design	Randomised controlled trial; prospective cohort study; case-control study; cross-sectional study.	Retrospective or prospective cohort study; case-control study; cross-sectional study.
Time frame	Follow-up restricted by funding; duration influenced by health outcomes occurrence in prospective studies.	Retrospective data available only since EHR implementation; low cost for additional years of follow-up.
Study population	Recruitment-based; fewer participants than EHR.	Based on patient use of a specific health system, therefore problems with representativeness; many participants potentially available.

Study feature	Traditional study	EHR study
Follow-up	Scheduled and with fixed, predefined times between visits.	Variable follow-up times, driven by health condition; more patients lost to follow-up.
Data collection and storage	Established protocol; standardised and robust approach to data collection; possibility of storing biological samples for future analysis.	Data collected at different levels of detail, based on provider practices; data stored: clinical diagnoses, laboratory test results, medication prescriptions, problem lists, and clinical narrative; biological samples rarely banked.
Conditions captured	Outcomes and severities are specified in the research study protocol at the beginning of the study.	Only outcomes that require care or are necessary to deliver care are available; presence of missing data for those outcomes that patients experienced but for which they did not go to the doctor.
Outcome ascertainment	Pre-specified variables.	Entire records are available; availability confounded by disease severity, socio-economic status and other factors.
Covariate ascertainment	Pre-specified variables; might be affected by recall bias (i.e. patients recalling events before being recruited)	Often missing data on social and behavioural domains.
External validity	Low: Difficult to obtain representative samples because patients must agree to participate and this might be influence by incentives or healthy volunteer effects.	Although not representative of the general population, high external validity brought by the real-world nature of the data (i.e. collected during routine care).

1.1.3 Challenges presented by EHR studies

From Table 1.2 we can identify three main challenges that EHR studies present in comparison to traditional studies.

First, research based on EHR data is observational in nature. Therefore, EHR data cannot be used to evaluate the intended effect of a treatment or a procedure, for which randomised controlled trials are the best approach [26]. Subject to assumptions (i.e. no unmeasured

confounding), EHR data can only be used to discover new associations and find explanations to phenomena [26].

Second, in contrast to traditional studies, the lack of standardised procedures for data collection has an impact on data quality and organisation [27–29]. Information is recorded during clinical care and GPs have the possibility of using two formats [14–16]. On the one hand, GPs can record information in an unstructured format, such as narrative or free-text entries. In this case, Natural Language Processing techniques (NLP) may be used to extract meaning from GPs' notes [30,31]. The main challenges in this area are represented by the human-like knowledge and understanding of context that machines need in order to resolve ambiguities in human language and infer the correct semantic [32]. This is even more difficult in the medical field, which is prone to using jargon, synonyms and abbreviations. Main advances in NLP have been observed in recent years [32], however, like for SIR, free-text is often omitted from research databases, as personal information might be still present even after de-identification efforts [25]. On the other hand, GPs can record information in a structured format by using a coding system, Read codes in the UK primary care [33]. This associates a unique code to every piece of information that GPs might need to record (i.e. diagnoses, symptoms, laboratory test results or medication prescriptions). In this instance, researchers develop lists of relevant codes that are used to ascertain outcomes and covariates from the EHR data [16]. Although UK primary care has adopted a single coding system for almost two decades, concerns were raised about the presence of different versions of the Read codes (i.e. 4-byte versus 5-byte and version 2 versus version 3) [14] and inconsistent data recording by GPs [16]. However, this was ameliorated by the introduction of QOF, which has been proven to improve data quality and completeness for patients living with the conditions within the incentives scheme [34–37].

Finally, generalizability of findings from EHR data is restricted to patients who had contact with the NHS (i.e. due to illness, monitoring visits or need of medical advice), rather than the general population [38]. For example, in the UK primary care databases we would have rich information about patients living with conditions in QOF, while having a lot of missing data on healthy people who do not interact often with the NHS. Although the issue of missing data has been treated extensively in the literature [39], little has been said in the literature about the "informative presence" of certain data in EHRs [40].

1.1.4 Opportunities presented by EHR studies

Despite the abovementioned challenges, EHR studies have advantages over traditional studies in terms of costs, data quantity, richness and heterogeneity. These translate into several opportunities for EHR data to be used for research [20,25,29,30,40,41].

First, due to the relatively low cost of aggregating large quantities of data together, once the infrastructure is in place, EHRs allow easy creation of research cohorts [25]. These can be specific to a particular disease, clinical context (i.e. as already said the UK is world leader in primary care research databases [19]) or integrate data from different domains. This permits the reuse of data for multiple projects and to validate findings from smaller studies.

Second, the heterogeneity and richness of EHR data provide ample opportunities to look at multiple outcomes, risk factors and particular subgroups [25]. This is particular important because EHR data allow more information to be gathered than traditional studies about rare outcomes, and facilitate the investigation of stigmatised conditions (i.e. HIV or mental health), for which recruitment and follow-up can be challenging [25]. Furthermore, EHR data, with their real-world nature [41], present unique opportunities to improve our medical knowledge on subgroups of patients who are poorly characterised in the literature, as often excluded by randomised clinical trials [42]. This is particularly relevant for multimorbid patients (i.e. patients living with multiple concurring conditions, see paragraph 1.5 for details), who are increasing in number [43,44] and for whom evidence-based clinical practice is lacking [45].

Finally, due to their size and heterogeneity EHR data have been extensively used in the last years to develop predictive models (i.e. tools to identify in advance patients with increased risk of mortality, developing a disease or experience an adverse event) [40], with some of them (e.g. QRisk [46,47] in the UK primary care) that have been already introduced in clinical practice.

1.1.5 Current challenges in predictive modelling using EHR data

In this thesis, among the aforementioned opportunities, we focused on developing predictive models using EHR data. As shown by Goldstein *et al.* [40] in their recent systematic review, this area of research, despite the increasing number of studies published in the literature, presents several examples where EHR data advantages are currently under-exploited.

First, although EHRs provide vast longitudinal information about patients, current EHR-based predictive models often involve a limited number of predictors and studies do not take advantage of the longitudinal nature of the data [40].

Second, although some sort of validation of the predictive models is carried out (i.e. cross validation or bootstrap), authors of current predictive models do not take advantage of EHR large data to perform external validation studies (i.e. testing performance of a predictive model outside the context the predictive model was developed in) [40,48–50], and when they do this is often not done properly [49]. External validation studies are essential to evaluate how

generalizable a model is and to what extent it could be used in clinical practice [40,48–50]. Furthermore, if previously published predictive models for the outcome under study are available, their performance should be tested in the clinical domain of interest instead of developing a new model. This is rarely done in the literature [49].

Third, EHR data not only provide the opportunity to perform external validation studies on large number of patients, they also allow performance to be calculated in relevant subgroups (e.g. male/female, people with particular risk factors) for whom a predictive model might be particularly helpful [40].

Finally, authors of current EHR-based predictive models focus their attention only on a limited number of techniques to develop the models and evaluating its performance. Particularly, logistic regression is mainly used to develop predictive models in the literature [40]. Although logistic regression is easy to implement and interpret (i.e. it is included in the group of the *white-boxes* methodologies), more effective, but less interpretable, techniques exist (called *black-boxes*) [51]. These are infrequently used in the literature [40,52]. Furthermore, although it is difficult to interpret and alternatives are available [53], the c-statistic is still the only reported metric in most predictive model studies [40].

1.2 PATIENT PORTALS

Patient access to their own EHR data has been identified as a key route to engage and activate patients in their care [54–59]. The informatics systems that patients use to access their records are called *patient portals* [12]. They also provide patients with access to basic activities such as booking appointments or communicating with physicians. Most platforms that allow patients access to their records come from the long-term conditions context [60,61]. Patients with long-term conditions undergo complex longitudinal follow-ups, and self-management is considered a key component of their care [62].

Despite the increasing availability of patient portals and reported high satisfaction among users [63–65], there are several challenges related to patient portals implementation.

First, overall uptake is still low [12], with many people stopping using patient portals shortly after their first login [66]. This potentially happens because with the introduction of patient portals the care business model did not change accordingly [67], therefore patients do not perceive any additional value to their current care [68]. Kaiser Permanente's patient portal [69], the most widely used privately owned patient portal in the world [70], showed how

patient-centeredness and making information actionable are the main components for a successful implementation [71].

Second, although patients mainly access their EHR data to check their laboratory test results [66,72–75], there are concerns on current patient portals' effectiveness in supporting patients in this task [60,76]. Several studies have shown that patients find it difficult to interpret and act on laboratory test results [77–80], however to date little is known on how to best support them.

Finally, the evidence about the impact of patient portals on health outcomes and care processes is inconsistent [67,81,82]. Several studies failed to show benefits from using patient portals [67,81,82], however the reasons for this lack of impact are still poorly understood.

The NHS is aiming to provide online access to primary care EHR data to all patients by 2018 [83]. This comes after a major failure like *HealthSpace* [18], which was a similar initiative suspended in 2010 for lack of uptake and impact on NHS costs and patient engagement [84]. In order to avoid further disappointments, it is particularly important to investigate how to overcome the abovementioned challenges.

1.3 MULTIMORBIDITY

Multimorbidity is defined as "any combination of chronic disease with at least one other disease (acute or chronic) or bio-psychosocial factor (associated or not) or somatic risk factor" [85]. With an aging population, the prevalence of patients with multiple conditions increases [43,44]. Particularly, population-based studies revealed the presence of at least one long-term condition in over a third of patients [44,86], with two thirds of those aged over 65 years and three quarters of those aged over 85 years having at least two concurring conditions [87]. This places extra demand on healthcare systems [88,89], especially in primary care, which is has a pivotal role in multimorbid patients management [90,91].

The combination of diseases that are commonly reported to concur in multimorbidity are [92–94]: cardiovascular diseases, diabetes mellitus, chronic kidney disease, chronic musculoskeletal disorders, chronic lung disorders, and mental health disorders (particularly dementia and depression). This interaction of conditions generates: (1) duplication of tests, (2) obstacles in the continuity of care, (3) confusing self-management information, and (4) medication errors. As they are often excluded from clinical trials [42], another major issue faced in multimorbidity is the lack of evidence-based clinical practice [42].

EHR data offer new information about multimorbid patients [95], with multimorbidity that is often measured with summary scores in research studies based on EHR data [96]. These summary scores range from a simple summation of the number of conditions a patient has to more complex scores that accounts for the prognostic impact of the different pathologies [97–101]. Although EHR systems can provide rich longitudinal information, a big limitation of most studies in the literature is that they do not exploit this data [102–104]. Particularly, they measure comorbidities at a single time point, disregarding the dynamic nature of the evolution of the comorbidity burden over time [102–104]. This seems counter intuitive as it is reasonable to hypothesise that those with rising comorbidity over time may have worse health outcomes [103].

In this thesis, we considered the Charlson Comorbidity Index (CCI) [101], which is one of the most widely used examples of comorbidity summary scores [96]. The CCI has different weights for several clinical conditions (i.e. myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, connective tissue disease, ulcer disease, liver disease [mild, moderate and severe], diabetes [without and with complications], hemiplegia, chronic renal disease, cancer, metastatic tumor, leukemia, lymphoma, and acquired immune deficiency syndrome) in relation to their impact on prognosis. Although originally developed to predict mortality risk after hospitalisation, the CCI has been shown to independently predict adverse outcomes across a broad spectrum of conditions [105–115].

1.4 CHRONIC KIDNEY DISEASE

Chronic Kidney Disease (CKD) is a long-term condition characterized by a gradual, irreversible loss of kidney function over a period of months or years.[116]. CKD has a substantial burden of disease worldwide [117–120], with kidney disease related mortality being the 9th leading cause of death in the United States [121]. An increasing number of people are diagnosed with CKD [122,123]. In the UK, a study of 2.8M UK adults in 2010 reported a 5.9% prevalence of stage 3-5 CKD [124], with costs related to CKD care in 2009-2010 estimated around £1.45 billion (1.3% of the National Health Service (NHS) budget) [125]. These costs are set to rise steeply [123,125].

The main determinants of kidney function are ethnicity, age and gender, with CKD mainly caused by conditions including hypertension, diabetes, nephritis, and use of nephrotoxic medications [126]. CKD causes abnormalities in the anatomical structure or function of kidneys [116]. Kidney function is measured by the glomerular filtration rate (GFR), which

describes the flow rate of filtered fluids through the kidneys [127]. GFR, in combination with albumin-creatinine ratio (ACR), is used to diagnose and stage CKD [116] (see Table 1.3 and Table 1.4). ACR aims at identifying proteinuria (i.e. large quantity of proteins in urine), which is a key sign of CKD [116]. Alternatives to ACR are also protein-creatinine ratio and 24h proteinuria [116].

Table 1.3: GFR categories in CKD [116].

GFR (ml/min per 1.73 m^2)	GFR category	Terms
≥ 90	G1	Normal or high
60-89	G2	Mildly decreased
45-59	G3a	Mildly to moderately decreased
30-44	G3b	Moderately to severely decreased
15-29	G4	Severely decreased
<15	G5	Kidney failure

Abbreviations: CKD: Chronic Kidney Disease; GFR: glomerular filtration rate.

Table 1.4: ACR categories in CKD [116].

ACR (mg/mmol)	ACR category	Terms
< 30	A1	Normal or mildly increased
30-300	A2	Mildly increased
>300	A3	Severely increased

Abbreviations: CKD: Chronic Kidney Disease; ACR: albumin-creatinine ratio.

In clinical practice, an estimated version of the GFR is used (called estimated GFR [eGFR]). This is based on serum creatinine, in combination with age, gender and ethnicity. There are different formulas to calculate eGFR, with the main ones that are the Modification of Diet in Renal Disease (MDRD) [128] and Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) [129] formulas. The latter represents an improvement on the MDRD formula but is not commonly used in clinical practice.

Despite efforts worldwide to improve early detection [130], CKD often remains undiagnosed in its early stages [122]. Although difficult to interpret and requiring a blood sample, most of current CKD clinical surveillance relies on eGFR [130]. Prediction models to identify patients at high risk of developing CKD can extend this clinical screening toolkit. However, only a minority of the numerous models available in the literature to predict CKD onset were externally validated [131,132].

Once diagnosed with CKD, as for many chronic conditions, patients undergo regular blood tests. This can range from yearly tests in the early stages to quarterly tests for stable patients

in the severe stages. For CKD patients in the UK there is patient portal called PatientView [64] that patients who are referred to a nephrologist in secondary care can use to access their laboratory test results. Although the platform has been established in 2004, overall uptake is still low [66].

1.5 RESEARCH QUESTIONS

The main aim of this thesis is to explore how to extract actionable information from routinely collected data in EHRs in different contexts – i.e. for supporting researchers, policy makers and EHR developers. Our objectives were to investigate the following research questions in two main research areas (i.e. predictive modelling and patient portals):

1. *RQ1:* How can we use clinical longitudinal information in EHRs more effectively to investigate multimorbidity?

2. *RQ2:* How can we use EHR data to externally validate existing predictive models?

3. *RQ3:* How can we use predictive modelling and interface design to enhance the presentation of clinical laboratory test results in EHRs and patient portals?

Each research question focused on, but did not limit to, providing actionable information within a specific context in Table 1.1. Particularly, RQ1 aimed at providing actionable information to researcher, RQ2 to health population management and RQ3 to technology development and EHR providers.

1.6 STRUCTURE OF THE THESIS

The results of this thesis are organised in "alternative format" and presented as a series of articles for publication in peer-reviewed journals.

RQ1 is addressed in Chapters 2 and 3. Chapter 2 is a systematic review that introduces the reader to the informatics challenges presented by concurring clinical conditions and highlights the importance of research using EHR data in this area. Chapter 3 explores how to better investigate multimorbidity by considering patient longitudinal information when predicting mortality in EHR studies.

Chapter 4 addresses RQ2, and explores how EHR data can be used to externally validate and compare existing prediction models. The clinical problem that was considered was CKD onset.

RQ3 is addressed in Chapters 5, 6, and 7. Chapter 5 describes a systematic review of the literature about the influence of patient portals on decision making. Chapter 6 presents a study where we evaluated the effect of different interface design techniques on the interpretation of laboratory test results of CKD patients. In chapter 7, we investigated how to develop a prediction model to produce dynamic patient-tailored reference ranges for laboratory test results. The model was applied to potassium data, which is one of the most important parameters in CKD management, and tested with GPs.

The thesis concludes with Chapter 8 that includes the discussion of the significance of the presented work and future directions.

Four of the articles included in the thesis have already been published in peer-reviewed journals (Chapter 2, Chapter 3, Chapter 4, and Chapter 7); while Chapter 5 and Chapter 6 are currently under peer review.

1.7 REFERENCES

1. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim J-L, Reuterwall C, et al. External review and validation of the Swedish national inpatient register. BMC Public Health. 2011;11: 450. doi:10.1186/1471-2458-11-450

2. Kushniruk AW, Bates DW, Bainbridge M, Househ MS, Borycki EM. National efforts to improve health information system safety in Canada, the United States of America and England. Int J Med Inform. 2013;82: e149–e160. doi:http://dx.doi.org/10.1016/j.ijmedinf.2012.12.006

3. Schoen C, Osborn R, Squires D, Doty M, Rasmussen P, Pierson R, et al. A survey of primary care doctors in ten countries shows progress in use of health information technology, less in other areas. Health Aff (Millwood). 2012;31: 2805–16. doi:10.1377/hlthaff.2012.0884

4. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med. 2010;363. doi:10.1056/NEJMp1006114

5. The UK Department of Health. Information for health: an information strategy for the modern NHS 1998–2005 [Internet]. London, United Kingdom; 1998. Available: http://webarchive.nationalarchives.gov.uk/20130107105354/http:/www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4014469.pdf

6. Roland M. Linking physicians' pay to the quality of care--a major experiment in the United kingdom. N Engl J Med. 2004;351: 1448–54. doi:10.1056/NEJMhpr041294

7. Friedman CP, Wong AK, Blumenthal D. Achieving a Nationwide Learning Health System. Sci Transl Med. 2010;2: 57cm29 LP-57cm29. Available: http://stm.sciencemag.org/content/2/57/57cm29.abstract

8. Wilson P, Sundgren M, Singleton P. Primary and secondary use of EHR systems: Enhancing clinical research for better health and high quality healthcare [Internet]. Brussels; 2007. Available: https://ec.europa.eu/digital-single-market/en/news/enhancing-clinical-research-better-health-and-high-quality-healthcare

9. Ramanathan T, Schmit C, Menon A, Fox C. The Role of Law in Supporting Secondary Uses of Electronic Health Information. J Law Med Ethics. 2015;43: 48–51. doi:10.1111/jlme.12215

10. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. J Am Med Inform Assoc. American Medical Informatics Association; 2007;14: 1–9. doi:10.1197/jamia.M2273

11. van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. A systematic review of barriers to data sharing in public health. BMC Public Health. London: BioMed Central; 2014;14: 1144. doi:10.1186/1471-2458-14-1144

12. Irizarry T, DeVito Dabbs A, Curran RC. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res. 2015;17: e148. doi:10.2196/jmir.4255

13. Mitus AJ, Coughlin L. The Value of Actionable Content in a Clinical Setting: Access to Better Information Facilitates Enhanced Cancer Care. Am Heal Drug Benefits. Engage Healthcare Communications, LLC; 2013;6: 104–106. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4046459/

14. de Lusignan S, et al. Routinely-collected general practice data are complex, but with systematic processing can be used for quality improvement and research. Inform Prim Care. 2006;14: 59–66.

15. de Lusignan S, van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. Fam Pract. 2006;23: 253–263. Available: http://fampra.oxfordjournals.org/content/23/2/253.abstract

16. Olier I, Springate DA, Ashcroft DM, Doran T, Reeves D, Planner C, et al. Modelling Conditions and Health Care Processes in Electronic Health Records: An Application to Severe Mental Illness with the Clinical Practice Research Datalink. PLoS One. Public Library of Science; 2016;11: e0146715. Available: http://dx.doi.org/10.1371%2Fjournal.pone.0146715

17. O'Dowd A. GP pilots are "last chance saloon" for care.data scheme. BMJ Br Med J. 2015;350. Available: http://www.bmj.com/content/350/bmj.h419.abstract

18. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. BMJ. 2010;341. Available: http://www.bmj.com/content/341/bmj.c5814.abstract

19. Shephard E, Stapley S, Hamilton W. The use of electronic databases in primary care research. Fam Pract. 2011;28: 352–354. Available: http://fampra.oxfordjournals.org/content/28/4/352.short

20. Kousoulis AA, Rafi I, de Lusignan S. The CPRD and the RCGP: building on research success by enhancing benefits for patients and practices. Br J Gen Pract. Royal College of General Practitioners; 2015;65: 54–55. doi:10.3399/bjgp15X683353

21. New JP, Bakerly ND, Leather D, Woodcock A. Obtaining real-world evidence: the Salford Lung Study. Thorax. 2014; Available: http://thorax.bmj.com/content/early/2014/03/06/thoraxjnl-2014-205259.abstract

22. The UK Department of Health. National Institute for Health and Care Excellence [Internet]. [cited 24 Oct 2016]. Available: https://www.nice.org.uk/

23. Payne TH, Detmer DE, Wyatt JC, Buchan IE. National-scale clinical information exchange in the United Kingdom: lessons for the United States. J Am Med Inform Assoc. BMA House, Tavistock Square, London, WC1H 9JR: BMJ Group; 2011;18: 91–98. doi:10.1136/jamia.2010.005611

24. NHS. A Guide to GP Systems of Choice [Internet]. Available: https://digital.nhs.uk/GP-Systems-of-Choice

25. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using Electronic Health Records for Population Health Research: A Review of Methods and Applications. Annu Rev Public Health. Annual Reviews; 2016;37: 61–81. doi:10.1146/annurev-publhealth-032315-021353

26. Vandenbroucke JP. Observational Research, Randomised Trials, and Two Views of Medical Science. PLoS Med. San Francisco, USA: Public Library of Science; 2008;5: e67. doi:10.1371/journal.pmed.0050067

27. Ancker JS, Shih S, Singh MP, Snyder A, Edwards A, Kaushal R, et al. Root Causes Underlying Challenges to Secondary Use of Data. AMIA Annu Symp Proc. American Medical Informatics Association; 2011;2011: 57–62. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3243274/

28. van Velthoven MH, Mastellos N, Majeed A, O'Donoghue J, Car J. Feasibility of extracting data from electronic medical records for research: an international comparative study. BMC Med Inform Decis Mak. London: BioMed Central; 2016;16: 90. doi:10.1186/s12911-016-0332-1

29. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. Summit on Translat Bioinforma. American Medical Informatics Association; 2010;2010: 1–5. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041534/

30. Karystianis G, Sheppard T, Dixon WG, Nenadic G. Modelling and extraction of variability in free-text medication prescriptions from an anonymised primary care electronic medical record research database. BMC Med Inform Decis Mak. 2016;16: 18. doi:10.1186/s12911-016-0255-x

31. MacRae J, Darlow B, McBain L, Jones O, Stubbe M, Turner N, et al. Accessing primary care Big Data: the development of a software algorithm to explore the rich content of consultation records. BMJ Open. 2015;5. Available: http://bmjopen.bmj.com/content/5/8/e008160.abstract

32. Hirschberg J, Manning CD. Advances in natural language processing. Science (80-). American Association for the Advancement of Science; 2015;349: 261–266. doi:10.1126/science.aaa8685

33. NHS England. Read Codes [Internet]. [cited 16 Jun 2014]. Available: http://www.connectingforhealth.nhs.uk/systemsandservices/data/uktc/readcodes

34. Kontopantelis E, Olier I, Planner C, Reeves D, Ashcroft DM, Gask L, et al. Primary care consultation rates among people with and without severe mental illness: a UK cohort study using the Clinical Practice Research Datalink. BMJ Open. 2015;5. Available: http://bmjopen.bmj.com/content/5/12/e008650.abstract

35. Reilly S, Olier I, Planner C, Doran T, Reeves D, Ashcroft DM, et al. Inequalities in physical comorbidity: a longitudinal comparative cohort study of people with severe mental illness in the UK. BMJ Open. 2015;5. Available: http://bmjopen.bmj.com/content/5/12/e009010.abstract

36. Taggar JS, Coleman T, Lewis S, Szatkowski L. The impact of the Quality and Outcomes Framework (QOF) on the recording of smoking targets in primary care medical records: cross-sectional analyses from The Health Improvement Network (THIN) database. BMC Public Health. BioMed Central; 2012;12: 329. doi:10.1186/1471-2458-12-329

37. Sutton M, Elder R, Guthrie B, Watt G. Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers. Health Econ. John Wiley & Sons, Ltd.; 2010;19: 1–13. doi:10.1002/hec.1440

38. Rusanov A, Weiskopf NG, Wang S, Weng C. Hidden in plain sight: bias towards sick patients when sampling patients with sufficient electronic health record data for research. BMC Med Inform Decis Mak. 2014;14: 51. doi:10.1186/1472-6947-14-51

39. Wells BJ, Chagin KM, Nowacki AS, Kattan MW. Strategies for Handling Missing

Data in Electronic Health Record Derived Data. eGEMs. AcademyHealth; 2013;1: 1035. doi:10.13063/2327-9214.1035

40. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Informatics Assoc. 2016; Available: http://jamia.oxfordjournals.org/content/early/2016/05/17/jamia.ocw042.abstract

41. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PRO, Bernstam E V, et al. Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research. Med Care. 2013;51: S30–S37. doi:10.1097/MLR.0b013e31829b1dbd

42. Van Spall HGC, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. JAMA. American Medical Association; 2007;297: 1233–40. doi:10.1001/jama.297.11.1233

43. Uijen AA, van de Lisdonk EH. Multimorbidity in primary care: prevalence and trend over the last 20 years. Eur J Gen Pract. 2008;14 Suppl 1: 28–32. doi:10.1080/13814780802436093

44. Koné Pefoyo AJ, Bronskill SE, Gruneir A, Calzavara A, Thavorn K, Petrosyan Y, et al. The increasing burden and complexity of multimorbidity. BMC Public Health. London: BioMed Central; 2015;15: 415. doi:10.1186/s12889-015-1733-2

45. CM B, Darer J, Boult C, LP F, Boult L, AW W. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: Implications for pay for performance. JAMA. 2005;294: 716–724. Available: http://dx.doi.org/10.1001/jama.294.6.716

46. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. BMJ. 2008;336: 1475–1482. doi:10.1136/bmj.39609.449676.25

47. Hippisley-Cox J, Coupland C, Brindle P. The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study. BMJ Open. 2014;4. Available: http://bmjopen.bmj.com/content/4/8/e005809.abstract

48. Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk

prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol. 2015;68: 25–34. doi:http://dx.doi.org/10.1016/j.jclinepi.2014.09.007

49. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol. 2014;14: 40. doi:10.1186/1471-2288-14-40

50. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016;353. Available: http://www.bmj.com/content/353/bmj.i3140.abstract

51. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction [Internet]. Springer; 1st ed. 2001. Corr. 3rd printing edition; 2008. Available: http://www.amazon.com/The-Elements-Statistical-Learning-Prediction/dp/0387952845

52. Baesens B, Martens D, Setiono R, Zurada JM. Guest Editorial White Box Nonlinear Prediction Models. IEEE Trans Neural Networks. 2011;22: 2406–2408. doi:10.1109/TNN.2011.2177735

53. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and ElaborationThe TRIPOD Statement: Explanation and Elaboration. Ann Intern Med. 2015;162: W1–W73. doi:10.7326/M14-0698

54. Pearce C, Bainbridge M. A personally controlled electronic health record for Australia. J Am Med Inform Assoc. 2014;21: 707–13. doi:10.1136/amiajnl-2013-002068

55. Institute of Medicine (USA). Achieve Meaningful use - Stage 2 [Internet]. Available: http://www.webcitation.org/6qBAzdht7

56. Urowitz S, Wiljer D, Apatu E, Eysenbach G, Delenardo C, Harth T, et al. Is Canada ready for patient accessible electronic health records? A national scan. BMC Med Inform Decis Mak. 2008;8: 33. doi:10.1186/1472-6947-8-33

57. Pagliari C, Shand T, Fisher B. Embedding online patient record access in UK primary care: a survey of stakeholder experiences. JRSM Short Rep. Royal Society of Medicine Press; 2012;3: 34. doi:10.1258/shorts.2012.012009

58. Fisher B. Patients' access to their electronic record: offer patients access as soon as you can. Br J Gen Pract. British Journal of General Practice; 2013;63: e423-5. doi:10.3399/bjgp13X668384

59. Adler-Milstein J, Sarma N, Woskie LR, Jha AK. A Comparison Of How Four Countries Use Health IT To Support Care For People With Chronic Conditions. Health Aff (Millwood). 2014;33: 1559–66. doi:10.1377/hlthaff.2014.0424

60. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and Provider Attitudes Toward the Use of Patient Portals for the Management of Chronic Disease: A Systematic Review. Brixey J, editor. J Med Internet Res. Toronto, Canada: JMIR Publications Inc.; 2015;17: e40. doi:10.2196/jmir.3703

61. Price M, Bellwood P, Kitson N, Davies I, Weber J, Lau F. Conditions potentially sensitive to a Personal Health Record (PHR) intervention, a systematic review. BMC Med Inform Decis Mak. 2015;15: 1–12. doi:10.1186/s12911-015-0159-1

62. Bayliss EA, Ellis JL, Steiner JF. Barriers to self-management and quality-of-life outcomes in seniors with multimorbidities. Ann Fam Med. 2007;5: 395–402. doi:10.1370/afm.722

63. Wiljer D, Leonard KJ, Urowitz S, Apatu E, Massey C, Quartey NK, et al. The anxious wait: assessing the impact of patient accessible EHRs for breast cancer patients. BMC Med Inform Decis Mak. 2010;10: 46. doi:10.1186/1472-6947-10-46

64. Bartlett C, Simpson K, Turner AN. Patient access to complex chronic disease records on the Internet. BMC Med Inform Decis Mak. 2012;12: 87. doi:10.1186/1472-6947-12-87

65. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med. American College of Physicians; 2013;159: 677–87. doi:10.7326/0003-4819-159-10-201311190-00006

66. Phelps GR, Taylor J, Simpson K, Samuel J, Turner NA. Patients' Continuing Use of an Online Health Record: A Quantitative Evaluation of 14,000 Patient Years of Access Data. J Med Internet Res. 2014;16: e241. doi:10.2196/jmir.3371

67. de Lusignan S, Mold F, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic interpretative review. BMJ Open. 2014;4: e006021. doi:10.1136/bmjopen-2014-006021

68. Hazara AM, Bhandari S. Barriers to patient participation in a self-management and education website Renal PatientView: A questionnaire-based study of inactive users. Int J Med Inform. 2016;87: 10–14. doi:http://dx.doi.org/10.1016/j.ijmedinf.2015.12.004

69. Silvestre A-L, Sue VM, Allen JY. If You Build It, Will They Come? The Kaiser Permanente Model Of Online Health Care. Health Aff. 2009;28: 334–344. doi:10.1377/hlthaff.28.2.334

70. Johnson LW, Garrido T, Christensen K, Handley M. Successful Practices in the Use of Secure E-mail. Perm J. 2014;18: 50–54.

71. Otte-Trojel T, Rundall TG, de Bont A, van de Klundert J, Reed ME. The organizational dynamics enabling patient portal impacts upon organizational performance and patient health: a qualitative study of Kaiser Permanente. BMC Health Serv Res. 2015;15: 1–12. doi:10.1186/s12913-015-1208-2

72. Cimino JJ, Patel VL, Kushniruk AW. The patient clinical information system (PatCIS): technical solutions for and experience with giving patients access to their electronic medical records. Int J Med Inform. 2002;68: 113–127. doi:10.1016/S1386-5056(02)00070-9

73. Shah SGS, Fittonb R, Hannanc A, Fisherd B, Younga T, Barnette J. Accessing personal medical records online: A means to what ends? Int J Med Inform. 2015;84: 111–8. doi:10.1016/j.ijmedinf.2014.10.005

74. Bhavnani V, Fisher B, Winfield M, Seed P. How patients use access to their electronic GP record--a quantitative study. Fam Pract. 2011;28: 188–94. doi:10.1093/fampra/cmq092

75. Young MJ, Scheinberg E, Bursztajn H. Direct-to-patient laboratory test reporting: balancing access with effective clinical communication. JAMA. American Medical Association; 2014;312: 127–8. doi:10.1001/jama.2014.5823

76. O'Kane M, Freedman D, Zikmund-Fisher BJ. Can patients use test results effectively if they have direct access? BMJ Br Med J. 2015;350. Available: http://www.bmj.com/content/350/bmj.h673.abstract

77. Taha J, Sharit J, Czaja SJ. Usability of an electronic personal health record (PHR) among a diverse group of adults. Proceedings of the Human Factors and Ergonomics Society. University of Miami, Miller School of Medicine, United States; 2014. pp. 619–623. doi:10.1177/1541931214581132

78. Brewer NT, Gilkey MB, Lillie SE, Hesse BW, Sheridan SL. Tables or bar graphs? Presenting test results in electronic medical records. Med Decis Making. United States; 2012;32: 545–553. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=22472914

79. Zikmund-Fisher BJ, Exe NL, Witteman HO. Numeracy and Literacy Independently Predict Patients' Ability to Identify Out-of-Range Test Results. J Med Internet Res. Journal of Medical Internet Research; 2014;16: e187. doi:10.2196/jmir.3241

80. Britto MT, Jimison HB, Munafo JK, Wissman J, Rogers ML, Hersh W. Usability testing finds problems for novice users of pediatric portals. J Am Med Inform Assoc. United States; 2009;16: 660–669. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=19567793

81. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res. Journal of Medical Internet Research; 2012;14: e162. doi:10.2196/jmir.2238

82. Davis Giardina T, Menon S, Parrish DE, Sittig DF, Singh H. Patient access to medical records and healthcare outcomes: a systematic review. J Am Med Inform Assoc. 2014;21: 737–41. doi:10.1136/amiajnl-2013-002239

83. Iacobucci G. Patients promised online access to their medical records by 2018. BMJ. 2015;351. Available: http://www.bmj.com/content/351/bmj.h4740.abstract

84. Wicks P. Abandoned HealthSpace. Sharing access might increase engagement. BMJ. England; 2010;341: c7209. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=21172980

85. Le Reste JY, Nabbe P, Manceau B, Lygidakis C, Doerr C, Lingner H, et al. The European General Practice Research Network presents a comprehensive definition of multimorbidity in family medicine and long term care, following a systematic review of relevant literature. J Am Med Dir Assoc. 2013;14: 319–25. doi:10.1016/j.jamda.2013.01.001

86. van Oostrom SH, Picavet HSJ, van Gelder BM, Lemmens LC, Hoeymans N, van Dijk CE, et al. Multimorbidity and comorbidity in the Dutch population – data from general practices. BMC Public Health. BioMed Central; 2012;12: 715. doi:10.1186/1471-

2458-12-715

87. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet. 2012;380: 37–43. doi:10.1016/S0140-6736(12)60240-2

88. Wallace E, Salisbury C, Guthrie B, Lewis C, Fahey T, Smith SM. Managing patients with multimorbidity in primary care. BMJ. 2015;350. Available: http://www.bmj.com/content/350/bmj.h176.abstract

89. Boyd CM, Fortin M. Future of multimorbidity research: How should understanding of multimorbidity inform health system design? Public Health Rev. Division of Geriatric Medicine and Gerontology, Department of Medicine, Johns Hopkins University School of Medicine, 5200 Eastern Avenue, Baltimore, MD 21224-2734, United States; 2011;33: 451–474. Available: http://www.webcitation.org/6NYkztx37

90. Salisbury C, Johnson L, Purdy S, Valderas JM, Montgomery AA. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. Br J Gen Pract. 2011;61: e12-21. doi:10.3399/bjgp11X548929

91. Haggerty JL. Ordering the chaos for patients with multimorbidity. BMJ. 2012;345: e5915. doi:10.1136/bmj.e5915

92. Zhang X, Decker FH, Luo H, Geiss LS, Pearson WS, Saaddine JB, et al. Trends in the prevalence and comorbidities of diabetes mellitus in nursing home residents in the United States: 1995-2004. J Am Geriatr Soc. 2010;58: 724–30. doi:10.1111/j.1532-5415.2010.02786.x

93. Stevens LA, Li S, Wang C, Huang C, Becker BN, Bomback AS, et al. Prevalence of CKD and comorbid illness in elderly patients in the United States: results from the Kidney Early Evaluation Program (KEEP). Am J Kidney Dis. 2010;55: S23-33. doi:10.1053/j.ajkd.2009.09.035

94. Yaffe K, Ackerson L, Kurella Tamura M, Le Blanc P, Kusek JW, Sehgal AR, et al. Chronic kidney disease and cognitive function in older adults: findings from the chronic renal insufficiency cohort cognitive study. J Am Geriatr Soc. 2010;58: 338–45. doi:10.1111/j.1532-5415.2009.02670.x

95. Fraccaro P, Arguello Casteleiro M, Ainsworth J, Buchan I. Adoption of Clinical Decision Support in Multimorbidity: A Systematic Review. JMIR Med Inf. 2015;3: e4. doi:10.2196/medinform.3503

96. Yurkovich M, Avina-Zubieta JA, Thomas J, Gorenchtein M, Lacaille D. A systematic review identifies valid comorbidity indices derived from administrative health data. J Clin Epidemiol. 2015;68: 3–14. doi:10.1016/j.jclinepi.2014.09.010

97. Clark DO, Korff M Von, Saunders K, Baluch WM, Simon GE. A Chronic Disease Score with Empirically Derived Weights. Med Care. Lippincott Williams & Wilkins; 1995;33: 783–795. doi:10.2307/3766468

98. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity Measures for Use with Administrative Data. Med Care. 1998;36. Available: http://journals.lww.com/lww-medicalcare/Fulltext/1998/01000/Comorbidity_Measures_for_Use_with_Administrative.4.aspx

99. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J-C, et al. Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. Med Care. 2005;43. Available: http://journals.lww.com/lww-medicalcare/Fulltext/2005/11000/Coding_Algorithms_for_Defining_Comorbidities_in.10.aspx

100. Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. J Clin Epidemiol. 1992;45: 197–203. doi:http://dx.doi.org/10.1016/0895-4356(92)90016-G

101. Charlson ME, Pompei P, Ales K, Mackenzie C. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. J Chronic Dis. 1987;40: 373–383. doi:10.1016/0021-9681(87)90171-8

102. Wang CY, Baldwin L-M, Saver BG, Dobie SA, Green PK, Cai Y, et al. The Contribution of Longitudinal Comorbidity Measurements to Survival Analysis. Med Care. 2009;47: 813–821. doi:10.1097/MLR.0b013e318197929c

103. Zeng C, Ellis JL, Steiner JF, Shoup JA, McQuillan DB, Bayliss EA. Assessment of Morbidity Over Time in Predicting Health Outcomes. Med Care. 2014;52. Available: http://journals.lww.com/lww-medicalcare/Fulltext/2014/03001/Assessment_of_Morbidity_Over_Time_in_Predicting.11.aspx

104. Strauss VY, Jones PW, Kadam UT, Jordan KP. Distinct trajectories of multimorbidity in primary care were identified using latent class growth analysis. J Clin Epidemiol. 2014;67: 1163–1171. doi:http://dx.doi.org/10.1016/j.jclinepi.2014.06.003

105. Reyes C, Estrada P, Nogués X, Orozco P, Cooper C, Díez-Pérez A, et al. The impact

of common co-morbidities (as measured using the Charlson index) on hip fracture risk in elderly men: a population-based cohort study. Osteoporos Int. Springer London; 2014;25: 1751–1758. doi:10.1007/s00198-014-2682-9

106. Huang Y, Gou R, Diao Y, Yin Q, Fan W, Liang Y, et al. Charlson comorbidity index helps predict the risk of mortality for patients with type 2 diabetic nephropathy . J Zhejiang Univ Sci B. Hangzhou: Zhejiang University Press; 2014;15: 58–66. doi:10.1631/jzus.B1300109

107. Ng ACC, Chow V, Yong ASC, Chung T, Kritharides L. Prognostic Impact of the Charlson Comorbidity Index on Mortality following Acute Pulmonary Embolism. Respiration. 2013;85: 408–416. Available: http://www.karger.com/DOI/10.1159/000342024

108. Wu C-C, Hsu T-W, Chang C-M, Yu C-H, Lee C-C. Age-Adjusted Charlson Comorbidity Index Scores as Predictor of Survival in Colorectal Cancer Patients Who Underwent Surgical Resection and Chemoradiation. Medicine (Baltimore). 2015;94. Available: http://journals.lww.com/md-journal/Fulltext/2015/01020/Age_Adjusted_Charlson_Comorbidity_Index_Scores_as.24.aspx

109. Dias-Santos D, Ferrone CR, Zheng H, Lillemoe KD, Fernández-del Castillo C. The Charlson age comorbidity index predicts early mortality after surgery for pancreatic cancer. Surgery. Elsevier; 2015;157: 881–887. doi:10.1016/j.surg.2014.12.006

110. Erickson SR, Cole E, Kline-Rogers E, Eagle KA. The Addition of the Charlson Comorbidity Index to the GRACE Risk Prediction Index Improves Prediction of Outcomes in Acute Coronary Syndrome. Popul Health Manag. Mary Ann Liebert, Inc., publishers; 2013;17: 54–59. doi:10.1089/pop.2012.0117

111. Lu KJ, Kearney LG, Ord M, Jones E, Burrell LM, Srivastava PM. Age adjusted Charlson Co-morbidity Index is an independent predictor of mortality over long-term follow-up in infective endocarditis. Int J Cardiol. Elsevier; 2015;168: 5243–5248. doi:10.1016/j.ijcard.2013.08.023

112. Rattanasompattikul M, Feroze U, Molnar MZ, Dukkipati R, Kovesdy CP, Nissenson AR, et al. Charlson comorbidity score is a strong predictor of mortality in hemodialysis patients. Int Urol Nephrol. 2012;44: 1813–1823. doi:10.1007/s11255-011-0085-9

113. Grossman R, Mukherjee D, Chang DC, Bennett R, Brem H, Olivi A, et al. Preoperative Charlson Comorbidity Score Predicts Postoperative Outcomes Among Older

Intracranial Meningioma Patients. World Neurosurg. Elsevier; 2015;75: 279–285. doi:10.1016/j.wneu.2010.09.003

114. Schmidt M, Jacobsen JB, Lash TL, Bøtker HE, Sørensen HT. 25 year trends in first time hospitalisation for acute myocardial infarction, subsequent short and long term mortality, and the prognostic impact of sex and comorbidity: a Danish nationwide cohort study. BMJ Br Med J. BMJ Publishing Group Ltd.; 2012;344: e356. doi:10.1136/bmj.e356

115. Daskivich TJ, Chamie K, Kwan L, Labo J, Dash A, Greenfield S, et al. Comorbidity and competing risks for mortality in men with prostate cancer. Cancer. Wiley Subscription Services, Inc., A Wiley Company; 2011;117: 4642–4650. doi:10.1002/cncr.26104

116. Levin A, Stevens PE. Summary of KDIGO 2012 CKD Guideline: behind the scenes, need for guidance, and a framework for moving forward. Kidney Int. International Society of Nephrology; 2014;85: 49–61. doi:10.1038/ki.2013.444

117. Meguid El Nahas A, Bello AK. Chronic kidney disease: the global challenge. Lancet. 2005;365: 331–40. doi:10.1016/S0140-6736(05)17789-7

118. Barsoum RS. Chronic Kidney Disease in the Developing World. N Engl J Med. Massachusetts Medical Society; 2006;354: 997–999. doi:10.1056/NEJMp058318

119. Schoolwerth AC, Engelgau MM, Hostetter TH, Rufo KH, Chianchiano D, McClellan WM, et al. Chronic kidney disease: a public health problem that needs a public health action plan. Prev Chronic Dis. 2006;3: A57. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1563984&tool=pmcentrez&rendertype=abstract

120. Mills KT, Xu Y, Zhang W, Bundy JD, Chen C-S, Kelly TN, et al. A systematic analysis of worldwide population-based data on the global burden of chronic kidney disease in 2010. Kidney Int. International Society of Nephrology; 2015;88: 950–957. Available: http://dx.doi.org/10.1038/ki.2015.230

121. Kenneth D, Kochanek MA, Sherry L, Murphy BS, Jiaquan X, Betzaida T-VMS, et al. National Vital Statistics Reports - Deaths: Final Data for 2014. 2016.

122. Coresh J, Selvin E, Stevens LA, Manzi J, Kusek JW, Eggers P, et al. Prevalence of chronic kidney disease in the United States. JAMA. American Medical Association; 2007;298: 2038–47. doi:10.1001/jama.298.17.2038

123. United States Renal Data System. USRDS 2013 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States [Internet]. Bethesda, MD: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; 2013. Available: http://www.usrds.org/atlas.aspx

124. Jameson K, Jick S, Hagberg KW, Ambegaonkar B, Giles A, O'Donoghue D. Prevalence and management of chronic kidney disease in primary care patients in the UK. Int J Clin Pract. 2014;68: 1110–21. doi:10.1111/ijcp.12454

125. Kerr M, Bray B, Medcalf J, O'Donoghue DJ, Matthews B. Estimating the financial cost of chronic kidney disease to the NHS in England. Nephrol Dial Transplant. 2012;27 Suppl 3: iii73-80. doi:10.1093/ndt/gfs269

126. Kazancioğlu R. Risk factors for chronic kidney disease: an update. Kidney Int Suppl. Nature Publishing Group; 2013;3: 368–371. doi:10.1038/kisup.2013.79

127. Wikipedia. Renal function [Internet]. Available: https://en.wikipedia.org/wiki/Renal_function#Glomerular_filtration_rate

128. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. Ann Intern Med. 1999;130: 461–70. Available: http://www.ncbi.nlm.nih.gov/pubmed/10075613

129. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI. A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009;150. doi:10.7326/0003-4819-150-9-200905050-00006

130. Radhakrishnan J, Remuzzi G, Saran R, Williams DE, Rios-Burrows N, Powe N, et al. Taming the chronic kidney disease epidemic: a global view of surveillance efforts. Kidney Int. 2014;86: 246–250. doi:10.1038/ki.2014.190

131. Collins GS, Omar O, Shanyinde M, Yu L-M. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. J Clin Epidemiol. 2013;66: 268–77. doi:10.1016/j.jclinepi.2012.06.020

132. Echouffo-Tcheugui JB, Kengne AP. Risk models to predict chronic kidney disease and its progression: a systematic review. Remuzzi G, editor. PLoS Med. Public Library of Science; 2012;9: e1001344. doi:10.1371/journal.pmed.1001344

Chapter 2

ADOPTION OF CLINICAL DECISION SUPPORT IN MULTIMORBIDITY: A SYSTEMATIC REVIEW

Paolo Fraccaro [1,2], Mercedes Arguello[3] , John Ainsworth [1,2], Iain E. Buchan[1,2]

[1] Health eResearch Centre, Farr Institute for Health Informatics Research, University of Manchester, UK.

[2] NIHR Greater Manchester Primary Care Patient Safety Translational Research Centre, Institute of Population Health, The University of Manchester, Manchester, United Kingdom.

[3] Bio-Health Informatics Group, School of Computer Science, The University of Manchester, Manchester, UK.

Contribution: PF (corresponding author) and IB designed the study. PF and MA (leading reviewer) retrieved all relevant publications. PF extracted all the data from the included studies. PF and MA drafted the manuscript . JA and IB critically edited the manuscript.

Based on: Fraccaro P, Arguello Casteleiro M, Ainsworth J, Buchan I. Adoption of Clinical Decision Support in Multimorbidity: A Systematic Review. JMIR Med Inf. 2015;3: e4. doi:10.2196/medinform.3503.

2.1 ABSTRACT

Background

Patients with multiple conditions have complex needs and are increasing in number as populations age. This multimorbidity is one of the greatest challenges facing health care. Having more than one condition generates (1) interactions between pathologies, (2) duplication of tests, (3) difficulties in adhering to often conflicting clinical practice guidelines, (4) obstacles in the continuity of care, (5) confusing self-management information, and (6) medication errors. In this context, clinical decision support (CDS) systems need to be able to handle realistic complexity and minimize iatrogenic risks.

Objective

The aim of this review was to identify to what extent CDS is adopted in multimorbidity.

Methods

This review followed PRISMA guidance and adopted a multidisciplinary approach. Scopus and PubMed searches were performed by combining terms from three different thesauri containing synonyms for (1) multimorbidity and comorbidity, (2) polypharmacy, and (3) CDS. The relevant articles were identified by examining the titles and abstracts. The full text of selected/relevant articles was analyzed in-depth. For articles appropriate for this review, data items were collected on clinical tasks, diseases, decision maker, methods, data input context, user interface considerations, and evaluation of effectiveness.

Results

A total of 50 articles were selected for the full in-depth analysis and 20 studies were included in the final review. Medication (n=10) and clinical guidance (n=8) were the predominant clinical tasks. Four studies focused on merging concurrent clinical practice guidelines. A total of 17 articles reported their CDS systems were knowledge-based. Most articles reviewed considered patients' clinical records (n=19), clinical practice guidelines (n=12), and clinicians' knowledge (n=10) as contextual input data. The most frequent diseases mentioned were cardiovascular (n=9) and diabetes mellitus (n=5). In all, 12 articles mentioned generalist doctor(s) as the decision maker(s). For articles reviewed, there were no studies referring to the active involvement of the patient in the decision-making process or to patient self-management. None of the articles reviewed adopted mobile technologies. There were no rigorous evaluations of usability or effectiveness of the CDS systems reported.

Conclusion

This review shows that multimorbidity is underinvestigated in the informatics of supporting clinical decisions. CDS interventions that systematize clinical practice guidelines without considering the interactions of different conditions and care processes may lead to unhelpful or harmful clinical actions. To improve patient safety in multimorbidity, there is a need for more evidence about how both conditions and care processes interact. The data needed to build this evidence base exist in many electronic health record systems and are underused.

Keywords: decision support systems, management; systematic review; multiple chronic diseases; multiple pathologies; multiple medications.

2.2 INTRODUCTION

2.2.1 Rationale

Patients affected by multiple diseases are acknowledged to be one of the greatest challenges for modern health care, especially as populations age [1]. Different terms have been used in the medical literature to refer to coexistent pathologies, the most accepted are [2] *comorbidity*, defined in 1970 as "any distinct additional clinical entity that has existed or may occur during the clinical course of a patient who has the index disease under study" [3], and *multimorbidity*, later defined as "the coexistence of two or more chronic conditions, where one is not necessarily more central than others" [4]. In this review, we look at the presence of simultaneous medical conditions as the decision-making context without emphasizing the prominence of any one condition, and we follow the European General Practice Research Network, which defines *multimorbidity* as "any combination of chronic disease with at least one other disease (acute or chronic) or bio-psychosocial factor (associated or not) or somatic risk factor" [5]. Here we use *multimorbidity* in a broad sense to infer comorbidity as well.

2.2.2 Impact of Multimorbidity on Public Health

Estimates of the prevalence of multimorbidity emanate from countries with detailed primary care records. A national population study carried out in the Netherlands estimated an overall prevalence of 29.7%, ranging from 10% in those younger than 20 years to 78% in those older than 80 [6]. Another population study in Scotland found out an overall prevalence of 23.2% [7]. The prevalence of multimorbidity in population increases with age [8]. Thus, a growing proportion of the population is affected by multimorbidity as populations age [9], particularly in countries with demographic patterns like the United Kingdom [10]. Previous studies [11–13] most commonly report the following disease groups as likely to concur: cardiovascular diseases, diabetes mellitus, chronic kidney disease, chronic musculoskeletal disorders, chronic lung disorders, and mental health disorders (particularly dementia and depression). There is also a greater burden of multimorbidity at younger ages (younger than 65 years) in deprived areas [7]. Thus, the public health and economic impact of multimorbidity is large [14]. In the United States, 84% of total health expenditure involves patients with more than one condition [15], whereas multimorbid patients in England accounted for the majority of primary care encounters [16] and this is expected to rise [15].

2.2.3 Patient-Centered Care and Iatrogenic Risks

The model of care in multimorbidity is changing, from a disease- and organization-centered approach [3] to patient-centered, holistic care [17]. Patient-centeredness considers psychological and physiological needs, the patient's concerns and priorities for care, self-care, and coordination between different professions and organizations, with primary care as an integrator [17]. Although patient-centered care is ideal for managing complex, chronic conditions, it is challenging to implement [5]; therefore, at present, patients with multimorbidity are commonly underserved by poorly integrated care systems [18,19]. This fragmentation reduces the safety, effectiveness, and efficiency of care [1]. A previous study reported that 10% to 20% of unscheduled care among older multimorbid adults is iatrogenic (e.g. medication-related harm) [20].

2.2.4 Self-Management and Continuity of Care

The presence of simultaneous care plans for multiple conditions leads to confusion and, in turn, generates safety hazards. Clear care plans, blending clinical care with self-management are essential in multimorbidity [21]. Such plans need to incorporate not only biomedical but also psychosocial factors, such as mood, informal care network, and patient income/finances [21]. Communication between patients/carers and health professionals over complex care plans can be challenging; therefore, self-care may be unreliable [21,22]. For example, it was estimated that in the United States an average Medicare patient with one chronic condition sees four different health care professionals in one year and this number increases to 14 in the presence of five different chronic conditions [22]. Increasing the number of health professionals involved creates a combinatorial explosion of communication interfaces and, for the patient, greater difficulty in understanding, remembering, and recalling guidance [22]. The most common problems arising from this miscommunication are duplication of tests and harmful decisions made on the basis of incomplete or incorrect information [23,24]. Primary care and general practitioners, in particular, are seen as a nexus of coordination for complex care such as this [24]. However, general practitioner workload is increasing beyond its capacity with the rising prevalence of chronic diseases and multimorbidity [25].

2.2.5 Clinical Practice Guidelines and Polypharmacy

Clinical research processes tend to focus narrowly on a single disease, mechanism, or treatment. This parsimony is reflected in the production of clinical practice guidelines; therefore, interactions between diseases are barely touched upon in care pathways (even if they are referred to as "integrated") [26]. More recently, organizations such as the National

Institute for Health and Care Excellence (NICE) have started to address multimorbidity explicitly [27], and a framework of principles for system-wide action to deal with comorbidities has been developed in England by the Department of Health and the National Health Service (NHS) [28]. Most current guidelines, however, do not consider interactions between diseases or between treatments [29]. Therefore, potential synergies or conflicts between different care pathways operating for the same patient may be missed [30]. For example, Boyd et al [29] applied clinical practice guidelines to a hypothetical case of a 79-year-old woman with multiple moderately severe chronic conditions (osteoporosis, osteoarthritis, diabetes mellitus, hypertension, and chronic obstructive pulmonary diseases). The guideline-derived treatment regimen was extremely complex and potentially harmful—comprising 14 nonpharmacologic treatments (i.e. self-monitoring, diet, exercise, health care visits, and laboratory testing) and 12 unique medications with 19 doses of medication per day [29]. Even in simpler cases, such as the presence of two diseases and two related treatments, researchers report 16 possible exposure patterns (half relevant for clinical practice guidelines) and four possible interaction combinations [26]. The two previous examples precipitate a "prescribing cascade" whereby drugs are prescribed to treat the adverse effects of other drugs, which is common in polypharmacy (the use of multiple medications) [31].

Even the most primary care–focused of health care systems, such as the NHS [27], do not deal safely, effectively, or efficiently with multimorbidity and polypharmacy [32]. In the future, with an aging population, most health care system resources will be stretched by the care needs of multimorbid patients [33].

2.2.6 Informatics Implications

Multimorbid health care requires complex communication, analysis, summarization, and presentation of heterogeneous clinical information from multiple sources. It is acknowledged that electronic health records (EHRs), especially in primary care, require enhanced functionality to support decisions in these complex care processes [34]. A clinical decision support (CDS) system provides "clinicians, patients or individuals with knowledge and person-specific or population information, intelligently filtered or presented at appropriate times, to foster better health processes, better individual patient care, and better population health." [35]. Despite notable failures [36], CDS systems have the potential to improve clinical outcomes [37,38]. Indeed, multimorbidity was defined as one of the "grand challenges in clinical decision support" by Sittig et al [39]; however, this area remains underinvestigated [40,41], with concerns raised over the unmet needs in primary care [40]. Some of the current challenges are lack of provision of integrated clinical practice guidelines, disease-centered rather than patient-centered approaches, difficulties in embedding CDS into clinical systems,

and lack of training to make best use of CDS [40]. EHRs and computerized physician order entry systems include rules that deal with drug-drug interactions; however, the whole patient context is not considered and the system may "overalert" physicians [42]. The overalert is another main risk in multimorbidity, which is known as *alert fatigue*: "the mental state provoked by managing too many irrelevant alerts from the system, which consume physical and psychological energies and lead the user to ignore also the relevant alerts resulting in potential harm for the patient" [43]. Prescribing alerts are especially important in polypharmacy, which has well-established risks of harm [44]. However, in some situations, multiple prescriptions are valid [30] and should not be dissuaded by inappropriate alerts. Context awareness, such as an "application's ability to adapt to changing circumstances and respond according to the context of use" [45], is crucial in decision support interventions [46], especially for multimorbidity where many variables are in place. However, a greater understanding of which information and sociotechnical factors of the context have to be taken into account in health care has still to be established [47].

Previous reviews have investigated specific aspects of CDS in multimorbidity; for example, prescribing in the elderly [48] and chronic disease management [49]. We could find no satisfactory review of CDS in multimorbidity from a technical/methodological perspective to guide the engineering of future systems. This interdisciplinary review plugs that gap.

2.2.7 Aim and Objectives

The aim was to review the current state of the art of CDS in multimorbidity. The objectives were to review the aspects of decision support target, contextual information about patients/practitioners/services, decision support technology, user interface considerations, decision maker(s), diseases, and evaluation. These aspects were analyzed to identify what works and what does not in CDS for multimorbidity, why systems failed to produce the expected outcomes, and what solutions might be adopted to address the problems.

2.3 METHODS

This review follows the guidelines from Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) framework [50]. PRISMA consists of a list of 27 items and a 4-phase flow diagram to complete that was identified as the optimal way to perform and report systematic reviews and meta-analyses about health care interventions by an experienced group of researchers and methodologists [50].

2.3.1 Eligibility

2.3.1.1 Inclusion Criteria

Studies that linked the concepts of multimorbidity, comorbidity, or polypharmacy to the concept of CDS, referring to the definitions provided previously, were selected from the literature.

The studies included in this literature review are papers about CDS systems that (1) address general issues about the multimorbid population, (2) support care for a particular subpopulation of multimorbid patients, (3) manage comorbidities related to a main disease, (4) deal with multiple concurrent medications in multimorbid population, and (5) describe statistical or machine-learning methods for clinical prediction in which the multimorbid patients' data feed the modeling/learning and a holistic approach is adopted.

2.3.1.2 Exclusion Criteria

Studies excluded from this literature review were about (1) CDS characteristics in general, without describing a CDS system in detail; (2) economic evaluations of CDS; (3) CDS systems in which multimorbidity was not a key feature; (4) social and operational research into CDS with no reference to clinical outcomes; (5) statistical or machine-learning approaches in which comorbidities were part of the model, but the patient-centered approach was not considered; and (6) systems that checked drug-drug interactions by means of simple rules, without taking into account multimorbidity or comorbidities.

2.3.2 Information Sources

MEDLINE and Scopus [51] were selected as the source indexes because they conform to the Cochrane requirement [52] of being "searched electronically both for words in the title or abstract and by using the standardized indexing terms, or controlled vocabulary, assigned to each record." We used the PubMed [53] interface to MEDLINE, which also includes up-to-date citations not yet indexed in MEDLINE [52]. In addition, Scopus can use Medical Subject Headings (MeSH) terms for structured queries [54].

Some target studies could only be found in the grey literature, such as theses and conference proceedings. Scopus allows search restrictions to some categories of grey literature, such as conference proceedings. This wider searching aimed to reduce publication bias.

The searches were performed in December 2013 and January 2014 without any restriction in the publication date.

2.3.3 Search

For the search, we followed three key points from the Cochrane Handbook [52]:

1. Searches should seek high sensitivity—this may result in poor precision.

2. Too many different search concepts should be avoided, but a wide variety of search terms should be combined with "or" within each concept.

3. Both free-text and subject headings should be used (e.g. MeSH) [55].

The focused clinical question that drove this systematic review was: What is the current level of adoption of CDS in multimorbidity? To answer this question, three different search concepts were selected:

1. Decision support: it has many related MeSH descriptors, such as "decision support systems, management" or "decision support techniques." Examples of individual hyponyms manually selected are "clinical decision support system," "decision support software," and "decision support tool."

2. Multimorbidity: it has zero related MeSH descriptors. Semantically, the closed concept comorbidity has one MeSH descriptor. Examples of synonyms manually selected are "concurrent conditions," "multiple chronic diseases," and "multiple pathologies."

3. Polypharmacy: it has just one MeSH descriptor and it should not be confused with the concept polypharmacology. Examples of synonyms manually selected are "several prescriptions," "poly-prescriptions," and "multiple medications."

In essence, the search created for the focused clinical question that drove this systematic review was based on three different search concepts and the hyponyms and synonym terms combined with "or." Conceptually, our clinical query was the following (see Supplementary Table 2.1 for full search):

< decision support > AND (<multimorbidity > OR < polypharmacy >)

In Scopus, the query created imposed that the relevant terms selected appear in the title, abstract, or keywords. The search yielded 954 articles (see Figure 2.1). Only literature from the social sciences, arts, and humanities was excluded from the search, and no restriction on the type of publication was imposed. Therefore, a wider selection of articles beyond the grey literature was retrieved.

Because multimorbidity is underrepresented in MeSH (i.e. no MeSH descriptor), we created a PubMed query that looked for the relevant terms selected in the title/abstract. The search created yielded 10,223 articles (i.e. 10 times more document results than in Scopus). We investigated the origin of this high number by looking at the query as it appeared under search details when using the PubMed search engine. Some of the synonyms manually selected for multimorbidity were not recognized; thus, they were split up automatically by PubMed [56]. Herein, the query as executed in PubMed contained overly general terms, such as "conditions," "diseases," and "pathologies." This severely affected the performance of the query. To further illustrate this, a subquery automatically generated by PubMed as part of the original query "decision support[Title/Abstract] AND conditions[Title/Abstract]" yielded 420 results. However, this subquery did not reflect our focused clinical question and it was very unlikely that it would retrieve the papers that we were interested in. Because the quality of any search depends on all constituents, we recognized that our original query was unsuitable for the PubMed search engine. More importantly, we became aware of the difficulties of constructing a PubMed query tailored to the medical question being investigated. Next, we tried to create more focused queries for the PubMed search engine, such as "multimorbidity[Title/Abstract] AND decision support[Title/Abstract]," which yielded only six articles. The low number of papers retrieved made us suspect that a substantive amount of papers were missing.

Knowing other researchers who were also conducting systematic reviews in the area of clinical decision support, we thought of a search intended for a global evidence map [57] (i.e. a search that sought to address broader questions about a particular area rather than focused clinical questions). It should be noted that global evidence maps are similar to systematic reviews because they are both conducted in a formal process; however, the time taken for a global evidence map is longer (in excess of two years [58]). We were interested in decision support related to electronic clinical documentation systems and safety surveillance, so we created a new PubMed query to provide a better context of the area under study where our clinical query should focus on. The new query as it appeared under search details when using the PubMed search engine was:

"decision support[Title/Abstract] OR (safety[Title/Abstract] AND surveillance[Title/Abstract]) OR electronic health record[Title/Abstract] OR electronic medical record[Title/Abstract] OR electronic patient record[Title/Abstract]"

This approach was adopted to guarantee the inclusion of all relevant papers even when CDS functionalities were described in studies about EHRs or safety surveillance systems without using CDS-related terms. To identify articles relevant to our focused clinical question, we

used automatic annotation of all papers' excerpts retrieved by the broader query using the hyponyms and synonym terms that appeared in the original clinical query for the three different search concepts originally selected (see Supplementary Table 2.1). For details, see the next subsection.

2.3.4 Study Selection

For the PubMed paper excerpts retrieved out of the broad query, we modified the manual approach to screening citations for systematic reviews and adopted some automation. In the area of automated document classification, there is an emerging body of research that uses machine-learning methods to help with the process of citation screening (e.g. [59]). We adopted a simpler, but well-founded, type of automation for prescreening PubMed paper excerpts, which did not classify paper excerpts as "relevant" or "irrelevant." We used automatic annotation of text (title and abstracts) based on a controlled vocabulary known beforehand and tailored to our study. This method is analogous to the bioinformatics practice of relating genes that have been annotated using a common schema, such as an ontology [60,61], which is directly relevant to systematic reviews [62,63]. We note that the Cochrane Collaboration is considering ontologies to support evidence synthesis [62].

The annotation was performed using a control vocabulary (i.e. the list of the hyponyms and synonym terms manually created for our clinical query). This annotation can coexist with native annotations from PubMed paper excerpts based on MeSH and/or authors' keywords. The concrete details of the annotation process are out of the scope of this paper. Once the annotation was performed, a selection of papers were selected based on our clinical query "< decision support > AND (<multimorbidity > OR < polypharmacy >)." Thus, only paper excerpts with at least one term in title/abstract related to decision support and at least one term in title/abstract related to multimorbidity or polypharmacy were identified as related to our clinical question.

Papers obtained by the preceding procedure were combined with the ones from the Scopus search and, after removing duplicates, screened on the basis of title and abstract. Relevant papers were assessed through full-text analysis to select the papers to be included in the systematic review.

2.3.5 Data Collection Process and Data Items

A careful selection of relevant features was agreed by the authors (PF, JA, and IB) and data on the following aspects were collected.

1. *Decision support target:* clinical tasks supported by the CDS system: prevention, diagnosis, care pathway guidance (i.e. management of patients according to clinical practice guidelines), medication (e.g. prescription, medication review), patient education, patient self-management, and care continuity (supporting communication between health care professionals involved in multimorbid patients).

2. *Contextual information:* Information regarding the context processed or taken into account by the system to provide support: patient clinical notes (i.e. demographics or family history), laboratory results, comorbidities, medications, clinical practice guidelines, and clinicians' knowledge.

3. *Decision support technology:*

    i. *Mode of delivery*: type of technical solution used to deliver the system: desktop application, Web application, and mobile application.

    ii. *Methodology:* methods used to perform the CDS intervention: data visualization techniques [64] (i.e. providing intuitive interfaces to minimize errors); social network techniques; international communication and coding standard, such as Health Level Seven International (HL7) (i.e. a set of International Organization for Standardization [ISO] approved framework to communicate information between healthcare information systems at the 7th layer of the Open Systems Interconnection model [65]), to develop sharable CDS solutions that can work across different systems and providers, and Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) [66] or International Classification of Diseases (ICD9-10) [67], to store data; machine-learning techniques [68]; natural language processing [69]; knowledge-based systems [70] (i.e. using rules or based on ontologies [71]); and mobile technologies.

    iii. *User interface considerations*: reported considerations about techniques to enhance and make easier user utilization of the system: interactivity, user-centered design, summarization, and workflow graphs.

    iv. *Decision maker(s):* user(s) of the CDS system: nurse, specialist doctor, pharmacist, generalist doctor (i.e. general practitioner or family doctor), and patient.

v. *Diseases/conditions:* CDS target conditions: obesity, diabetes mellitus, cardiovascular diseases, chronic respiratory diseases, chronic kidney disease, neurological conditions, mental health disorders, chronic musculoskeletal diseases, etc.

vi. *Evaluation:* type of evaluation of system's effectiveness: uncontrolled impact studies (e.g. surveys or health services measurements before/after CDS), controlled comparisons (e.g. comparing new vs old/no CDS), and no evaluation.

A summary was generated for each data item and study.

2.3.6 Synthesis of Results

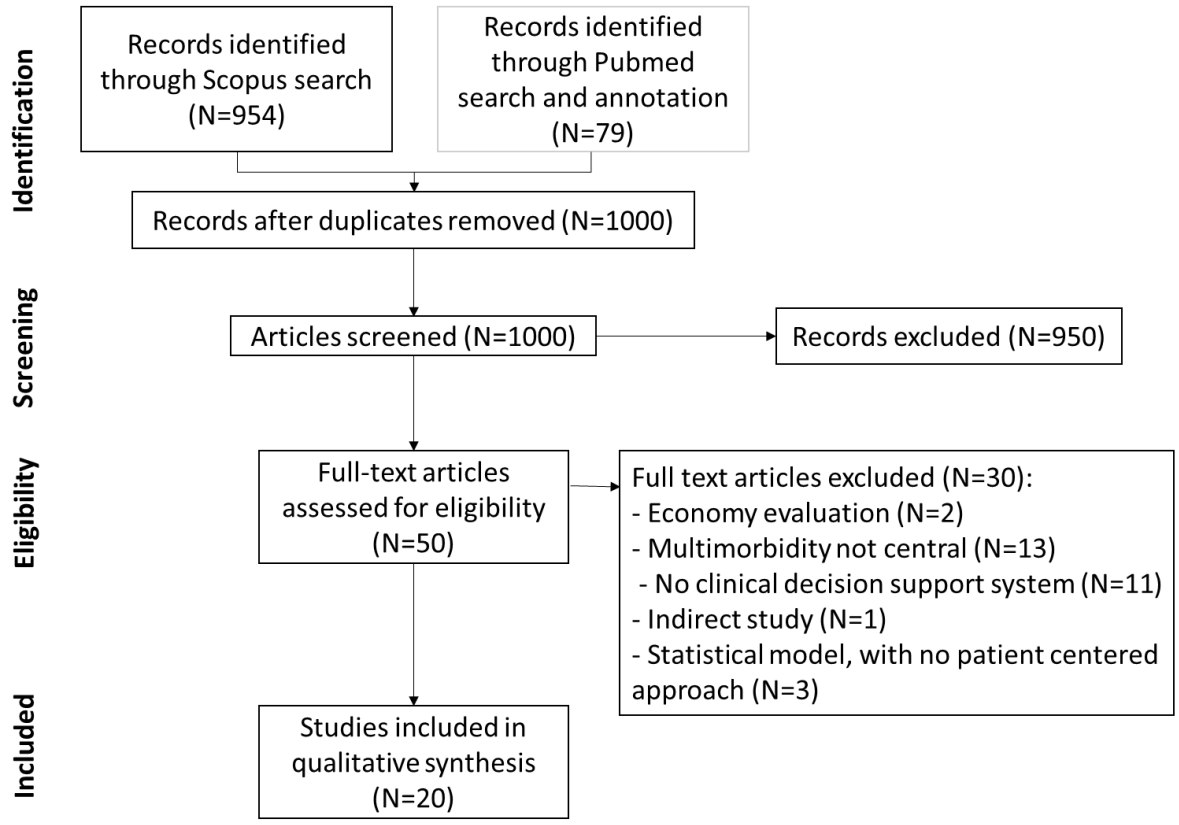
The results of the review are summarized in a table. The table is organized such that the aspects of CDS defined previously and provides a qualitative summary for each included study. An additional quantitative summary to highlight general trends over time and patterns of evidence is also provided.

2.4 RESULTS

2.4.1 Study Selection

The search via Scopus retrieved 954 articles. We retrieved 17,145 articles via PubMed by using the broad search introduced previously; 79 results were recalled after applying the programmatic filtering. After screening the title and abstract and removing duplicates, 50 articles were selected for in-depth analysis of the full text. A total of 20 studies were included in the review. The PRISMA process was followed and is reported in Figure 2.1.

Figure 2.1: Review flow diagram.



2.4.2 Study Characteristics

Table 2.1 contains the summary of all data items collected for each study included in the review along with its reference, while Table 2.2 shows the frequency distribution of the categories of aspects of CDS reported.

Table 2.1: Summary of collected items for included studies. * *No information was found about a particular data item.*

Authors	Decision support target	Contextual information	Decision support methods/delivery	User interface considerations	Decision maker/diseases	Evaluation
Abidi [72]	Pathways (merging clinical practice guidelines for different diseases into one personalized guideline)	Patient clinical notes & clinical practice guidelines	Knowledge-based system (ontology) & international standards/ *	*	*/*	*
Abidi et al [73]	Diagnosis & pathways (alignment of care pathways in a patient-specific comorbid combination) & patient education	Patient clinical notes & clinical practice guidelines	Knowledge-based system (ontology based)/ desktop application	Interactivity & summarization	Generalist doctor / chronic cardiovascular diseases	Controlled comparison-expert panel (revision by 2 generalist doctors and one specialist doctor)
Bindoff et al [74]	Medication (review)	Patient clinical notes & medications & laboratory results	Knowledge-based system (rule based)/ *	*	Pharmacists/*	Controlled comparison-human vs system comparison (system identified more problems)
Dassen et al [75]	Medication (prescription)	Patient clinical notes & medications & clinical practice guidelines & comorbidities & laboratory results	Knowledge-based system (ontology based) & international standards/desktop application	Interactivity & workflow graphs	Specialist doctor/cardiovascular diseases	*
de Wit et al [76]	Medication (review)	Patient clinical notes & clinical practice guidelines & clinician knowledge & laboratory results	Knowledge-based system (rule based)/*	*	Nurses/other (home care for the elderly)	No evaluation
Duke et al [77]	Medication (review)	Medications & clinician knowledge	Knowledge-based system & data visualization techniques & natural language processing/Web platform	Interactivity & summarization	Specialist doctor & generalist doctor/*	Controlled comparison-new vs old system (same accuracy but decreasing in time of 60%)

Authors	Decision support target	Contextual information	Decision support methods/delivery	User interface considerations	Decision maker/diseases	Evaluation
Farkas et al [78]	Diagnosis (comorbidities)	Patient clinical notes	Natural language processing/*	*	*/obesity	Controlled comparison-simulations (Fβ=1 score of 97% for classification based on textual evidence and 96% for intuitive judgments; Fβ=1 score of 76% for classification based on textual evidence and 67% for intuitive judgments)
Georg et al [79]	Medication (prescription)	Patient clinical notes & clinical practice guidelines	Knowledge-based system (rule based)/*	*	Generalist doctor/cardiovascular diseases	*
Grando et al [80]	Medication (prescription)	Patient clinical notes & clinical practice guidelines	Knowledge-based system (ontology based)/*	*	Generalist doctor/chronic respiratory diseases & diabetes & cardiovascular diseases & chronic musculoskeletal diseases & others	*
Jafarpour et al [81]	Pathways (merging clinical practice guidelines for different diseases into one personalized guideline)	Patient clinical notes & clinical practice guidelines & clinician knowledge	Knowledge-based system (ontology based)/*	*	Generalist doctor/cardiovascular diseases	No evaluation
Martínez-García et al [82]	Care continuity & pathways	Patient clinical notes & clinical practice guidelines & clinician knowledge	International standards & social network techniques/Web application (linked to electronic health record)	*	Nurse, generalist doctor, specialist doctor/*	Controlled comparison-survey (positively judged)

Authors	Decision support target	Contextual information	Decision support methods/delivery	User interface considerations	Decision maker/diseases	Evaluation
Michel et al [83]	Medication (prescription)	Patient clinical notes & clinical practice guidelines & clinician knowledge & medications & laboratory results & comorbidities	Knowledge-based system & data visualization techniques & international standards/desktop application (linked to electronic health record)	Summarization	Generalist doctor/chronic pain (opioid treated)	*
Naureckas et al [84]	Diagnosis & pathways	Patient clinical notes & clinical practice guidelines	Knowledge-based system & data visualization techniques/desktop application (linked to electronic health record)	User-centered design	Generalist doctor/child obesity and related diseases (e.g. diabetes, cardiovascular diseases, chronic kidney disease)	Impact evaluation-service performance metrics & survey
Riaño et al [85]	Diagnosis & medication (prescription) & pathways (developing a personalized treatment) & prevention	Patient clinical notes & clinical practice guidelines & clinician knowledge	Knowledge-based system (ontology based) & international standards/desktop application (linked to electronic health record)	*	Generalist doctor/home care in long-term conditions (e.g. obesity, diabetes, cardiovascular diseases, chronic respiratory diseases, chronic kidney disease, neurological conditions, mental health disorders, chronic musculoskeletal diseases)	Controlled comparison-survey (positively judged)
Riaño et al [86]	Medication (prescription)	Patient clinical notes & clinician knowledge	Knowledge-based system (rule based)/*	*	Generalist doctor/cardiovascular diseases & diabetes	Controlled comparison-expert panel (results validated by a generalist doctor)

Authors	Decision support target	Contextual information	Decision support methods/delivery	User interface considerations	Decision maker/diseases	Evaluation
Suojanen et al [87]	Diagnosis	Patient clinical notes & clinician knowledge	Machine learning/*	*	Specialist doctor/chronic neurological diseases	Controlled comparison-simulation (out of 30 cases: false positive rate=19%; false negative rate=23%)
Vallverdú et al [88]	Medication (prescription)	Patient clinical notes & clinician knowledge	Knowledge-based system (rule based)/desktop application	*	Generalist doctor/cardiovascular diseases & diabetes	Controlled comparison-expert panel (agreement with output from the system 100%-20/20)
Wicht et al [89]	Diagnosis (comorbidities)	Patient clinical notes & clinician knowledge	Knowledge-based system + data visualization techniques/Web platform	Interactivity	Specialist doctor/other (cancer)	Controlled comparison-expert panel (agreement with output from the system 84%-26/31)
Wilk et al [90]	Pathways (merging clinical practice guidelines for different diseases into one personalized guideline)	Patient clinical notes & clinical practice guidelines & clinician knowledge	Knowledge-based system (rule-based constraint logic programming)/*	Workflow graphs	Generalist doctor/other (duodenal ulcer, transient ischemic attack)	*
Wilk et al [91]	Pathways (alerting physicians about possible adverse interactions between 2 concurrent clinical practice guidelines)	Patient clinical notes & clinical practice guidelines	Knowledge-based system (rule-based constraint logic programming [92])/*	*	Specialist doctor & generalist doctor/chronic neurological & gastrointestinal diseases	*

Table 2.2: Synthesis of occurrences' numbers and references for collected data items.

Theme and category	Frequency	References
Decision support task		
Prevention	1	[85]
Diagnosis	6	[73,78,84,85,87,89]
Pathway	8	[72,73,81,82,84,85,90,91]
Medication	10	[74–77,79,80,83,85,86,88]
Patient education	1	[73]
Continuity of care	1	[82]
Self-management	0	—
Decision support technology		
Data visualization techniques	4	[77,83,84,89]
Social network techniques	1	[82]
International standards	5	[73,75,82,83,85]
Machine learning	1	[87]
Natural language processing	2	[77,78]
Knowledge-based system	17	[72–77,79–81,83–86,88–91]
Mobile technologies	0	—
Contextual information		
Patient clinical notes	19	[72–76,78–91]
Laboratory results	4	[74–76,83]
Comorbidities	2	[76,84]
Medications	4	[74,75,77,83]
Clinician knowledge	11	[76,77,81–83,85–90]
Clinical practice guidelines	13	[72,73,75,76,79–85,90,91]
Decision maker(s)		
Nurse	2	[76,82]
Specialist doctor	6	[75,77,82,87,89,91]
Generalist doctor	13	[73,77,79–86,88–91]
Pharmacist	1	[74
Patient	0	—
Not specified	2	[72,78]
Diseases		
Obesity	3	[78,84,85]
Diabetes	5	[80,84–86,88]
Cardiovascular diseases	9	[73,75,79–81,84–86,88]
Chronic respiratory diseases	2	[80,85]
Chronic kidney diseases	2	[84,85]
Chronic neurological conditions	3	[85,87,91]
Mental health disorders	1	[85]
Chronic musculoskeletal diseases	2	[80,85]
Other	8	[76,80,83–85,89,90]
Not specified	4	[72,74,77,82]

Theme and category	Frequency	References
User interface considerations		
Interactivity	4	[73,75,77,89]
User-centered design	1	[84]
Summarization	3	[73,77,83]
Workflow graphs	2	[75,90]
Not specified	13	[73,74,76,78–82,85–88,91]
Evaluation		
Impact evaluation (service performance metrics)	1	[84]
Impact evaluation (survey)	1	[84]
Controlled comparison (expert panel)	4	[72,86,88,89]
Controlled comparison (survey)	2	[82,85]
Controlled comparison (simulation)	2	[78,87]
Controlled comparison (human vs system)	1	[74]
Controlled comparison (new vs old system)	1	[77]
No evaluation	2	[76,81]
Not specified	7	[72,75,79,80,83,90,91]

2.4.3 Results of Individual Studies

Most of the papers reviewed focused on one of three clinical tasks: medication (n=10), clinical guidance (n=8), and diagnosis (n=6). From a methodological point of view, knowledge-based systems were the most frequently used (n=17). To further illustrate this, Riaño et al [85] described a CDS system that targets three decisions and uses knowledge-based systems. The authors developed a system that (1) provided patient-centered recommendations to better manage chronic diseases in the home setting and (2) used EHRs to refine an ontology, which described relevant concepts from clinical practice guidelines and the literature for 19 chronic diseases. The goal of this study was a patient-tailored ontology that contained patient-specific concepts that could be used to verify the diagnosis entered into the system. Starting from the personalized ontology, general treatment plans and patient management instructions could be combined into an individual plan. For multimorbid patients, a semiautomatic procedure applied, which involves the system's end-user. The system was able also to identify preventive opportunities by looking for anomalous circumstances, such as diagnosis inconsistent with other information or information missing which should always be presented alongside other information.

Abidi et al [73] presented a system that helped doctors in diagnosis and management of patients (two decision support targets) and used an ontology (knowledge centric), which was able to align clinical pathways for the multimorbid patient.

In the papers reviewed, medication was the main theme by far. This clinical task had the most contextualized input data and appeared as prescription (n=7) and medication review (n=3). Michel et al [84] developed a system that aimed to guide the generalist doctor through a summary of comprehensive relevant information (patient information, patient medication, patient laboratory results, and patient comorbidities) and suggested the optimal opioid treat for chronic pain. Dassen et al [75] developed a system, along the lines of Michel at al [83], considering comprehensive relevant information (patient information, patient medication, patient laboratory results, and patient comorbidities) and used an ontology to support cardiologists' prescriptions according to clinical practice guidelines. de Wit et al [76] focused on medication review and their system was intended to support safer care for the elderly. The system was capable of processing extracts of clinical data from electronic prescribing systems and EHRs (containing patient medication, patient conditions, and patient laboratory results) and alerted nurses about potentially harmful situations.

Another prevalent theme was the possible interaction between concurrent clinical practice guidelines for multimorbid patients. For example, Abidi et al [73] and Jafarpour et al [81] used ontologies to develop systems to merge two concurrent clinical practice guidelines into a comorbid personalized guideline. Jafarpour et al [81] carried out this task by creating an ontology that collected merging criteria obtained from clinical experts. Wilk et al [90,91] used constraint logic programming to identify and mitigate possible adverse interactions between clinical practice guidelines. The system described in [91] alerted doctors about possible hazards and suggested how to mitigate them. Martinez-Garcia et al [82] developed a system that improved clinical guidance by providing health care professionals with relevant information from clinical practice guidelines, and also supported communication between health care professionals. Their system (1) was directly linked to the EHR through HL7—an international standard for interoperability in health care and (2) adopted social networking techniques to enhance the continuity of care through a Web platform—it provided relevant patient information and performed safety checks according to clinical practice guidelines.

Some studies addressed the diagnosis of comorbidities for patients affected by an index condition/disease. For example, Farkas et al [78] used natural language processing applied to clinical notes to diagnose comorbidities in obese patients. Suojanen et al [87] used machine learning (causal Bayesian networks) for diagnosis of multiple concurrent neuropathies.

For the decision makers, generalist doctors were the most cited users of the CDS systems (n=13), followed by specialist doctors (n=6). No articles reported the patient as the decision maker. The system that appeared to involve the largest number of decision makers was

described by Martinez-Garcia et al [82], where nurses, specialist doctors, and generalist doctors were end users.

For disease, many papers considered multiple diseases (e.g. [80,84–86,91]), with Riano et al [85] reporting 19 chronic conditions.

For user interface considerations, most (n=13) articles did not provide details about the user interface. Where this information was provided, interactivity (n=4) [73,75,77,89] and summarization (n=3) [73,77,83] were the most cited features, whereas workflow graphs [75,90] were seldom mentioned. Only Naureckas et al [84] presented a CDS system that adopted a user-centered design with prompts and forms that helped generalist doctors to develop more effective behaviors for supporting diagnosis, management, and screening of comorbidities for children with obesity.

Regarding type of evaluation, some articles reported effectiveness objectively, including controlled comparisons (n=9) or impact evaluations (n=1). The articles that conducted surveys about their systems achieved positive judgments about the outcome provided [82,85]. In terms of accuracy, many studies reported good performance [87–89]. Duke et al [77] compared UpToDate [93] with a new system that had the same accuracy, but improved (by 60%) timeliness of decision. Bindoff et al [74] compared a CDS system with expert pharmacists when performing a medication review; overall, the system identified more potential problems than the human experts.

2.5 DISCUSSION

2.5.1 Summary of Evidence

This literature review found a modest number of papers addressing CDS and multimorbidity—an evidence base disproportionately small in comparison to the need for decision support.

2.5.2 The Lack of Patient-Centered Approaches

Most of the papers dealt with CDS targets that (1) are narrowly defined in terms of comorbidities around an index condition or (2) consider patient comorbidities only during prescription for a specific condition. Thereby, only a few of the studies reviewed refer to multimorbidity using a patient-centered approach, which is the ideal [5]. Riano et al [85] adopt a comprehensive approach to integrated care; however, user intervention is necessary to personalize treatments when multimorbidity is present.

2.5.3 Combination of Clinical Practice Guidelines

An important challenge of multimorbidity in CDS is the combination of clinical practice guidelines in a nonharmful way [39]. We found some studies that address this explicitly. An interesting solution is the one introduced by Jafarpour et al [81], which created an ontology with "merging criteria" provided by experts. Although rigorous evidence is lacking, to exploit physicians' "clinical mind-lines," such as "tacit guidelines that are internalized and collectively reinforced from the experience and discussion with colleagues and patients to embody the complex and flexible knowledge needed in practice" [94], seems the only solution. However, all systems described in the articles reviewed tend to simplify the analysis by referring to only two concurrent clinical practice guidelines. This scenario is too simplistic for the current reality because multimorbid patients often face more than two simultaneous pathologies [29].

2.5.4 Continuity of Care

Discontinuity of care between different health professionals is an important source of safety problems, which is highly relevant to multimorbidity considering the large numbers of professionals involved. Yet only 1 article [82] considered this aspect. Prevalent technologies such as social media may foster communication across different clinical settings. There is a notable gap in the evidence base here.

2.5.5 No Self-Management Interventions

Self-management is key in multimorbidity [21]. In the articles reviewed, no CDS interventions for multimorbid patient self-management were found. Similarly, we noticed the absence of mobile technologies for CDS in multimorbidity.

2.5.6 Methodological Considerations

From a methodological point of view, knowledge-based systems were most commonly reported. Data-driven methods, such as machine-learning techniques, were barely used in the reviewed studies, with just one study [87] adopting them.

2.5.7 The Technological Interoperability Shortfalls

Multimorbidity is composed of interacting variables; therefore, systems need to be aware of as many contextual factors as possible to deliver relevant support and information [95]. Emerging international standards, such as HL7, are supposed to enable interoperability in health care; however, only one article reviewed uses HL7, the system developed by Martinez-Garcia et al [83].

2.5.8 The Need for More Rigorous Evaluations

Evaluations of usability and effectiveness of systems are key to avoiding patient harm and waste in health care systems [96]. The so called "e-iatrogenesis" [97] arising from information systems has more potential pitfalls when there are multiple conditions. Rigorous evaluations are needed to test systems before and after their deployment to guarantee patient safety [98,99] [100,101]. We found a lack of rigorous evaluations of effectiveness and usability here, which is consistent with the overall state of CDS [36] research. Patient safety needs to be assured by rigorous evaluation, not only of the underlying software/technologies but also of their real-world interaction with users [100]. The expected approaches to evaluating human-computer interaction [101,102] were not found in the articles we examined.

2.5.9 Limitations

This review has several limitations. First, only Scopus and PubMed sources were searched—other relevant material may exist in the grey literature. Second, the titles and abstracts of the papers selected are anchored to the terms included in the three thesauri —some papers may have been missed if other synonyms were used. Third, it was not possible to find studies covering all aspects of CDS we considered—some aspects, such as the evaluation of the effectiveness and usability, were quite sparsely covered, but this is a general weakness of the CDS literature [36]. Finally, we did not follow the traditional systematic review process for all searches. However, we are confident that our strategy guaranteed the inclusion of all relevant papers about the topic. There is an ongoing discussion of what should and should not be automated in systematic reviews, particularly to strike the right balance between depth and timeliness [103]. Here we took the middle ground, using computational methods to make a more "concept-complete" search tractable. Therefore, this review may contribute to the ongoing discussion about semiautomated prescreening of medical literature while preserving rigorous methods of evidence synthesis.

2.5.10 Implications for Future Research and Conclusions

This review shows how multimorbidity is understudied in CDS, yet this is an area of public health and clinical importance that should be a prime target for CDS research.

There are already many technologies in health care and industry relevant to dealing with the complexity of multimorbid decision support. Kawamoto et al [104] argue that wider adoption of international terminologies (e.g. SNOMED CT) and electronic health record standards can lead to better CDS, tapping into the vast amount of data produced in routine clinical practice for multimorbid patients. Moreover, technical frameworks [105] were already proposed for a "shared and informed decision making" in industry that with appropriate adjustments could

be used to enhance continuity of care in multimorbidity. In addition, the absence of any substantial papers dealing with self-care for people affected by multiple conditions was remarkable given the rapid growth in connected/consumer health and its inevitable influence on CDS in the future.

Multimorbidity is a relatively new field of clinical research and more evidence is needed to support CDS in this area. This underpinning knowledge is, however, challenging. For example, patients with multiple conditions or on multiple medications are often excluded from clinical trials [106]. However, EHRs afford the possibility of observational studies important for understanding multimorbid disease risks, care processes, and care outcomes. Such observational data have established value in decreasing the prescribing cascade and other iatrogenesis [107]. Automation of care pathways/processes that are poorly understood, such as merging guidelines [30], may lead to unhelpful or harmful clinical actions. The informatics challenge herein is to build the evidence base about multimorbid care while engineering more supportive/directive clinical information systems incrementally. The clinical epidemiology and health services research must be interwoven with the systems development. Gathering more clinical evidence and getting more involvement from patients and health professionals is central to finding a technological approach to managing multimorbidity and enhancing patient safety. At the same time, rigorous evaluation of all sociotechnical and human-computer interaction aspects of produced CDS interventions is certainly a priority for the future.

Patients with multiple conditions are one of the most important groups for health care systems to understand and evolve to serve [33]. There are multiple dynamics in which CDS and health informatics can contribute in meeting this challenge: (1) using EHR data to understand multimorbidity and plug a relatively sparse evidence base, (2) coproducing care decisions between patients and practitioners in the face of complexity and uncertainty, and (3) blending n-of-1 patient experiments/experience with evidence about the "average patient like Mrs X..." It is hard to conceive of such complexity being tamed by today's EHR interfaces, punctuated by blizzards of alerts and dashboards. Future CDS may be part of an integrated health avatar [108]: "the electronic representation of an individual's health as directly measured or inferred by statistical models or clinicians." To achieve such integration, however, there is a pressing need for more realistically complex CDS research, particularly in multimorbidity.

2.6 REFERENCES

1. Boyd CM, Fortin M. Future of multimorbidity research: How should understanding of multimorbidity inform health system design? Public Health Rev. Division of Geriatric Medicine and Gerontology, Department of Medicine, Johns Hopkins University School of Medicine, 5200 Eastern Avenue, Baltimore, MD 21224-2734, United States; 2011;33: 451–474. Available: http://www.webcitation.org/6NYkztx37

2. Almirall J, Fortin M. The coexistence of terms to describe the presence of multiple concurrent diseases [Internet]. Journal of Comorbidity. 2013. pp. 4–9. Available: http://www.webcitation.org/6OcFK4U6O

3. Feinstein AR. The pre-therapeutic classification of co-morbidity in chronic disease. J Chronic Dis. 1970;23: 455–468. Available: http://www.webcitation.org/6NYS3ligc

4. Fortin M, Soubhi H, Hudon C, Bayliss EA, van den Akker M. Multimorbidity's many challenges. BMJ. 2007;334: 1016–7. doi:10.1136/bmj.39201.463819.2C

5. Le Reste JY, Nabbe P, Manceau B, Lygidakis C, Doerr C, Lingner H, et al. The European General Practice Research Network presents a comprehensive definition of multimorbidity in family medicine and long term care, following a systematic review of relevant literature. J Am Med Dir Assoc. 2013;14: 319–25. doi:10.1016/j.jamda.2013.01.001

6. van den Akker M, Buntinx F, Metsemakers JF, Roos S, Knottnerus JA. Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases. J Clin Epidemiol. 1998;51: 367–75. Available: http://www.webcitation.org/6NYSQImWv

7. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet. 2012;380: 37–43. doi:10.1016/S0140-6736(12)60240-2

8. Fortin M, Hudon C, Haggerty J, Akker M van den, Almirall J. Prevalence estimates of multimorbidity: a comparative study of two sources. BMC Health Serv Res. 2010;10: 111. doi:10.1186/1472-6963-10-111

9. Uijen AA, van de Lisdonk EH. Multimorbidity in primary care: prevalence and trend over the last 20 years. Eur J Gen Pract. 2008;14 Suppl 1: 28–32. doi:10.1080/13814780802436093

10. Dunnell K. The changing demographic picture of the UK: national statistician's annual article on the population. Popul Trends. 2007; 9–21. Available: http://www.webcitation.org/6NYkEaxw9

11. Zhang X, Decker FH, Luo H, Geiss LS, Pearson WS, Saaddine JB, et al. Trends in the prevalence and comorbidities of diabetes mellitus in nursing home residents in the United States: 1995-2004. J Am Geriatr Soc. 2010;58: 724–30. doi:10.1111/j.1532-5415.2010.02786.x

12. Stevens LA, Li S, Wang C, Huang C, Becker BN, Bomback AS, et al. Prevalence of CKD and comorbid illness in elderly patients in the United States: results from the Kidney Early Evaluation Program (KEEP). Am J Kidney Dis. 2010;55: S23-33. doi:10.1053/j.ajkd.2009.09.035

13. Yaffe K, Ackerson L, Kurella Tamura M, Le Blanc P, Kusek JW, Sehgal AR, et al. Chronic kidney disease and cognitive function in older adults: findings from the chronic renal insufficiency cohort cognitive study. J Am Geriatr Soc. 2010;58: 338–45. doi:10.1111/j.1532-5415.2009.02670.x

14. Taylor AW, Price K, Gill TK, Adams R, Pilkington R, Carrangis N, et al. Multimorbidity - not just an older person's issue. Results from an Australian biomedical study. BMC Public Health. 2010;10: 718. doi:10.1186/1471-2458-10-718

15. Anderson G. Chronic Care: Making the Case for Ongoing Care [Internet]. Robert Wood Johnson Foundation; 2010. Available: http://www.webcitation.org/6NYkTSQ5Z

16. Salisbury C, Johnson L, Purdy S, Valderas JM, Montgomery AA. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. Br J Gen Pract. 2011;61: e12-21. doi:10.3399/bjgp11X548929

17. Schofield T. Patient-Centered Medicine: Transforming the Clinical Method. BMJ. 1995;311: 1580–1580. doi:10.1136/bmj.311.7019.1580

18. Vogeli C, Shields AE, Lee TA, Gibson TB, Marder WD, Weiss KB, et al. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. J Gen Intern Med. 2007;22 Suppl 3: 391–5. doi:10.1007/s11606-007-0322-1

19. Zulman DM, Asch SM, Martins SB, Kerr EA, Hoffman BB, Goldstein MK. Quality

of Care for Patients with Multiple Chronic Conditions: The Role of Comorbidity Interrelatedness. J Gen Intern Med. 2013; doi:10.1007/s11606-013-2616-9

20. Budnitz DS, Shehab N, Kegler SR, Richards CL. Medication use leading to emergency department visits for adverse drug events in older adults. Ann Intern Med. 2007;147: 755–65. Available: http://www.webcitation.org/6NYl4U9bV

21. Bayliss EA, Ellis JL, Steiner JF. Barriers to self-management and quality-of-life outcomes in seniors with multimorbidities. Ann Fam Med. 2007;5: 395–402. doi:10.1370/afm.722

22. Eichner J, Blumenthal D. Medicare in the 21st Century: Building a Better Chronic Care System. SSRN Electron J. 2004; doi:10.2139/ssrn.555632

23. Kuperman GJ. Health-information exchange: why are we doing it, and what are we doing? J Am Med Informatics Assoc. 2011;18: 678 LP-682. Available: http://jamia.oxfordjournals.org/content/18/5/678.abstract

24. Haggerty JL. Ordering the chaos for patients with multimorbidity. BMJ. 2012;345: e5915. doi:10.1136/bmj.e5915

25. Østbye T, Yarnall KSH, Krause KM, Pollak KI, Gradison M, Michener JL. Is there time for management of patients with chronic diseases in primary care? Ann Fam Med. American Academy of Family Physicians; 2005;3: 209–14. doi:10.1370/afm.310

26. Weiss CO, Varadhan R, Puhan MA, Vickers A, Bandeen-Roche K, Boyd CM, et al. Multimorbidity and Evidence Generation. J Gen Intern Med. 2014; doi:10.1007/s11606-013-2660-5

27. The UK Department of Health. National Institute for Health and Care Excellence [Internet]. [cited 24 Oct 2016]. Available: https://www.nice.org.uk/

28. Department of Health, NHS England, PHE. Comorbidities: A framework of principles for system-wide action [Internet]. 2014. Available: http://www.webcitation.org/6PF1bbuVL

29. CM B, Darer J, Boult C, LP F, Boult L, AW W. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: Implications for pay for performance. JAMA. 2005;294: 716–724. Available: http://dx.doi.org/10.1001/jama.294.6.716

30. Guthrie B, Payne K, Alderson P, McMurdo MET, Mercer SW. Adapting clinical guidelines to take account of multimorbidity. BMJ. 2012;345: e6341. doi:10.1136/bmj.e6341

31. Koper D, Kamenski G, Flamm M, Böhmdorfer B, Sönnichsen A. Frequency of medication errors in primary care patients with polypharmacy. Fam Pract. 2013;30: 313–9. doi:10.1093/fampra/cms070

32. Gaal S, Verstappen W, Wolters R, Lankveld H, van Weel C, Wensing M. Prevalence and consequences of patient safety incidents in general practice in the Netherlands: a retrospective medical record review study. Implement Sci. 2011;6: 37. doi:10.1186/1748-5908-6-37

33. Dobrow M. Book review: Caring for people with chronic conditions: a health system perspective [Internet]. International Journal of Integrated Care. Igitur Publishing and Archiving Services; Available: http://www.webcitation.org/6NYmDYDdq

34. Krist AH, Beasley JW, Crosson JC, Kibbe DC, Klinkman MS, Lehmann CU, et al. Electronic health record functionality needed to better support primary care. J Am Med Inform Assoc. 2014; amiajnl-2013-002229-. doi:10.1136/amiajnl-2013-002229

35. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. J Am Med Inform Assoc. 2007;14: 141–5. doi:10.1197/jamia.M2334

36. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. Djulbegovic B, editor. PLoS Med. Public Library of Science; 2011;8: e1000387. doi:10.1371/journal.pmed.1000387

37. Cantrill MD, FACEP S V. Computers in Patient Care: The Promise and the Challenge. Queue. ACM; 2010;8: 20. doi:10.1145/1839572.1841832

38. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med. 2012;157: 29–43. doi:10.7326/0003-4819-157-1-201207030-00450

39. Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, et al. Grand challenges in clinical decision support. J Biomed Inform. 2008;41: 387–92. doi:10.1016/j.jbi.2007.09.003

40. Grace A, Mahony C, O'Donoghue J, Heffernan T, Molony D, Carroll T. Evaluating the effectiveness of clinical decision support systems: the case of multimorbidity care. J Decis Syst. Taylor & Francis; 2013;22: 97–108. doi:10.1080/12460125.2013.780320

41. Smith SM, Soubhi H, Fortin M, Hudon C, O'Dowd T. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. Cochrane database Syst Rev. 2012;4: CD006560. doi:10.1002/14651858.CD006560.pub2

42. Kuperman GJ, Bobb A, Payne TH, Avery AJ, Gandhi TK, Burns G, et al. Medication-related clinical decision support in computerized provider order entry systems: a review. J Am Med Inform Assoc. 2007;14: 29–40. doi:10.1197/jamia.M2170

43. Sijs IH van der (Heleen). Drug safety alerting in computerized physician order entry: Unraveling and counteracting alert fatigue. 2009; Available: http://www.webcitation.org/6NYq64xRe

44. O'Brien R, Wyke S, Guthrie B, Watt GCM, Mercer SW. An "endless struggle": a qualitative study of general practitioners' and practice nurses' experiences of managing multimorbidity in socio-economically deprived areas of Scotland [Internet]. Chronic Illness. Sage Publications Ltd.; 2011. doi:10.1177/1742395310382461 <http://dx.doi.org/10.1177/1742395310382461 >)

45. Kjeldskov J, Skov M. Supporting Work Activities in Healthcare by Mobile Electronic Patient Records. In: Masoodian M, Jones S, Rogers B, editors. Computer Human Interaction SE - 20. Springer Berlin Heidelberg; 2004. pp. 191–200. doi:10.1007/978-3-540-27795-8_20

46. Burstein F, Holsapple CW. Decision support systems in context. Inf Syst E-bus Manag. 2008;6: 221–223. doi:10.1007/s10257-008-0085-1

47. Bricon-Souf N, Newman CR. Context awareness in health care: a review. Int J Med Inform. 2007;76: 2–12. doi:10.1016/j.ijmedinf.2006.01.003

48. Yourman L, Concato J, Agostini J V. Use of computer decision support interventions to improve medication prescribing in older adults: A systematic review. Am J Geriatr Pharmacother. 2008;6: 119–129. Available: http://www.webcitation.org/6NYn2eOiy

49. Roshanov PS, Misra S, Gerstein HC, Garg AX, Sebaldt RJ, Mackay JA, et al. Computerized clinical decision support systems for chronic disease management: a

decision-maker-researcher partnership systematic review. Implement Sci. 2011;6: 92. doi:10.1186/1748-5908-6-92

50. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. Public Library of Science; 2009;6: e1000100. doi:10.1371/journal.pmed.1000100

51. Elevier. Scopus [Internet]. Available: https://www.scopus.com/

52. Lefebvre C, Manheimer E, Glanville J. Searching for studies. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 510. The Cochrane Collaboration; 2011.

53. Home - PubMed - NCBI [Internet]. Available: http://www.webcitation.org/6NYnlPJeE

54. Sullo E. Scopus. J Med Libr Assoc. Medical Library Association; 2007;95: 367–368. doi:10.3163/1536-5050.95.3.367

55. Lowe HJ, Barnett GO. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. JAMA. 1994;271: 1103–8. Available: http://www.ncbi.nlm.nih.gov/pubmed/8151853

56. An Introduction to PubMed [Internet]. Available: http://www.webcitation.org/6Sp4XaoEJ

57. Bragge P, Clavisi O, Turner T, Tavender E, Collie A, Gruen RL. The Global Evidence Mapping Initiative: scoping research in broad topic areas. BMC Med Res Methodol. 2011;11: 92. doi:10.1186/1471-2288-11-92

58. Bero L, Busuttil G, Farquhar C, Koehlmoos TP, Moher D, Nylenna M, et al. Measuring the performance of the Cochrane library. Cochrane database Syst Rev. 2012;12: ED000048. doi:10.1002/14651858.ED000048

59. Wallace BC, Small K, Brodley CE, Lau J, Trikalinos TA. Deploying an interactive machine learning system in an evidence-based practice center. Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI '12. New York, New York, USA: ACM Press; 2012. p. 819. doi:10.1145/2110363.2110464

60. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID:

Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 2003;4: P3. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3720094&tool=pmcentrez&rendertype=abstract

61. Huntley RP, Harris MA, Alam-Faruque Y, Blake JA, Carbon S, Dietze H, et al. A method for increasing expressivity of Gene Ontology annotations using a compositional approach. BMC Bioinformatics. 2014;15: 155. doi:10.1186/1471-2105-15-155

62. Mavergames C, Oliver S, Becker L. Systematic Reviews as an interface to the web of (trial) data: Using PICO as an ontology for knowledge synthesis in evidence-based healthcare research. SePublica. 2013; 22–26.

63. Huang Z, ten Teije A, van Harmelen F. SemanticCT: A Semantically-Enabled System for Clinical Trials. In: Riaño D, Lenz R, Miksch S, Peleg M, Reichert M, ten Teije A, editors. Process Support and Knowledge Representation in Health Care SE - 2. Springer International Publishing; 2013. pp. 11–25. doi:10.1007/978-3-319-03916-9_2

64. Card S. The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications [Internet]. Lawrence Erlbaum Assoc Inc; 2007. Available: http://www.webcitation.org/6NYny57Bu

65. Health Level Seven International - Homepage [Internet]. Available: http://www.webcitation.org/6NYo2EQsR

66. SNOMED Clinical Terms® (SNOMED CT®) [Internet]. U.S. National Library of Medicine; Available: http://www.webcitation.org/6NYo9VKom

67. WHO | International Classification of Diseases (ICD). World Health Organization; Available: http://www.webcitation.org/6NYoG44MR

68. Mehryar M, Afshin R, Ameet T. Foundations of Machine Learning [Internet]. The MIT press, editor. Cambridge, Massachussets; London, England: The MIT Press; 2012. Available: http://www.webcitation.org/6NYoL0o7A

69. Jurafsky D, Martin JH. Speech And Language Processing: An Introduction to Natural Language Processing , Computational Linguistics, and Speech Recognition [Internet]. Pearson Prentice Hall/Pearson education international; 2009. Available:

http://www.webcitation.org/6NYoR2I09

70. Brachman RJ, Levesque HJ. Knowledge representation and reasoning [Internet]. Elsevier/Morgan Kaufmann Publishers; 2004. Available: http://www.webcitation.org/6NYodAVnC

71. Suárez-Figueroa MC, Gómez-Pérez A, Motta E, Gangemi A. Ontology Engineering in a Networked World [Internet]. Springer; 2012. Available: http://www.webcitation.org/6NYoxjEXv

72. Abidi S. A Conceptual Framework for Ontology Based Automating and Merging of Clinical Pathways of Comorbidities. In: Riaño D, editor. Knowledge Management for Health Care Procedures SE - 5. Springer Berlin Heidelberg; 2009. pp. 55–66. doi:10.1007/978-3-642-03262-2_5

73. Abidi S, Cox J, Shepherd M, Abidi SSR. Using OWL Ontologies for Clinical Guidelines Based Comorbid Decision Support. 2012 45th Hawaii International Conference on System Sciences. IEEE; 2012. pp. 3030–3038. doi:10.1109/HICSS.2012.629

74. Bindoff I, Stafford A, Peterson G, Kang BH, Tenni P. The potential for intelligent decision support systems to improve the quality and consistency of medication reviews. J Clin Pharm Ther. 2012;37: 452–8. doi:10.1111/j.1365-2710.2011.01327.x

75. Dassen WRM, Gorgels APM, Berendsen A, Dijk WA, de Clercq PA, Hasman A, et al. Guideline assessment and implementation in congestive heart failure. Computers in Cardiology, 2003. IEEE; 2003. pp. 331–334. doi:10.1109/CIC.2003.1291158

76. de Wit HAJM, Mestres Gonzalvo C, Hurkens KPGM, Mulder WJ, Janknegt R, Verhey FR, et al. Development of a computer system to support medication reviews in nursing homes. Int J Clin Pharm. 2013;35: 668–72. doi:10.1007/s11096-013-9827-3

77. Duke JD, Li X, Grannis SJ. Data visualization speeds review of potential adverse drug events in patients on multiple medications. J Biomed Inform. 2010;43: 326–31. doi:10.1016/j.jbi.2009.12.001

78. Farkas R, Szarvas G, Hegedus I, Almási A, Vincze V, Ormándi R, et al. Semi-automated construction of decision rules to predict morbidities from clinical texts. J Am Med Inform Assoc. 2009;16: 601–5. doi:10.1197/jamia.M3097

79. Georg G, Séroussi B, Bouaud J. Synthesis of elementary single-disease recommendations to support guideline-based therapeutic decision for complex polypathological patients. Stud Health Technol Inform. 2004;107: 38–42. Available: http://www.webcitation.org/6NX2FnFrI

80. Grando A, Farrish S, Boyd C, Boxwala A. Ontological approach for safe and effective polypharmacy prescription. AMIA Annu Symp Proc. 2012;2012: 291–300. Available: http://www.webcitation.org/6NX2M5qoI

81. Jafarpour B, Abidi S. Merging Disease-Specific Clinical Guidelines to Handle Comorbidities in a Clinical Decision Support Setting. In: Peek N, Marín Morales R, Peleg M, editors. Artificial Intelligence in Medicine SE - 5. Springer Berlin Heidelberg; 2013. pp. 28–32. doi:10.1007/978-3-642-38326-7_5

82. Martínez-García A, Moreno-Conde A, Jódar-Sánchez F, Leal S, Parra C. Sharing clinical decisions for multimorbidity case management using social network and open-source tools. J Biomed Inform. 2013;46: 977–84. doi:10.1016/j.jbi.2013.06.007

83. Michel M, Trafton J, Martins S, Wang D, Tu S, Johnson N, et al. Improving Patient Safety Using ATHENA-Decision Support System Technology: The Opioid Therapy for Chronic Pain Experience. Adv Patient Saf New Dir Altern Approaches (Vol 4 Technol Medicat Safety). Agency for Healthcare Research and Quality (US); 2008;4. Available: http://www.webcitation.org/6NX2ok7D7

84. Naureckas SM, Zweigoron R, Haverkamp KS, Kaleba EO, Pohl SJ, Ariza AJ. Developing an electronic clinical decision support system to promote guideline adherence for healthy weight management and cardiovascular risk reduction in children: a progress update. Transl Behav Med. 2011;1: 103–7. doi:10.1007/s13142-011-0019-1

85. Riaño D, Real F, López-Vallverdú JA, Campana F, Ercolani S, Mecocci P, et al. An ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients. J Biomed Inform. 2012;45: 429–46. doi:10.1016/j.jbi.2011.12.008

86. Riaño D, Collado A. Model-Based Combination of Treatments for the Management of Chronic Comorbid Patients. In: Peek N, Marín Morales R, Peleg M, editors. Artificial Intelligence in Medicine SE - 2. Springer Berlin Heidelberg; 2013. pp. 11–16. doi:10.1007/978-3-642-38326-7_2

87. Suojanen M, Andreassen S, Olesen KG. A method for diagnosing multiple diseases in MUNIN. IEEE Trans Biomed Eng. 2001;48: 522–32. doi:10.1109/10.918591

88. López-Vallverdú J, Riaño D, Collado A. Rule-Based Combination of Comorbid Treatments for Chronic Diseases Applied to Hypertension, Diabetes Mellitus and Heart Failure. In: Lenz R, Miksch S, Peleg M, Reichert M, Riaño D, Teije A, editors. Process Support and Knowledge Representation in Health Care SE - 2. Springer Berlin Heidelberg; 2013. pp. 30–41. doi:10.1007/978-3-642-36438-9_2

89. Wicht A, Wetter T, Klein U. A web-based system for clinical decision support and knowledge maintenance for deterioration monitoring of hemato-oncological patients. Comput Methods Programs Biomed. 2013;111: 26–32. doi:10.1016/j.cmpb.2013.02.007

90. Wilk S, Michalowski M, Michalowski W, Hing MM, Farion K. Reconciling pairs of concurrently used clinical practice guidelines using Constraint Logic Programming. AMIA Annu Symp Proc. 2011;2011: 944–53. Available: http://www.webcitation.org/6NX3RIELG

91. Wilk S, Michalowski W, Michalowski M, Farion K, Hing MM, Mohapatra S. Mitigation of adverse interactions in pairs of clinical practice guidelines using constraint logic programming. J Biomed Inform. 2013;46: 341–53. doi:10.1016/j.jbi.2013.01.002

92. Balduccini M, Son TC, editors. Logic Programming, Knowledge Representation, and Nonmonotonic Reasoning [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. doi:10.1007/978-3-642-20832-4

93. Evidence-Based Clinical Decision Support at the Point of Care | UpToDate [Internet]. Available: http://www.webcitation.org/6NYpIR20k

94. Gabbay J, May A le. Practice-based Evidence for Healthcare: Clinical Mindlines [Internet]. Taylor & Francis; 2010. Available: http://books.google.co.uk/books/about/Practice_based_Evidence_for_Healthcare.html?id=gafpKvpl5iMC&pgis=1

95. Riedmann D, Jung M, Hackl WO, Stühlinger W, van der Sijs H, Ammenwerth E. Development of a context model to prioritize drug safety alerts in CPOE systems. BMC Med Inform Decis Mak. 2011;11: 35. doi:10.1186/1472-6947-11-35

96. Magrabi F, Ong M, Runciman W, Coiera E. Patient safety problems associated with heathcare information technology: an analysis of adverse events reported to the US Food and Drug Administration. AMIA Annu Symp Proc. 2011;2011: 853–7. Available: http://www.webcitation.org/6NYpU7QLY

97. Weiner JP, Kfuri T, Chan K, Fowles JB. "e-Iatrogenesis": the most critical unintended consequence of CPOE and other HIT. J Am Med Inform Assoc. 2007;14: 387–8; discussion 389. doi:10.1197/jamia.M2338

98. Denham CR, Classen DC, Swenson SJ, Henderson MJ, Zeltner T, Bates DW. Safe use of electronic health records and health information technology systems: trust but verify. J Patient Saf. 2013;9: 177–89. doi:10.1097/PTS.0b013e3182a8c2b2

99. Kilbridge PM, Welebob EM, Classen DC. Development of the Leapfrog methodology for evaluating hospital implemented inpatient computerized physician order entry systems. Qual Saf Health Care. 2006;15: 81–4. doi:10.1136/qshc.2005.014969

100. Coiera E. Interaction design theory. Int J Med Inform. 2003;69: 205–222. Available: http://www.webcitation.org/6NYplIayC

101. Horsky J, Schiff GD, Johnston D, Mercincavage L, Bell D, Middleton B. Interface design principles for usable decision support: a targeted review of best practices for clinical prescribing interventions. J Biomed Inform. 2012;45: 1202–16. doi:10.1016/j.jbi.2012.09.002

102. Phansalkar S, Edworthy J, Hellier E, Seger DL, Schedlbauer A, Avery AJ, et al. A review of human factors principles for the design and implementation of medication safety alerts in clinical information systems. J Am Med Inform Assoc. 2010;17: 493–501. doi:10.1136/jamia.2010.005264

103. Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E. Systematic review automation technologies. Syst Rev. BioMed Central Ltd; 2014;3: 74. doi:10.1186/2046-4053-3-74

104. Kawamoto K, Del Fiol G, Lobach DF, Jenders R a. Standards for scalable clinical decision support: need, current and emerging standards, gaps, and proposal for progress. Open Med Inform J. 2010;4: 235–44. doi:10.2174/1874431101004010235

105. Karacapilidis N. An Overview of Future Challenges of Decision Support Technologies. Intelligent Decision-making Support Systems SE - 20. Springer

London; 2006. pp. 385–399. doi:10.1007/1-84628-231-4_20

106. Van Spall HGC, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. JAMA. American Medical Association; 2007;297: 1233–40. doi:10.1001/jama.297.11.1233

107. Vogt-Ferrier N. Older patients, multiple comorbidities, polymedication… should we treat everything? Eur Geriatr Med. 2011;2: 48–51. doi:10.1016/j.eurger.2010.11.011

108. Buchan I, Winn J, Bishop CM. A Unified Modeling Approach to Data-Intensive Healthcare. The fourth paradigm. 2009. pp. 91–97. Available: http://www.webcitation.org/6NYqlYcap

2.7 SUPPLEMENTARY MATERIAL

Supplementary Table 2.1: Search terms and Boolean logic applied in the Scopus search and automatic annotation on the results from the PubMed query.

Search terms and boolean logic
(“clinical decision support” OR “clinical decision support review” OR “clinical decision support system” OR “clinical decision support systems” OR “clinical decision support tool” OR “clinical decision support tools” OR “computer decision support” OR “computerised clinical decision support” OR “computerised clinical decision support system” OR “computerised clinical decision support systems” OR “computerised decision support” OR “computerized clinical decision support” OR “computerized clinical decision support system” OR “computerized clinical decision support systems” OR “computerized decision support” OR “decision support” OR “decision support software” OR “decision support system” OR “decision support systems” OR “decision support tool” OR “decision support tools” OR “electronic decision support” OR “medical decision support” OR “safety surveillance” OR “surveillance safety” OR “system decision support” OR “alert” OR “alerts”) AND (("co-occurrence condition" OR "co-occurrence conditions" OR "co-occurrence condition" OR "co-occurrence conditions" OR "co-occurrence disease" OR "co-occurrence diseases" OR "co-occurrence disorder" OR "co-occurrence disorders" OR "co-occurrence illness" OR "co-occurrence pathologies" OR "co-occurrence pathology” OR “co-occurring condition” OR “co-occurring conditions” OR “co-occurring disease” OR “co-occurring diseases” OR “co-occurring disorder” OR

“co-occurring disorders” OR
“co-occurring illness” OR
“co-occurring pathologies” OR
“co-occurring pathology” OR
“comorbidities” OR
“comorbidity” OR
“concurrent condition” OR
“concurrent conditions” OR
“concurrent disease” OR
“concurrent diseases” OR
“concurrent disorder” OR
“concurrent disorders” OR
“concurrent illness” OR
“concurrent pathologies” OR
“concurrent pathology” OR
“cooccurence condition” OR
“cooccurence conditions” OR
“cooccurence disease” OR
“cooccurence diseases” OR
“cooccurence disorder” OR
“cooccurence disorders” OR
“cooccurence illness” OR
“cooccurence pathologies” OR
“cooccurence pathology” OR
“cooccuring condition” OR
“cooccuring conditions” OR
“cooccuring disease” OR
“cooccuring diseases” OR
“cooccuring disorder” OR
“cooccuring disorders” OR
“cooccuring illness” OR
“cooccuring pathologies” OR
“cooccuring pathology” OR
“cumulative condition” OR
“cumulative conditions” OR
“cumulative disease” OR
“cumulative diseases” OR
“cumulative disorder” OR
“cumulative disorders” OR
“cumulative illness” OR
“cumulative pathologies” OR
“cumulative pathology” OR
“multi-morbidities” OR
“multi-morbidity” OR
“multimorbidities” OR
“multimorbidity” OR
“multiple acute” OR
“multiple chronic” OR
“multiple chronic condition” OR
“multiple chronic conditions” OR
“multiple chronic disease” OR
“multiple chronic diseases” OR

“multiple chronic disorder” OR
“multiple chronic disorders” OR
“multiple chronic illness” OR
“multiple chronic pathologies” OR
“multiple chronic pathology” OR
“multiple chronical” OR
“multiple condition” OR
“multiple conditions” OR
“multiple disease” OR
“multiple diseases” OR
“multiple disorder” OR
“multiple disorders” OR
“multiple illness” OR
“multiple pathologies” OR
“multiple pathology” OR
“polymorbidities” OR
“polymorbidity” OR
“polypathologies” OR
“polypathology” OR
“polypathy”)
OR
(“many medication” OR
“many medications” OR
“many pharmaceutical preparation” OR
“many pharmaceutical preparations” OR
“many pharmacologic substance” OR
“many pharmacologic substances” OR
“many prescription” OR
“many prescriptions” OR
“multiple medication” OR
“multiple medications” OR
“multiple pharmaceutical preparation” OR
“multiple pharmaceutical preparations” OR
“multiple pharmacologic substance” OR
“multiple pharmacologic substances” OR
“multiple prescription” OR
“multiple prescriptions” OR
“poly-pharmacies” OR
“poly-pharmacy” OR
“poly-prescription” OR
“poly-prescriptions” OR
“polypharmacies” OR
“Polypharmacy” OR
“several medication” OR
“several medications” OR
“several pharmaceutical preparation” OR
“several pharmaceutical preparations” OR
“several pharmacologic substance” OR
“several pharmacologic substances” OR
“several prescription” OR
“several prescriptions”))

Chapter 3

PREDICTING MORTALITY FROM CHANGE-OVER-TIME IN THE CHARLSON COMORBIDITY INDEX: A RETROSPECTIVE COHORT STUDY IN A DATA-INTENSIVE UK HEALTH SYSTEM

Paolo Fraccaro [1,2], Evangelos Kontopantelis [1,6], Matthew Sperrin [1], Niels Peek [1,2], Christian Mallen [3], Philip Urban [4], Iain E. Buchan [1,2], Mamas A. Mamas [1,5]

[1] Health eResearch Centre, Farr Institute for Health Informatics Research, University of Manchester, UK.

[2] NIHR Greater Manchester Primary Care Patient Safety Translational Research Centre, Institute of Population Health, The University of Manchester, Manchester, United Kingdom.

[3] Research Institute for Primary Care & Health Sciences, Arthritis Research UK Primary Care Centre, Keele University, Keele, Staffordshire, UK.

[4] Cardiovascular Department, Hôpital de La Tour, Geneva, Switzerland.

[5] Keele Cardiovascular Research Group, Keele University Stoke-on-Trent and Royal Stoke Hospital, University Hospital North Midlands, Stoke-on-Trent, UK.

[6]NIHR School for Primary Care Research, University of Manchester, Manchester, UK.

Contribution: PF (corresponding author), MS, EK, NP and MM designed the study. PF extracted the data from all sources and performed the analyses. PF, MS, EK, NP and MM wrote the manuscript. CM, PU and IB critically edited the manuscript. PF and EK replied to the reviewers' comments during the review process.

Based on: Fraccaro P, Kontopantelis E, Sperrin M, Peek N, Mallen C, et al. (2016) Predicting mortality from change-over-time in the Charlson Comorbidity Index: A retrospective cohort study in a data-intensive UK health system. Medicine (Baltimore) 95. doi:10.1097/MD.0000000000004973.

3.1 ABSTRACT

Background

Multimorbidity is common among older people and presents a major challenge to health systems worldwide. Metrics of multimorbidity are, however, crude: focusing on measuring comorbid conditions at single time-points rather than reflecting the longitudinal and additive nature of chronic conditions. In this paper, we explore longitudinal comorbidity metrics and their value in predicting mortality.

Methods

Using linked primary and secondary care data, we conducted a retrospective cohort study on adults in Salford, UK from 2005-2014 (n=287,459). We measured multimorbidity with the Charlson Comorbidity Index (CCI) and quantified its changes in various time windows. We used survival models to assess the relationship between CCI changes and mortality, controlling for gender, age, baseline CCI and time-dependent CCI. Goodness-of-fit was assessed with the Akaike Information Criterion and discrimination with the c-statistic.

Results

Overall, 15.9% patients experienced a change in CCI after 10 years, with a mortality rate of 19.8%. The model that included gender and time-dependent age, CCI, and CCI change across consecutive time windows had the best fit to the data but equivalent discrimination to the other time-dependent models. The absolute CCI score gave a constant hazard ratio (HR) of around 1.3 per unit increase, while CCI *change* afforded greater prognostic impact, particularly when it occurred in shorter time windows (maximum HR value for the 3-month time window, with 1.63 and 95% confidence interval 1.59-1.66).

Conclusions

Change over time in comorbidity is an important but overlooked predictor of mortality, which should be considered in research and care quality management.

Keywords: Charlson comorbidity index; survival analysis; comorbidity; multimorbidity; prognostic impact; Salford Integrated Record; retrospective cohort study; risk stratification.

3.2 INTRODUCTION

In 'ageing' populations the prevalence of patients with multiple conditions increases [1,2] placing extra demands on healthcare systems [3,4]. Population-based studies have revealed the presence of at least one long-term condition in over a third of patients [2,5], with two thirds of those aged over 65 years and three quarters of those aged over 85 years having at least two concurring conditions [6].

Linked electronic health records (EHRs) may offer new information about multimorbidity [7]. Some EHRs hold comorbidity scores [8], ranging from simple summation of the number of conditions to more complex scores that assign different weights to diseases in respect of their prognoses [9–13]. Although EHRs can provide rich longitudinal information most studies use the data available at a single time-point to measure comorbidity, which treats it as a static phenomenon when it is logically dynamic [14–16]. Similarly in prognostic studies, only those comorbid conditions present at baseline are commonly considered, while new conditions arising may affect the outcome of interest [14–16]. While it is reasonable to hypothesise that those with rising comorbidity over time may have worse health outcomes [15] this group of patients are poorly characterised in the literature.

This study aimed to characterise the distribution, and changes over time, of comorbidities, as measured by the Charlson Comorbidity Index (CCI) [13], in a UK population with high-quality EHRs. We also sought to investigate different ways to account for longitudinal patterns of comorbidity in survival analyses and see if this enhanced the prediction of mortality.

3.3 METHODS

3.3.1 Data source

Data were extracted from the Salford Integrated Record (SIR) – an anonymised extract of linked data from all 53 primary care providers and one secondary care provider in the UK City of Salford (population in Census 2011 of ~235k [17]). The data in SIR includes all primary care and secondary care records (i.e. focused on long-term conditions management) as well as all results from biochemical testing across primary and secondary care. Data are stored as Read codes v2 and v3 [18].

Salford is a relatively deprived area, with almost a third of neighbourhoods in the most deprived tenth for England [19]. In terms of multimorbidity burden, Salford is in the 61st

centile, as measured by England's primary care Quality and Outcomes Framework (QOF) [20].

3.3.2 Study period and population

The study period was from 1 April 2005 to 31 December 2014. As QOF has been proven to influence general practitioners data recording behaviours and improve data quality on included clinical conditions [21–24], we focused on the period after QOF was introduced and used its financial years (1 April to 31 March). We used an open cohort design and included all patients aged 18 years or older, registered in one of the SIR primary care practices. Patients were considered as participating in the study until death or migration out of the area.

3.3.3 Comorbidity burden measurement: Charlson comorbidity index calculation

We measured comorbidity burden by using the CCI [13] – a widely-used score [8], which has different weights for 22 clinical conditions in relation to their impact on prognosis. Although originally developed to predict mortality risk after hospitalisation, it has been shown to independently predict adverse outcomes across a broad spectrum of conditions [25–35].

We calculated the CCI on the basis of the work of Khan and colleagues [36], who provided a list of validated Read diagnostic codes for calculating it in UK primary care. Every time a relevant Read diagnostic code was found for a patient, the CCI was updated using the weights for the related disease category. Age was modelled separately and not included in the CCI calculation.

Although part of the original version of the CCI, we were not able to include HIV/AIDS and dementia in our study due to privacy restrictions on access to data about sexual or mental health illness in the SIR.

In addition to the original CCI definition, we stratified the disease categories into cardiovascular (i.e. myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, diabetes mellitus, renal disease) and non-cardiovascular (i.e. peptic ulcer disease, cancer, metastatic disease, hemiplegia, liver disease, Chronic pulmonary disease) diseases. We then repeated the process explained above and obtained two individual scores (cardiovascular CCI and non-cardiovascular CCI).

3.3.4 Data analysis

To investigate the proportion of patients experiencing changes in comorbidities during follow-up, we calculated the difference between patient CCI values at baseline, then at one, five and

10 years. For each follow-up period we next calculated the overall proportion of patients that had a CCI change and their mortality rates. We repeated this analysis by stratifying for the CCI value at baseline (i.e. 0, 1, 2, >=3) and reported separately proportion of change and crude mortality rate for CCI changes of 0, 1, 2, and >=3.

To evaluate the prognostic importance of comorbidity burden changes over time and the time period over which changes occur, we performed survival analyses using Cox regression models [37] with time to death from any cause as the outcome. We built three different datasets by discretising time into 3-, 6- and 12-month time windows (see Supplementary Figure 3.1 and Supplementary Table 3.1) and implemented different models by increasing the level of model's complexity. The models considered:

1. Age, gender and CCI at baseline (*model 1*).
2. Gender and time-dependent age and CCI (*model 2*).
3. Gender and CCI at baseline as well as time-dependent age and CCI (*model 3*).
4. Gender and baseline CCI value in addition to time-dependent age and cumulative CCI change from baseline (*model 4*).
5. Gender and time-dependent age, CCI and CCI change over consecutive time windows (*model 5*).

For both the non-stratified and cardiovascular stratified analyses, time-dependent covariates were modelled by updating their values at the beginning of each time window (see Supplementary Table 3.1).

We used the Akaike Information Criterion (AIC) to assess model goodness-of-fit [38]. For each model, we also assessed discrimination with 95% confidence intervals for the c-statistic by calculating c-index over100 bootstrap iterations. Finally, we calculated models' Variance Inflation Factors (VIF), which assesses collinearity between covariates, and checked the proportional hazards assumption.

3.3.5 Sensitivity analyses

We performed several sensitivity analyses. First, we evaluated possible clustering effects related to the different primary care practices from which the data arose by repeating our main analysis with the addition of a random intercept at practice level. Second, since the currency of the original CCI disease weights is under debate, we repeated all analyses with an updated version of the CCI [39]. Thirdly, we only considered the patients that experienced a change in CCI during the follow-up. Fourthly, we repeated all analyses by categorising both CCI and CCI change as 0, 1, 2, >=3 and assessing interaction terms between CCI value and CCI change.

Finally, as c-statistic to compare different prediction models has been criticised [40,41], we also compared the simplest (model 1) and most complex (model 5) of the models we tested in terms of Net Reclassification Improvement (NRI) and Integrated Discrimination Improvement (IDI) [40,41] to quantify differences in predictive ability. We based our analysis on Wong at al [42] who calculated IDI and NRI to compare a time-fixed and time-dependent model in a survival analysis. Particularly, we calculated IDI as the difference between the mean predicted risk in patients who died and patients who did not die for both models. As in the context of our analysis there are no clear risk categories to which patients are assigned, we implemented a category-less NRI and calculated the proportion of the correct (i.e. model 5 predicted higher risk than model 1 for patients who died) minus incorrect predictions in the events plus the proportions of correct (i.e. model 5 predicted lower risk than model 1 for patients who did not die) minus incorrect predictions for non-events. For both IDI and NRI values above 0 indicate better performance. We calculated 95% Confidence Intervals for both IDI and NRI for each time point over 100 bootstrap iterations.

3.4 RESULTS

3.4.1 Study population characteristics

A total of 357,829 patients were recorded in the SIR database during the study period. We excluded 65,182 patients because they were under the age of 18 years and 5,188 patients because of conflicting registration data (such as temporary residents). A total of 287,459 patients were included in the analysis, with a mortality rate of 5.7% (N=16,452) recorded during the study period. Table 3.1 shows patient characteristics at baseline. The proportion of women was 49.3% and mean age at baseline was 38.3 years (Standard deviation [SD] 18.8), with a mean follow-up time of 7.9 years (SD 2.8). Mean deprivation as measured by the Townsend score[43] which incorporates four variables (i.e. unemployment, non-car ownership, non-home ownership, and household overcrowding) to calculate material deprivation within a population, was 1.9 (SD 3.4). The majority of patients were Caucasian (85.8%). Mean body mass index (BMI) at baseline was 25.8 kg/m^2 (SD 5.7). The prevalence of CCI disease categories at baseline varied from 0.1% to 13.2%, with chronic pulmonary disease having the highest prevalence, followed by diabetes (3.5%). Prevalence rates for cancer, cerebrovascular disease, myocardial infarction, peptic ulcer disease and musculoskeletal disease varied between 1% and 2%, whilst for all other comorbidities prevalence rates were below 1%.

Table 3.1: Patient characteristics at baseline. Abbreviation: CCI, Charlson comorbidity index; N/A, Not applicable.

Parameters		**Values at entry date**	
			Missing
Included patients		287,459	None
Patients dead during study period		16,452 (5.7)	None
Follow-up (mean,SD)		7.9 (2.8)	None
Patients with increase in CCI during study period (%)		45,779 (15.9)	None
Age (mean,SD)		38.3 (18.8)	None
Female (%)		141,794 (49.3)	None
Townsend index (mean,SD)		1.9 (3.4)	None
Ethnicity	White	144,011 (85.8)	119,665 (41.6)
	Indian	2,698 (1.6)	
	Pakistani	1,775 (1.1)	
	Bangladeshi	379 (0.2)	
	Other Asian	2,152 (1.3)	
	Black Caribbean	318 (0.2)	
	Black African	5,461 (3.3)	
	Chinese	2,180 (1.3)	
	Other	8,820 (5.3)	
Smoking	Non-smoker	99,303 (55.4)	108,285 (37.7)
	Ex-smoker	32,378 (18.1)	
	Light smoker [1-9 cg/day]	13,938 (7.8)	
	Moderate smoker [10-19 cg/day]	19,458 (10.9)	
	Heavy smoker [>=20 cg/day]	14,097 (7.9)	
BMI [kg/m^2] (mean,SD)		25.8 (5.7)	137,931 (48)
Cancer (%)		5,878 (2)	N/A
Cerebrovascular disease (%)		4,449 (1.5)	N/A
Chronic pulmonary disease (%)		37,842 (13.2)	N/A
Congestive heart disease (%)		2,311 (0.8)	N/A
Diabetes (%)		9,152 (3.2)	N/A
Diabetes with complications (%)		1,116 (0.4)	N/A
Hemiplegia (%)		473 (0.2)	N/A
Metastatic tumour (%)		122 (0)	N/A
Mild liver disease (%)		370 (0.1)	N/A
Mod liver disease (%)		149 (0.1)	N/A
Myocardial infarction (%)		4,277 (1.5)	N/A
Peptic ulcer disease (%)		3,449 (1.2)	N/A
Peripheral vascular disease (%)		2,126 (0.7)	N/A
Renal disease (%)		1,827 (0.6)	N/A
Rheumatological disease (%)		2,886 (1)	N/A

Table 3.2 reports trends over time of prevalence of the CCI diseases categories (see Supplementary Figure 3.2 for graphical representation). Prevalence rates for cancer, chronic pulmonary disease, and diabetes increased during the study period, while they decreased for myocardial infarction, peptic ulcer and musculoskeletal disease. Renal disease prevalence peaked in financial years 2009/10 and 2010/11 and then slightly decreased. All the other disease categories remained stable.

Table 3.2: Charlson comorbidity index disease categories trend over the study period (on QOF financial years, such as 1st of April to 31st March of the next year) in terms of number of patients affected and prevalence.

Diseases	Year 2005 (N=199,043)	Year 2006 (N=206,392)	Year 2007 (N=214,911)	Year 2008 (N=226,001)	Year 2009 (N=234,904)	Year 2010 (N=243,104)	Year 2011 (N=250,958)	Year 2012 (N=257,777)	Year 2013 (N=264,369)	Year 2014 (N=271,007)
Cancer (%)	6,289 (3.2)	6,868 (3.3)	7,525 (3.5)	8,416 (3.7)	8,923 (3.8)	9,396 (3.9)	9,813 (3.9)	10,194 (4)	10,479 (4)	10,682 (3.9)
Cerebrovascular disease (%)	4,612 (2.3)	4,817 (2.3)	5,059 (2.4)	5,361 (2.4)	5,615 (2.4)	5,823 (2.4)	5,966 (2.4)	6,028 (2.3)	6,046 (2.3)	6,131 (2.3)
Chronic pulmonary disease (%)	29,319 (14.7)	31,069 (15.1)	32,827 (15.3)	34,929 (15.5)	36,618 (15.6)	38,164 (15.7)	39,520 (15.7)	40,766 (15.8)	41,685 (15.8)	42,544 (15.7)
Congestive heart disease (%)	2,466 (1.2)	2,526 (1.2)	2,748 (1.3)	3,097 (1.4)	3,192 (1.4)	3,331 (1.4)	3,415 (1.4)	3,448 (1.3)	3,473 (1.3)	3,496 (1.3)
Diabetes (%)	9,095 (4.6)	9,669 (4.7)	10,376 (4.8)	11,376 (5)	12,024 (5.1)	12,652 (5.2)	13,133 (5.2)	13,600 (5.3)	14,046 (5.3)	14,244 (5.3)
Diabetes with complications (%)	1,277 (0.6)	1,472 (0.7)	1,623 (0.8)	1,759 (0.8)	1,897 (0.8)	2,123 (0.9)	2,396 (1)	2,513 (1)	2,680 (1)	2,772 (1)
Hemiplegia (%)	432 (0.2)	459 (0.2)	472 (0.2)	488 (0.2)	499 (0.2)	516 (0.2)	530 (0.2)	555 (0.2)	563 (0.2)	571 (0.2)
Metastatic tumour (%)	133 (0.1)	150 (0.1)	170 (0.1)	204 (0.1)	233 (0.1)	247 (0.1)	273 (0.1)	304 (0.1)	303 (0.1)	305 (0.1)
Mild liver disease (%)	362 (0.2)	394 (0.2)	430 (0.2)	497 (0.2)	552 (0.2)	624 (0.3)	684 (0.3)	758 (0.3)	811 (0.3)	865 (0.3)
Mod liver disease (%)	159 (0.1)	166 (0.1)	191 (0.1)	221 (0.1)	234 (0.1)	262 (0.1)	279 (0.1)	304 (0.1)	313 (0.1)	319 (0.1)
Myocardial infarction (%)	4,410 (2.2)	4,533 (2.2)	4,691 (2.2)	4,928 (2.2)	4,986 (2.1)	5,047 (2.1)	5,102 (2)	5,145 (2)	5,170 (2)	5,218 (1.9)
Peptic ulcer disease (%)	3,490 (1.8)	3,601 (1.7)	3,737 (1.7)	3,891 (1.7)	3,940 (1.7)	3,949 (1.6)	3,957 (1.6)	3,959 (1.5)	3,913 (1.5)	3,883 (1.4)
Peripheral vascular disease (%)	2,259 (1.1)	2,364 (1.1)	2,517 (1.2)	2,673 (1.2)	2,749 (1.2)	2,790 (1.1)	2,882 (1.1)	2,954 (1.1)	3,048 (1.2)	3,076 (1.1)
Renal disease (%)	3,005 (1.5)	8,976 (4.3)	10,983 (5.1)	11,714 (5.2)	12,306 (5.2)	12,388 (5.1)	12,340 (4.9)	12,318 (4.8)	11,965 (4.5)	11,781 (4.3)
Rheumatological disease (%)	2,906 (1.5)	2,972 (1.4)	3,043 (1.4)	3,197 (1.4)	3,273 (1.4)	3,340 (1.4)	3,401 (1.4)	3,435 (1.3)	3,456 (1.3)	3,498 (1.3)

3.4.2 Comorbidities change and mortality

Over the study period, we observed a change in CCI at 1 year for 5,533 (1.9%) patients, with a crude mortality rate documented within this group of 3.1%. The number of patients for whom we observed CCI changes after five and 10 years were 30,025 (10.4%) and 45,096 (15.9%), with a respective crude mortality rate of 10.0% and 19.8%. When comparing mortality between the group of patients that had a change in CCI and those that did not we found odds ratios of 8.8 (95% confidence interval [CI] 7.5-10.4), 6.6 (95% CI 6.3-6.9), and 7.8 (95% CI 7.5-8.0) at the three time points, respectively.

Table 3.3 reports the mortality odds ratios associated with a change in CCI of 1, 2, and equal or more than 3 units, respectively (see Supplementary Figure 3.3, Supplementary Figure 3.4 and Supplementary Figure 3.5 for details about prevalence of CCI change and related mortality). Overall, the odds ratios increased for bigger CCI changes and decreased for longer follow-up times and higher baseline CCI values. All comparisons were statistically significant (P-values lower than 0.05), with the exception of some analyses for baseline CCI 2 and 3.

Table 3.3: Odds ratio of mortality for group of patients that had a change in Charlson Comorbidity Index (CCI) and the patients that did not have it for different baseline CCI values across the study.***Value not statistically significant at a 0.05 level.**

Baseline CCI	CCI change	1-year follow-up Odds ratio [95% CI]	5-year follow-up Odds ratio [95% CI]	10-year follow-up Odds ratio [95% CI]
0	1	6.35 [4.33,9.32]	3.58 [3.16,4.05]	3.68 [3.41,3.97]
	2	18.18 [13.09,25.25]	10.34 [9.42,11.34]	11.31 [10.64,12.02]
	>=3	70.12 [36.98,132.95]	17.58 [15.45,20.01]	19.33 [17.98,20.78]
1	1	7.59 [4.95,11.65]	2.88 [2.43,3.42]	3.24 [2.91,3.61]
	2	10.68 [7.1,16.07]	5.82 [5.17,6.55]	7.26 [6.67,7.91]
	>=3	12.73 [3.94,41.18]	8.42 [7.24,9.78]	9.34 [8.51,10.25]
2	1	0.59***** [0.19,1.87]	1.12***** [0.89,1.39]	1.24 [1.07,1.44]
	2	1.91 [0.96,3.77]	1.69 [1.44,1.98]	2.19 [1.94,2.48]
	>=3	7.93 [2.75,22.86]	2.24 [1.83,2.74]	2.85 [2.5,3.24]
>=3	1	1.69***** [0.92,3.08]	1.09***** [0.86,1.37]	1.29 [1.09,1.54]
	2	1.27***** [0.62,2.63]	1.22 [1.03,1.45]	1.96 [1.7,2.27]
	>=3	5.54 [1.87,16.4]	1.37 [1.08,1.74]	1.88 [1.6,2.21]

3.4.3 Regression analyses

Table 3.4 summarises covariates prognostic impact (per unit increase), AIC and c-statistic for the 6-month time window analysis. These are reported separately for the non-stratified and cardiovascular-stratified analyses.

We observed the same prognostic impact (HR 1.50 95% CI 1.49-1.51) in model 3 for the time-dependent CCI value and model 4 for the CCI cumulative change over study period, which in both cases was much greater than the baseline CCI value (HR 0.81 95% CI 0.80-0.82 in model 3 and HR 1.21 95% CI 1.20-1.22 in model 4). In addition, it can be seen that longitudinal changes in CCI provide additional prognostic information (HR 1.51 95% CI 1.48-1.54) to that provided by the absolute CCI score (HR 1.30 95% CI 1.29-1.31) also when looking at changes across different time windows (i.e. model 5). Interestingly, longitudinal changes in the non-cardiovascular components of CCI provide a much greater hazard for mortality (HR 1.66 95% CI 1.62-1.70) than the cardiovascular components (HR 1.17 95% CI 1.12-1.22).

Increasing the model complexity from model 1 to model 5 led to a better fit of the models to the data, as witnessed by a decrease in AIC, but not a substantial improvement in the c-statistic. Model 3 and model 4 were equivalent in terms of goodness-of-fit (i.e. same AIC value). VIF values were lower than 2 for all included variables across all models, showing no indication of collinearity between the covariates. Finally, there was no evidence to reject the hypothesis of proportional hazards.

Results from the analyses with 3- and 12-month time windows showed similar findings to those undertaken with 6-month time windows.

Table 3.5 shows hazard ratios for model 5, such as the model that obtained the best AIC values, across all the three different time windows (i.e. 12-, 6-, and 3-month time windows). Looking at the non-stratified analysis, CCI (per unit increase) had similar prognostic impact across all analyses, whilst longitudinal changes in CCI hazard ratio (per unit increase) augmented with shorter time windows. The cardiovascular-stratified analysis showed similar figures, whilst the non-cardiovascular CCI score had bigger prognostic impact in shorter time periods.

Table 3.4: Results for the 6-month time windows in terms of AIC and hazard ratios. Abbreviation: CCI, Charlson comorbidity index; CI, confidence interval.

	Model	AIC	c-statistic 95% CI	Variable	Hazard Ratio [95% CI]
NON-STRATIFIED ANALYSIS	1	362,230	0.90 - 0.90	Baseline age (per year)	1.08 [1.08,1.09]
				Gender (M vs F)	1.26 [1.22,1.30]
				Baseline CCI (per unit)	1.23 [1.21,1.24]
	2	358,054	0.91 - 0.91	Age (per year)	1.08 [1.08,1.08]
				Gender (M vs F)	1.21 [1.18,1.25]
				CCI (per unit)	1.35 [1.34,1.36]
	3	357,290	0.91 - 0.91	Age (per year)	1.08 [1.08,1.08]
				Gender (M vs F)	1.24 [1.20,1.28]
				Baseline CCI (per unit)	0.81 [0.80,0.82]
				CCI (per unit)	1.50 [1.49,1.51]
	4	357,290	0.91 - 0.91	Age (per year)	1.08 [1.08,1.08]
				Gender (M vs F)	1.24 [1.20,1.28]
				Baseline CCI (per unit)	1.21 [1.20,1.22]
				CCI cumulative change (per unit)	1.50 [1.49,1.51]
	5	357,000	0.91 - 0.91	Age (per year)	1.08 [1.08,1.08]
				Gender (M vs F)	1.22 [1.20,1.26]
				CCI (per unit)	1.30 [1.29,1.31]
				CCI change (per unit)	1.51 [1.48,1.54]
CARDIOVASCULAR STRTIFIED ANALYSIS	1	362,169	0.90 - 0.90	Baseline age (per year)	1.08 [1.08,1.09]
				Gender (M vs F)	1.25 [1.21,1.29]
				Baseline cardiovascular CCI (per unit)	1.29 [1.27,1.31]
				Baseline non-cardiovascular CCI (per unit)	1.16 [1.15,1.18]
	2	357,952	0.91 - 0.91	Age (per year)	1.08 [1.08,1.08]
				Gender (M vs F)	1.22 [1.18,1.26]
				Cardiovascular CCI (per unit)	1.30 [1.29,1.31]
				Non-cardiovascular CCI (per unit)	1.42 [1.40,1.43]
	3	356,818	0.91 - 0.91	Age (per year)	1.08 [1.08,1.08]
				Gender (M vs F)	1.22 [1.180,1.26]
				Baseline cardiovascular CCI (per unit)	0.93 [0.91,0.95]
				Cardiovascular CCI (per unit)	1.34 [1.32,1.36]
				Baseline non-cardiovascular CCI (per unit)	0.71 [0.70,0.72]
				Non-cardiovascular CCI (per unit)	1.69 [1.66,1.71]
	4	356,818	0.91 - 0.91	Age (per year)	1.08 [1.08,1.08]
				Gender (M vs F)	1.22 [1.18,1.26]
				Baseline cardiovascular CCI (per unit)	1.24 [1.22,1.26]
				Cardiovascular CCI cumulative change (per unit)	1.33 [1.32,1.36]
				Baseline non-cardiovascular CCI index (per unit)	1.20 [1.18,1.22]
				Non-cardiovascular CCI cumulative change (per unit)	1.69 [1.67,1.71]
	5	356,735	0.91 - 0.91	Age (per year)	1.08 [1.08,1.08]
				Gender (M vs F)	1.22 [1.18,1.26]
				Cardiovascular CCI (per unit)	1.29 [1.27,1.30]
				Cardiovascular CCI change (per unit)	1.17 [1.12,1.22]
				Non-cardiovascular CCI index (per unit)	1.32 [1.30,1.34]
				Non-cardiovascular CCI change (per unit)	1.66 [1.62,1.70]

Table 3.5: Hazard ratios for model 5 across the different time windows analyses. Abbreviation: CCI, Charlson comorbidity index.

		Model 5		
	Variable	**12-month time windows Hazard ratio [95% confidence interval]**	**6-month time windows Hazard ratio [95% confidence interval]**	**3-month time windows Hazard ratio [95% confidence interval]**
NON-STRATIFIED ANALYSIS	Age (per year)	1.08 [1.08,1.08]	1.08 [1.08,1.08]	1.08 [1.08,1.08]
	Gender (M vs F)	1.22 [1.20,1.26]	1.22 [1.20,1.26]	1.21 [1.18,1.25]
	CCI (per unit)	1.28 [1.26,1.29]	1.30 [1.30,1.31]	1.33 [1.32,1.34]
	CCI change (per unit)	1.41 [1.38,1.44]	1.51 [1.48,1.54]	1.63 [1.59,1.66]
CARDIOVASCULAR STRATIFIED ANALYSIS	Age (per year)	1.08 [1.08,1.08]	1.08 [1.08,1.08]	1.08 [1.08,1.08]
	Gender (M vs F)	1.21 [1.17,1.25]	1.22 [1.18,1.26]	1.21 [1.17,1.25]
	Cardiovascular CCI (per unit)	1.28 [1.270,1.30]	1.29 [1.28,1.30]	1.29 [1.28,1.30]
	Cardiovascular CCI change (per unit)	1.12 [1.09,1.16]	1.18 [1.12,1.22]	1.32 [1.25,1.39]
	Non-cardiovascular CCI (per unit)	1.26 [1.24,1.28]	1.32 [1.30,1.34]	1.38 [1.36,1.40]
	Non-cardiovascular CCI change (per unit)	1.60 [1.56,1.63]	1.66 [1.62,1.70]	1.68 [1.64,1.72]

The sensitivity analysis that focused only on patients that experienced CCI changes gave comparable results to the main analysis, with the exception of an increased difference in c-statistic, for which we observed values for the 95% CI ranging from 0.79-0.79 for model 1to 0.83-0.84 for model 5 in the 6-month time window analysis. The sensitivity analysis that compared model 1 and 5 in terms of IDI and NRI showed that for most time points the two models had the same predictive ability (see Supplementary Figure 3.6), with model 5 that was slightly better in IDI and model 1 in NRI. Sensitivity analyses, exploring variations by care provider, disease weighting schemes and score categorisations, showed similar results to our main analysis.

3.5 DISCUSSION

Our population-based analysis in over a quarter of a million people with detailed primary care records suggests that comorbidity is a dynamic process, with one in 10 patients showing a change in CCI over five years. This longitudinal pattern of comorbidity was associated with increased mortality risk, where change over time in CCI was a stronger predictor than CCI at baseline. In addition, the more rapid changes in CCI posed a greater mortality risk.

This study confirms that as populations 'age', a significant proportion of patients will experience comorbidity changes over time, and that longitudinal uses of metrics like CCI hold important prognostic information. Specifically, using Cox regression models with time-dependent CCI and including CCI-change provides additional prognostic information that should be considered when studying long-term outcomes in EHRs.

Regarding the choice of variables and how they are included in a regression model, we did not observe much variation in terms of discriminatory ability, with models 2 to 5 being almost equivalent. To increase the contrasts between models we restricted the discrimination comparison to the 15.9% of the population with comorbidity changes, however, the differences did not greatly increase. Similar predictive performance between models was also found in the IDI and NRI sensitivity analysis.

Since many of the comorbidities comprising the CCI score are heterogeneous, we stratified our analysis by two broad condition groups: with separate cardiovascular and non-cardiovascular variables. In predicting mortality: for the *cardiovascular CCI*, absolute score had a greater impact than its change over time, whilst for the *non-cardiovascular CCI* change over time was more predictive than absolute value. A possible explanation is that cardiovascular diseases are yoked in common pathophysiological mechanisms with similar

progression over time and share overlapping treatment strategies, hence longitudinal changes in the burden of these conditions is less likely to be as important when one already has an existing cardiovascular disorder. In contrast, non-cardiovascular CCI encompasses a heterogeneous group of diseases such as cancer, peptic ulcers, pulmonary pathology or liver diseases with separate pathophysiological mechanisms and treatment strategies whose prognostic impact is likely to be additive hence why dynamic changes in non-cardiovascular CCI has such important prognostic implications.

Although multimorbidity is often considered a static process [14–16], studies have analysed three main aspects of longitudinal changes in comorbidities [14–16,44–46]: finding trajectories of comorbidity evolution over time [16,44,46]; investigating the best way of longitudinally modelling comorbidities when making predictions [14,15]; and assessing if prognostic impact of comorbidities is temporary or persistent [45]. Comorbidities were mostly defined as counts of diseases [16,44–46], with CCI used in just two studies [14,15] and no comparisons made between the cardiovascular and non-cardiovascular components of CCI. Whilst CCI 'weights' clinical conditions by their prognostic impact, a simple count of co-morbidities would be confounded by the fact that different clinical conditions will impact on prognosis differently. Therefore, a change in the number of comorbid conditions will have very different prognostic implications depending on which conditions have changed.

To date, the studies of Aarts *et al.* [45] and Strauss *et al.* [16] are the only ones to report the numbers of patients experiencing change in comorbidities over time. Aarts *et al.* [45] followed 1,184 patients aged 24-81 for six years in a Dutch prospective study and reported that 16.4% with changes in comorbidities. Strauss *et al.* [16] studied 24,641 people aged >50 for three years in a UK primary care setting and reported 60% of these older patients had changes in comorbidities. Our study was most comparable with that of Aarts *et al.* [45] as we considered similar age groups, and our results for overall change over time in comorbidity were very similar.

Four studies have related advancing multimorbidity to worse health outcomes [14–16,45], and we extended the methodologies used. Aarts *et al.* [45] and Strauss *et al.* [16], used latent class analysis to identify different multimorbid trajectories in primary care data, and found worse self-reported health among patients with greater (especially steeper) changes in comorbidities. Zeng *et al*. [15] associated steeper CCI yearly change with worse general health in older (>65 years) patients with at least three comorbidities (N=~15,000) over a 10-year period. Finally, Wang and colleagues [14], reported much greater prognostic impact for time-dependent CCI levels compared to CCI values at baseline amongst a population of United States Medicare patients older than 65 years (N=50,000). Our Model 3 reflected that of Wang et al. and we found it fitted less well than the model that explicitly considered CCI changes over different

time windows, and identified baseline CCI value as a protective factor, which is counter-intuitive. These results suggest that the explicit inclusion of CCI changes allowed us to better capture and describe the complexity of comorbidity burden evolution over time.

Strauss *et al.* [16], Lappenschaar *et al.* [44] and Quiñones *et al.* [46] have also looked at defining longitudinal trajectories of comorbidity burden. Lappenschaar *et al.* [44] used a Bayesian network to find associations between diseases and health risks to predict evolution of comorbidities over time (e.g. diabetic retinopathy and hypertension). Quiñones *et al.* [46] estimated ethnicity-specific comorbidity trajectories for white Americans, black Americans and Mexicans. These studies did not consider the association between longitudinal comorbidity burden and outcome, they focused on deriving trajectories that predict how quickly patients encounter new comorbidities, which is similarly important.

To our knowledge, this is the first study to report the prognostic impact of comorbidity burden evolution, as measured by CCI, in a natural/geographical population, and to consider the discrete contributions of cardiovascular vs. non-cardiovascular conditions.

Yet our analysis has several limitations. Firstly, the SIR database relies on clinicians' observations and entry of relevant codes into EHRs, which may be an incomplete or inaccurate representation of patients' health. Most of the conditions in CCI, however, are recorded well in English primary care because they are part of the QOF pay-for-performance scheme. We observed a direct example of the effect of QOF in our dataset for renal disease. Particularly, prevalence in the Salford population raised from 1.5% in 2005 to 4.3% in 2006, and this is likely to be related to the introduction of renal disease in QOF during that year. Some cases will remain undiagnosed but this would be unusual given the serious nature of the CCI conditions, also dissent rates (i.e. refusal of assessment or treatment by the patient) in this context are low [47]. Secondly, like most other investigators we considered CCI as a 'rolling' measure, cumulating comorbidity burden until death. However, a limited number of the CCI disease categories (i.e. peptic ulcer or cancer) might be cured (a state not recorded in our EHR). Thirdly, due to data-reuse restrictions, we did not have information about two CCI disease categories: HIV and dementia. We note, however, that HIV and dementia prevalence in our population is 4 in 1,000 [48] and less than 1%, respectively [49]. Given that the included data cover most of the disease burden of the population and the principal determinants of their outcomes we do not feel that the inclusion of HIV and dementia data would lead to substantially different findings. Fourthly, general practices in Salford use two different EHRs, which can have a small influence over the data captured [50], but this is unlikely to be substantial given the incentivised data capture for the CCI conditions. Finally, although our analyses focused only on the city of Salford, our cohort was composed of ~280,000 patients

and the data were collected from all 53 primary care practices in Salford. It is true that almost a third of neighbourhoods in Salford are in the most deprived tenth of England. However, in terms of multimorbidity Salford is in the 61st centile. Therefore we expect that our results would be generalizable to other areas in England and UK. Finally, although important from a methodological perspective, our results are difficult to interpret from a clinical perspective, as after certain thresholds (e.g. CCI greater than 5) stratification of risk becomes ineffective and mortality risk in a primary care context is difficult to be translated into action. Future studies should repeat our analyses by using different outcomes, such as loss of function, disability, hospitalisation, and healthcare resource utilisation. These outcomes are more relevant to the primary care context and, in case of good model's performance, would provide useful information to clinicians in clinical practice.

Comorbidity burden is a dynamic process, with one in 10 patients in our study of British adults experiencing at least one change in comorbidity as measured by the CCI over a period of five years. Longitudinal models that include time-dependent CCI level and CCI change appear to be the most successful in capturing the effect of comorbidity burden on mortality and should be considered in survival analyses using EHR data – for research or for care quality management.

3.6 REFERENCES

1. Uijen AA, van de Lisdonk EH (2008) Multimorbidity in primary care: prevalence and trend over the last 20 years. Eur J Gen Pract 14 Suppl 1: 28–32. Available: http://www.webcitation.org/6NYkJQDj3. Accessed 13 January 2014.

2. Koné Pefoyo AJ, Bronskill SE, Gruneir A, Calzavara A, Thavorn K, et al. (2015) The increasing burden and complexity of multimorbidity. BMC Public Health 15: 415. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4415224/.

3. Wallace E, Salisbury C, Guthrie B, Lewis C, Fahey T, et al. (2015) Managing patients with multimorbidity in primary care. BMJ 350. Available: http://www.bmj.com/content/350/bmj.h176.abstract.

4. Boyd CM, Fortin M (2011) Future of multimorbidity research: How should understanding of multimorbidity inform health system design? Public Health Rev 33: 451–474. Available: http://www.webcitation.org/6NYkztx37.

5. van Oostrom SH, Picavet HSJ, van Gelder BM, Lemmens LC, Hoeymans N, et al. (2012) Multimorbidity and comorbidity in the Dutch population – data from general practices. BMC Public Health 12: 715. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3490727/.

6. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, et al. (2012) Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 380: 37–43. Available: http://www.webcitation.org/6NYk4ABPt. Accessed 22 January 2014.

7. Fraccaro P, Arguello Casteleiro M, Ainsworth J, Buchan I (2015) Adoption of Clinical Decision Support in Multimorbidity: A Systematic Review. JMIR Med Inf 3: e4. Available: http://medinform.jmir.org/2015/1/e4/.

8. Yurkovich M, Avina-Zubieta JA, Thomas J, Gorenchtein M, Lacaille D (2015) A systematic review identifies valid comorbidity indices derived from administrative health data. J Clin Epidemiol 68: 3–14. Available: http://www.sciencedirect.com/science/article/pii/S0895435614003576. Accessed 5 February 2015.

9. Clark DO, Korff M Von, Saunders K, Baluch WM, Simon GE (1995) A Chronic Disease Score with Empirically Derived Weights. Med Care 33: 783–795. Available: http://www.jstor.org/stable/3766468.

10. Elixhauser A, Steiner C, Harris DR, Coffey RM (1998) Comorbidity Measures for Use with Administrative Data. Med Care 36. Available: http://journals.lww.com/lww-medicalcare/Fulltext/1998/01000/Comorbidity_Measures_for_Use_with_Administrative.4.aspx.

11. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, et al. (2005) Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. Med Care 43. Available: http://journals.lww.com/lww-medicalcare/Fulltext/2005/11000/Coding_Algorithms_for_Defining_Comorbidities_in.10.aspx.

12. Von Korff M, Wagner EH, Saunders K (1992) A chronic disease score from automated pharmacy data. J Clin Epidemiol 45: 197–203. Available: http://www.sciencedirect.com/science/article/pii/089543569290016G.

13. Charlson ME, Pompei P, Ales K, Mackenzie C (1987) A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. J Chronic Dis 40: 373–383. doi:10.1016/0021-9681(87)90171-8.

14. Wang CY, Baldwin L-M, Saver BG, Dobie SA, Green PK, et al. (2009) The Contribution of Longitudinal Comorbidity Measurements to Survival Analysis. Med Care 47: 813–821. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2701975/.

15. Zeng C, Ellis JL, Steiner JF, Shoup JA, McQuillan DB, et al. (2014) Assessment of Morbidity Over Time in Predicting Health Outcomes. Med Care 52. Available: http://journals.lww.com/lww-medicalcare/Fulltext/2014/03001/Assessment_of_Morbidity_Over_Time_in_Predicting.11.aspx.

16. Strauss VY, Jones PW, Kadam UT, Jordan KP (2014) Distinct trajectories of multimorbidity in primary care were identified using latent class growth analysis. J Clin Epidemiol 67: 1163–1171. Available: http://www.sciencedirect.com/science/article/pii/S0895435614002182.

17. Office for National Statistics (UK) (2011) Census 2011. Available: https://www.salford.gov.uk/census2011.htm.

18. NHS England (n.d.) Read Codes. Available: http://www.connectingforhealth.nhs.uk/systemsandservices/data/uktc/readcodes. Accessed 16 June 2014.

19. Department for Communities and Local Government (2015) English indices of deprivation 2015. Available: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015.

20. Roland M (2004) Linking physicians' pay to the quality of care--a major experiment in the United kingdom. N Engl J Med 351: 1448–1454. Available: http://www.ncbi.nlm.nih.gov/pubmed/15459308. Accessed 13 November 2014.

21. Sutton M, Elder R, Guthrie B, Watt G (2010) Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers. Health Econ 19: 1–13. Available: http://dx.doi.org/10.1002/hec.1440.

22. Taggar JS, Coleman T, Lewis S, Szatkowski L (2012) The impact of the Quality and Outcomes Framework (QOF) on the recording of smoking targets in primary care medical records: cross-sectional analyses from The Health Improvement Network (THIN) database. BMC Public Health 12: 329. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4104830/.

23. Olier I, Springate DA, Ashcroft DM, Doran T, Reeves D, et al. (2016) Modelling Conditions and Health Care Processes in Electronic Health Records: An Application to Severe Mental Illness with the Clinical Practice Research Datalink. PLoS One 11: e0146715. Available: http://dx.doi.org/10.1371%2Fjournal.pone.0146715.

24. Kontopantelis E, Reeves D, Valderas JM, Campbell S, Doran T (2013) Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study. BMJ Qual Saf 22: 53–64. Available: http://qualitysafety.bmj.com/content/early/2012/08/21/bmjqs-2012-001033.full. Accessed 16 June 2014.

25. Reyes C, Estrada P, Nogués X, Orozco P, Cooper C, et al. (2014) The impact of common co-morbidities (as measured using the Charlson index) on hip fracture risk in elderly men: a population-based cohort study. Osteoporos Int 25: 1751–1758. Available: http://dx.doi.org/10.1007/s00198-014-2682-9.

26. Huang Y, Gou R, Diao Y, Yin Q, Fan W, et al. (2014) Charlson comorbidity index helps predict the risk of mortality for patients with type 2 diabetic nephropathy . J Zhejiang Univ Sci B 15: 58–66. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3891119/.

27. Ng ACC, Chow V, Yong ASC, Chung T, Kritharides L (2013) Prognostic Impact of

the Charlson Comorbidity Index on Mortality following Acute Pulmonary Embolism. Respiration 85: 408–416. Available: http://www.karger.com/DOI/10.1159/000342024.

28. Wu C-C, Hsu T-W, Chang C-M, Yu C-H, Lee C-C (2015) Age-Adjusted Charlson Comorbidity Index Scores as Predictor of Survival in Colorectal Cancer Patients Who Underwent Surgical Resection and Chemoradiation. Medicine (Baltimore) 94. Available: http://journals.lww.com/md-journal/Fulltext/2015/01020/Age_Adjusted_Charlson_Comorbidity_Index_Scores_as.24.aspx.

29. Dias-Santos D, Ferrone CR, Zheng H, Lillemoe KD, Fernández-del Castillo C (2015) The Charlson age comorbidity index predicts early mortality after surgery for pancreatic cancer. Surgery 157: 881–887. Available: http://dx.doi.org/10.1016/j.surg.2014.12.006.

30. Erickson SR, Cole E, Kline-Rogers E, Eagle KA (2013) The Addition of the Charlson Comorbidity Index to the GRACE Risk Prediction Index Improves Prediction of Outcomes in Acute Coronary Syndrome. Popul Health Manag 17: 54–59. Available: http://dx.doi.org/10.1089/pop.2012.0117.

31. Lu KJ, Kearney LG, Ord M, Jones E, Burrell LM, et al. (2015) Age adjusted Charlson Co-morbidity Index is an independent predictor of mortality over long-term follow-up in infective endocarditis. Int J Cardiol 168: 5243–5248. Available: http://dx.doi.org/10.1016/j.ijcard.2013.08.023.

32. Rattanasompattikul M, Feroze U, Molnar MZ, Dukkipati R, Kovesdy CP, et al. (2012) Charlson comorbidity score is a strong predictor of mortality in hemodialysis patients. Int Urol Nephrol 44: 1813–1823. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3595168/.

33. Grossman R, Mukherjee D, Chang DC, Bennett R, Brem H, et al. (2015) Preoperative Charlson Comorbidity Score Predicts Postoperative Outcomes Among Older Intracranial Meningioma Patients. World Neurosurg 75: 279–285. Available: http://dx.doi.org/10.1016/j.wneu.2010.09.003.

34. Schmidt M, Jacobsen JB, Lash TL, Bøtker HE, Sørensen HT (2012) 25 year trends in first time hospitalisation for acute myocardial infarction, subsequent short and long term mortality, and the prognostic impact of sex and comorbidity: a Danish nationwide cohort study. BMJ Br Med J 344: e356. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3266429/.

35. Daskivich TJ, Chamie K, Kwan L, Labo J, Dash A, et al. (2011) Comorbidity and competing risks for mortality in men with prostate cancer. Cancer 117: 4642–4650. Available: http://dx.doi.org/10.1002/cncr.26104.

36. Khan NF, Perera R, Harper S, Rose PW (2010) Adaptation and validation of the Charlson Index for Read/OXMIS coded databases. BMC Fam Pract 11: 1. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2820468&tool=pmcentrez&rendertype=abstract. Accessed 11 October 2013.

37. Cox DR (1972) Regression Models and Life-Tables. J R Stat Soc Ser B 34: 187–220. Available: http://www.jstor.org/stable/2985181.

38. Akaike H (1974) A new look at the statistical model identification. Autom Control IEEE Trans 19: 716–723. doi:10.1109/TAC.1974.1100705.

39. Quan H, Li B, Couris CM, Fushimi K, Graham P, et al. (2011) Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. Am J Epidemiol 173: 676–682. Available: http://aje.oxfordjournals.org/content/173/6/676.long. Accessed 20 March 2014.

40. Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS (2008) Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. Stat Med 27: 157–172. Available: http://dx.doi.org/10.1002/sim.2929.

41. Pencina MJ, Steyerberg EW, D'Agostino RB (2011) Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 30: 11–21. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3341973/.

42. Wong J, Taljaard M, Forster AJ, Escobar GJ, van Walraven C (2013) Addition of time-dependent covariates to a survival model significantly improved predictions for daily risk of hospital death. J Eval Clin Pract 19: 351–357. Available: http://dx.doi.org/10.1111/j.1365-2753.2012.01832.x.

43. Townsend P, Phillimore P, Beattie A (1988) Health and deprivation: inequality and the North. Routledge.

44. Lappenschaar M, Hommersom A, Lucas PJF, Lagro J, Visscher S, et al. (2015) Multilevel temporal Bayesian networks can model longitudinal change in multimorbidity. J Clin Epidemiol 66: 1405–1416. Available: http://dx.doi.org/10.1016/j.jclinepi.2013.06.018.

45. Aarts S, den Akker M van, Bosma H, Tan F, Verhey F, et al. (2015) The effect of

multimorbidity on health related functioning: Temporary or persistent? Results from a longitudinal cohort study. J Psychosom Res 73: 211–217. Available: http://dx.doi.org/10.1016/j.jpsychores.2012.05.014.

46. Quiñones AR, Liang J, Bennett JM, Xu X, Ye W (2011) How Does the Trajectory of Multimorbidity Vary Across Black, White, and Mexican Americans in Middle and Old Age? Journals Gerontol Ser B Psychol Sci Soc Sci 66B: 739–749. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3198247/.

47. Doran T, Kontopantelis E, Fullwood C, Lester H, Valderas JM, et al. (2012) Exempting dissenting patients from pay for performance schemes: retrospective analysis of exception reporting in the UK Quality and Outcomes Framework. BMJ 344. Available: http://www.bmj.com/content/344/bmj.e2405.abstract.

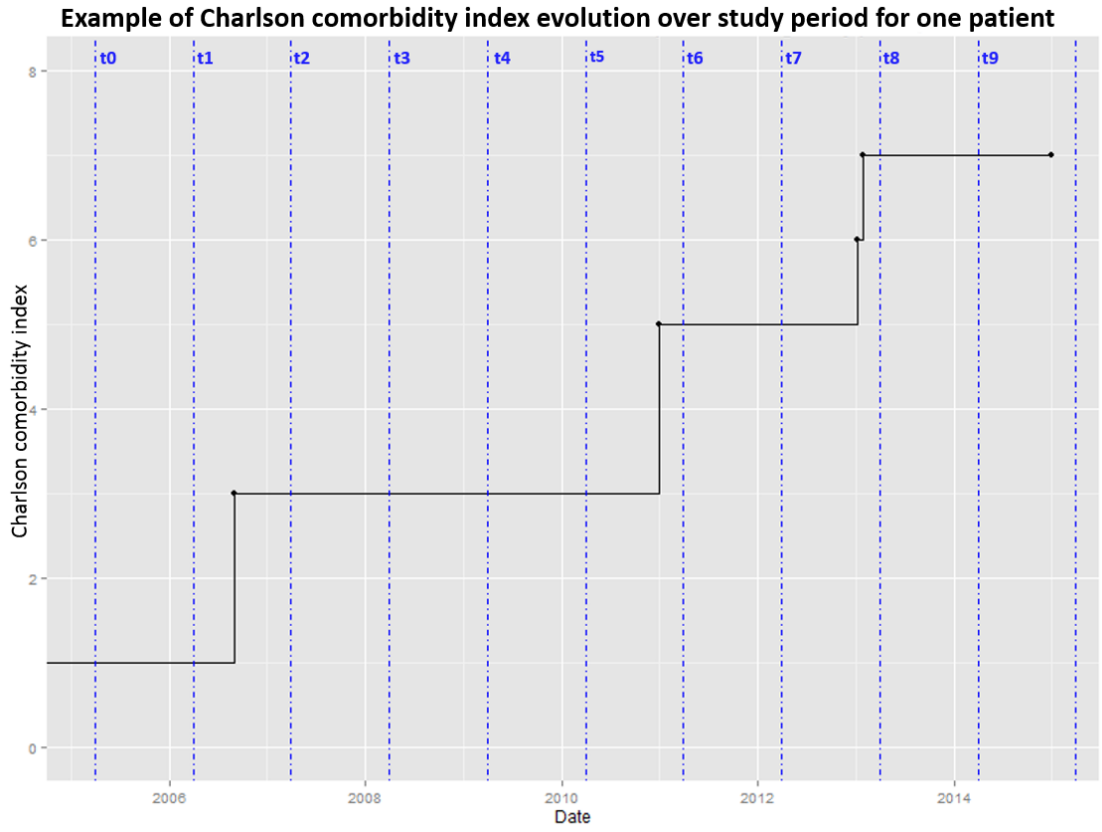
48. Salford City Partnership (2012) Sexual health in Salford. Available: http://www.partnersinsalford.org/2765.htm.

49. Health and Social Care Information Centre (2015) Quality and Outcomes Framework (QOF) - 2014-15. Available: http://www.hscic.gov.uk/catalogue/PUB18887.

50. Kontopantelis E, Buchan I, Reeves D, Checkland K, Doran T (2013) Relationship between quality of care and choice of clinical computing system: retrospective analysis of family practice performance under the UK's quality and outcomes framework. BMJ Open 3. Available: http://bmjopen.bmj.com/content/3/8/e003190.abstract.

3.7 SUPPLEMENTARY MATERIAL

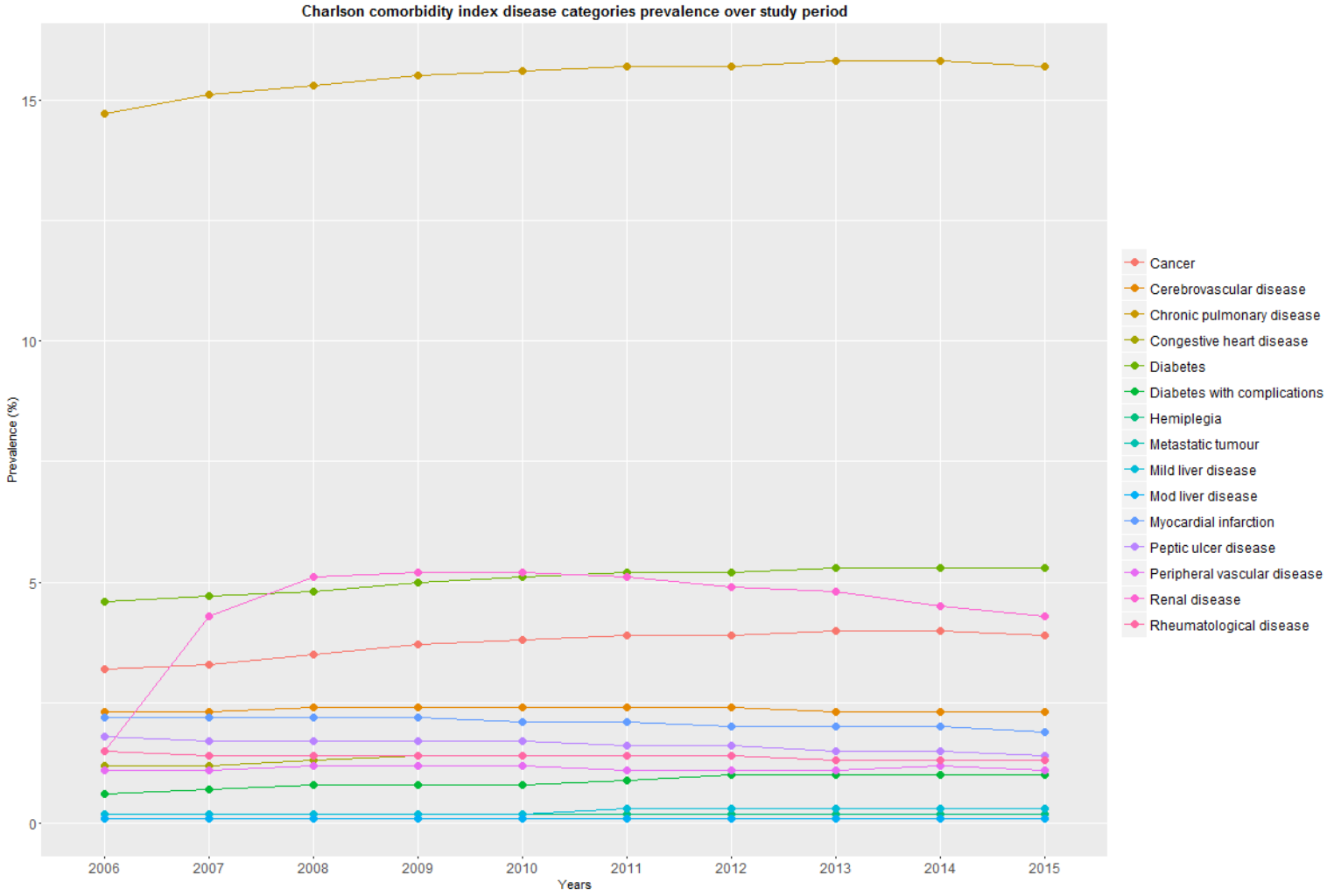
Supplementary Figure 3.1: Example of the process used to calculate build our datasets applied to a patient's data. The vertical blue dotted lines represent 12-month time windows.



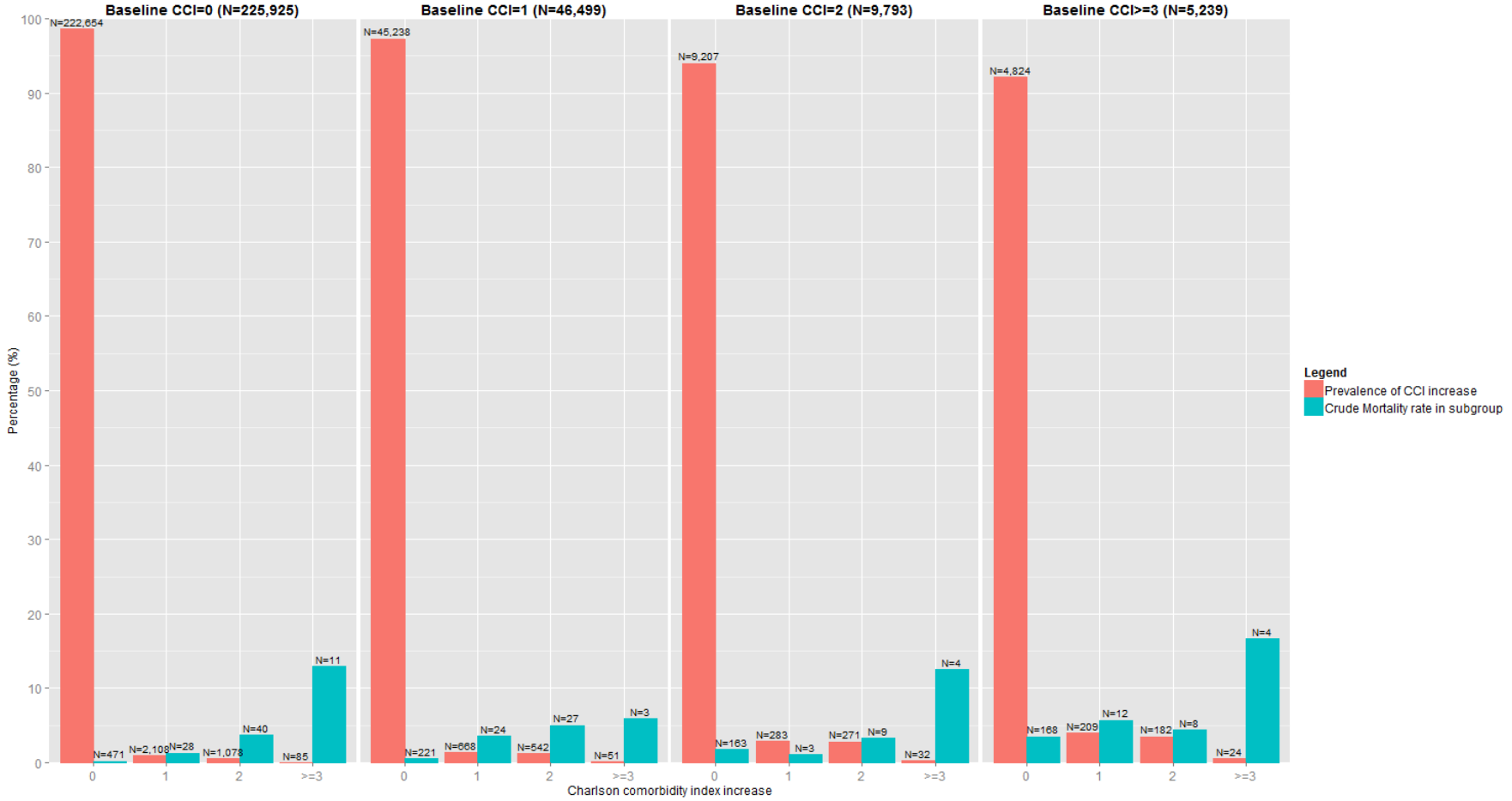
Supplementary Table 3.1: Information recorded for each time window for the patient in Supplementary Figure 3.1: age; gender; Charlson comorbidity index at baseline (i.e. t0); Charlson comorbidity index; Charlson comorbidity index cumulative change (i.e. change between Charlson comorbidity index at baseline and current Charlson comorbidity index value); change between current Charlson comorbidity index value and the one at the previous time window; and a binary variable that tells if the patient died in the period of time between the current and the next time-window. **Abbreviation:** CCI: Charlson comorbidity index.

Variable	**t0**	**t1**	**t2**	**t3**	**t4**	**t5**	**t6**	**t7**	**t8**	**t9**
Age	58	59	60	61	62	63	64	65	66	67
Gender	M	M	M	M	M	M	M	M	M	M
CCI at baseline	1	1	1	1	1	1	1	1	1	1
CCI	1	1	3	3	3	3	5	5	7	7
CCI cumulative change	0	0	2	2	2	2	4	4	6	6
CCI change	0	0	2	0	0	0	2	0	2	0
Death	NO	NO	NO	NO	NO	NO	NO	NO	NO	YES

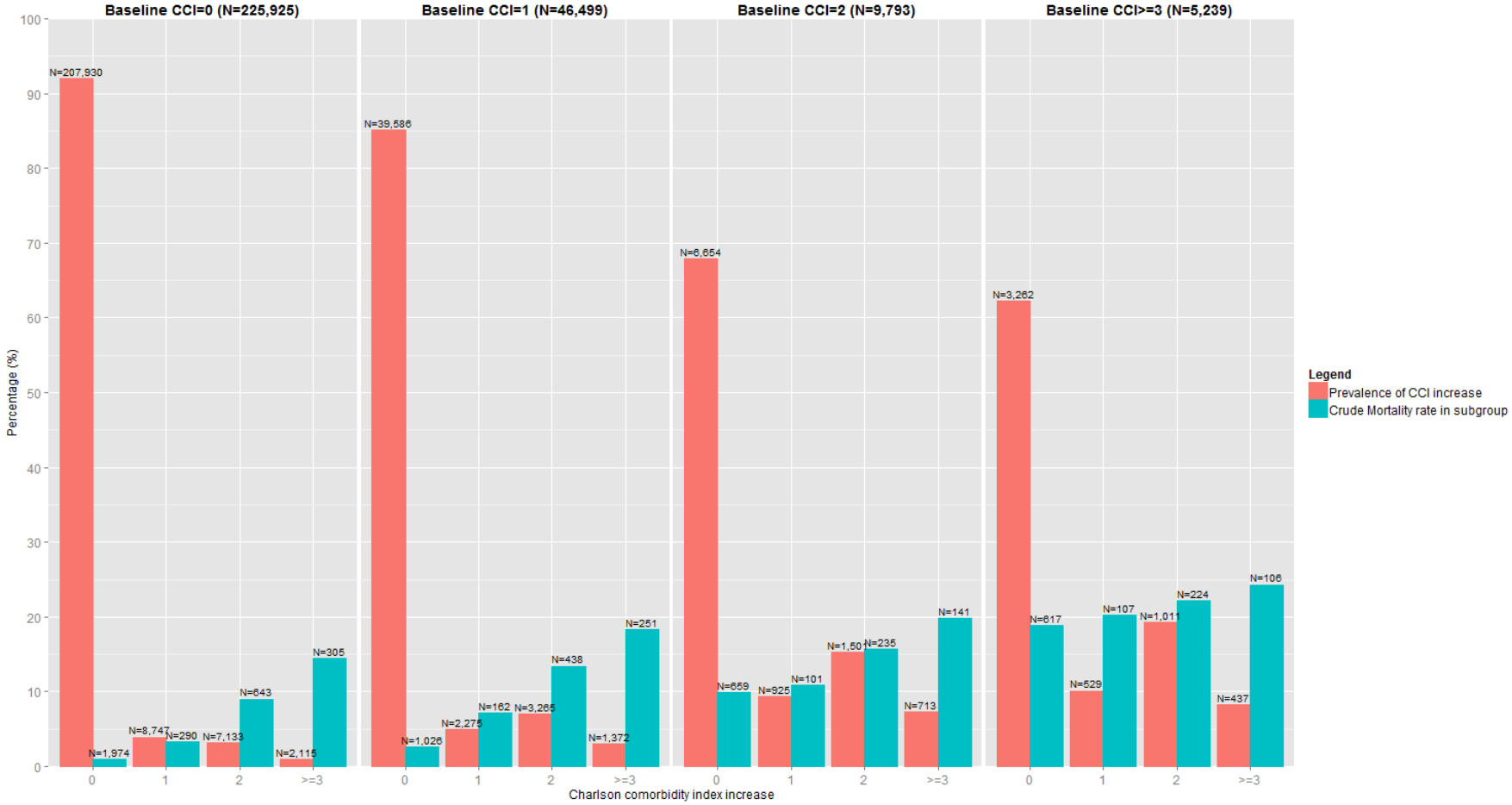
Supplementary Figure 3.2: Prevalence of the Charlson comorbidity index disease categories over the study period.



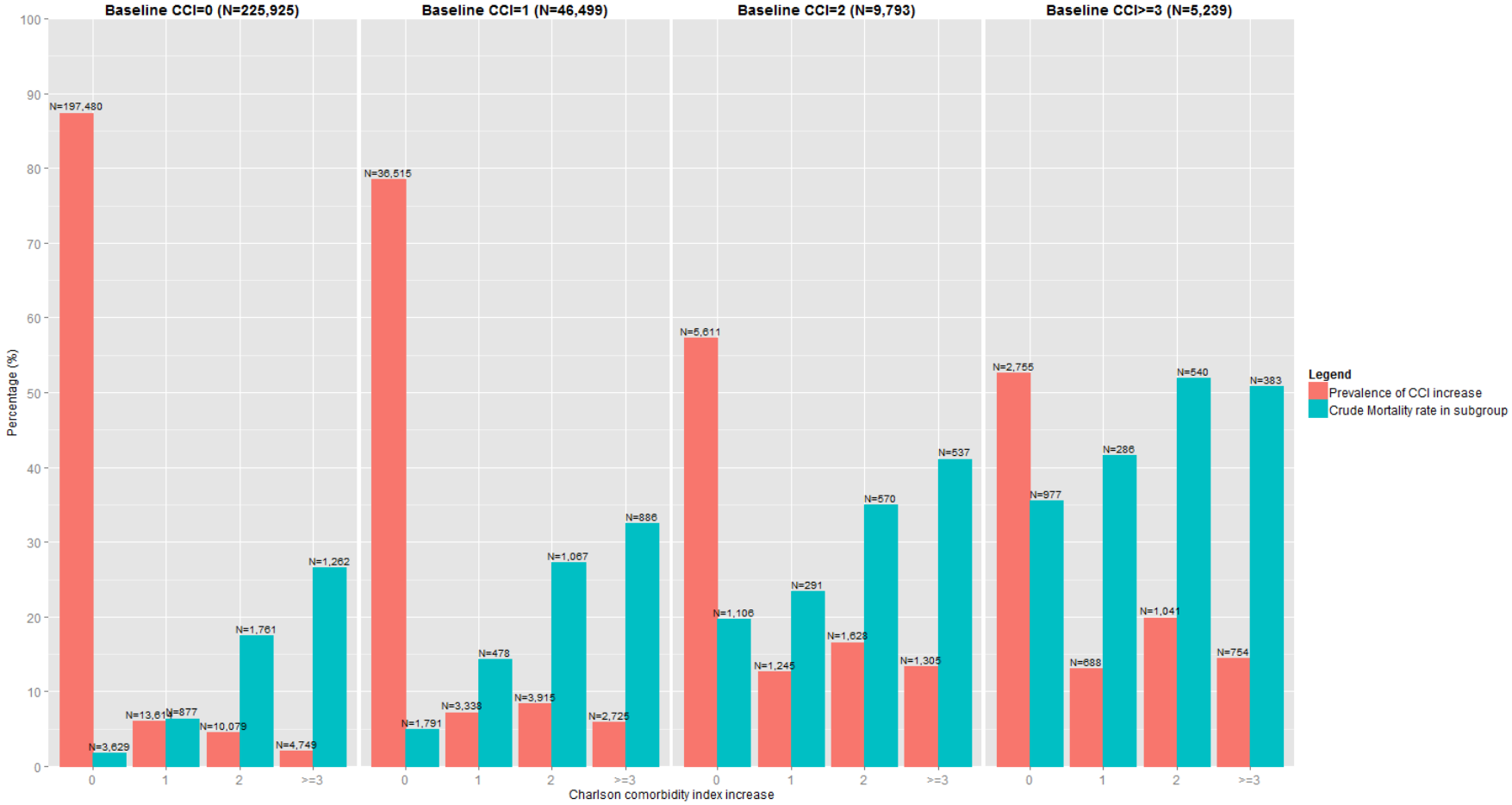
Supplementary Figure 3.3: Prevalence of Charlson comorbidity index increase (i.e. 0,1,2,>=3) in the SIR cohort after 1-year follow-up. Prevalence of increase is calculated on different subgroups on the basis of Charlson comorbidity index value at baseline (i.e. 0,1,2,>=3). For each subgroup, crude mortality is reported. Abbreviation: CCI, Charlson comorbidity index.



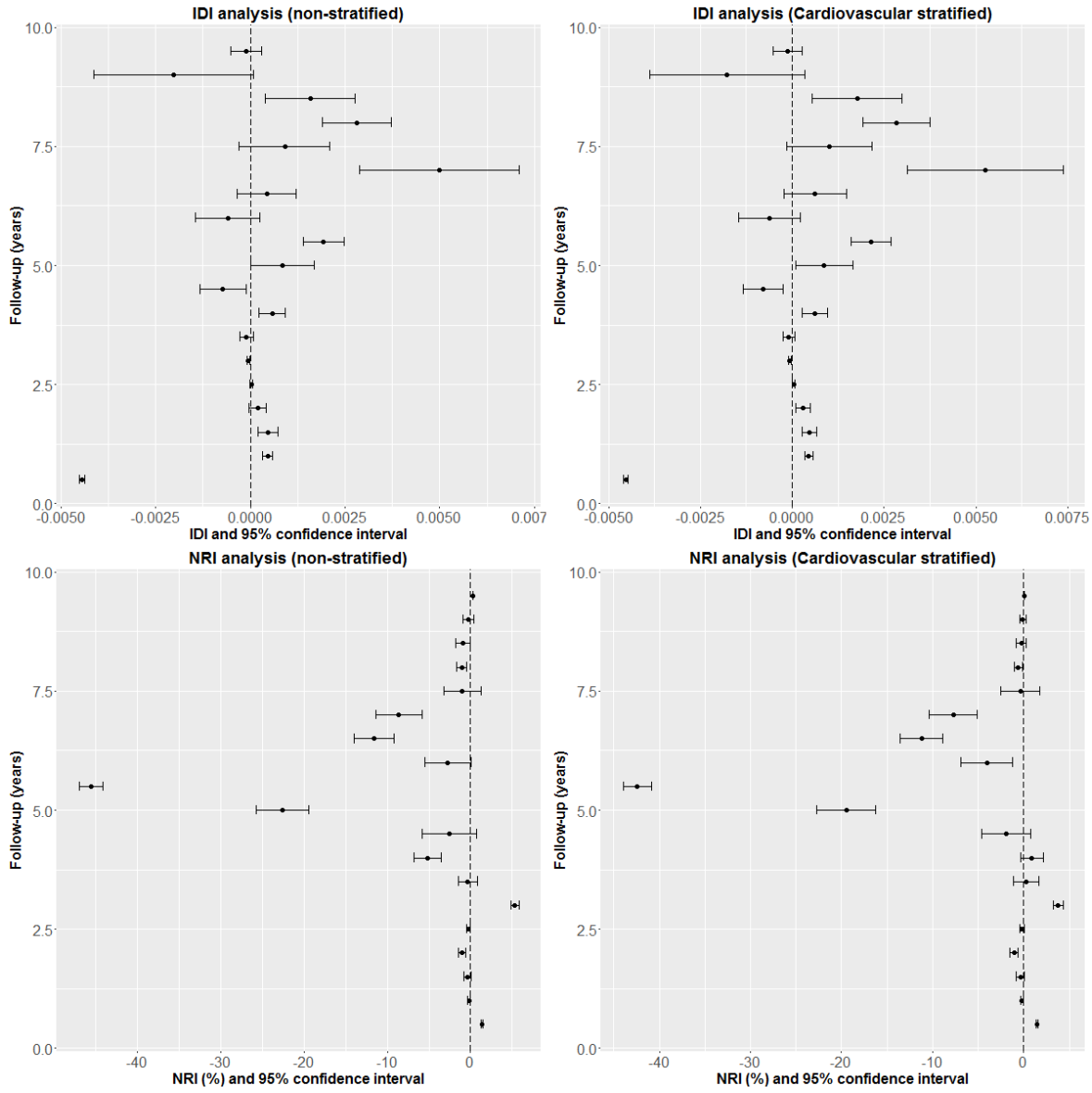
Supplementary Figure 3.4: Prevalence of Charlson comorbidity index increase (i.e. 0,1,2,>=3) in the SIR cohort after 5-years follow-up. Prevalence of increase is calculated on different subgroups on the basis of Charlson comorbidity index value at baseline (i.e. 0,1,2,>=3). For each subgroup, crude mortality is reported. Abbreviation: CCI, Charlson comorbidity index.



Supplementary Figure 3.5: Prevalence of Charlson comorbidity index increase (i.e. 0,1,2,>=3) in the SIR cohort over study period. Prevalence of increase is calculated on different subgroups on the basis of Charlson comorbidity index value at baseline (i.e. 0,1,2,>=3). For each subgroup, crude mortality is reported. Abbreviation: CCI, Charlson comorbidity index.



Supplementary Figure 3.6: Integrated Discrimination Index (IDI) and Net Reclassification Index (NRI) analysis to compare model 1 and model 5 for the 6-month time windows analysis. 95% confidence intervals are calculated from 100 bootstraps iterations.



Chapter 4

AN EXTERNAL VALIDATION OF MODELS TO PREDICT THE ONSET OF CHRONIC KIDNEY DISEASE USING POPULATION-BASED ELECTRONIC HEALTH RECORDS FROM SALFORD, UK

Paolo Fraccaro[1,2,3], Sabine van der Veer[2,3], Benjamin Brown[1,2,3], Mattia Prosperi[2,3,4], Donal O'Donoghue[5], Gary S. Collins[6], Iain Buchan[1,2,3], Niels Peek[1,2,3]

[1] NIHR Greater Manchester Primary Care Patient Safety Translational Research Centre, Institute of Population Health, The University of Manchester, Manchester, UK.

[2] Health eResearch Centre, Farr Institute for Health Informatics Research, UK.

[3] Centre for Health Informatics, Institute of Population Health, The University of Manchester, Manchester, UK.

[4] Department of Epidemiology, University of Florida, Gainesville, FL, USA.

[5] Renal Clinic, Salford Royal NHS Trust, Salford, UK.

[6] Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Oxford, UK

Contribution: PF, MP, NP (corresponding author) and GC designed the study. PF extracted the data from all sources and performed the analyses. PF, SV, GC and NP wrote the manuscript. BB, MP, DO and IB critically edited the manuscript.

Based on: Fraccaro P, van der Veer S, Brown B, Prosperi M, O'Donoghue D, Collins GS, et al. An external validation of models to predict the onset of chronic kidney disease using population-based electronic health records from Salford, UK. BMC Med. 2016;14: 1–15. doi:10.1186/s12916-016-0650-2.

4.1 ABSTRACT

Background

Chronic kidney disease (CKD) is a major and increasing constituent of disease burdens worldwide. Early identification of patients at increased risk of developing CKD can: guide interventions to slow disease progression; initiate timely referral to appropriate kidney care services; and support targeting of care resources. Risk prediction models can extend laboratory-based CKD screening to earlier stages of disease, but to-date only a few of them have been externally validated or head-to-head compared outside development populations. Our objective was to validate published CKD prediction models applicable in primary care.

Methods

We synthesised two recent systematic reviews of CKD risk prediction models and externally validated selected models for a five-year horizon of disease onset. We used linked, anonymised, structured (coded) primary and secondary care data from patients resident in Salford (population ~234k), UK. All adult patients with at least one record in 2009, were followed-up until: the end of 2014; death; or CKD onset (n=178,399). CKD onset was defined as repeated impaired eGFR measures over a period of at least three months, or physician diagnosis of CKD Stage 3-5. For each model we assessed discrimination, calibration, and decision curve analysis.

Results

Seven relevant CKD risk prediction models were identified. Five models also had an associated simplified scoring system. All models discriminated well between patients developing CKD or not, with c-statistics around 0.90. Most of the models were poorly calibrated to our population, substantially over-predicting risk. The two models that did not require recalibration were also the ones that had the best performance in the decision curve analysis.

Conclusions

Included CKD prediction models showed good discriminative ability but over-predicted the actual 5-year CKD risk in English primary care patients. QKidney, the only UK-developed model, outperformed the others. Clinical prediction models should be (re)calibrated for their intended uses.

Keywords: chronic kidney disease; clinical prediction models; eGFR; decision support; electronic health records; model validation; model calibration.

4.2 INTRODUCTION

Chronic kidney disease (CKD) present a substantial burden of disease worldwide [1–4], with an increasing number of people being diagnosed [5,6]. A 2010 study of 2.8M UK adults reported a 5.9% prevalence of stage 3-5 CKD [7]. In the UK, costs related to CKD care in 2009-2010 were estimated around £1.45 billion (1.3% of the National Health Service (NHS) budget) [8] – costs that are set to rise steeply [6,8].

Early detection of CKD, and identification of patients at increased risk of developing CKD, can improve care by: guiding preventive measures to slow disease progression; initiating timely referral to nephrology care; and supporting better allocation of resources [9]. Yet, despite efforts worldwide to improve detection [10], CKD often remains undiagnosed in its early stages [5]. Currently, most CKD clinical surveillance relies on estimated Glomerular Filtration Rate (eGFR) from serum creatinine testing [10]. In the UK, national clinical practice guidelines recommend systematic monitoring, in the primary care setting, of eGFR in patients with CKD risk factors (i.e. diabetes, hypertension, cardiovascular diseases, or use of particular medications) [11]. In addition, eGFR has been calculated routinely in UK NHS laboratories since 2006 where at least age, sex and creatinine variables are available – so CKD may be picked up in a variety of clinical contexts. The value of universal clinical/opportunistic screening for CKD remains unclear [12].

Risk prediction models can extend the clinical screening toolkit from measured to predicted disease, affording more timely intervention, for example to reduce risk factors [13]. Several models have been developed to predict CKD onset but most have not been validated outside the setting in which they were developed [14,15]. So the portability of these models to other populations, risk environments and healthcare settings has yet to be demonstrated. Furthermore, comprehesive head-to-head comparisons of these purportedly alternative models are lacking in the literature [14–16]. Only one comparison of two CKD prediction models in a small cohort was published to date [17].

The aim of this study was to externally validate and compare the performance of previously published models for predicting 5-year CKD risk using routine healthcare records from a UK population with well-studied, high quality electronic health records.

4.3 METHODS

4.3.1 Reporting

The reporting of this external validation study follows the TRIPOD statement [18,19], which is a set of recommendations for the reporting of studies describing the development, validation, or updating of prediction models [18,19].

4.3.2 Literature review

Two recent systematic reviews, by Collins *et al*.[14], and Echouffo-Tcheugui and Kengne [15], identified prediction models on CKD onset and CKD progression [14,15]. From these reviews we selected models predicting CKD onset that could be used in primary care. Models were excluded if:

1. they were developed for a specific subpopulation (e.g. HIV patients [20]);
2. the covariates coefficients and regression formula were not reported in the original study; or
3. with more than one predictor not routinely collected in UK primary care (more than one predictor for which we had >70% missing data in our dataset).

Where available, we included simplified scoring systems accompanying the included prediction models. Such systems typically produce an integer score for each patient, where higher scores represent higher predicted risk but there is no relationship with absolute risk.

4.3.3 Validation cohort

4.3.3.1 Outcome

The outcome of interest was onset of CKD within five years. Existing models employ various definitions of CKD [14,15]. For our study, we followed international guidelines [21] and considered a recent study [7] reporting UK CKD prevalence based on primary care records. We defined CKD as:

a) the presence of at least two consecutive eGFR values below 60 ml/min/1.73 m^2, as calculated with the Modification of Diet in Renal Disease (MDRD) formula [22] (i.e. the formula used by biochemical laboratories serving most of UK primary care practices, with results stored in primary care databases under the "451E." Read code), over a period of three months or longer; or

b) the presence of a CKD Stage 3-5 diagnostic code.

We were unable to incorporate albumin-creatinine ratio (ACR, a predictor of kidney damage [23] noted in international guidelines [21]) because ACR data are available only for selected groups of patients at risk of CKD, such as those on diabetes care pathways.

4.3.3.2 Data source

We used linked, anonymised data from the Salford Integrated Record (SIR) up to the end of 2014. SIR is an EHR that has been overlain on primary and secondary care clinical information systems for over ten years in the city of Salford (population 234k) – an early-adopter site of healthcare IT in the UK. SIR includes patient records submitted by all 53 primary care providers and the one secondary care provider for this population, stored as Read codes versions 2 and 3 [24]. The data cover all primary care, some of secondary care – focused on long-term conditions management – and all results from biochemical testing across primary and secondary care.

4.3.3.3 Study population

Salford is a relatively deprived population with a high burden of disease, where the EHR data have been used extensively to study the population's health and care. Like all English localities, Salford's primary care is measured and remunerated under the Quality and Outcomes Framework (QOF), including counts of the mean number of conditions per registered patient, where Salford falls in 61st centile [25].

We included all adults (aged 18 years or older) registered with a Salford practice with at least one record in SIR between 1st April 2009 and 31st March 2010 – the financial year. We looked at the financial rather than calendar year to take account of the QOF, which might have influenced the quality of data recorded by GPs [26,27]. For all retrieved patients the entry date was the date of the first record in the financial year 2009. Included patients were followed until 31st December 2014, or censored when they moved outside of Salford or died.

We excluded patients with CKD stage 3-5 before study entry, which was determined by diagnostic codes and eGFR measurements (following our definition of CKD onset).

We also defined a cohort of patients with complete follow-up data, consisting of patients who either developed CKD in the study period or had at least five years of follow-up. We used this cohort to validate models derived with logistic regression, which requires complete follow-up data.

4.3.3.4 Predictors and missing data

We used Read codes retrieved from *clinicalcodes.org* [28] to extract clinical and laboratory variables from the SIR database. *Clinicalcodes.org* is a repository of Read codes used in previously published articles; we used Read codes from five studies [29–33] (the full list of codes is available in Supplementary File 1 at https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-016-0650-2). For comorbidities, such as hypertension and peripheral vascular disease, we identified any related diagnostic Read code before the patient's study entry date. If the type of diabetes was not specified in the diagnostic code or contradicting codes were present (i.e. diabetes type 1 and type 2 for the same patient), we assigned 'type 1' to patients with the first diabetes code before 35 years of age, and 'type 2' to all other diabetes patients. For medications, such as nonsteroidal anti-inflammatory drugs (NSAIDs) or hypertensive medications, we looked for at least two prescriptions in the six months prior to entry date. Finally, for laboratory tests, we selected the most recent result within 12 months before the entry date.

Since more than 90% of the population in Salford is of White British ethnicity [34], we considered patients without a recorded ethnicity code as White British. We imputed values for predictors using multiple imputation by chained equations with 10 iterations to minimise the effect of selectively ignoring those with any missing data (using the *mice* package in R [35]).

4.3.4 Data analysis

We implemented models developed by logistic and Cox proportional hazards (CPH) regression formulas using published coefficients and intercept or baseline hazard provided. For the QKidney models [36] we used the information from svn.clinrisk.co.uk/opensource/qkidney – a web-based calculator written in C was re-coded in R language as per Supplementary File 2 (available at https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-016-0650-2). For simplified scoring systems, we computed the total simplified score for each patient in our dataset. In addition, if the original model was a logistic regression and the intercept was not reported, we estimated it from information about CKD prevalence and predictors summary measures (mean for continuous variables and prevalence for binary variables) in the development population.

We assessed the performance of the models and the associated simplified scoring systems in terms of discrimination and calibration. Discrimination is the ability of a model to distinguish between patients who develop and patients who do not develop CKD. Discrimination was assessed by calculating the area under receiving operating characteristic curve (AUC) and Harrell's c-index [37–39]. 95% confidence intervals for the AUC and c-index were calculated from 500 bootstrap

iterations. We evaluated calibration by calculating the mean absolute prediction error (MAPE), calibration slope, and by calibration plots. MAPE is the average difference in predicted and observed onset of CKD, and expressed by a number between 0 and 1, with values closer to 0 indicating better performance [40].

$$\boldsymbol{MAPE} = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} |\boldsymbol{\hat{y}_i} - \boldsymbol{y_i}|$$

Where n is the total number of patients $\boldsymbol{\hat{y}_i}$ is the predicted risk of CKD onset for the ith patient and $\boldsymbol{y_i}$ is the observed outcome (CKD onset or not) for the same patient.
Calibration slopes are regression slopes of predicted probabilities fitted to the external validation dataset [41]. The optimal value is 1, with values smaller than 1 reflecting overfitting of the model. Calibration plots compare mean predicted risks with mean observed outcomes for subgroups with similar predicted risks. A model is considered to be well calibrated if the plot follows the 45° line from lower left corner to upper right corner of the plot. In our analysis, we created calibration plots using the R package PredictABEL [42].

For the simplified scoring systems, we compared sensitivity, specificity and positive predictive value (PPV) obtained by using the decision-making threshold that was reported in the original publications, as well as using the optimal threshold for our study population as calculated with Youden's method [43]. If a study did not present any risk score or we could not use the proposed simplified score because of more than one missing predictor in our dataset, sensitivity, specificity and PPV was evaluated for the full model instead.

To interpret the performance of included models we used the framework for external validation from Debray et al. [44]. Therefore we assessed the extent to which case-mix of the development datasets and our validation dataset were similar, by comparing the mean linear predictor of models in the two cohorts. Since individual patient data of the development datasets were not publicly available, the mean linear predictor was calculated as the sum of the intercept and the product of models coefficients and predictors' prevalence (for binary variables) or mean (for continuous variables) provided in summary statistics of original studies. In order to assess how accurate the mean linear predictor calculation based on the summary statistics was, in our validation dataset we also calculated the mean linear predictor by calculating the mean and standard deviation (SD) of the linear predictor from the individual patient data.

Finally, to evaluate the clinical impact of implementing the models in practice as screening tools, we performed two analyses. First, we performed decision curve analysis that evaluates how

different threshold probabilities alter the false-positive and false-negative rate expressed in terms of net benefit [45]. When carrying out a head-to-head comparison of different prediction models on the same population, the interpretation is straightforward: at each clinically relevant probability threshold, the model that has the highest net benefit is preferable. Models are also compared to the extreme choices of designating *all* and *no* patients at high risk of developing CKD. Second, for each model we evaluated the potential implementation of a CKD prevention high-risk approach [46] based on model's prediction by calculating the proportion of observed CKD cases in our dataset within the highest tenth of predicted risk (i.e. the 10% of patients with highest predicted risks).

Data manipulation and statistical analyses were performed using R software (www.r-project.org).

4.3.5 Sensitivity analyses

We performed several sensitivity analyses. First, since the risk of developing CKD in the asymptomatic general population is low [47], we also validated each of the models in patients with established CKD risk factors at entry date. Following the UK National Institute for Clinical Excellence (NICE) guidelines on early detection of CKD [11], these risk factors were: use of calcineurin inhibitor drugs, lithium, or nonsteroidal anti-inflammatory drugs (NSAIDs); diabetes mellitus; hypertension; acute kidney injury (AKI) in the previous two years; history of cardiovascular disease (CVD), renal calculi or prostatic hypertrophy, systemic lupus erythematosus, or haematuria; and family history of kidney disease. Second, as most models in our study used a single measured renal impairment to define CKD, we repeated the analysis while using a more inclusive definition of CKD onset as the presence of a CKD 3-5 diagnostic code or a single eGFR measurement below 60 ml/min/1.73 m^2. Third, we considered patients who died during follow-up as if they developed CKD, because mortality is frequently attributable to CKD and most risk prediction models do not account for death as a competing risk. Fourth, we calculated eGFR by using the CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration) formula [48] and repeated our main analysis (e.g. CKD defined as impaired eGFR for at least three months or CKD 3-5 diagnostic code). Fifth, we repeated our main analysis by using a prediction horizon of four instead of five years. Finally, we repeated the analyses omitting individuals with any missing observation.

4.4 RESULTS

4.4.1 CKD prediction models included for external validation

Figure 4.1 depicts the model inclusion process. Of the 29 models identified by Collins *et al*. [14] and Echouffo-Tcheugui and Kengne [15], 18 were developed with the aim of predicting CKD onset. We excluded three models because of incomplete reporting of regression models (regression coefficients not fully reported) in the original paper [49] and one model because it was developed in a specific sub-population (namely HIV patients) [20]. We excluded a further seven models for which we had more than one missing predictor in our dataset, including Halbesma et al [50] for missing data for eGFR, urinary excretion, and c-reactive protein; Chien et al [51] for missing post-pondrial glucose, proteinuria and uric acid; O'seaghdha *et al.* [52] for missing eGFR and quantitative albuminuria, and finally, we excluded two models by Kshirsagar *et. al.* [53] and O'seaghdha *et al.* [52] because of missing low high-density lipoprotein cholesterol level and eGFR, respectively. The final set consisted of seven models (five logistic regression models and two CPH regression models) and five simplified scoring systems [36,51–56]. Table 4.1 describes the details of the included models, and Supplementary Table 4.1, Supplementary Table 4.2, and Supplementary Table 4.3 detail the population characteristics of the development datasets, the regression coefficients, and the simplified scoring systems.

Figure 4.1: Procedure to identify and select CKD prediction models.

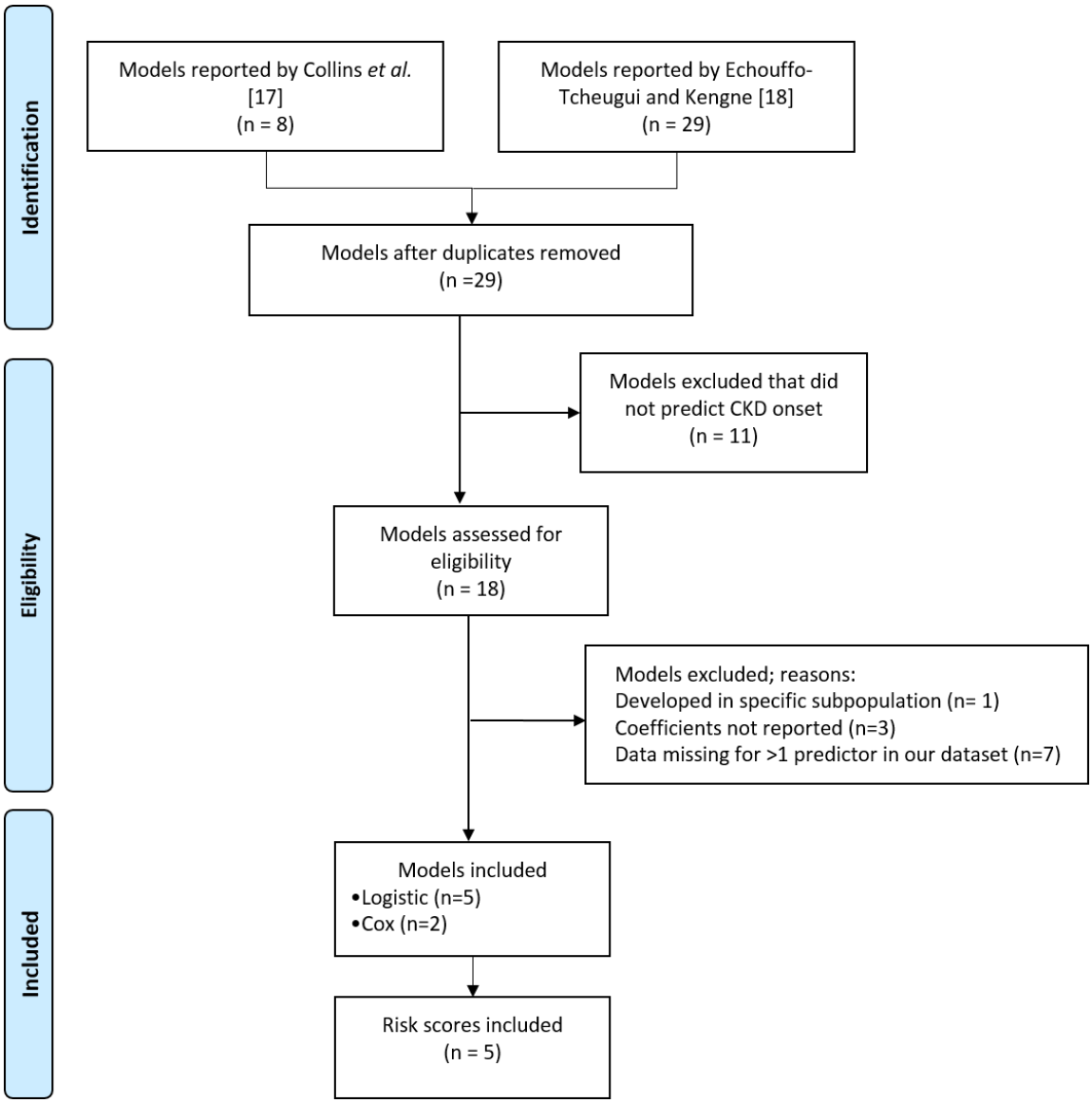


Table 4.1: Details of studies developing CKD prediction models that were included for external validation.

Authors [ref] **Publication year**	**Study design/Study context** **Study period**	**Ethnicity** **Age range**	**Population size** **Number (%) CKD cases**	**Type of models** **Time horizon**	**Handling of missing values** **Method of internal validation**	**Definition of CKD**	**Predictors in model**
Bang et al. [54] 2007	Cross-sectional population-based survey/ Screening programme 1999-2002	US-mixed 20-85 years	8,530 601 (7.5%)	Logistic 2 years	Excluded Random split-sample	At least one eGFR measurement <60 [a)]	Age, gender, anaemia, proteinuria [a)], hypertension, diabetes mellitus, history of cardiovascular disease, history of heart failure, peripheral vascular disease
Chien et al [51] 2010	Prospective cohort study/ Secondary care 2003	Taiwan-Chinese 51.2 years (mean)	5,168 190 (3.7)	Cox 4 years	NR NR	At least one eGFR measurement <60 [a)]	Age, BMI, diastolic blood pressure, type 2 diabetes, history of stroke
Hippisley-Cox and Coupland (QKidney®) [36] 2010	Prospective cohort population based/ Primary care 2002-2008	UK-mixed 35-74 years	1,591,884 23,786 (1.5%)	Cox 5 years	Multiple imputation Random split-sample	At least one eGFR measurement <45[a)], kidney transplant; dialysis; nephropathy diagnosis; proteinuria	Age, ethnicity, deprivation, smoking, BMI, systolic blood pressure, diabetes mellitus, rheumatoid arthritis, cardiovascular disease, treated hypertension, congestive cardiac failure; peripheral vascular disease, NSAID use and family history of kidney disease.
Kshirsagar et al. [53] 2008	Prospective cohort study/ Community-based 1987-1989	US-white and black 45-64 years	9,470 1605 (16.9%)	Logistic 9 years	NR Random split sample	At least one eGFR measurement <60 [a)]	Age, gender, anaemia, hypertension, type 2 diabetes mellitus, history of cardiovascular disease, history of heart failure, peripheral vascular disease
Kwon et al. [55] 2012	Cross-sectional survey/ Population-based 2007-2009	Korean-Asian >=19 years	6,565 100 (1.5%)	Logistic 1-year	Excluded Split sample	At least one eGFR measurement <60 [a)]	Age, gender, anaemia, proteinuria [a)], hypertension, type 2 diabetes mellitus, history of cardiovascular disease
O'seaghdha et al. [52] 2011	Prospective cohort study/ Population-based 1995-2008	US white 45-64 years	2,490 229 (9.2%)	Logistic 10 years	Excluded Bootstrap	At least one eGFR measurement <60 [a)]	Age, hypertension, diabetes mellitus
Thakkinstian et al. [56] 2011	Cross-sectional survey/ Community-based NR	Thailand-Asian >=18 years	3,459 606 (17.5%)	Logistic 1-year	NR Bootstrap	At least one eGFR measurement <90 [a)]	Age, hypertension, diabetes mellitus, kidney stones

Abbreviations: BMI, body mass index; CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate, NR, not reported; US, United States.

a) Predictor not included in external validation due to missing data in our dataset.

All models were developed outside the UK, with the exception of QKidney® [36] (www.qkidney.org), which was developed on a large population from England and Wales selected from general practices using the EMIS EHR. All included models used a different definition of CKD, but the majority used an older definition based only on one impaired eGFR measurement. Time horizons in original papers were different to our 5-year definition, with the exception of QKidney® [36], which however allowed other time horizons options (1-, 2-, 3- and 4- year). For three models, the prediction time horizon was not specified [54–56]. However, we could derive from study duration and data collection procedures in the original publications that the time horizon was one [56], two [54] and nine [54] years, respectively. For the remaining models the reported time horizons were between four and ten years [51,52,54].

Predictors included in the models were largely based on known CKD risk factors (hypertension, diabetes mellitus, or history of cardiovascular disease). The only biomarkers included were systolic and diastolic blood pressure, and body mass index (BMI). Multiple imputation of missing values was applied to these variables, along with deprivation, haemoglobin (i.e. to calculate presence of anemia) and smoking. In these predictors, missing values ranged from 1.8% to 70.0%, with a median value of 46.0%. Conversely, we excluded proteinuria as predictor from our analyses due to 94.6% of missing data (see Table 4.2), therefore the models by Bang et al [54] and Kwon et al [55] had one missing predictor. Finally, three of the included models, which derived a simplified scoring system, [53,55,57] did not report the intercept of their underpinning logistic regression model, therefore we estimated the intercepts from the prevalence of CKD and predictors' summary statistics in original studies.

4.4.2 Study population characteristics

Figure 4.2 shows the cohort selection process. There were 187,533 adult patients with at least one record in the financial year 2009 in our database, of which 178,399 remained after applying our exclusion criteria, with 6,941 patients (3.9%) that died before developing CKD. There were 162,653 patients (91.2%) who had complete follow-up data. Overall, there were 6,038 incident cases of CKD during the study period. Table 4.2 and Table 4.3 describe the characteristics of cohorts with complete and incomplete follow-up.

Figure 4.2: Validation Cohort selection.

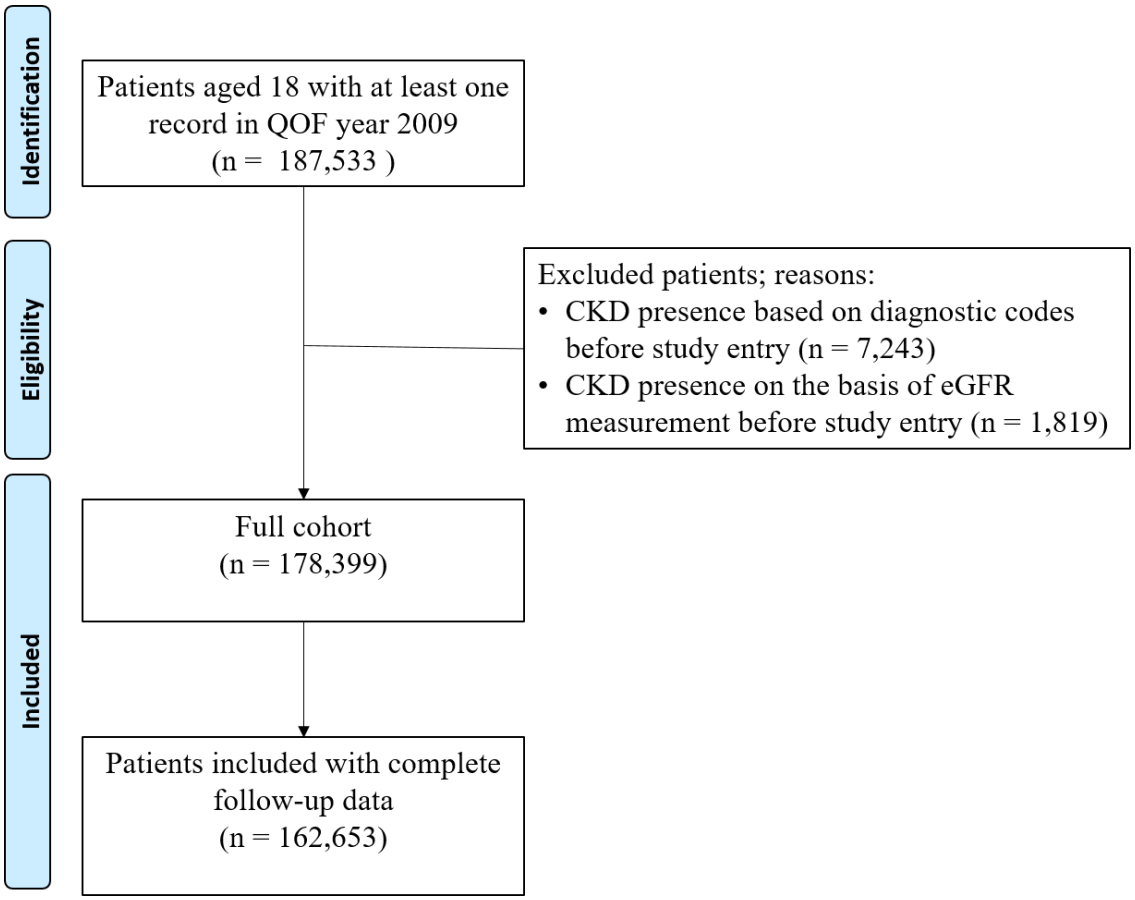


Table 4.2: Patients with complete and incomplete follow-up data stratified for CKD onset; values are numbers (%) unless indicated otherwise.

Parameters		No CKD				CKD	
		Patients with complete follow-up		Patients with incomplete follow-up			
			Missing		Missing		Missing
Included patients		156,615	None	172,361	None	6,038	None
Died before developing CKD		719 (0.5)	None	6,941 (4)	None	/	None
Follow-up (mean,SD)		5.6 (0.2)	None	5.4 (0.7)	None	2.6 (1.7)	None
Age (mean,SD)		42.1 (16.7)	None	42.7 (17.3)	None	70.3 (12.5)	None
Female Gender		82,883 (52.9)	None	89,389 (51.9)	None	3,452 (57.2)	None
Townsend index (mean,SD) [e]		1.6 (3.5)	2,900 (1.9)	1.6 (3.4)	3,244 (1.9)	1.4 (3.4)	47 (0.8)
Ethnicity	Not recorded	55,586 (35.6)	Not applicable [d]	61,220 (35.6)	Not applicable [d]	2,014 (33.4)	Not applicable [d]
	White	90,443 (57.8)		99,243 (57.7)		3,889 (64.5)	
	Other	10,586 (6.8)		11,898 (6.9)		135 (2.2)	
Smoking [e]	Non-smoker	66,769 (48.8)	19,901 (12.7)	72,137 (48.4)	23,296 (13.5)	2,167 (37.7)	292 (4.8)
	Ex-smoker	29,980 (21.9)		33,097 (22.2)		2,475 (43.1)	
	Light smoker [1-9 cg/day]	11,072 (8.1)		12,128 (8.1)		344 (6)	
	Moderate smoker [10-19 cg/day]	16,951 (12.4)		18,472 (12.4)		413 (7.2)	
	Heavy smoker [>=20 cg/day]	11,942 (8.7)		13,231 (8.9)		347 (6)	
BMI [kg/m^2] (mean,SD) [e]		26.6 (6)	33,717 (21.5)	28 (6.1)	38,628 (22.4)	28.4 (6)	518 (8.6)
Diastolic blood pressure [mmHg] (mean,SD) [e]		76.9 (9.8)	75,616 (48.3)	78.9 (10.2)	85,075 (49.4)	75.8 (10.2)	1,164 (19.3)
Systolic blood pressure [mmHg] (mean,SD) [e]		128.2 (15.8)	75,602 (48.3)	130.5 (16.7)	85,058 (49.3)	136.3 (16.7)	1,166 (19.3)
eGFR [mL/min/1.73 m^2] (mean,SD)		83.7 (9.4)	118,912 (75.9)	82.5 (9.4)	131,103 (76.1)	69.4 (11.3)	1,828 (30.3)
Hb [g/dl] [e]		13.9 (1.6)	110,723 (70.7)	13.8 (1.6)	122,430 (71)	13.4 (1.7)	2,530 (41.9)
Proteinuria [a] [b]		751 (0.5)	149,234 (95.3)	18 (0.2)	164,097 (95.2)	236 (3.9)	4,665 (77.3)
Quantitative albuminuria [b] [c]		129 (0.1)	152,266 (97.2)	4 (0)	167,482 (97.2)	62 (1)	5,167 (85.6)
HDL cholesterol level [b] [mg/dl] (mean, SD)		25.9 (7.9)	122,477 (78.2)	26.7 (7.9)	135,066 (78.4)	25.7 (7.8)	2,413 (40)

Abbreviations: BMI, body mass index; CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; HDL, high-density lipoprotein; MDRD, modification of diet in renal disease; SD, standard deviation.

[a] Albumin:creatinine ratio >30mg/mmol or albumin concentration >200mg/l, or diagnostic code

[b] Variable excluded as predictor from external validation due to >70% missing values

[c] Albumin:creatinine ratio >30mg/mmol

[d] Patients without recorded ethnicity were considered as white (see Methods section)

[e] Multiple imputation applied to missing values.

Table 4.3: Prevalence of CKD risk factors (as expressed in NICE guidelines) stratified for CKD onset; values are numbers (%) unless indicated otherwise.

CKD risk factors	No CKD		CKD (N=6,038)
	Patients with complete follow-up (N=156,615)	Patients with incomplete follow-up (N=172,361)	
Hypertension [a)]	22,074 (14.1)	24,971 (14.5)	3,554 (58.9)
Hypertensive treatment [b)]	22,122 (14.1)	24,769 (14.4)	3,655 (60.5)
Type 1 Diabetes Mellitus [a)]	703 (0.4)	740 (0.4)	36 (0.6)
Type 2 Diabetes Mellitus [a)]	5,574 (3.6)	6,383 (3.7)	1,221 (20.2)
History of cardiovascular disease [a)]	11,096 (7.1)	13,407 (7.8)	2,182 (36.1)
History of heart failure [a)]	743 (0.5)	1,088 (0.6)	387 (6.4)
History of stroke [a)]	1,875 (1.2)	2,538 (1.5)	509 (8.4)
Peripheral vascular disease [a)]	2,127 (1.4)	2,532 (1.5)	331 (5.5)
Kidney stones [a)]	751 (0.5)	814 (0.5)	64 (1.1)
Rheumatoid arthritis [a)]	1,321 (0.8)	1,512 (0.9)	142 (2.4)
Systemic lupus erythematosus [a)]	99 (0.1)	104 (0.1)	8 (0.1)
Family history of kidney disease [a)]	25 (0)	28 (0)	3 (0)
NSAIDS use [b)]	5,101 (3.3)	5,389 (3.1)	402 (6.7)
Acute kidney injury in the last 2 years	1,975 (1.3)	2,633 (1.5)	413 (6.8)
Prostatic hypertrophy [a)]	967 (0.6)	1,143 (0.7)	173 (2.9)
Haematuria [a)]	3,176 (2)	3,574 (2.1)	341 (5.6)
Lithium use [b)]	150 (0.7)	219 (0.1)	52 (0.9)
Tacrolimus use [b)]	4 (0)	5 (0)	2 (0)
Cyclosporin use [b)]	12 (0.1)	20 (0)	6 (0.1)

Abbreviations: NSAIDs, Non-steroidal anti-inflammatory drugs; SD, standard deviation.

[a)] Based on diagnostic Read codes
[b)] At least two prescriptions in the six months before entry date.

4.4.3 External validation

Table 4.4 presents the results of the external validation: discrimination and calibration. AUC values ranged from 0.892 (95% Confidence Interval [CI] 0.888 to 0.985) to 0.910 (95% CI 0.907 to 0.913) for patients with complete follow-up data, and the c-index values for the two CPH models on the full cohort were 0.888 (95% CI 0.885 to 0.892) [51]and 0.900 (95% CI 0.897 to 0.903) [36] respectively. Simplified scores showed similar performance to the models from which they were derived. MAPE was below 0.1 for all models, with the only exception of Thakkinstian *et al* [56] for which the MAPE was 0.179 (Standard Deviation [SD] 0.161). Calibration plots (see Figure 4.3) and related calibration slopes (see Table 4.4) on the complete follow-up data showed similar figures to the MAPE analysis. Thakkinstian *et al* [56] confirmed a tendency for over-predicting risk with a calibration slope of 0.44 (95% CI 0.43 to 0.45). Conversely, the only models that were well-calibrated to our population were the ones by Bang *et al.* [54] and QKidney® [36] with calibration slope values of 0.97 (95% CI 0.96 to 0.98) and 1.02 (95% CI 1.01 to 1.04), respectively. All other models over predicted risks (calibration slopes ranging between 0.53 (95% CI 0.52 to 0.53) and 0.68 (95% CI 0.67 to 0.69), with the exception of Kshirsagar *et al.* [53] that predicted lower risk and had a calibration slope of 1.74 (95% CI 1.72 to 1.76).

Table 4.4: Discrimination, MAPE and calibration slopes of included models in patients with complete follow-up data (all models and risk scores) and in the full validation cohort (Cox proportional hazards regression models only).

	Study	Patients with complete follow-up (N= 162,653)			Full validation cohort (N= 178,399)	
		AUC [95% CI]	MAPE (SD) [a)]	Calibration slope [CI]	c-index [95% CI]	MAPE (SD) [a)]
MODELS	Bang *et al.* [54]	0.899 [0.895,0.903]	0.063 (0.162)	0.97 [0.96,0.98]	NA	NA
	Chien *et al* [51][b)]	0.898 [0.895,0.901]	0.081 (0.162)	0.65 [0.64,0.65]	0.888 [0.885,0.892]	0.085 (0.166)
	QKidney® [36][b)]	0.910 [0.907,0.913]	0.050 (0.166)	1.02 [1.01,1.04]	0.900 [0.897,0.903]	0.052 (0.165)
	Kshirsagar *et. al.* [53]	0.896 [0.892,0.900]	0.068 (0.164)	1.74 [1.72,1.76]	NA	NA
	Kwon *et al.* [55]	0.899 [0.895,0.902]	0.086 (0.158)	0.68 [0.67,0.69]	NA	NA
	O'seaghdha *et al.* [52]	0.907 [0.904,0.911]	0.089 (0.169)	0.53 [0.52,0.53]	NA	NA
	Thakkinstian *et al* [56]	0.892 [0.888,0.985]	0.179 (0.161)	0.44 [0.43,0.45]	NA	NA
SIMPLIFIED SCORES	Bang *et al.* [54]	0.895 [0.891,0.899]	NA	NA	NA	NA
	Chien *et al* - [51]	0.880 [0.876,0.883]	NA	NA	NA	NA
	Kshirsagar et. al. [53]	0.891 [0.887,0.895]	NA	NA	NA	NA
	Kwon *et al.* [55]	0.895 [0.891,0.898]	NA	NA	NA	NA
	Thakkinstian *et al* [56]	0.869 [0.864,0.873]	NA	NA	NA	NA

Abbreviations: AUC, area under receiver operating characteristic curve; eGFR, estimated glomerular filtration rate; NA, not applicable; SD, standard deviation; CI, confidence interval.

a) Calculated as mean difference between observed and predicted CKD cases

b) Cox proportional hazard regression model

Figure 4.3: Calibration plot of predicted and observed risk for the cohort of patients with complete follow-up. On the bottom a rug plot in the form of histogram shows predicted values distribution.

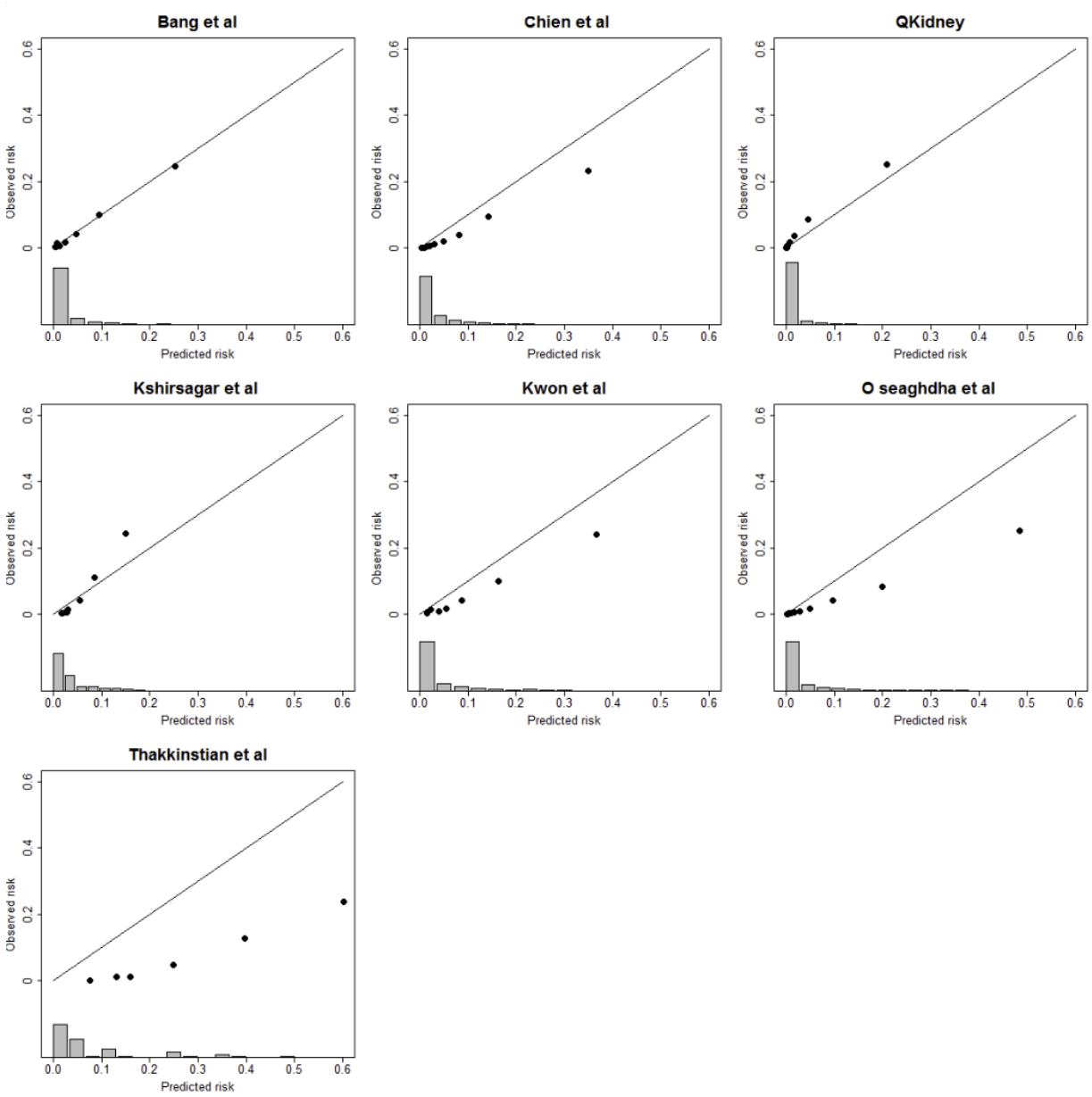


Table 4.5 reports positive predictive value, sensitivity and specificity for each of the simplified scoring systems. In this analysis we included the full QKidney® [36] model as it does not have an associated simplified scoring system. We also included the full O'seaghdha *et al.* [52] model because we could not implement their scoring system: multiple predictors had 70% or more missing values in our dataset. For two scoring systems (Chien *et al* [51] and Thakkinstian *et al* [56]) the best threshold in our population was different that the threshold proposed in the development study. For QKidney® [36] and O'seaghdha *et al.* [52], who did not report a threshold in the development study, the optimal threshold in our population was 0.017 (SD 0.002) and 0.086 (SD 0.010), respectively. In terms of performance, models showed similar performance, with PPV around 0.145, sensitivity around 0.86 and specificity around 0.80.

The distributions of the linear predictors in the development datasets and the validation dataset, calculated as proposed by Debray et al [44], are shown in Table 4.6. For all models, the mean of the linear predictor in the validation dataset was lower than in the development datasets: we found mean differences between 0.2 and 0.6, except for the model of O'seaghdha *et al.* [52] which had a difference of 1.5. There were few differences between the mean linear predictors computed on our dataset using summary statistics compared with individual patient data.

The threshold probability associated with the highest tenth of predicted risk varied from 0.0692, for QKidney® [36], to 0.4256, for the model developed by Thakkinstian et *al.* [56]. When applying these thresholds to select the 10% of patients with highest predicted risks, QKidney® [36] identified 64.5% of all patients that developed CKD during the study period. Proportions for the other models ranged from 55.2% for the model from Thakkinstian et *al.* [56], to 64.0% for the model of O'seaghdha *et al.* [52].

Table 4.5: Positive predictive value, sensitivity and specificity for simplified scoring systems when applying to the threshold that was proposed in the development study and best threshold on our dataset, calculated using the Youden's method [43]. As QKidney® does not have any associated score in the original publication, we reported results for the full model. O'seaghdha *et al.* [52] reported a simplified score system, however this could not be used in our population because of missing predictors. Therefore, we calculated performance for the full model instead.

Study		**Threshold (SD)**	**PPV (SD)**	**Sensitivity (SD)**	**Specificity (SD)**
Bang *et al.* [54]	Proposed	4	0.146 (0.002)	0.865 (0.004)	0.805 (0.001)
	Best	4	0.146 (0.002)	0.865 (0.004)	0.805 (0.001)
Chien *et al* [51]	Proposed	7	0.106 (0.001)	0.916 (0.003)	0.701 (0.001)
	Best	8	0.133 (0.002)	0.863 (0.004)	0.783 (0.001)
QKidney® [36]	Proposed	NR	NA	NA	NA
	Best	0.017 (0.002)	0.147 (0.006)	0.870 (0.012)	0.805 (0.012)
Kshirsagar et. al. [53]	Proposed	3	0.143 (0.002)	0.872 (0.004)	0.799 (0.001)
	Best	3	0.143 (0.002)	0.872 (0.004)	0.799 (0.001)
Kwon *et al.* [55]	Proposed	4	0.147 (0.002)	0.862 (0.004)	0.807 (0.001)
	Best	4	0.147 (0.002)	0.862 (0.004)	0.807 (0.001)
O'seaghdha *et al.* [52]	Proposed	NA	NA	NA	NA
	Best	0.086 (0.010)	0.138 (0.007)	0.885 (0.015)	0.786 (0.015)
Thakkinstian *et al* [56]	Proposed	5	0.071 (0.001)	0.936 (0.003)	0.529 (0.001)
	Best	6	0.140 (0.002)	0.861 (0.004)	0.796 (0.001)

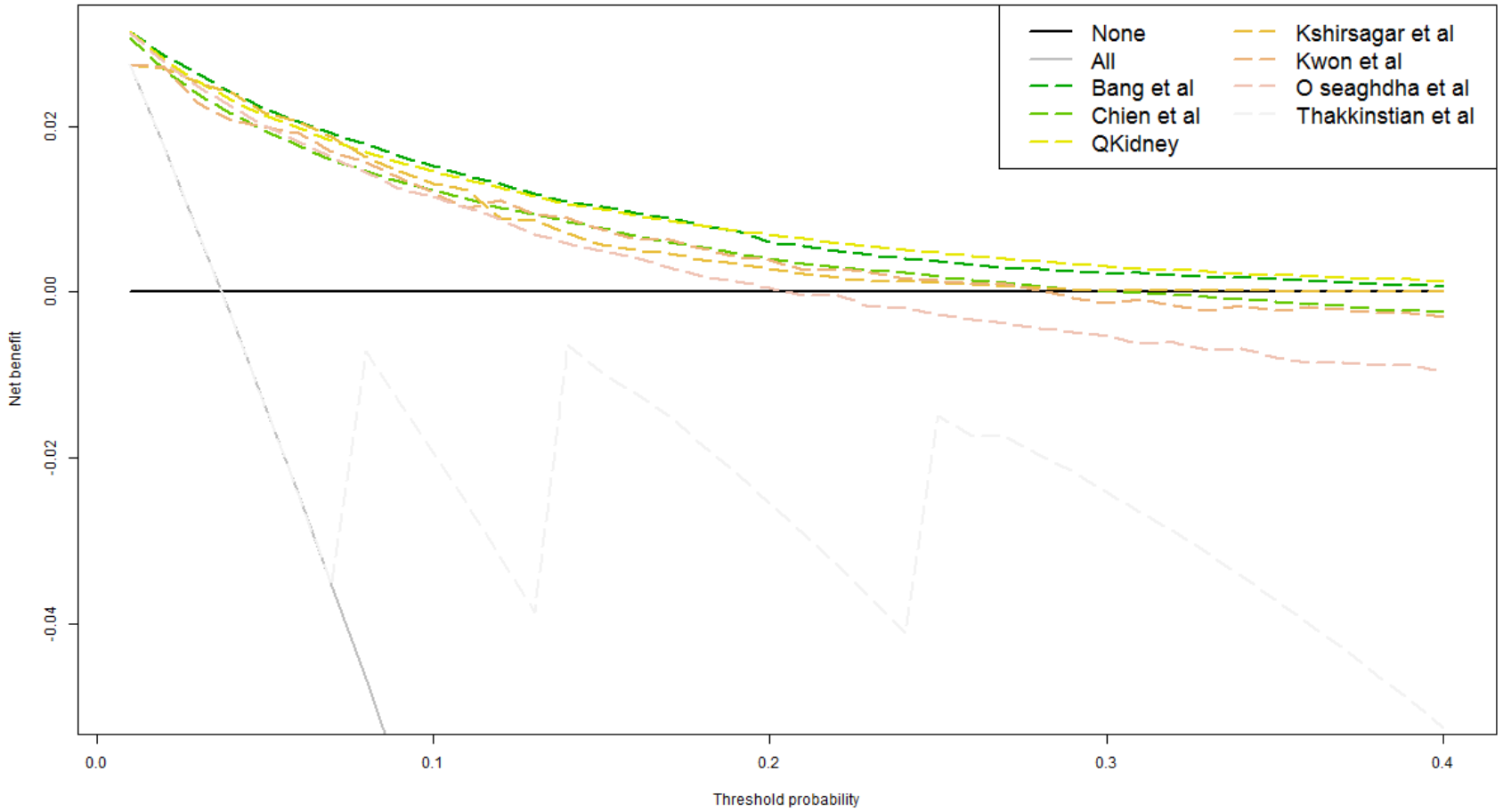
Abbreviations: PPV, positive predictive value; NR, Not reported; NA, not applicable; SD, standard deviation.

Table 4.6: Mean linear predictor, calculated in development datasets and in our validation dataset (patients with complete follow-up data only).

	Study	Development dataset	Validation dataset, patients with complete follow-up (N= 162,653)	
		Mean linear predictor (from summary statistics)	Mean linear predictor (from summary statistics)	Mean linear predictor (SD) (from Individual Patient Data)
MODELS	Bang *et al.* [54]	-3.9	-4.2	-4.2 (1.4)
	Chien *et al* [51][b)]	0.1	-0.5	-0.5 (1.5)
	QKidney® [36][b)]	-0.1	-0.3	-0.1 (1.9)
	Kshirsagar *et. al.* [53]	-3.0	-3.5	-3.5 (0.8)
	Kwon *et al.* [55]	-3.0	-3.4	-3.3 (1.2)
	O'seaghdha *et al.* [52]	-2.3	-3.8	-3.8 (1.9)
	Thakkinstian *et al* [56]	-1.6	-1.8	-1.8 (0.9)

Decision curves for the cohort of patients with complete follow-up are presented in Figure 4.4. The models by Bang *et al.* [54] and QKidney® [36] had the best performance. At predicted probability thresholds lower than 0.5 their net benefit was greater than all other models and greater than strategies labelling all patients at high risk (black line) or none at high risk (grey line). For predicted probability thresholds greater than 0.5 Bang *et al.* [54] and QKidney® [36] were equivalent to the choice of not labelling any patient as high CKD risk (grey line).

Figure 4.4: Decision curve analysis for the cohort of patients with complete follow-up.



4.4.4 Sensitivity analyses

The sensitivity analysis conducted on patients with CKD risk factors showed comparable calibration and MAPE (Bang *et al.* [54] and QKidney® [36] were the only well-calibrated models), with an overall decrease in discrimination of about 0.1 (see Supplementary Table 4.4) compared to our main analysis. Specifically, AUC values on patients with complete follow-up ranged from 0.756 (95% CI 0.749 to 0.762) to 0.801 (95% CI 0.795 to 0.808) while the c-index values for the two Cox regression models were 0.755 (95% CI 0.749 to 0.761) [51] and 0.775 (95% CI 0.769 to 0.781) [36], respectively. The performance of the simplified scoring systems was worse compared to the models from which they were derived.

The sensitivity analysis in which CKD was defined by the presence of only one eGFR measurement lower than 60 ml/min/1.73 m^2 or a diagnostic code for CKD 3-5 led to a higher prevalence of CKD onset (6.6%, n=11,351), with overall predictive model performance that slightly decreased (see Supplementary Table 4.5), especially in terms of calibration. CKD onset prevalence was also higher (3.9%, n= 6,988) when we calculated eGFR by using the CKD-EPI formula, with an increase in absolute numbers of ~1,000 cases and an average age in this group of 76 years (SD 8.1). Overall performance was similar to our main analysis, and only the model by Bang *et al.* [54] was well-calibrated in this sensitivity analysis (see Supplementary Table 4.6). As expected, we witnessed an increase in CKD onset prevalence (7.6%, n= 13,652) when we counted patients that died during follow-up as if they developed CKD, however that did not lead to changes in discriminative performance of the models (see Supplementary Table 4.7). Conversely, calibration improved for all models that were over-predicting CKD in our main analysis. In the analysis restricted to patients with complete data on all predictors we found an overall decrease in performance of about 0.08 for AUCs and c-index (see Supplementary Table 4.8), while the sensitivity analysis that used a 4-year time horizon showed similar discriminative performance to our main analysis, but worse calibration for all models but QKidney® (see Supplementary Table 4.9).

4.5 DISCUSSION

We externally validated and compared seven published models for the prediction of CKD onset [14,15], using a recent five-year window with well-studied EHR data, typical of UK NHS primary care and chronic disease management. All models discriminated well between patients who developed CKD compared with those who did not. Five models had an associated simplified scoring system, each of which had a similar performance to its parent model. Only two models

were well-calibrated to the risk levels in our population [36,54]. Among the 10% of patients with highest predicted risks, 48.0% to 64.5% actually developed CKD.

Two key strengths of this study are: 1) its large sample size; and 2) its cohort being based on a geographically-defined population rather than tied to a particular EHR, which minimizes selection bias at enrolment. In addition, whilst five out of seven models had already been externally validated [17,36,51,54,55,58] and two had been mutually compared [17], our study is the first comprehensive head-to-head comparison of multiple CKD prediction models on a large independent population.

Three previous UK-based studies [36,58,59] have externally validated QKidney® [36] and reported: a c-statistic of 0.87; good calibration; and similar proportions of identified CKD cases among the 10% of patients with highest predicted risks. Although each study externally validated QKidney® [36] using UK primary care EHR data our study extended the validation. Collins *et al.* [59] and Hippisley-Cox and Coupland [36,58] adopted the same inclusion criteria (i.e. patients aged between 35-74 years), CKD definition (i.e. eGFR measurement <45 ml/min/1.73 m^2, kidney transplant; dialysis; nephropathy diagnosis and proteinuria) and stratification by gender adopted in the original development study [36]. Whereas, our study included all adults (aged 18 years and over) and used a more robust definition of the outcome.

A previous study compared the models from Chien *et al*. [51] and Thakkinstian *et al.* [56] in mixed-ancestry South Africans [17]. This study found that the models underestimated CKD risk in this population, while in our external validation both models over-predicted CKD risk. A likely explanation is the difference in CKD onset prevalence between the development cohorts, the cohort from Mogue *et al.* [17] dataset, and our cohort. Specifically, the study population from Mogue *et al.* [17] had a much higher prevalence of CKD cases than these development cohorts while our study population had a lower prevalence.

The included prediction models and simplified scoring systems had remarkably good discriminative ability in our dataset, with better performance than in most of the original studies. This is, on the one hand, surprising because models usually perform similarly or worse in external validation. On the other hand, we used a more robust definition of CKD, requiring impaired eGFR (eGFR<60 ml/min/1.73 m^2) for at least three months, rather than the one that was used in most of the original studies [51–55], which looks at CKD measurements in isolation. The latter definition inflates incidence of CKD diagnosis [36] and therefore leads to a poorer signal-to-noise ratio and a decrease in model performance [61], as shown in our sensitivity analysis (see Supplementary Table 4.5). Another advantage of our definition, which is based on the international KDIGO guidelines [62], is that it is closer to the definition of CKD that is currently used in UK clinical

practice. Along the same lines, we used the MDRD formula to calculate eGFR, which is currently used in UK clinical practice. We also performed a sensitivity analysis to investigate whether using the CKD-EPI formula [48] would have led to different results, which confirmed the findings from Carter *et al.* [63] that the CKD-EPI formula calculates lower eGFR values than the MDRD formula for older patients. Rather than looking at repeated measurements over time, a potential alternative to improve the signal-to-noise ratio would be to use an eGFR threshold lower than 60 ml/min/1.73 m^2. For example, a threshold of 45 ml/min/1.73 m^2 (i.e. the one used for the development of QKidney® [36]) would certainly increase the likelihood of detecting patients with moderate to severe CKD cases. However, only a very small proportion of patients with mild CKD (i.e. eGFR between 45 and 60 ml/min/1.73 m^2) will actually progress towards moderate to severe stages, and models predictions often do not predict health outcomes with very high accuracy in practice. Therefore, there is the risk of focusing on a too small population to have an effective high risk approach to disease prevention [64].

In the complete case analysis, and in the analysis restricted to patients with established CKD risk factors, there was a marked decrease in discriminative performance. In both cases, further to the decrease in sample size, a plausible explanation is that these analyses increased the differences in case-mix between development and validation datasets. The complete case analysis considers only patients without missing predictors, who are more likely to have had healthcare contacts related to their disease. As in the cohort with established CKD risk factors, this excludes a large group of healthy patients, and thus leads a quite different population than the one for which the models were developed. Based on our findings it seems that a different model is needed for patients with established CKD risk factors. Such a model could use other information that is not routinely available in the majority of the low-risk population, like creatinine levels.

We observed an over-prediction of CKD risk by the majority of models, which can be explained largely by differences in case-mix between our validation cohort and the development populations. First, the incidence of CKD in most development datasets was higher than in our validation cohort. As a consequence, the baseline CKD risks calculated (i.e. model intercepts) in the development datasets were too high for our population. Furthermore, as the mean linear predictor analysis showed, our population appeared to be healthier (i.e. lower mean predictor values) than the populations that were used in the development studies. We also found in some models unexpectedly large coefficients for some covariates. For example, three of the included models [53–55] had coefficients for covariates such as anaemia or peripheral vascular disease that were either comparable or larger than more well-established CKD risk factors like diabetes or hypertension. Finally, another possible explanation of models' poor calibration is the adoption of

a slightly different definition for some predictors in this study, in concordance to the ones used in the NHS, when compared to the original studies.

No calibration problems were found for the models by Bang *et al.* [54] and QKidney® [36]. However, we left out an important predictor from the model by Bang *et al.* [54], proteinuria, because it was missing from our dataset. Because the model is well calibrated now, we expect that it would have overpredicted risks if proteinuria had been present. QKidney® [36] was originally developed in the UK primary care (England and Wales), and it was the only model for which the analysed time horizon (five years) was the same as in the development paper. Therefore a good calibration was expected. This was confirmed by the fact that we obtained similar mean linear predictors in our dataset to the ones reported in the original development study (Table 4.6).

Overall, the only model that could conceivably be applied in our population without recalibration was QKidney® [36]. QKidney® consistently outperformed all the other models in terms of both discrimination and calibration, and its performance is comparable to existing validation studies [36,58,59]. The model could be used via the web calculator (www.qkidney.org) or directly integrated into EHRs.

From a methodological perspective, there is room for improvement in CKD prediction modelling. First, future studies should consider to use the CKD definition provided by the international KDIGO guidelines [62]. This should also be used to re-estimate the CKD risk prediction models already available. Second, none of the models included in our analysis accounted for death as a competing risk. We recommend that authors of future models use methodologies [65,66] to do so. Third, authors should take advantage of the new opportunities offered by EHR databases to develop and validate future CKD predicton models [67]. Particularly, besides the possibility of accessing larger sample sizes and to have more predictors, EHRs give the opportunity of observing repeated measurements and account for changes over time of patient's relevant conditions and biomarkers [67,68]. This is particularly important in CKD, where comorbidities and biomerkers like creatinine play a key role. Finally, authors of CKD prediction models should consider the possiblity of using aggregation methods [69,70], which, by considering previous evidence and combining strengths from different prediction models, have been proven to be more effective than new model development [69,70].

Our study has several limitations. First, we excluded eleven models that we identified from the two reviews [14,15] because they included variables not present in our data. However these models were qualitatively less applicable to our prediction population/context than those included. Second, we removed proteinuria from the models by Bang *et al* [54] and Kwon *et al.* [55] because proteinuria was rarely available for patients in our dataset, and this has likely

impaired the estimated performance of these models. Third, we could not reproduce the exact KDIGO definition of CKD because ACR is not routinelly collected in UK primary care. Again these limitations are unlikely to influence the implications of our findings for current practice. Finally, we had missing values for ethnicity and considered patients for which there was no ethnicity information recorded as if they were of White British ethnicity. Poor recording of ethnicity is an acknowledged issue in the NHS [71]. However, because of the regional nature of our data, which covers only the city of Salford (England, UK) where white prevalence is higher than 90% [34], we believe that this did not affect our findings.

4.6 CONCLUSION

To conclude, we have provided an independent, external validation of CKD prediction models with data that will soon be available in most parts of the UK. All included models had good discriminative performance, but most of them were poorly calibrated. Although no model was ideal, QKidney® [36] performed best, and could support a high-risk approach to CKD prevention in primary care. This study underlines the need for ongoing (re)calibration of clinical prediction models in their contexts of use.

4.7 REFERENCES

1. Meguid El Nahas A, Bello AK (2005) Chronic kidney disease: the global challenge. Lancet 365: 331–340. Available: http://www.sciencedirect.com/science/article/pii/S0140673605177897. Accessed 11 June 2014.

2. Barsoum RS (2006) Chronic Kidney Disease in the Developing World. N Engl J Med 354: 997–999. Available: http://dx.doi.org/10.1056/NEJMp058318.

3. Schoolwerth AC, Engelgau MM, Hostetter TH, Rufo KH, Chianchiano D, et al. (2006) Chronic kidney disease: a public health problem that needs a public health action plan. Prev Chronic Dis 3: A57. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1563984&tool=pmcentrez&rendertype=abstract. Accessed 12 June 2014.

4. Mills KT, Xu Y, Zhang W, Bundy JD, Chen C-S, et al. (2015) A systematic analysis of worldwide population-based data on the global burden of chronic kidney disease in 2010. Kidney Int 88: 950–957. Available: http://dx.doi.org/10.1038/ki.2015.230.

5. Coresh J, Selvin E, Stevens LA, Manzi J, Kusek JW, et al. (2007) Prevalence of chronic kidney disease in the United States. JAMA 298: 2038–2047. Available: http://jama.jamanetwork.com/article.aspx?articleid=209357. Accessed 12 June 2014.

6. United States Renal Data System (2013) USRDS 2013 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States. Available: http://www.usrds.org/atlas.aspx.

7. Jameson K, Jick S, Hagberg KW, Ambegaonkar B, Giles A, et al. (2014) Prevalence and management of chronic kidney disease in primary care patients in the UK. Int J Clin Pract 68: 1110–1121. Available: http://www.ncbi.nlm.nih.gov/pubmed/24852335. Accessed 25 November 2014.

8. Kerr M, Bray B, Medcalf J, O'Donoghue DJ, Matthews B (2012) Estimating the financial cost of chronic kidney disease to the NHS in England. Nephrol Dial Transplant 27 Suppl 3: iii73-80. Available: http://ndt.oxfordjournals.org/content/27/suppl_3/iii73.long. Accessed 24 May 2014.

9. Grams ME, Coresh J (2013) Assessing risk in chronic kidney disease: a methodological review. Nat Rev Nephrol 9: 18–25. Available: http://dx.doi.org/10.1038/nrneph.2012.248.

Accessed 23 September 2014.

10. Radhakrishnan J, Remuzzi G, Saran R, Williams DE, Rios-Burrows N, et al. (2014) Taming the chronic kidney disease epidemic: a global view of surveillance efforts. Kidney Int 86: 246–250. Available: http://www.ncbi.nlm.nih.gov/pubmed/24897034. Accessed 28 July 2014.

11. National Institute for Healrh and Care Excellence (NICE) UK (2014) Chronic kidney disease: early identification and management of chronic kidney disease in adults in primary and secondary care. Available: https://www.nice.org.uk/guidance/cg182.

12. Glassock RJ, Winearls CG (2008) Routine reporting of estimated glomerular filtration rate: not ready for prime time. Nat Clin Pr Neph 4: 422–423. Available: http://dx.doi.org/10.1038/ncpneph0860.

13. Steyerberg EW (2009) Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Springer. doi:10.1007/978-0-387-77244-8.

14. Collins GS, Omar O, Shanyinde M, Yu L-M (2013) A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. J Clin Epidemiol 66: 268–277. Available: http://www.sciencedirect.com/science/article/pii/S0895435612002636. Accessed 8 May 2014.

15. Echouffo-Tcheugui JB, Kengne AP (2012) Risk models to predict chronic kidney disease and its progression: a systematic review. PLoS Med 9: e1001344. Available: http://dx.plos.org/10.1371/journal.pmed.1001344. Accessed 28 April 2014.

16. Collins GS, Moons KGM (2012) Comparing risk prediction models. BMJ 344: e3186. Available: http://www.bmj.com/content/344/bmj.e3186. Accessed 13 January 2015.

17. Mogueo A, Echouffo-Tcheugui JB, Matsha TE, Erasmus RT, Kengne AP (2015) Validation of two prediction models of undiagnosed chronic kidney disease in mixed-ancestry South Africans. BMC Nephrol 16: 1–14. Available: http://dx.doi.org/10.1186/s12882-015-0093-6.

18. Collins GS, Reitsma JB, Altman DG, Moons KGM (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med 13: 1–10. Available: http://dx.doi.org/10.1186/s12916-014-0241-z.

19. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, et al. (2015) Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and ElaborationThe TRIPOD Statement: Explanation and Elaboration. Ann Intern Med 162: W1–W73. Available: +.

20. Ando M, Yanagisawa N, Ajisawa A, Tsuchiya K, Nitta K (2011) A simple model for predicting incidence of chronic kidney disease in HIV-infected patients. Clin Exp Nephrol 15: 242–247. Available: http://www.ncbi.nlm.nih.gov/pubmed/21246239. Accessed 5 August 2014.

21. Stevens PE, Levin A (2013) Evaluation and Management of Chronic Kidney Disease: Synopsis of the Kidney Disease: Improving Global Outcomes 2012 Clinical Practice Guideline. Ann Intern Med 158: 825–830. Available: http://dx.doi.org/10.7326/0003-4819-158-11-201306040-00007.

22. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, et al. (1999) A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. Ann Intern Med 130: 461–470. Available: http://www.ncbi.nlm.nih.gov/pubmed/10075613. Accessed 4 November 2014.

23. Bello A, Thompson S, Lloyd A, Hemmelgarn B, Klarenbach S, et al. (2012) Multiple versus single and other estimates of baseline proteinuria status as predictors of adverse outcomes in the general population. Am J Kidney Dis 59: 364–371. Available: http://www.sciencedirect.com/science/article/pii/S0272638611013254. Accessed 5 March 2015.

24. NHS England (n.d.) Read Codes. Available: http://www.connectingforhealth.nhs.uk/systemsandservices/data/uktc/readcodes. Accessed 16 June 2014.

25. Roland M (2004) Linking physicians' pay to the quality of care--a major experiment in the United kingdom. N Engl J Med 351: 1448–1454. Available: http://www.ncbi.nlm.nih.gov/pubmed/15459308. Accessed 13 November 2014.

26. Sutton M, Elder R, Guthrie B, Watt G (2010) Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers. Health Econ 19: 1–13. Available: http://dx.doi.org/10.1002/hec.1440.

27. Taggar JS, Coleman T, Lewis S, Szatkowski L (2012) The impact of the Quality and Outcomes Framework (QOF) on the recording of smoking targets in primary care medical

records: cross-sectional analyses from The Health Improvement Network (THIN) database. BMC Public Health 12: 329. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4104830/.

28. Springate DA, Kontopantelis E, Ashcroft DM, Olier I, Parisi R, et al. (2014) ClinicalCodes: an online clinical codes repository to improve the validity and reproducibility of research using electronic medical records. PLoS One 9: e99825. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4062485&tool=pmcentrez&rendertype=abstract. Accessed 12 November 2014.

29. Doran T, Kontopantelis E, Valderas JM, Campbell S, Roland M, et al. (2011) Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. BMJ 342: d3590. Available: http://www.bmj.com/content/342/bmj.d3590. Accessed 28 November 2014.

30. Kontopantelis E, Springate D, Reeves D, Ashcroft DM, Valderas JM, et al. (2014) Withdrawing performance indicators: retrospective analysis of general practice performance under UK Quality and Outcomes Framework. BMJ 348: g330. Available: http://www.bmj.com/content/348/bmj.g330?ijkey=AwxEb4LKHleAKNk&keytype=ref. Accessed 18 December 2014.

31. Fairhurst C, Watt I, Martin F, Bland M, Brackenbury WJ (2014) Exposure to sodium channel-inhibiting drugs and cancer survival: protocol for a cohort study using the QResearch primary care database. BMJ Open 4: e006604. Available: http://bmjopen.bmj.com/content/4/11/e006604.full. Accessed 18 December 2014.

32. Nicholson A, Ford E, Davies KA, Smith HE, Rait G, et al. (2013) Optimising use of electronic health records to describe the presentation of rheumatoid arthritis in primary care: a strategy for developing code lists. PLoS One 8: e54878. Available: http://dx.plos.org/10.1371/journal.pone.0054878. Accessed 3 January 2015.

33. Crooks CJ, West J, Card TR (2013) Comorbidities affect risk of nonvariceal upper gastrointestinal bleeding. Gastroenterology 144: 1384–1393, 1393-2-9. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3776918&tool=pmcentrez&rendertype=abstract. Accessed 16 April 2014.

34. Salford City Council (2011) BME communities. Available: http://www.salford.gov.uk/bmecommunities.htm. Accessed 21 January 2015.

35. Buuren van S, Groothuis-Oudshoorn K (2011) mice: Multivariate Imputation by Chained Equations in R. J Stat Softw. Available: http://doc.utwente.nl/78938/1/Buuren11mice.pdf. Accessed 9 June 2014.

36. Hippisley-Cox J, Coupland C (2010) Predicting the risk of chronic Kidney Disease in men and women in England and Wales: prospective derivation and external validation of the QKidney Scores. BMC Fam Pract 11: 49. Available: http://www.biomedcentral.com/1471-2296/11/49. Accessed 8 May 2014.

37. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, et al. (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12: 77. Available: http://www.biomedcentral.com/1471-2105/12/77. Accessed 23 May 2014.

38. Newson R (2006) Confidence intervals for rank statistics: Somers' D and extensions. Stata J 6: 309–334. Available: http://www.stata-journal.com/article.html?article=snp15_6.

39. Jr Harrell FE (2014) Package "Hmisc." Available: http://cran.r-project.org/web/packages/Hmisc/Hmisc.pdf.

40. Verburg IWM, de Keizer NF, de Jonge E, Peek N (2014) Comparison of Regression Methods for Modeling Intensive Care Length of Stay. PLoS One 9: e109684. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4215850/.

41. Cox DR (1958) Two Further Applications of a Model for Binary Regression. Biometrika 45: 562–565 CR–Copyright © 1958 Biometrika Tru. Available: http://www.jstor.org/stable/2333203.

42. Kundu S, Aulchenko YS, van Duijn CM, Janssens ACJW (2011) PredictABEL: an R package for the assessment of risk prediction models. Eur J Epidemiol 26: 261–264. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3088798&tool=pmcentrez&rendertype=abstract. Accessed 10 June 2014.

43. Youden WJ (1950) Index for rating diagnostic tests. Cancer 3: 32–35. Available: http://doi.wiley.com/10.1002/1097-0142(1950)3:1%3C32::AID-CNCR2820030106%3E3.0.CO;2-3. Accessed 4 June 2014.

44. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, et al. (2015) A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol 68: 279–289. Available:

http://www.sciencedirect.com/science/article/pii/S0895435614002753.

45. Vickers AJ, Elkin EB (2006) Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 26: 565–574. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2577036/.

46. Rose G (2001) Sick individuals and sick populations. Int J Epidemiol 30: 427–432. Available: http://ije.oxfordjournals.org/content/30/3/427.abstract.

47. Moyer VA (2012) Screening for chronic kidney disease: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 157: 567–570. Available: http://annals.org/article.aspx?articleid=1355168. Accessed 21 January 2015.

48. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, et al. (2009) A new equation to estimate glomerular filtration rate. Ann Intern Med 150. Available: http://dx.doi.org/10.7326/0003-4819-150-9-200905050-00006.

49. Fox CS, Gona P, Larson MG, Selhub J, Tofler G, et al. (2010) A multi-marker approach to predict incident CKD and microalbuminuria. J Am Soc Nephrol 21: 2143–2149. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3014027&tool=pmcentrez&rendertype=abstract. Accessed 1 May 2014.

50. Halbesma N, Jansen DF, Heymans MW, Stolk RP, de Jong PE, et al. (2011) Development and validation of a general population renal risk score. Clin J Am Soc Nephrol 6: 1731–1738. Available: http://cjasn.asnjournals.org/content/6/7/1731.long. Accessed 7 May 2014.

51. Chien K-L, Lin H-J, Lee B-C, Hsu H-C, Lee Y-T, et al. (2010) A prediction model for the risk of incident chronic kidney disease. Am J Med 123: 836–846.e2. Available: http://www.sciencedirect.com/science/article/pii/S0002934310004602. Accessed 8 May 2014.

52. O'Seaghdha CM, Lyass A, Massaro JM, Meigs JB, Coresh J, et al. (2012) A risk score for chronic kidney disease in the general population. Am J Med 125: 270–277. Available: http://www.sciencedirect.com/science/article/pii/S0002934311008370. Accessed 8 May 2014.

53. Kshirsagar A V, Bang H, Bomback AS, Vupputuri S, Shoham DA, et al. (2008) A simple algorithm to predict incident kidney disease. Arch Intern Med 168: 2466–2473. Available:

http://archinte.jamanetwork.com/article.aspx?articleid=773460. Accessed 8 May 2014.

54. Bang H, Vupputuri S, Shoham DA, Klemmer PJ, Falk RJ, et al. (2007) SCreening for Occult REnal Disease (SCORED): a simple prediction model for chronic kidney disease. Arch Intern Med 167: 374–381. Available: http://archinte.jamanetwork.com/article.aspx?articleid=411731. Accessed 8 May 2014.

55. Kwon K-S, Bang H, Bomback AS, Koh D-H, Yum J-H, et al. (2012) A simple prediction score for kidney disease in the Korean population. Nephrology (Carlton) 17: 278–284. Available: http://www.ncbi.nlm.nih.gov/pubmed/22171932. Accessed 8 May 2014.

56. Thakkinstian A, Ingsathit A, Chaiprasert A, Rattanasiri S, Sangthawan P, et al. (2011) A simplified clinical prediction score of chronic kidney disease: a cross-sectional-survey study. BMC Nephrol 12: 45. Available: http://www.biomedcentral.com/1471-2369/12/45. Accessed 8 May 2014.

57. O'Seaghdha CM, Yang Q, Wu H, Hwang S-J, Fox CS (2012) Performance of a genetic risk score for CKD stage 3 in the general population. Am J Kidney Dis 59: 19–24. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3242901&tool=pmcentrez&rendertype=abstract. Accessed 5 August 2014.

58. Hippisley-Cox J, Coupland C, Brindle P (2014) The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study. BMJ Open 4. Available: http://bmjopen.bmj.com/content/4/8/e005809.abstract.

59. Collins G, Altman D (2012) Predicting the risk of chronic kidney disease in the UK: an evaluation of QKidney® scores using a primary care database. Br J Gen Pract 62: e243-50. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3310030&tool=pmcentrez&rendertype=abstract. Accessed 12 June 2014.

60. de Lusignan S, Tomson C, Harris K, van Vlymen J, Gallagher H (2011) Creatinine fluctuation has a greater effect than the formula to estimate glomerular filtration rate on the prevalence of chronic kidney disease. Nephron Clin Pract 117: c213-24. Available: http://www.karger.com/Article/FullText/320341. Accessed 10 September 2014.

61. Walsh C, Hripcsak G (2014) The effects of data sources, cohort selection, and outcome definition on a predictive model of risk of thirty-day hospital readmissions. J Biomed

Inform 52: 418–426. Available: http://www.sciencedirect.com/science/article/pii/S1532046414001877.

62. Levin A, Stevens PE (2014) Summary of KDIGO 2012 CKD Guideline: behind the scenes, need for guidance, and a framework for moving forward. Kidney Int 85: 49–61. Available: http://dx.doi.org/10.1038/ki.2013.444. Accessed 25 November 2014.

63. Carter JL, Stevens PE, Irving JE, Lamb EJ (2011) Estimating glomerular filtration rate: comparison of the CKD-EPI and MDRD equations in a large UK cohort with particular emphasis on the effect of age. QJM 104: 839–847. Available: http://qjmed.oxfordjournals.org/content/104/10/839.abstract.

64. Rose G (2001) Sick individuals and sick populations. Int J Epidemiol 30: 427–432. Available: http://dx.doi.org/10.1093/ije/30.3.427.

65. Satagopan JM, Ben-Porat L, Berwick M, Robson M, Kutler D, et al. (2004) A note on competing risks in survival data analysis. Br J Cancer 91: 1229–1235. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2410013/.

66. Putter H, Fiocco M, Geskus RB (2007) Tutorial in biostatistics: competing risks and multi-state models. Stat Med 26.

67. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA (2016) Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Informatics Assoc. Available: http://jamia.oxfordjournals.org/content/early/2016/05/17/jamia.ocw042.abstract.

68. Akbarov A, Williams R, Brown B, Mamas M, Peek N, et al. (2015) A Two-stage Dynamic Model to Enable Updating of Clinical Risk Prediction from Longitudinal Health Record Data: Illustrated with Kidney Function. Stud Health Technol Inform 216: 696–700. doi:10.3233/978-1-61499-564-7-696.

69. Debray TPA, Koffijberg H, Vergouwe Y, Moons KGM, Steyerberg EW (2012) Aggregating published prediction models with individual participant data: a comparison of different approaches. Stat Med 31: 2697–2712.

70. Martin GP, Mamas MA, Peek N, Buchan I, Sperrin M (2017) Clinical prediction in defined populations: a simulation study investigating when and how to aggregate existing models. BMC Med Res Methodol 17: 1. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5217317/.

71. Hull SA, Mathur R, Badrick E, Robson J, Boomla K (2011) Recording ethnicity in primary care: assessing the methods and impact. Br J Gen Pract 61: e290-4. Available: http://bjgp.org/content/61/586/e290.abstract. Accessed 13 January 2015.

4.8 SUPPLEMENTARY MATERIAL

Supplementary Table 4.1: Population characteristics in included models' original studies.

	Bang *et al.* [54]	**Chien *et al*** [51]			**Kshirsagar *et. al.*** [53]		**Kwon *et al.*** [55]	**O'seaghdha *et al.*** [52]	**Thakkinstian *et al*** [56]	**QKidney** [36]
Parameter	**Total**	**Total**	**No CKD**	**CKD**	**No CKD**	**CKD**	**Total**	**Total**	**Total**	**Total**
Number of patients	8,530	4,978	5,168	190	9,470	1,605	6,565	2,490	3,459	1,574,749
Number of cases (%)	601 (7.5)	190 (3.7)	/	/	/	/	100 (1.5)	229 (9.2)	606 (17.5)	23,786 (1.5)
Age (mean, SD)	46 (31.4)	51.2 (10.5)	50.8 (10.3)	60.9 (9.4)	57 (9)	62 (9)	44.2 (0.4)	57.1 (8.9)	45.2 (0.79)	47.3 (11.1)
Gender - Female (%)	52%	36.7%	36.9%	32.1%	56%	59%	50.0%	53.0%	54.5%	49.22%
Townsend score (mean, SD)										-0.5 (3.4)
Ethnicity	White 72%	Chinese 100%	Chinese 100%	Chinese 100%	White 78%	White 83%	Asian 100%	White 100%	Asian 100%	White 95%
Smoking status										Non-smoker 51.11% Ex-smoker 18.21 Light smoker 6.43% Moderate smoker 7.89% Heavy smoker 6.41%
BMI [kg/m^2] (mean, SD)	28.0 (12.9)	23.9 (3.1)	23.9 (3.1)	24.7 (3.0)			23.6 (0.1)		24 (0.2)	26.7 (4.7)
Diastolic blood pressure [mmHg] (mean, SD)		73.7(10.5)	73.5(10.5)	77.6 (11.0)	73 (11)	73 (12)				
Systolic blood pressure [mmHg] (mean, SD)		123.2 (15.8)	122.9 (15.7)	131.3 (16.2)	122 (19)	132 (22)		126 (18)		133.1 (19.5)
eGFR [mL/min/1.73 m^2] (mean, SD)	94 (48.9)	82.2 (13.1)	82.7(13.0)	68.3(7.0)			85.9 (0.7)	92 (23)		
Anemia (%)	2.7%	2.9%	2.9%	1.6%	2%	3%	8.1%	7.4%		
Postpondrial glucose [mg/dl] (mean, SD)		118.4(52.5)	118.0 (51.2)	130.9 (77.7)						
HbA1c [%](mean, SD)		5.52 (0.71)	5.51(0.70)	5.78 (0.98)						
Uric acid [mg/dl] (mean, SD)		6.11(1.46)	6.09 (1.46)	6.55 (1.47)						
CRP [mg/dl] (mean, SD)		0.15 (0.37)	0.15 (0.35)	0.23 (0.69)						
Anemia (%)	2.7%	2.9%	2.9%	1.6%	2%	3%	8.1%	7.4%		
Proteinuria (%)	10%						10.3%	17.1%		
Quantitative albuminuria (%)								9.2%		
High-density lipoprotein cholesterol level [mg/dl] (mean, SD)	51.3 (34.2)									
Hypertension (%)	34%	23.9%	23.0%	47.9%	36%	55%	22.5%	35.3%	27.5%	
Hypertensive treatment (%)										9.94%
Diabetes Mellitus (%)	8%	Type 2 12.2%	Type 2 11.8%	Type 2 23.7%	9%	17%	8.3%		11.9%	Type 1 0.28% Type 2 3.12%
History of cardiovascular disease (%)					8%	17%	3.2%	8.3%		4.5%
History of heart failure (%)					0.7	2.3				0.5%
History of stroke (%)		0.4%	0.3%	2.1%						
Peripheral vascular disease (%)	2.7%				4%	9%				1%
Kidney stones (%)									5.0%	0.68%
Rheumatoid arthritis (%)										0.76%
Systemic lupus erythematosus (%)										0.07%
NSAIDs use (%)										27.04%
Family history of kidney disease (%)										0.05

Supplementary Table 4.2: Coefficients of validated models. QKidney included also Townsend score, fractional polynomial terms for age, body mass index and systolic blood pressure as well as interactions between the age terms and type 1 diabetes, type 2 diabetes and treated hypertension (full details in supplementary file 1). *Cox proportional hazard model (baseline hazard=0.9632).**Intercept estimated from summary measures and CKD prevalence from development dataset.

Parameter	**Bang *et al.*** [54]	**Chien *et al**** [51]	**Kshirsagar *et. al.*** [53]	**Kwon *et al.*** [55]	**O'seaghdha *et al.*** [52]	**Thakkinstian *et al*** [56]	**QKidney®** [36]		
Intercept	-5.4	-6.8	-4.1**	-4.4**	-8.3**	-2.8	**F**	**M**	**Strata**
Age	1.55 [50-59] 2.31 [60-69] 3.23 [>=70]	0.08 [per year]	0.63 [50-59] 1.33 [60-69] 1.46 [>70]	1.16 [50-59] 1.91 [60-69] 2.71 [>70]	0.05 [per year]	0.6 [40-59] 1.4 [60-69] 2.1 [>70]	Included	Included	
Gender - Female	0.29		0.13	0.4					
Anaemia	0.93		0.48	0.94					
Hypertension (y/n)	0.45		0.55	0.48	0.32	0.8			
Type 1 Diabetes Mellitus (y/n)	0.44		0.33	0.73		0.9	2.10	2.51	
Type 2 Diabetes Mellitus (y/n)	0.44	0.37	0.33	0.73	0.29	0.9	1.50	1.80	
History of cardiovascular disease (y/n)	0.59		0.26	0.60			0.32	0.33	
History of heart failure (y/n)	0.45		0.50				0.82	1.04	
Hystory of stroke (y/n)		1.24							
Peripheral vascular disease (y/n)	0.74		0.41				0.30	0.38	
Proteinuria	0.83			0.48					
BMI		0.06 [per kg/m^2]					Included	Included	
Diastolic blood pressure		0.02 [per mmHg]							
Kidney stones						1			
Ethnicity							0.30 0.44 0.40 0.15 -0.73 -0.57 0.12 0.20	0.15 0.69 0.30 0.36 -0.19 0.16 0.31 0.29	[Indian] [Pakistani] [Bangladeshi] [Other Asian] [Black Caribbean] [Black African] [Chinese] [Other ethnic group]
Treated hypertension (yes/no)							0.91	1.02	
NSAID use (yes/no)							0.26	0.25	
Family history of kidney disease (y/n)							0.75	1.27	
Rheumatoid arthritis (y/n)							0.48		
Systemic lupus erythematosus (y/n)							0.88		
Smoking							0.17 0.27 0.24 0.36	0.12 0.14 0.21 0.22	[Ex-smoker] [Light smoker] [Moderate smoker] [Heavy smoker]
Systolic blood pressure							Included	Included	
Townsend score							Included	Included	

Supplementary Table 4.3: Weights of validated simplified scoring systems.

Parameter	**Bang *et al.*** [54]	**Chien *et al*** [51]	**Kshirsagar *et. al.*** [53]	**Kwon *et al.*** [55]	**Thakkinstian *et al*** [56]
Age	2 [50-59] 3 [60-69] 4 [>=70]	3 [45-54] 5 [55-64] 8 [>=65]	1 [50-59] 2 [60-69] 3 [>=70]	2 [50-59] 3 [60-69] 4 [>=70]	1 [<40] 2 [40-59] 4 [60-69] 8 [>=70]
Gender - Female	1		1	1	
Anaemia	1		1	1	
Hypertension (yes/no)	1		1	1	1 [No] 2 [Yes]
Type 1 Diabetes Mellitus (yes/no)	1		1	1	1 [No] 3 [Yes]
Type 2 Diabetes Mellitus (yes/no)	1	1	1	1	1 [No] 3 [Yes]
History of cardiovascular disease (yes/no)	1		1	1	
History of heart failure (yes/no)	1		1		
Hystory of stroke (yes/no)		4			
Peripheral vascular disease (yes/no)	1		1		
Proteinuria	1			1	
BMI		1 [21-25] 2 [>=26]			
Diastolic blood pressure		1 [66-79] 2 [>=80]			
Kidney stones					1 [No] 3 [Yes]

Supplementary Table 4.4: Discrimination, MAPE and calibration slopes of included models in patients with established risk factors for CKD at inclusion, computed in patients with complete follow-up data (all models and risk scores) and in the full validation cohort (Cox proportional hazards regression models only).

	Study	**Patients with complete follow-up (N= 44,183)**			**Full validation cohort (N= 49,002)**	
		AUC [CI]	**MAPE (SD) [a)]**	**Calibration slope [CI]**	**c-index [CI]**	**MAPE (SD) [a)]**
MODELS	Bang *et al.* [54]	0.795 [0.798,0.802]	0.166 (0.24)	0.93 [0.91,0.95]	NA	NA
	Chien *et al* [51][b)]	0.781 [0.775,0.787]	0.196 (0.232)	0.64 [0.62,0.66]	0.755[0.749,0.761]	0.166 (0.24)
	QKidney® [36][b)]	0.801 [0.795,0.808]	0.144 (0.255)	0.93 [0.91,0.95]	0.775[0.769,0.781]	0.196 (0.232)
	Kshirsagar *et. al.* [53]	0.779 [0.773,0.786]	0.162 (0.252)	1.81 [1.77,1.86]	NA	NA
	Kwon *et al.* [55]	0.794 [0.788,0.800]	0.212 (0.211)	0.69 [0.67,0.71]	NA	NA
	O'seaghdha *et al.* [52]	0.796 [0.790,0.803]	0.207 (0.221)	0.54 [0.53,0.56]	NA	NA
	Thakkinstian *et al* [56]	0.756 [0.749,0.762]	0.23 (0.227)	0.49 [0.47,0.50]	NA	NA
SIMPLIFIED SCORES	Bang *et al.* [54]	0.786 [0.780,0.792]	NA	NA	NA	NA
	Chien *et al* - [51]	0.743 [0.736,0.749]	NA	NA	NA	NA
	Kshirsagar et. al. [53]	0.785 [0.779,0.790]	NA	NA	NA	NA
	Kwon *et al.* [55]	0.784 [0.778,0.790]	NA	NA	NA	NA
	Thakkinstian *et al* [56]	0.763 [0.756,0.770]	NA	NA	NA	NA

Abbreviations: AUC, area under receiver operating characteristic curve; eGFR, estimated glomerular filtration rate; NA, not applicable; SD, standard deviation; CI, 95% confidence interval.

a) Calculated as mean difference between observed and predicted CKD cases

b) Cox proportional hazard regression model

c) Difference in lower

Supplementary Table 4.5: Models and scores' AUC and c-index discrimination as well as MAPE by adopting the CKD definition of eGFR<60 ml/min/1.73 m2 on a single occasion as well as a CKD 3-5 diagnostic code.

	Study	Patients with complete follow-up (N= 159,593)			Patients with incomplete follow-up (N= 172,907)	
		AUC [CI]	MAPE (SD) [a)]	Calibration slope [CI]	c-index [CI]	MAPE (SD) [a)]
MODELS	Bang *et al.* [54]	0.855 [0.851,0.859]	0.090 (0.218)	1.43 [1.41,1.44]	NA	NA
	Chien *et al* [51][b)]	0.868 [0.865,0.871]	0.101 (0.210)	1.04 [1.03,1.05]	0.856 [0.853,0.859]	0.100 (0.207)
	QKidney® [36][b)]	0.879 [0.876,0.882]	0.077 (0.228)	1.55 [1.53,1.57]	0.867 [0.864,0.870]	0.074 (0.222)
	Kshirsagar *et. al.* [53]	0.856 [0.852,0.859]	0.097 (0.224)	2.70 [2.67,2.73]	NA	NA
	Kwon *et al.* [55]	0.852 [0.848,0.855]	0.110 (0.204)	1.02 [1.01,1.03]	NA	NA
	O'seaghdha *et al.* [52]	0.876 [0.873,0.879]	0.104 (0.205)	0.84 [0.83,0.84]	NA	NA
	Thakkinstian *et al* [56]	0.860 [0.856,0.863]	0.186 (0.178)	0.72 [0.71,0.73]	NA	NA
SIMPLIFIED SCORES	Bang *et al.* [54]	0.851 [0.847,0.855]	NA	NA	NA	NA
	Chien *et al* - [51]	0.851 [0.847,0.854]	NA	NA	NA	NA
	Kshirsagar et. al. [53]	0.841 [0.837,0.845]	NA	NA	NA	NA
	Kwon *et al.* [55]	0.851 [0.847,0.855]	NA	NA	NA	NA
	Thakkinstian *et al* [56]	0.834 [0.830,0.838]	NA	NA	NA	NA

Abbreviations: AUC, area under receiver operating characteristic curve; eGFR, estimated glomerular filtration rate; NA, not applicable; SD, standard deviation; CI, 95% confidence interval.

a) Calculated as mean difference between observed and predicted CKD cases

b) Cox proportional hazard regression model

Supplementary Table 4.6: Models and scores' AUC and c-index discrimination as well as MAPE by calculating eGFR with the CKD-EPI formula.

	Study	Patients with complete follow-up (N=161,949)			Patients with incomplete follow-up (N=177,102)	
		AUC [CI]	MAPE (SD) [a)]	Calibration slope [CI]	c-index [CI]	MAPE (SD) [a)]
MODELS	Bang *et al.* [54]	0.908 [0.905,0.911]	0.069 (0.17)	1.05 [1.04,1.06]	NA	NA
	Chien *et al* [51][b)]	0.910 [0.907,0.913]	0.082 (0.168)	0.8 [0.79,0.8]	0.9[0.897,0.903]	0.085 (0.169)
	QKidney® [36][b)]	0.920 [0.918,0.922]	0.053 (0.176)	1.26 [1.25,1.27]	0.91[0.907,0.913]	0.054 (0.173)
	Kshirsagar *et. al.* [53]	0.905 [0.902,0.908]	0.074 (0.174)	1.96 [1.94,1.98]	NA	NA
	Kwon *et al.* [55]	0.905 [0.902,0.908]	0.093 (0.165)	0.74 [0.74,0.75]	NA	NA
	O'seaghdha *et al.* [52]	0.920 [0.917,0.923]	0.088 (0.169)	0.66 [0.65,0.66]	NA	NA
	Thakkinstian *et al* [56]	0.902 [0.899,0.905]	0.177 (0.159)	0.54 [0.53,0.54]	NA	NA
SIMPLIFIED SCORES	Bang *et al.* [54]	0.895 [0.891,0.899]	NA	NA	NA	NA
	Chien *et al* - [51]	0.880 [0.876,0.884]	NA	NA	NA	NA
	Kshirsagar et. al. [53]	0.891 [0.887,0.895]	NA	NA	NA	NA
	Kwon *et al.* [55]	0.895 [0.891,0.899]	NA	NA	NA	NA
	Thakkinstian *et al* [56]	0.869 [0.864,0.874]	NA	NA	NA	NA

Abbreviations: AUC, area under receiver operating characteristic curve; eGFR, estimated glomerular filtration rate; NA, not applicable; SD, standard deviation; CI, 95% confidence interval.

a) Calculated as mean difference between observed and predicted CKD cases

b) Cox proportional hazard regression model

Supplementary Table 4.7: Models and scores' AUC and c-index discrimination as well as MAPE for the sensitivity analysis by considering patients who died as if they developed CKD.

	Study	Patients with complete follow-up (N=169,548)			Patients with incomplete follow-up (N=179,072)	
		AUC [CI]	MAPE (SD) [a)]	Calibration slope [CI]	c-index [CI]	MAPE (SD) [a)]
MODELS	Bang *et al.* [54]	0.886 [0.883,0.889]	0.098 (0.219)	1.39 [1.38,1.41]	NA	NA
	Chien *et al* [51][b)]	0.888 [0.885,0.891]	0.109 (0.212)	1.05 [1.04,1.06]	0.875 [0.873,0.878]	0.105 (0.207)
	QKidney® [36][b)]	0.899 [0.896,0.902]	0.084 (0.230)	1.48 [1.47,1.49]	0.886 [0.884,0.8819]	0.079 (0.225)
	Kshirsagar *et. al.* [53]	0.881 [0.878,0.884]	0.107 (0.232)	2.72 [2.70,2.74]	NA	NA
	Kwon *et al.* [55]	0.881 [0.878,0.884]	0.119 (0.205)	1.02 [1.01,1.03]	NA	NA
	O'seaghdha *et al.* [52]	0.900 [0.897,0.903]	0.112 (0.204)	0.87 [0.86,0.88]	NA	NA
	Thakkinstian *et al* [56]	0.876 [0.873,0.879]	0.196 (0.181)	0.76 [0.75,0.77]	NA	NA
SIMPLIFIED SCORES	Bang *et al.* [54]	0.877 [0.874,0.880]	NA	NA	NA	NA
	Chien *et al* - [51]	0.866 [0.863,0.869]	NA	NA	NA	NA
	Kshirsagar et. al. [53]	0.868 [0.865,0.871]	NA	NA	NA	NA
	Kwon *et al.* [55]	0.876 [0.873,0.879]	NA	NA	NA	NA
	Thakkinstian *et al* [56]	0.868 [0.864,0.872]	NA	NA	NA	NA

Abbreviations: AUC, area under receiver operating characteristic curve; eGFR, estimated glomerular filtration rate; NA, not applicable; SD, standard deviation; CI, 95% confidence interval.

a) Calculated as mean difference between observed and predicted CKD cases

b) Cox proportional hazard regression model

Supplementary Table 4.8: Models and scores' AUC and c-index discrimination as well as MAPE by focusing only on patients without missing values in any predictor (complete cases).

	Study	Patients with complete follow-up (N=36,092)			Patients with incomplete follow-up (N=39,283)	
		AUC [CI]	MAPE (SD) [a)]	Calibration slope [CI]	c-index [CI]	MAPE (SD) [a)]
MODELS	Bang *et al.* [54]	0.840 [0.834,0.846]	0.13 (0.223)	0.95 [0.925,0.97]	NA	NA
	Chien *et al* [51][b)]	0.828 [0.822,0.834]	0.158 (0.219)	0.65 [0.632,0.67]	0.804[0.798,0.81]	0.168 (0.222)
	QKidney® [36][b)]	0.846 [0.840,0.852]	0.113 (0.234)	0.93 [0.908,0.95]	0.821[0.815,0.827]	0.12 (0.233)
	Kshirsagar *et. al.* [53]	0.831 [0.825,0.837]	0.128 (0.232)	1.82 [1.773,1.87]	NA	NA
	Kwon *et al.* [55]	0.839 [0.833,0.845]	0.164 (0.21)	0.7 [0.685,0.72]	NA	NA
	O'seaghdha *et al.* [52]	0.840 [0.834,0.846]	0.18 (0.222)	0.54 [0.526,0.55]	NA	NA
	Thakkinstian *et al* [56]	0.816 [0.809,0.823]	0.278 (0.202)	0.47 [0.461,0.49]	NA	NA
SIMPLIFIED SCORES	Bang *et al.* [54]	0.895 [0.891,0.899]	NA	NA	NA	NA
	Chien *et al* - [51]	0.880 [0.876,0.884]	NA	NA	NA	NA
	Kshirsagar et. al. [53]	0.891 [0.887,0.895]	NA	NA	NA	NA
	Kwon *et al.* [55]	0.895 [0.891,0.899]	NA	NA	NA	NA
	Thakkinstian *et al* [56]	0.869 [0.864,0.874]	NA	NA	NA	NA

Abbreviations: AUC, area under receiver operating characteristic curve; eGFR, estimated glomerular filtration rate; NA, not applicable; SD, standard deviation; CI, 95% confidence interval.

a) Calculated as mean difference between observed and predicted CKD cases

b) Cox proportional hazard regression model

Supplementary Table 4.9: Models and scores' AUC and c-index discrimination as well as MAPE by using a time-horizon of 4-year.

	Study	Patients with complete follow-up (N=172,984)			Patients with incomplete follow-up (N=178,399)	
		AUC [CI]	MAPE (SD) [a)]	Calibration slope [CI]	c-index [CI]	MAPE (SD) [a)]
MODELS	Bang *et al.* [54]	0.902 [0.897,0.907]	0.059 (0.143)	0.67 [0.66,0.68]	NA	NA
	Chien *et al* [51][b)]	0.901 [0.898,0.904]	0.075 (0.149)	0.49 [0.48,0.49]	0.891[0.888,0.895]	0.081 (0.155)
	QKidney® [36][b)]	0.914 [0.91,0.917]	0.038 (0.146)	1 [0.99,1.02]	0.904[0.9,0.907]	0.041 (0.148)
	Kshirsagar *et. al.* [53]	0.9 [0.896,0.904]	0.062 (0.14)	1.26 [1.25,1.28]	NA	NA
	Kwon *et al.* [55]	0.9 [0.896,0.904]	0.087 (0.148)	0.47 [0.47,0.48]	NA	NA
	O'seaghdha *et al.* [52]	0.911 [0.908,0.914]	0.085 (0.161)	0.4 [0.39,0.41]	NA	NA
	Thakkinstian *et al* [56]	0.896 [0.892,0.890]	0.176 (0.158)	0.33 [0.33,0.34]	NA	NA
SIMPLIFIED SCORES	Bang *et al.* [54]	0.895 [0.891,0.899]	NA	NA	NA	NA
	Chien *et al* - [51]	0.88 [0.876,0.884]	NA	NA	NA	NA
	Kshirsagar et. al. [53]	0.891 [0.887,0.895]	NA	NA	NA	NA
	Kwon *et al.* [55]	0.895 [0.891,0.899]	NA	NA	NA	NA
	Thakkinstian *et al* [56]	0.869 [0.864,0.874]	NA	NA	NA	NA

Abbreviations: AUC, area under receiver operating characteristic curve; eGFR, estimated glomerular filtration rate; NA, not applicable; SD, standard deviation; CI, 95% confidence interval.

a) Calculated as mean difference between observed and predicted CKD cases

b) Cox proportional hazard regression model

Chapter 5

METHODS FOR EVALUATING USABILITY, USAGE AND INFLUENCE ON DECISION-MAKING OF PATIENT PORTALS: A SYSTEMATIC REVIEW

Paolo Fraccaro[1,2], Markel Vigo[3] , Panagiotis Balatsoukas[2], Iain E. Buchan[1,2], Niels Peek[1,2], Sabine N. van der Veer[2]

[1] Health eResearch Centre, Farr Institute for Health Informatics Research, University of Manchester, UK.

[2] NIHR Greater Manchester Primary Care Patient Safety Translational Research Centre, Institute of Population Health, The University of Manchester, Manchester, United Kingdom.

[3] School of Computer Science, University of Manchester, Manchester, UK.

Contribution: PF (corresponding author), SvdV and NP conceived and designed the study. PF, PB, MV and SvdV were involved in data collection and synthesis. PF, MV, IB, NP and SvdV drafted the manuscript. All authors critically revised the work for important intellectual content, and approved the final version of the manuscript for publication.

Reference: The paper will be submitted to the International Journal of Medical Informatics.

5.1 ABSTRACT

Background

Patient portals are considered valuable conduits for supporting patients' self-management. However, it is still unknown why they often fail to impact on outcomes. One reason may be that many evaluation studies disregard the complex process that leads to improved outcomes, which starts with an effective interaction with the system (step 1) in order to receive information (step 2), which might influence users' decision-making (step 3). We systematically reviewed the literature to identify methodological approaches used to evaluate these aspects in patient portals.

Methods

We used Coiera's information value chain as a theoretical framework to structure our search and synthesis. We searched MEDLINE and Scopus by combining terms related to patient portals and evaluation methodologies. Two reviewers selected relevant papers through duplicate screening, and one extracted data from the included papers.

Results

We included 115 articles. The large majority (n=104) evaluated aspects related to interaction with patient portals. Usage was most often assessed (n=61), mainly by analysing system interaction data (n=50). Overall usability (n=57) was commonly assessed through non-validated questionnaires (n=44). 57 studies investigated the information received from patient portals, primarily by analysing interaction data to evaluate usage of specific system functionalities (n=34). Only eleven studies explicitly assessed the influence of patient portals on patients' and clinicians' decisions.

Conclusions

The information received from patient portals and their influence on decision-making are underinvestigated. They should be more thoroughly addressed as they might conceal some of the reasons for current portals' lack of impact on health outcomes.

Keywords: user-computer interface; computers/utilization; decision making; personal health record; patient access to records; patient portals; decision support systems, clinical.

5.2 INTRODUCTION

Patient portals are seen as a key route to engage patients in care [1–6], and as a valuable conduit to support them with self-managing their health and conditions [6–8]. Most patient portals provide individuals with access to their health records, with laboratory test results as the most accessed piece of information [9–13]. Further, they support basic activities such as booking appointments online, recording symptoms, or communication with healthcare providers [14]. Many portals target people living with chronic conditions, including asthma, cancer, diabetes, and multiple sclerosis [7,15].

Despite the increasing availability of patient portals and reported high satisfaction among users [14,16,17], overall uptake remains low [18]. Additionally, many people stop using patient portals shortly after their first login [12], potentially because they do not consider it of additional value to their care [19]. Furthermore, systematic reviews found no established evidence of improved health outcomes related to patient portals use [20–24]. Although some have reported positive effects on patient engagement and satisfaction [20,21,23,24], their findings regarding impact on care processes and quality were inconsistent [20,21,24]. The reasons for such lack of impact are poorly understood [20,21]. In order to increase this understanding, others have advocated that more studies should focus on or take into account the complex processes whereby health information systems –in our case: patient portals— might lead to improved care and outcomes [25,26].

Coiera proposed in his 'information value chain' [26] that for a health information system to have impact, users first need to effectively interact with it in order to receive information, which might then influence their decision-making. This process can be evaluated from different perspectives, which involve usability, usage and decision-making. We systematically reviewed the literature to identify studies that evaluated these aspects of patient portals, and to investigate the methodological approaches they used. We expect our review to provide a useful guide for future studies, both in terms of the most appropriate methodologies to adopt, as well as the aspects that warrant more research.

5.3 METHODS

We followed the PRISMA statement [27] to design and report our systematic review, where applicable.

5.3.1 Theoretical framework

We used Coiera's 'information value chain' as the theoretical framework [26] to guide our study selection and data synthesis. In this framework, Coiera explains in five steps how the use of health information systems might lead to a change in health outcomes [26]. A main feature of the information value chain is that each step can be evaluated and quantified on its own, with positive results in one step increasing the likelihood of obtaining improvements in the next steps. If we apply Coiera's framework to patient portals, the chain starts with patients interacting with the system (step 1), which for example can be evaluated in terms of usability or usage (e.g. if and how often patients logged into the system). From some interactions, patients will receive information from the system (step 2). The amount and type of information received will depend on which patient portal functionalities patients accessed. This could be, for example, viewing a medication or problem lists, or laboratory results. Where the portal functionality allows patients to record information such as symptoms, the quantity and accuracy of data logged into the system can be evaluated. Step 3 will then focus on whether this information led to patients and clinicians making or changing a decision. For example, patients could decide to contact their healthcare provider if they are worried about an out-of-range laboratory result, or notify their general practitioner of an incorrect medication entry in their health record. At the same time, information recorded by patients through the portal might lead to a clinician requesting an extra laboratory test, or updating the medication list. In both cases, one can count the number of decisions that changed and evaluate their appropriateness. Ultimately, these decisions may alter the process of care (step 4), such as a change in utilisation of the health care resources, patient activation or medication prescriptions. In some cases, such changes will lead to better health outcomes (step 5), such as improvements in blood sugar control or quality of life.

In our review we focused on identifying studies that evaluated aspects of patient portals related to the first three steps of Coiera's information value chain (i.e. interaction with the system, receipt of information, and influence on decision making) [26].

5.3.2 Search strategy

In compliance with guidance from the Cochrane collaboration [28], we searched MEDLINE via Ovid [29] and Scopus [30] for articles in English using both words in title, abstract, or keywords as well as standardized indexing terms. We combined terms referring to patient portals with terms pertaining to evaluations of system usage, usability and decision-making that reflected Coiera's information value chain steps 1 to 3; Supplementary Table 5.1 contains the search syntax for both databases. The searches were performed on the 18th of July 2016, without limits on year of publication.

5.3.2.1 Selection of relevant studies

In our review, we were interested in studies that adhered to the following criteria:

- Evaluating a patient portal, following the definition of patient portals from Irizarry *et al.* [18]. This included systems that were either "tethered" or "untethered" to an Electronic Health Record (EHR), as well as prototypes or mock-ups of patient portals. We focused on systems that gave users access to (part of) their medical records (e.g. laboratory test results, medications or problem lists), allowed them to enter health data, or share it with healthcare professionals. We excluded systems that only provided patients with educational material, or online booking or secure messaging functionalities;
- Having patients, carers, or healthy volunteers from the general population as the study sample, as they are the people most commonly targeted by patient portals;
- Reporting findings on patient portal use, i.e. related to at least one of the first three steps of Coiera's information value chain [26]. We excluded studies that only evaluated the intention to use patient portals, as well as those solely reporting on the impact on care processes or health outcomes;
- Systematic reviews or original articles in English, reporting on experiments in controlled laboratory settings, as well as on field studies in a real world context. Included studies could focus on a specific system or more than one system at the same time. We included full papers published in conference proceedings, while excluding conference abstracts. Narrative reviews, editorials, view point papers and grey literature were also excluded.

After removing duplicates from the MEDLINE and Scopus searches, the principal reviewer (PF) independently screened the titles and abstracts of all studies, whereas two others (PB; SvdV) each did half. For studies considered potentially relevant, we retrieved the full papers to decide on final inclusion, which was also done independently and in duplicate by two reviewers (PF; SvdV). At both stages, disagreement was solved through discussion.

5.3.3 Data extraction and synthesis

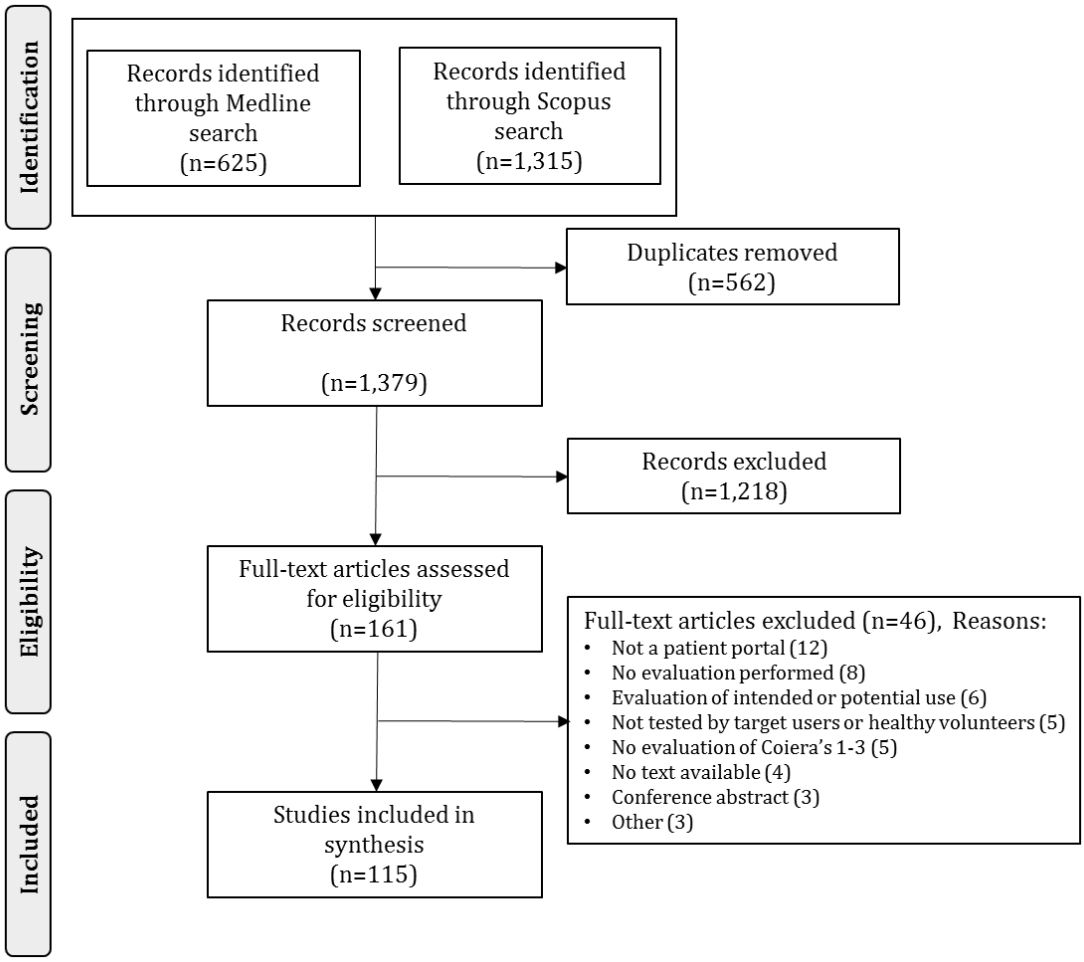
We developed a data abstraction form building on previous reviews of patient portals [23,31] and usability evaluations of health information systems [25,32,33]. We pilot-tested the form among the authors (PF, SvdV, MV, NP) for clarity and completeness. The final form included items related to general study characteristics; study population; type of patient portal (tethered or untethered); patient portal functionalities (access to records; data recording; data sharing); study design (within-subject; between-subject; mixed); ecology (naturalistic or controlled); setting (field, laboratory or remote); guided by theoretical framework (yes or no); evaluation methods; and metrics reported. For studies that used questionnaires as an evaluation method, we also recorded if the adopted instrument had been previously validated, based on previous work, or developed for the specific study (i.e. ad-hoc questionnaire). Examples of validated instruments are the System Usability Scale (SUS) [34]; Computer System Usability Questionnaire (CSUQ) [35] for usability; and the Questionnaire for User Interface Satisfaction (QUIS) [36]. One author (PF) extracted the data for all studies. Uncertainties during the data extraction process were discussed and resolved by discussion with a second member of the research team (SvdV).

We performed a qualitative synthesis of the extracted data and organised our results according to the steps in Coiera's information value chain. For each step, we highlighted the different aspects there were evaluated, the method used, and commonly reported metrics.

5.4 RESULTS

The searches from MEDLINE and Scopus yielded 1,379 potentially relevant articles. Of those, 161 were selected for full-text screening, resulting in a total of 115 studies to be finally included in the review (Figure 5.1) [9,11,37–149].

Figure 5.1: Flow diagram of screening and inclusion of relevant studies.



5.4.1 Study characteristics

Table 5.1, Table 5.2 and Table 5.3 display a summary of the information extracted on the characteristics and design of the 115 included studies, and the patient portals they evaluated; full details of each included study are provided in Supplementary Table 5.2.

5.4.1.1 General study characteristics

Out of all studies we found 111 unique projects. Some studies reported different analyses from the same project [66,110,126,149], whereas others reported preliminary [67,117] and final results [9,65] of the same project separately. Table 5.1 shows their characteristics. The majority were performed in the United States (n=81; 74%), were published after 2010 (n=73; 66%), had patients with a specific condition as their study population (n=39; 34%), and had more than 100 participants (n=59; 53%).

Table 5.1: General characteristics of included unique projects (total n=111).

General study characteristics	N (%)	References
Year of publication		
Before 2005	7 (7)	[9,44,68,83,84,87,88,117]
From 2005 to 2010	31 (27)	[38,41,42,45,47,49,52,53,56,59,61,73,76,80,85,97,102,103,108,111–114,116,118,120,127,130,135,147,148]
After 2010	73 (66)	[11,37,39,40,43,46,48,50,51,54,55,57,58,60,62–67,69–72,74,75,77–79,81,82,86,89–96,98–101,104–107,109,110,115,119,121–126,128,129,131–134,136–146,149]
Georgraphical location		
United States	81 (74)	[9,38–40,43–49,52–58,61,64–70,73–75,77,80–85,88–90,92–95,98–101,103–105,107–112,117,120–125,127–135,137–142,144–148]
Europe	17 (15)	[11,37,42,50,60,62,72,76,79,86,87,96,102,106,113,114,116]
Canada	8 (7)	[41,59,63,71,78,91,97,115]
Other	5 (4)	[51,119,126,136,143,149]
Study population		
People living with a specific condition	39 (34)	[38,40,43,44,48,50,53,56,57,60,62–64,67,71,72,75,78,79,81,83,86,88,92–94,96,97,105,106,108,116,124,127,128,133,136,143,145]
General population	34 (32)	[9,45–47,49,55,58,65,66,68,69,73,74,77,84,89,95,99,100,102,104,110,112,113,115,117–120,122,123,126,129,137,142,144,149]
Primary care patients	25 (22)	[11,37,42,52,54,61,70,87,90,98,101,107,108,114,121,125,132,134,135,138–140,146–148]
Carers	8 (7)	[39,80,82,85,91,103,111,131]
Other	5 (4)	[41,51,59,76,130]
Sample size		
<30	28 (26)	[9,37,39,42,43,45,46,50,51,53,55–57,65,67,75,77,78,83,85,91,107–109,111,114–117,143]
Between 30 and 100	20 (18)	[38,47,59,60,62–64,68,73,82,87,88,90,99,100,102,105,112,113,137]
More than 100	59 (53)	[11,40,41,48,49,52,54,58,61,66,69–72,74,76,79–81,84,86,89,94–98,101,103,106,110,118–136,138–142,144,146–149]
Not reported/unclear	4 (3)	[44,92,93,104]

5.4.1.2 Type of patient portals evaluated

The 115 included studies reported evaluations of a total of 80 different patient portals. Eighteen were tested in more than one study, with a median of two studies (range, 2 to 10) reporting on the same patient portal. MyChart (n=10) [48,66,84,92,99,110,125,129,131,141] and Patient Gateway (n=9) [52,61,98,134,135,138,146–148] were the patient portals that appeared in most studies. The majority of the tested patient portals were tethered (n=58, 73%), and 32 (40%) were prototypes (Table 5.2). Data recording functionalities and access to records were provided by 53 (66%) and 58 (73%) of patient portals, respectively.

Table 5.2: Type of patient portal evaluated (total n=80).

Type of patient portal evaluated	N (%)	References
Patient portal type		
Tethered	58 (73)	[9,11,39–42,44,48,50,52,54,56,58,60–72,74,76–79,81,82,84,86–101,103–113,115–119,121,124,125,127–135,137–142,145–148]
Untethered	22 (28)	[37,38,43,45–47,49,51,53,55,57,59,73,80,83,85,102,114,120,122,123,126,136,143,144,149]
Prototype [a)]		
Yes	32 (40)	[37,39,43,53,55–57,62,63,68,69,75,80,81,83,85,88,91,95,102,108,109,111,112,114–116,124,143–145]
No	51 (64)	[9,11,38,40–42,44–52,54,58–61,63–65,67,70–74,76–79,82,84,86,87,89,90,92–94,96–101,103–107,110,113,117–123,125–142,146–149]
Main functionalities [a)]		
Access to records	58 (73)	[9,11,39–42,44,48,50,52,54,58,60–72,74–79,81,82,84,86–101,103–113,115–119,121,123–125,127–135,137–142,145–148]
Data recording	53 (66)	[9,37–40,42–62,66,69–71,73–75,77,79,80,82–86,88,89,92,96–100,102,106,110,112–123,125–130,133–138,140,141,143,144,146–148]
Data sharing	30 (38)	[9,37–39,44–47,49,50,52,56,61,62,70,71,73,74,79,82,88,98,100,102,112–114,116–119,121,123,127,130,133–136,138,146–149]

[a)] Categories are not mutually exclusive.

5.4.1.3 Evaluation characteristics

Most of the 115 included studies (n=93; 81%) followed a within-subject design, were longitudinal (n=69; 40%), and were carried out in the field (n=84; 73%); the most frequently used ecology was naturalistic (n=76; 66%) (Table 5.3). In a minority of studies (n=14; 12%), authors explicitly referred to a theoretical framework they used to guide their analysis, with the Technology Acceptance Model [150] and Nielsen's usability heuristics [151] being the most frequently cited.

Table 5.3: Evaluation characteristics of the included studies (total n=115).

Evaluation characteristics	N (%)	References
Study design		
Within-subject	93 (81)	[9,11,37,39,40,42–49,51–61,65–69,71–80,83–88,90–94,96–112,114–120,122,123,125–132,134,137,140–144,146,147,149]
Between-subject	20 (18)	[38,41,50,62–64,81,82,89,113,121,124,131,133,135,136,138,139,145,148]
Mixed	2 (2)	[70,95]
Temporal factor		
Cross-sectional	45 (39)	[11,39,45–48,50,53,55,56,64–67,69,73–75,77,80,81,83–85,87,91,95,99,100,105,107–111,114,115,120,122,124,128–130,134,143]
Longitudinal	69 (60)	[9,37,38,40–44,49,51,52,54,58–63,68,70–72,76,78,79,82,86,88–90,92–94,96–98,101–104,106,112,113,116–119,121,123,125–127,131–133,135–142,144–149]
Mixed	1 (1)	[57]
Study setting [a)]		
Laboratory	30 (26)	[39,45–47,53,55–57,65–67,73,77,83,85,87,91,95,99,100,107–112,114,115,143]
Field	84 (73)	[9,11,37,38,40–44,49–52,54,57–64,68–72,74–76,78,79,82,84,86,88–90,92–94,96–98,101–106,113,116–142,144–149]
Remote	4 (4)	[48,57,80,81]
Ecology [a)]		
Controlled	41 (36)	[38,39,41,45,46,53,55–57,62,65–67,70,73,77,80–83,85,91,95,99,100,107–111,113–115,121,124,130,135,138,143,145,148]
Naturalistic	76 (66)	[9,11,37,40,42–44,47–52,54,57–61,63,64,68–72,74–76,78,79,84,86–90,92–94,96–98,101–106,112,116,118–120,122,123,125–129,131–134,136,137,139–142,144,146,147,149]
Theoretical framework used		
Yes	14 (12)	[39,43,48,64,75,80,93,107,108,113,114,146,149]
No	101 (88)	[9,11,37,38,40–42,44–46,49–63,65–74,76–79,81–92,94–106,109–112,115–125,127–145,147,148]

[a)] Some publications reported on different phases of the study, reporting multiple study settings and ecologies.

5.4.2 Evaluation methods

Table 5.4 displays the methods that studies applied to evaluate the different aspects of patient portal use. Overall, interaction with patient portals (Coiera's step 1), users receiving information (step 2), and influence of patient portals on patients' and clinicians' decisions (step 3) were evaluated in 104 (90%), 57 (50%) and 11 (10%) studies, respectively. Overall, 49 studies (43%) looked at two steps out of the first three in Coiera's information value chain [9,37,38,40,41,43,44,48,49,52,54,57–62,70,72,74,75,81,88–90,92,94,96,97,101–103,112,116–119,121,123,125,127,130,131,133,137–139,141,146]. Only four studies (4%) [11,126,134,136] considered all first three steps.

Of the 80 unique systems we found, step 1 was evaluated in 77 (96%) cases, step 2 in 41 (51%), and step 3 in nine (11%) systems. Overall, 39 systems (49%) had only one of the three steps evaluated, while 35 (44%) had two, and six systems had all of the first three Coiera's steps evaluated.

Table 5.4: Synthesis of methods used to evaluate Coiera's steps 1 to 3. [a)] Percentages are calculated for each aspect separately.

Evaluated aspects	Evaluation methods	N (%) [a)]	References
Step 1: Aspects related to the interaction with the system (n=104)			
Overall usability (n=57, 55%)	Questionnaire	55 (96)	[37,39–41,43,44,48–50,52,56,57,59–65,67–70,72,73,78,80–82,84,87,88,91,95,97,99,100,107–111,113,114,116,117,120,123,124,127,134,143,145,146]
	Interview or focus groups	9 (16)	[9,53,60,63,75,84,110,117,145]
Usability – Identifying user interface issues (n=15, 15%)	Think-aloud	11 (73)	[39,45,55,75,85,107,109,111,114,115,143]
	Interviews or focus groups	6 (40)	[46,75,77,107–109]
	Video recording or observations	4 (27)	[46,53,75,77]
	Questionnaire	1 (7)	[45]
Usability – Effectiveness (n=18, 17%)	Tasks analysis	17 (94)	[39,45,47,65,66,73,80,81,83,91,95,99,100,109–111,143]
	System interaction logs analysis	1 (6)	[60]
Usability – Efficiency (n=12, 12%)	Task analysis	11 (92)	[45,46,65,67,77,85,87,95,111,114,143]
	User interface Event Mining	1 (8)	[45]
	Questionnaire	1 (8)	[85]
	System interaction logs analysis	1 (8)	[88]
Usage (n=61, 59%)	System interaction logs analysis	50 (83)	[9,40–43,49,52,54,57–63,68,70,71,75,78,79,86,89,90,92–94,96,98,101,103,104,106,112,113,117–119,121,125,127,129,131–133,136,139,142,144,148]
	Questionnaire	10 (15)	[11,48,72,74,82,105,124,126,128,149]
	Data log analysis	1 (2)	[38]
Step 2: Aspects related to receiving information from the system (n=57)			
Usage of different functionalities (n=47, 82%)	System interaction logs analysis	34 (72)	[9,40,41,43,44,51,58–60,62,70,75,88–90,92,94,96,101,103,112,117–119,125,127,131,133,135,138–142]
	Questionnaire	9 (19)	[11,48,49,72,74,97,122,126,146]
	Data log analysis	5 (11)	[40,52,61,136,137]
Record completeness, quality or quantity (n=11, 19%)	Data log analysis	10 (91)	[37,38,52,59,61,102,116,130,135,138]
	Questionnaire	1 (9)	[133]
Records shared with health professionals (n=4, 7%)	System interaction logs analysis	3 (75)	[61,127,138]
	Data log analysis	1 (25)	[52]
System resource utilisation (n=4, 7%)	System interaction logs analysis	3 (75)	[44,57,76]
	Google Analytics	1 (25)	[114]
Behavioural styles (n=3, 5%)	System interaction logs analysis	2 (67)	[90,101]
	Factor analysis	1 (33)	[76]
	Interviews	1 (33)	[90]
	Hierarchical clustering	1 (33)	[101]
Step 3: Aspects related to the influence of system use on decisions (n=11)			
Patient's' decisions (n=5, 45%)	Questionnaire	5 (100)	[11,81,126,134,136]
Clinician's' decisions (n=6, 55%)	Data log analysis	6 (100)	[102,121,130,137,138,147]
	Interviews	2 (33)	[121,138]

5.4.2.1 Step 1: Interaction (n=104)

5.4.2.1.1 Usability

Usability was the most studied aspect related to interaction, with four different themes that emerged from our synthesis of the included studies.

The first theme included studies that analysed the overall usability of the system with questionnaires (n=55). These were mostly ad-hoc questionnaires that included questions aimed at assessing the user experience and satisfaction with the system. Only in 11 cases [43,48,63,69,97,107,109,111,113,120,143] the questionnaires were validated or based on previous work: the SUS questionnaire [34] was used in two studies, and the IBM Computer Usability Satisfaction Questionnaire [35] and QUIS [36] were used once.

The second theme is represented by studies that aimed at identifying usability issues with the user interface. Within this theme, the most common method was the think-aloud technique (n=11), where participants verbalise their thoughts and actions while using the system.

The third and fourth usability theme focused on the effectiveness and efficiency of the interaction with the patient portal. Task analysis was the main method adopted, with 17 and 11 studies applying this approach to evaluate effectiveness and efficiency, respectively. Of these, 15 were laboratory-based, while two were performed remotely, with participants who completed the simulated tasks via an online system (i.e. without having to attend any laboratory session). Percentage of successful tasks (n=10) [45,66,80,81,83,99,109–111,143] and time to complete a task (n= 12) [45,46,65,67,77,85,87,95,111,114,143] were the most common reported metrics for effectiveness and efficiency, respectively. Four studies applied cognitive task analysis and evaluated effectiveness and efficiency in relation to the cognitive processes of users while using the system [65,66,73,100], while four provided an overall effectiveness score for each participant [66,99,100,110]. This was calculated as a weighted sum of all points that users got during each task multiplied by coefficients that the researchers assigned to reflect the difficulty of each task. Efficiency was also assessed once through User Interface event mining (i.e. reporting the number of clicks or keystrokes to accomplish a task), and with a validated questionnaire (i.e. National Aeronautical and Space Administration Task Load Index [152]) that aimed to measure users' burden when performing the tasks.

5.4.2.1.2 Usage

Usage was another commonly evaluated aspect of interaction with patient portals. The analysis of system interaction logs that was most frequently used (83%), with the percentage of active users within the study population being the main metric reported (n= 47)

[9,11,38,42,48,49,52,54,57,58,60–62,68,70–72,74,75,79,86,89,93,94,96,98,101,103,105,106,117,118,121,124–129,131,132,136,139,142,144,148,149]. Whereas most studies defined participants as an active user if they logged into the portal at least once, four studies applied logging in at least twice [60,62,93,101]; in one case, utilisation of the main patient portal functionality was an additional criterion for being qualified as a user [60]. Often, authors reported the frequency of use over the study period as an additional metric (n=43) [9,11,40,41,43,58–60,62,63,70–72,74,78,79,82,86,90,92,93,96,98,101,103,105,106,112,113,117–119,124–129,133,136,139,142,149].

5.4.2.2 Step 2: Receiving information (n=57)

Usage of the specific patient portal functionalities was the main aspect that was evaluated in relation to receiving information, with similar methods and metrics being reported as for overall system usage in step 1. Most authors considered participants an active user of a specific functionality if they accessed it at least once, with some reporting also the frequency of use or number of actions per session (n=11) [9,49,59,74,75,101,103,117–119,133]. In studies where the main purpose of the system was data recording (e.g. daily activities or symptoms), data completeness, accuracy, quality and quantity were often additionally evaluated through analysing data logs (n=10).

System resources utilisation (e.g. number of pages viewed, or number of server requests) was also evaluated (n= 4), while three studies derived behavioural styles of users based on the type and frequency of information they accessed.

5.4.2.3 Step 3: Effect on decision-making (n=11)

Studies that evaluated the influence of patient portal use on decision-making frequently employed questionnaires to assess the effect on patients' decisions (n=5). The reported metrics were percentage of patients deciding to implement lifestyle changes (n=4) [11,126,134,136], percentage of patients who would act on their laboratory results by calling their doctor immediately [81]**,** and percentage of patients who contacted their healthcare professionals after seeing their problem list [134]. Clinicians' decisions were mostly assessed by data logs analysis (n=6), reporting the following metrics: adverse drug events that clinicians could have identified on the basis of information provided by patients via the patient portals [121,137,138]; appropriate medication prescriptions and requested diagnostic tests [130]; additional diagnoses [102]; clinicians' updates of the medication regimen [147].

5.5 DISCUSSION

5.5.1 Summary of findings

We performed a systematic review of the literature on evaluations of patient portal usability, usage and influence on decision-making. Our review was guided by Coiera's information value chain as a theoretical framework. Almost all included studies that evaluated patient portal usage or usability did so with a focus on interaction, using system interaction logs analysis, questionnaires, think-aloud techniques and task analysis as the main evaluation methods. Only half of included studies looked at the information that was received from the portal, mainly assessing the usage of different system functionalities through the analysis of interaction logs. Finally, 10% of studies explicitly assessed the influence of the patient portal on patients' or clinicians' decisions.

5.5.2 Relation to other studies

Although many systematic reviews have been recently published on patient portals [8,14,20–22,24,31,153], our systematic review is the first that focused comprehensively on usability, usage and decisions-making from a methodological perspective. Our findings are consistent with those reported by Irizarry *et al.* [18]. Their realist review aimed to identify factors influencing patients' engagement with patient portals. They found 120 studies around five major topics (i.e. patient adoption, provider endorsement, health literacy, usability, and utility). Although they followed a different search strategy and their aim was not to review the methodologies used in the included studies, they also reported on many studies where patients were considered as active users if they logged in at least once. Furthermore, they identified 20 studies that considered usability of patient portals, most of which used questionnaires.

Our findings are also in line with two methodological systematic reviews that looked at usability studies of Electronic Health Record systems [33] and current practice in usability studies [32], respectively. They also found that usability effectiveness and efficiency was infrequently evaluated [32,33]. This is surprising given that effectiveness and efficiency are essential usability qualities defined by standardization documents including the ISO 9241-11 [154]. Furthermore, Hornbæk *et al.* [32] concluded already in 2006 that most of the studies evaluating usability through questionnaires used non-validated instruments that were often designed ad hoc for the study in question. Our review showed that that ten years after Hornbæk's review this methodological issue still often goes unaddressed in patient portal evaluations.

5.5.3 What is the meaning of the findings and what are their implications?

To better understand how patient portals can change decision making, each of the steps along Coiera's information value chain should be evaluated. Ideally this happens for the same system within a single study context. However, less than half of the portals had two steps in the information value chain addressed, and there were only six for which all three steps had been evaluated. Furthermore, we observed that the number of studies that focussed on the first three steps of Coiera's information value chain decreased steeply along the chain: whereas the interaction with the system was addressed in 104 studies, receiving information from patient portals and their influence on decision-making was only evaluated 57 and 11 times, respectively. As a chain can only be as strong as its weakest link, some of the unknown reasons for patient portals' lack of impact might be concealed by these under-investigated aspects. The influence of patient portals on decisions is less straightforward to measure than the other steps in Coiera's information value chain, as this often involves looking at variables that are not routinely and consistently collected (e.g. being part of the discussion between clinicians and patients or patient's life style). However, we still were able to identify some studies that looked at the influence of patient portals on decision making by patients. The opportunities to measure patients' decisions will increase in the future, as more behaviours in everyday life are routinely captured through digital media [155].

From a methodological perspective, there may be room for improvement in patient portal evaluation studies. First, questionnaires to evaluate usage could be combined with analysis of system interaction logs, as relying only on questionnaires might be prone to recall mistakes [25] and lack in objectivity [32,156]. Furthermore, considering people who logged into a system once to be active users is likely to be oversimplistic. As with any new technology, people are likely to try out a portal before they decide to actually use it. More complex and comprehensive metrics that combine frequency of system and functionality usage may be more appropriate for usage evaluations [157]. Second, authors of future studies should consider using validated questionnaires to evaluate usability, user satisfaction, and user experience [32,34–36,152,156]. This could address concerns about the validity of the results, as well as enhance the comparability of these results across different studies [32,156].

5.5.4 Limitations

This systematic review has several limitations. First, we adopted a specific perspective to investigate our research question by choosing Coiera's information value chain. Although it is an established theoretical framework that effectively describes the evaluation of informatics systems, our findings have been biased towards the steps of that chain. This implies that other

aspects that are likely to influence whether patient portals lead to improved health outcomes, such as socio-economic factors and health system characteristics, were out of scope for our review. Second, some of the steps in the chain may have been easier to search the literature for than others. Whilst we used a comprehensive strategy in which all steps were addressed by multiple search terms, we cannot exclude that our search strategy was more sensitive towards steps with more available text words and MeSH terms. Third, we did not perform the data extraction in duplicate. Therefore, this phase of the reviewing process might have been more prone to errors than the others; it was up to the discretion of the primary reviewer to discuss items that were less straightforward to extract and required more interpretation.

5.6 CONCLUSIONS

Our systematic review highlights important weaknesses and gaps in the evidence base on patient portals; weaknesses in methods, with the common use of single log-in studies and non-validated questionnaires; and major gaps in research into the amount of information received via portals and its influence on the decisions made by patients and clinicians. Addressing these gaps with appropriate methodology is likely to uncover important reasons for why current patient portals seem to have such little impact on health outcomes.

5.8 REFERENCES

1. Pearce C, Bainbridge M (2014) A personally controlled electronic health record for Australia. J Am Med Inform Assoc 21: 707–713. Available: http://jamia.bmj.com/content/21/4/707. Accessed 10 October 2014.

2. Institute of Medicine (USA) (n.d.) Achieve Meaningful use - Stage 2. Available: http://www.webcitation.org/6qBAzdht7.

3. Urowitz S, Wiljer D, Apatu E, Eysenbach G, Delenardo C, et al. (2008) Is Canada ready for patient accessible electronic health records? A national scan. BMC Med Inform Decis Mak 8: 33. Available: http://www.biomedcentral.com/1472-6947/8/33. Accessed 23 October 2014.

4. Pagliari C, Shand T, Fisher B (2012) Embedding online patient record access in UK primary care: a survey of stakeholder experiences. JRSM Short Rep 3: 34. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3365791/.

5. Fisher B (2013) Patients' access to their electronic record: offer patients access as soon as you can. Br J Gen Pract 63: e423-5. Available: http://bjgp.org/content/63/611/e423.abstract. Accessed 23 October 2014.

6. Adler-Milstein J, Sarma N, Woskie LR, Jha AK (2014) A Comparison Of How Four Countries Use Health IT To Support Care For People With Chronic Conditions. Health Aff (Millwood) 33: 1559–1566. Available: http://content.healthaffairs.org/content/33/9/1559. Accessed 18 September 2014.

7. Kruse CS, Argueta DA, Lopez L, Nair A (2015) Patient and Provider Attitudes Toward the Use of Patient Portals for the Management of Chronic Disease: A Systematic Review. J Med Internet Res 17: e40. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4376181/.

8. Price M, Bellwood P, Kitson N, Davies I, Weber J, et al. (2015) Conditions potentially sensitive to a Personal Health Record (PHR) intervention, a systematic review. BMC Med Inform Decis Mak 15: 32. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4411701/.

9. Cimino JJ, Patel VL, Kushniruk AW (2002) The patient clinical information system (PatCIS): technical solutions for and experience with giving patients access to their electronic medical records. Int J Med Inform 68: 113–127. Available:

http://www.sciencedirect.com/science/article/pii/S1386505602000709. Accessed 23 October 2014.

10. Shah SGS, Fittonb R, Hannanc A, Fisherd B, Younga T, et al. (2015) Accessing personal medical records online: A means to what ends? Int J Med Inform 84: 111–118. Available: http://www.sciencedirect.com/science/article/pii/S1386505614001932. Accessed 19 January 2015.

11. Bhavnani V, Fisher B, Winfield M, Seed P (2011) How patients use access to their electronic GP record--a quantitative study. Fam Pract 28: 188–194. Available: http://fampra.oxfordjournals.org/content/28/2/188. Accessed 22 October 2014.

12. Phelps GR, Taylor J, Simpson K, Samuel J, Turner NA (2014) Patients' Continuing Use of an Online Health Record: A Quantitative Evaluation of 14,000 Patient Years of Access Data. J Med Internet Res 16: e241. Available: http://www.jmir.org/2014/10/e241/.

13. Young MJ, Scheinberg E, Bursztajn H (2014) Direct-to-patient laboratory test reporting: balancing access with effective clinical communication. JAMA 312: 127–128. Available: http://jama.jamanetwork.com/article.aspx?articleid=1882585. Accessed 21 October 2014.

14. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, et al. (2013) Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 159: 677–687. Available: http://annals.org/article.aspx?articleid=1770672. Accessed 9 October 2014.

15. Price M, Bellwood P, Kitson N, Davies I, Weber J, et al. (2015) Conditions potentially sensitive to a Personal Health Record (PHR) intervention, a systematic review. BMC Med Inform Decis Mak 15: 1–12. Available: http://dx.doi.org/10.1186/s12911-015-0159-1.

16. Wiljer D, Leonard KJ, Urowitz S, Apatu E, Massey C, et al. (2010) The anxious wait: assessing the impact of patient accessible EHRs for breast cancer patients. BMC Med Inform Decis Mak 10: 46. Available: http://www.biomedcentral.com/1472-6947/10/46. Accessed 13 October 2014.

17. Bartlett C, Simpson K, Turner AN (2012) Patient access to complex chronic disease records on the Internet. BMC Med Inform Decis Mak 12: 87. Available:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3438097&tool=pmcentrez&rendertype=abstract. Accessed 9 October 2014.

18. Irizarry T, DeVito Dabbs A, Curran RC (2015) Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 17: e148. Available: http://www.jmir.org/2015/6/e148/.

19. Hazara AM, Bhandari S (2016) Barriers to patient participation in a self-management and education website Renal PatientView: A questionnaire-based study of inactive users. Int J Med Inform 87: 10–14. Available: http://www.sciencedirect.com/science/article/pii/S138650561530071X.

20. Ammenwerth E, Schnell-Inderst P, Hoerbst A (2012) The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 14: e162. Available: http://www.jmir.org/2012/6/e162/. Accessed 11 October 2014.

21. Davis Giardina T, Menon S, Parrish DE, Sittig DF, Singh H (2014) Patient access to medical records and healthcare outcomes: a systematic review. J Am Med Inform Assoc 21: 737–741. Available: http://jamia.bmj.com/content/21/4/737.long. Accessed 2 September 2014.

22. de Lusignan S, Mold F, Sheikh A, Majeed A, Wyatt JC, et al. (2014) Patients' online access to their electronic health records and linked online services: a systematic interpretative review. BMJ Open 4: e006021. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4158217&tool=pmcentrez&rendertype=abstract. Accessed 11 September 2014.

23. Mold F, de Lusignan S, Sheikh A, Majeed A, Wyatt JC, et al. (2015) Patients' online access to their electronic health records and linked online services: a systematic review in primary care. Br J Gen Pract 65: e141–e151. Available: http://bjgp.org/content/65/632/e141.abstract.

24. Kruse CS, Bolton K, Freriks G (2015) The effect of patient portals on quality outcomes and its implications to meaningful use: A systematic review. J Med Internet Res 17. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84925476798&partnerID=40&md5=54e8b8eb73ece83b848fd6cc5d2ee382.

25. Kushniruk AW, Patel VL (2004) Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 37: 56–76. Available: http://www.sciencedirect.com/science/article/pii/S1532046404000206.

26. Coiera E (2015) Guide to Health Informatics. CRC Press. 710 p.

27. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, et al. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 6: e1000100. Available: http://www.webcitation.org/6NYnckiur. Accessed 20 January 2014.

28. Lefebvre C, Manheimer E, Glanville J (2011) Searching for studies. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. The Cochrane Collaboration.

29. U.S. National Library of Medicine (n.d.) MEDLINE. Available: https://www.nlm.nih.gov/bsd/pmresources.html.

30. Elevier (n.d.) Scopus. Available: https://www.scopus.com/.

31. Osborn CY, Mayberry LS, Mulvaney SA, Hess R (2010) Patient Web Portals to Improve Diabetes Outcomes: A Systematic Review. Curr Diab Rep 10: 422–435. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3086814/.

32. Hornbæk K (2006) Current practice in measuring usability: Challenges to usability studies and research. Int J Hum Comput Stud 64: 79–102. Available: http://www.sciencedirect.com/science/article/pii/S1071581905001138.

33. Yen P-Y, Bakken S (2012) Review of health information technology usability study methodologies. J Am Med Inform Assoc 19: 413–422. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3341772/.

34. Brooke J (1996) SUS-A quick and dirty usability scale. Usability Eval Ind 189: 4–7.

35. Lewis JR (1995) IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. Int J Hum-Comput Interact 7: 57–78. Available: http://dx.doi.org/10.1080/10447319509526110.

36. Chin JP, Diehl VA, Norman KL (1988) Development of an Instrument Measuring User Satisfaction of the Human-computer Interface. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '88. New York, NY, USA: ACM. pp. 213–218. Available: http://doi.acm.org/10.1145/57167.57203.

37. Eccher C, Botteri M, Ortolani D, Forti S, Piras EM (2014) A mobile logbook to diagnose masked hypertension: a pilot application. Stud Health Technol Inform 205:

363–367. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=25160207.

38. Lieberman DZ, Kelly TF, Douglas L, Goodwin FK (2010) A randomized comparison of online and paper mood charts for people with bipolar disorder. J Affect Disord 124: 85–89. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=19896202.

39. Fiks AG, Mayne S, Karavite DJ, Debartolo E, Grundmeier RW (2014) A Shared e-decision support portal for pediatric asthma. J Ambul Care Manage 37: 120–126. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84896268834&partnerID=40&md5=803f4d305b9d80584111fd78e9cad0e5.

40. Dalal AK, Dykes PC, Collins S, Lehmann LS, Ohashi K, et al. (2016) A web-based, patient-centered toolkit to engage patients and caregivers in the acute care setting: A preliminary evaluation. J Am Med Informatics Assoc 23: 80–87. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84959461614&partnerID=40&md5=af8e549b8cf13a105922cd4bdf1fcbf8.

41. Shaw E, Howard M, Chan D, Waters H, Kaczorowski J, et al. (2008) Access to web-based personalized antenatal health records for pregnant women: a randomized controlled trial. J Obstet Gynaecol Can 30: 38–43. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-41149135017&partnerID=40&md5=da43ffba3e9d10cdea08bdb05ac74f51.

42. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J (2010) Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. BMJ 341: c5814. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2982892&tool=pmcentrez&rendertype=abstract. Accessed 22 October 2014.

43. Connelly K, Siek KA, Chaudry B, Jones J, Astroth K, et al. (2012) An offline mobile nutrition monitoring intervention for varying-literacy patients receiving hemodialysis: a pilot study examining usage and usability. J Am Med Inform Assoc 19: 705–712. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=22582206.

44. Kim E-H, Wang M, Lau C, Kim Y (2004) Application and evaluation of personal health information management system. Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings. Department of Bioengineering, University of Washington, Seattle, WA 98195, United States, Vol. 26 V. pp. 3159–3162. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-11144315968&partnerID=40&md5=77c265c0fc150d9a6c69177394cd95b7.

45. Gonzales P, Tran BQ (2010) Assessing the usability of web-based personal health records. IFMBE Proceedings. Department of Electrical Engineering and Computer Science, Catholic University of America, Washington, DC, United States, Vol. 32 IFMBE. pp. 217–220. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-78049370150&partnerID=40&md5=a9b1f8e7443cb019d31f35ec23a4c935.

46. Liu LS, Shih PC, Hayes GR (2011) Barriers to the adoption and use of personal health record systems. ACM International Conference Proceeding Series. Department of Informatics, Donald Bren School of Information and Computer Sciences, University of California, Irvine, CA 92697-3440, United States. pp. 363–370. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-79952606419&partnerID=40&md5=7cb4d67cd9bb27c002dcda46f3c099cb.

47. Lober WB, Zierler B, Herbaugh A, Shinstrom SE, Stolyar A, et al. (2006) Barriers to the use of a personal health record by an elderly population. AMIA Annu Symp Proc: 514–518. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-34748879952&partnerID=40&md5=862ad2915bfe68f226991bc2c5743c1d.

48. Ronis SD, Baldwin CD, McIntosh S, McConnochie K, Szilagyi PG, et al. (2015) Caregiver Preferences Regarding Personal Health Records in the Management of ADHD. Clin Pediatr (Phila) 54: 765–774. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=25567294.

49. Kim E-H, Stolyar A, Lober WB, Herbaugh AL, Shinstrom SE, et al. (2009) Challenges to using an electronic personal health record by a low-income elderly population. J Med Internet Res 11: e44. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=19861298.

50. Ennis L, Robotham D, Denis M, Pandit N, Newton D, et al. (2014) Collaborative development of an electronic Personal Health Record for people with severe and

enduring mental health problems. BMC Psychiatry 14: 1–7. Available: http://dx.doi.org/10.1186/s12888-014-0305-9.

51. Lau AYS, Parker A, Early J, Sacks G, Anvari F, et al. (2012) Comparative usage of a web-based personally controlled health management system and normal support: A case study in IVF. Electron J Heal Informatics 7. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84867708846&partnerID=40&md5=80fa9b3af6cae9c025c212b587de4bcb.

52. Schnipper JL, Gandhi TK, Wald JS, Grant RW, Poon EG, et al. (2008) Design and implementation of a web-based patient portal linked to an electronic health record designed to improve medication safety: the Patient Gateway medications module. Inform Prim Care 16: 147–155. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=18713531.

53. Luckmann R, Vidal A (2010) Design of a handheld electronic pain, treatment and activity diary. J Biomed Inform 43: S32-6. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=20937483.

54. Krist AH, Peele E, Woolf SH, Rothemich SF, Loomis JF, et al. (2011) Designing a patient-centered personal health record to promote preventive care. BMC Med Inform Decis Mak 11: 1–11. Available: http://dx.doi.org/10.1186/1472-6947-11-73.

55. Siek KA, Khan DU, Ross SE, Haverhals LM, Meyers J, et al. (2011) Designing a personal health application for older adults to manage medications: a comprehensive case study. J Med Syst 35: 1099–1121. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=21562730.

56. Tani S, Marukami T, Matsuda A, Shindo A, Takemoto K, et al. (2010) Development of a health management support system for patients with diabetes mellitus at home. J Med Syst 34: 223–228. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=20503606.

57. Samtani G, Maimoon L, Chuang J, Nybroe C, Liu X, et al. (2016) Diabeticlink: An internationally collaborative cyber-enabled patient empowerment platform. Lect Notes

Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 9545: 299–310. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84958531029&partnerID=40&md5=ea8d2c7cf628745d558f33465b34b25b.

58. Smith SG, O'Conor R, Aitken W, Curtis LM, Wolf MS, et al. (2015) Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. J Am Med Inform Assoc 22: 888–895. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84964693876&partnerID=40&md5=98e5d99251014c82c8cf8deb237c8c32.

59. Lam J, Barr RG, Catherine N, Tsui H, Hahnhaussen CL, et al. (2010) Electronic and paper diary recording of infant and caregiver behaviors. J Dev Behav Pediatr 31: 685–693. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=20975577.

60. Robotham D, Mayhew M, Rose D, Wykes T (2015) Electronic personal health records for people with severe mental illness; a feasibility study. BMC Psychiatry 15: 192. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=26245805.

61. Poon EG, Wald J, Schnipper JL, Grant R, Gandhi TK, et al. (2007) Empowering patients to improve the quality of their care: design and implementation of a shared health maintenance module in a US integrated healthcare delivery network. Stud Health Technol Inform 129: 1002–1006. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=17911866.

62. Chiche L, Brescianini A, Mancini J, Servy H, Durand J-M (2012) Evaluation of a prototype electronic personal health record for patients with idiopathic thrombocytopenic purpura. Patient Prefer Adherence 6: 725–734. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84885975571&partnerID=40&md5=f8dfd3f1d5581dd6eeb5908e92d703c3.

63. Guy S, Ratzki-Leewing A, Gwadry-Sridhar F (2012) Evaluation of a web-based patient portal for chronic disease management. Lect Notes Inst Comput Sci Soc Telecommun Eng 91 LNICST: 114–121. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-

84869597771&partnerID=40&md5=0ad7f66c2a5a9f8f79d3416b09c3459d.

64. Odlum M, Gordon P, Camhi E, Valdez E, Bakken S (2014) Exploring factors related to the adoption and acceptance of an internet-based electronic personal health management tool (EPHMT) in a low income, special needs population of people living with HIV and AIDS in New York City. Stud Health Technol Inform 201: 145–152. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=24943537.

65. Barron J, Bedra M, Wood J, Finkelstein J (2014) Exploring three perspectives on feasibility of a patient portal for older adults. Studies in Health Technology and Informatics. Johns Hopkins University, School of Medicine, Baltimore, MD, United States, Vol. 202. pp. 181–184. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84904181964&partnerID=40&md5=0162d556d4a324b4713779bc02422291.

66. Taha J, Czaja SJ, Sharit J, Morrow DG (2013) Factors affecting usage of a personal health record (PHR) to manage health. Psychol Aging 28: 1124–1139. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84891093159&partnerID=40&md5=20da3b22aa901738a93de0b6aa855ddb.

67. Barron JS, Finkelstein J (2013) Feasibility of providing personalized health information to older adults and their caregivers. IEEE EMBS Special Topic Conference on Point-of-Care (POC) Healthcare Technologies: Synergy Towards Better Global Healthcare, PHT 2013. Division of Geriatric Medicine and Gerontology, Johns Hopkins University, School of Medicine, Baltimore, MD 21224, United States. pp. 232–235. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84874570891&partnerID=40&md5=e1a5dcfcd42f67a3997870c6860b797e.

68. Masys D, Baker D, Butros A, Cowles KE (2002) Giving patients access to their medical records via the internet: The PCASSO experience. J Am Med Informatics Assoc 9: 181–191. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0036490572&partnerID=40&md5=64fbca7b6bcca419686f7e9f95005127.

69. Botts NE, Horan TA, Thoms BP (2011) HealthATM: personal health cyberinfrastructure for underserved populations. Am J Prev Med 40: S115-22. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N

&AN=21521584.

70. Nagykaldi Z, Aspy CB, Chou A, Mold JW (2012) Impact of a wellness portal on the delivery of patient-centered preventive care. J Am Board Fam Med 25: 158–167. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84859324804&partnerID=40&md5=3b398c0f644afa0af8a557dda67b75f4.

71. Lau M, Campbell H, Tang T, Thompson DJS, Elliott T (2014) Impact of patient use of an online patient portal on diabetes outcomes. Can J diabetes 38: 17–21. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=24485208.

72. van der Vaart R, Drossaert CHC, Taal E, Drossaers-Bakker KW, Vonkeman HE, et al. (2014) Impact of patient-accessible electronic medical records in rheumatology: use, satisfaction and effects on empowerment among patients. BMC Musculoskelet Disord 15: 102. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=24673997.

73. Peters KA, Green TF, Schumacher RM (2009) Improving the user interface and adoption of online personal health records. Proceedings of the Human Factors and Ergonomics Society. User Centric, Inc., Oakbrook Terrace, IL, United States, Vol. 2. pp. 704–708. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-77951579399&partnerID=40&md5=f7b46e56e1c8e0bf537a0bedbf227eaa.

74. Tom JO, Mangione-Smith R, Solomon C, Grossman DC (2012) Integrated personal health record use: association with parent-reported care experiences. Pediatrics 130: e183-90. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=22689872.

75. Wilcox L, Woollen J, Prey J, Restaino S, Bakken S, et al. (2016) Interactive tools for inpatient medication tracking: A multi-phase study with cardiothoracic surgery patients. J Am Med Informatics Assoc 23: 144–158. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84955492503&partnerID=40&md5=db78d99aa13b54ec19a90ba5ebf462c0.

76. Tuil WS, Verhaak CM, De Vries Robbé PF, Kremer JAM (2008) IVF patients show three types of online behaviour. Hum Reprod 23: 2501–2505. Available:

https://www.scopus.com/inward/record.uri?eid=2-s2.0-54149111172&partnerID=40&md5=59d038a4b7cb286ca3d04b69e611d8df.

77. Haggstrom DA, Saleem JJ, Russ AL, Jones J, Russell SA, et al. (2011) Lessons learned from usability testing of the VA's personal health record. J Am Med Informatics Assoc 18: i13–i17. Available: http://jamia.oxfordjournals.org/content/18/Supplement_1/i13.abstract.

78. Pai HH, Lau F, Barnett J, Jones S (2013) Meeting the health information needs of prostate cancer patients using personal health records. Curr Oncol 20: e561–e569. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84890055563&partnerID=40&md5=f6a7af5e4224ed0a4977e574a3242d9c.

79. Cunningham SG, Wake DJ, Waller A, Morris AD, Walker J (2013) My Diabetes My Way: An electronic personal health record for diabetes. Br J Diabetes Vasc Dis 13: 143–149. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84880753542&partnerID=40&md5=4f2b839a1043c6026109d67ec6bb063a.

80. Slagle JM, Gordon JS, Harris CE, Davison CL, Culpepper DK, et al. (2010) MyMediHealth - designing a next generation system for child-centered medication management. J Biomed Inform 43: S27-31. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=20937481.

81. Zikmund-Fisher BJ, Exe NL, Witteman HO (2014) Numeracy and Literacy Independently Predict Patients' Ability to Identify Out-of-Range Test Results. J Med Internet Res 16: e187. Available: http://www.jmir.org/2014/8/e187/. Accessed 25 August 2014.

82. Fiks AG, Mayne SL, Karavite DJ, Suh A, O'Hara R, et al. (2015) Parent-reported outcomes of a shared decision-making portal in asthma: a practice-based RCT. Pediatrics 135: e965-73. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=25755233.

83. Kim MI, Johnson KB (2004) Patient entry of information: evaluation of user interfaces. J Med Internet Res 6: e13. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15249262.

84. Hassol A, Walker JM, Kidder D, Rokita K, Young D, et al. (2004) Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. J Am Med Inform Assoc 11: 505–513. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15299001.

85. Sox CM, Gribbons WM, Loring BA, Mandl KD, Batista R, et al. (2010) Patient-Centered Design of an Information Management Module for a Personally Controlled Health Record. J Med Internet Res 12: e36.

86. Phelps RG, Taylor J, Simpson K, Samuel J, Turner AN (2014) Patients' continuing use of an online health record: a quantitative evaluation of 14,000 patient years of access data. J Med Internet Res 16: e241. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=25344621.

87. Pyper C, Amery J, Watson M, Crook C (2004) Patients' experiences when accessing their on-line electronic patient records in primary care. Br J Gen Pract 54: 38–43. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-0347946744&partnerID=tZOtx3y1.

88. Wang M, Lau C, Matsen FA 3rd, Kim Y (2004) Personal health information management system and its application in referral management. IEEE Trans Inf Technol Biomed 8: 287–297. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15484434.

89. Chrischilles EA, Hourcade JP, Doucette W, Eichmann D, Gryzlak B, et al. (2014) Personal health records: a randomized trial of effects on elder medication safety. J Am Med Inform Assoc 21: 679–686. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=24326536.

90. Tulu B, Bar-On I, Strong D, Trudel J, Johnson S, et al. (2011) Personal Health Records: Identifying utilization patterns from system use logs and patient interview. Proceedings of the Annual Hawaii International Conference on System Sciences. Worcester Polytechnic Institute, United States. pp. 2716–2725. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84857927594&partnerID=40&md5=5accb1d175b80622dbfe71abc7ce8a5c.

91. Frize M, Bariciak E, Gilchrist J (2013) PPADS: Physician-PArent Decision-Support for Neonatal Intensive Care. Stud Health Technol Inform 192: 23–27. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=23920508.

92. Gerber DE, Laccetti AL, Chen B, Yan J, Cai J, et al. (2014) Predictors and intensity of online access to electronic medical records among patients with cancer. J Oncol Pract 10: e307–e312. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84911931112&partnerID=40&md5=85d8b70d97c8100ff4ffc4af0c11d4da.

93. Gordon P, Camhi E, Hesse R, Odlum M, Schnall R, et al. (2012) Processes and outcomes of developing a continuity of care document for use as a personal health record by people living with HIV/AIDS in New York City. Int J Med Inform 81: e63-73. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medc&NEWS=N&AN=22841825.

94. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, et al. (2011) Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 18: 318–321. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=21262921.

95. Brewer NT, Gilkey MB, Lillie SE, Hesse BW, Sheridan SL (2012) Tables or bar graphs? Presenting test results in electronic medical records. Med Decis Making 32: 545–553. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=22472914.

96. Sieverink F, Kelders SM, Braakman-Jansen LMA, van Gemert-Pijnen JEWC (2014) The Added Value of Log File Analyses of the Use of a Personal Health Record for Patients With Type 2 Diabetes Mellitus: Preliminary Results. J Diabetes Sci Technol 8: 247–255. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4455425/.

97. Wiljer D, Leonard KJ, Urowitz S, Apatu E, Massey C, et al. (2010) The anxious wait: assessing the impact of patient accessible EHRs for breast cancer patients. BMC Med Inform Decis Mak 10: 46. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=20809950.

98. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, et al. (2011) The digital divide in adoption and use of a personal health record. Arch Intern Med 171: 568–574. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-79955130331&partnerID=40&md5=7545678b39df477df6156d36995af0c5.

99. Taha J, Sharit J, Czaja SJ (2014) The Impact of Numeracy Ability and Technology Skills on Older Adults' Performance of Health Management Tasks Using a Patient Portal. J Appl Gerontol 33: 416–436. Available: http://jag.sagepub.com/content/33/4/416.abstract.

100. Czaja SJ, Zarcadoolas C, Vaughon WL, Lee CC, Rockoff ML, et al. (2015) The usability of electronic personal health record systems for an underserved adult population. Hum Factors 57: 491–506. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=25875437.

101. Jones JB, Weiner JP, Shah NR, Stewart WF (2015) The Wired Patient: Patterns of Electronic Patient Portal Use Among Patients With Cardiac Disease or Diabetes. J Med Internet Res 17: e42. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4376207/.

102. Rubel P, Fayn J, Nollo G, Assanelli D, Li B, et al. (2005) Toward personal eHealth in cardiology. Results from the EPI-MEDICS telemedicine project. J Electrocardiol 38: 100–106. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-26444439044&partnerID=40&md5=f624b202e6a182169ca7ee4c77f6d40a.

103. Burke RP, Rossi AF, Wilner BR, Hannan RL, Zabinsky JA, et al. (2010) Transforming patient and family access to medical information: utilisation patterns of a patient-accessible electronic health record. Cardiol Young 20: 477–484. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=20456816.

104. Rosenbloom ST, Daniels TL, Talbot TR, McClain T, Hennes R, et al. (2012) Triaging patients at risk of influenza using a patient portal. J Am Med Informatics Assoc 19: 549–554. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84865169730&partnerID=40&md5=b3f9aa5afbd4363d9aac836189d54e05.

105. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA (2013) Understanding patient portal use: implications for medication management. J Med Internet Res 15:

e133. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medc&NEWS=N&AN=23823974.

106. Van Vugt M, De Wit M, Sieverink F, Roelofsen Y, Hendriks SH, et al. (2016) Uptake and Effects of the e-Vita Personal Health Record with Self-Management Support and Coaching, for Type 2 Diabetes Patients Treated in Primary Care. J Diabetes Res 2016. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84959387318&partnerID=40&md5=ee3df7d2e5c26f338c78b9ceaf063dc8.

107. Ant Ozok A, Wu H, Garrido M, Pronovost PJ, Gurses AP (2014) Usability and perceived usefulness of Personal Health Records for preventive health care: a case study focusing on patients' and primary care providers' perspectives. Appl Ergon 45: 613–628. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medc&NEWS=N&AN=24119975.

108. Ozok AA, Gurses AP, Wu H, Nelson M, Moen D, et al. (2009) Usability and user acceptance for personal health records: A perspective from healthcare citizens. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 5621 LNCS: 690–699. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-76649096312&partnerID=40&md5=9ba429700e11c7e808b58571a84ff5b6.

109. Segall N, Saville JG, L'Engle P, Carlson B, Wright MC, et al. (2011) Usability evaluation of a personal health record. AMIA Annu Symp Proc 2011: 1233–1242. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=22195184.

110. Taha J, Sharit J, Czaja SJ (2014) Usability of an electronic personal health record (PHR) among a diverse group of adults. Proceedings of the Human Factors and Ergonomics Society. University of Miami, Miller School of Medicine, United States, Vol. 2014–Janua. pp. 619–623. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84957706961&partnerID=40&md5=12abec12644461f4a234e66b3599b435.

111. Britto MT, Jimison HB, Munafo JK, Wissman J, Rogers ML, et al. (2009) Usability Testing Finds Problems for Novice Users of Pediatric Portals. J Am Med Inform Assoc 16: 660–669. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2744717/.

112. Kim EH, Stolyar A, Lober WB, Herbaugh AL, Shinstrom SE, et al. (2007) Usage patterns of a personal health record by elderly and disabled users. AMIA Annu Symp Proc: 409–413. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-56149093601&partnerID=40&md5=28530f29d8074dc7db2b16553301c70f.

113. Ertmer A, Uckert F (2005) User acceptance of and satisfaction with a personal electronic health record. Stud Health Technol Inform 116: 861–866. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=16160366.

114. Forti S, Purin B, Eccher C, Luppino N, Busolin C, et al. (2008) Using interaction design to improve usability of a PHR user interface based on visual elements. Patient-Centered E-Health. e-Health Applied Research Unit, FBK- Fondazione Bruno Kessler, Trento, Italy. pp. 127–142. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84876489032&partnerID=40&md5=ffba93237c6a21e3a4cb8667bcc9da2d.

115. Price M, Bellwood P, Davies I (2015) Using Usability Evaluation to Inform Alberta's Personal Health Record Design. Stud Health Technol Inform 208: 314–318. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84952064891&partnerID=40&md5=60475f6833f6d65a5e1c8d613550a0e5.

116. Cruz-Correia R, Fonseca J, Lima L, Araújo L, Delgado L, et al. (2007) Web-based or paper-based self-management tools for asthma-patients' opinions and quality of data in a randomized crossover study. Studies in Health Technology and Informatics. Department of Biostatistics and Medical Informatics, Faculty of Medicine of University of Porto, Al. Prof. Hernâni Monteiro, 4200-319, Porto, Portugal, Vol. 127. pp. 178–189. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-35549008417&partnerID=40&md5=583f5a99f430bf3bb90900d82b44abc8.

117. Cimino JJ, Patel VL, Kushniruk AW (2001) What do patients do with access to their medical records? Studies in Health Technology and Informatics. Department of Medical Informatics, Columbia University, 622 West 168th Street, VC-5, New York, NY 10032, United States, Vol. 84. pp. 1440–1444. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-77957200480&partnerID=40&md5=b7aeb9216aafe6cabfd672c00219d83b.

118. Weingart SN, Rind D, Tofias Z, Sands DZ (2006) Who uses the patient internet portal? The PatientSite experience. J Am Med Inform Assoc 13: 91–95. Available:

http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=16221943.

119. Lee G, Park JY, Shin S-Y, Hwang JS, Ryu HJ, et al. (2016) Which Users Should Be the Focus of Mobile Personal Health Records? Analysis of User Characteristics Influencing Usage of a Tethered Mobile Personal Health Record. Telemed e-Health 22: 419–428. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84968752863&partnerID=40&md5=0e51b318c2b524fa390843f5d6cd1deb.

120. Nazi KM (2010) Veterans' voices: use of the American Customer Satisfaction Index (ACSI) Survey to identify My HealtheVet personal health record users' characteristics, needs, and preferences. J Am Med Inform Assoc 17: 203–211. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=20190065.

121. Weingart SN, Carbo A, Tess A, Chiappetta L, Tutkus S, et al. (2013) Using a patient internet portal to prevent adverse drug events: A randomized, controlled trial. J Patient Saf 9: 169–175. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84883625156&partnerID=40&md5=6f6c50dcce37d28dae9af3f1fc009003.

122. Turvey CL, Zulman DM, Nazi KM, Wakefield BJ, Woods SS, et al. (2012) Transfer of information from personal health records: a survey of veterans using My HealtheVet. Telemed J E Health 18: 109–114. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=22304439.

123. Do N V, Barnhill R, Heermann-Do KA, Salzman KL, Gimbel RW (2011) The military health system's personal health record pilot with Microsoft HealthVault and Google Health. J Am Med Inform Assoc 18: 118–124. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3116258/.

124. O'Leary KJ, Lohman ME, Culver E, Killarney A, Smith Jr. GR, et al. (2016) The effect of tablet computers with a mobile patient portal application on hospitalized patients' knowledge and activation. J Am Med Informatics Assoc 23: 159–165. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84959486518&partnerID=40&md5=f03e781db2210fd1f47c63645b227e5c.

125. Tenforde M, Nowacki A, Jain A, Hickner J (2012) The association between personal health record use and diabetes quality measures. J Gen Intern Med 27: 420–424.

Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=22005937.

126. Lau AY, Dunn AG, Mortimer N, Gallagher A, Proudfoot J, et al. (2013) Social and self-reflective use of a Web-based personally controlled health management system. J Med Internet Res 15: e211. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=24064188.

127. Kahn JS, Hilton JF, Van Nunnery T, Leasure S, Bryant KM, et al. (2010) Personal health records in a public hospital: Experience at the HIV/AIDS clinic at San Francisco General Hospital. J Am Med Informatics Assoc 17: 224–228. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-77953169561&partnerID=40&md5=0c25dcebec7b02cfe5e3e36d79eef07b.

128. Keith McInnes D, Shimada SL, Rao SR, Quill A, Duggal M, et al. (2013) Personal health record use and its association with antiretroviral adherence: survey and medical record data from 1871 US veterans infected with HIV. AIDS Behav 17: 3091–3100. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=23334359.

129. Ancker JS, Osorio SN, Cheriff A, Cole CL, Silver M, et al. (2015) Patient activation and use of an electronic patient portal. Inform Health Soc Care 40: 254–266. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=24786648.

130. Fine AM, Kalish LA, Forbes P, Goldmann D, Mandl KD, et al. (2009) Parent-driven technology for decision support in pediatric emergency care. Jt Comm J Qual Patient Saf 35: 307–315. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=19565690.

131. Thompson LA, Martinko T, Budd P, Mercado R, Schentrup AM (2016) Meaningful use of a confidential adolescent patient portal. J Adolesc Heal 58: 134–140. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84975491538&partnerID=40&md5=bc43b550f97231ceb20fd756a6331a93.

132. Walker J, Leveille SG, Ngo L, Vodicka E, Darer JD, et al. (2011) Inviting patients to read their doctors' notes: patients and doctors look ahead: patient and physician surveys. Ann Intern Med 155: 811–819. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=22184688.

133. Scott Nielsen A, Halamka JD, Philip Kinkel R (2012) Internet portal use in an academic multiple sclerosis center. J Am Med Informatics Assoc 19: 128–133. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84856803646&partnerID=40&md5=b1aaa3e6cd8e4758682217208f45d9e4.

134. Wright A, Feblowitz J, Maloney FL, Henkin S, Ramelson H, et al. (2014) Increasing patient engagement: patients' responses to viewing problem lists online. Appl Clin Inform 5: 930–942. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=25589908.

135. Wald JS, Businger A, Gandhi TK, Grant RW, Poon EG, et al. (2010) Implementing practice-linked pre-visit electronic journals in primary care: patient and physician use and satisfaction. J Am Med Inform Assoc 17: 502–506. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=20819852.

136. Chen L, Chuang L-M, Chang CH, Wang C-S, Wang I-C, et al. (2013) Evaluating self-management behaviors of diabetic patients in a telehealthcare program: Longitudinal study over 18 months. J Med Internet Res 15. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84891798346&partnerID=40&md5=c73b9e4fa7d769ea837bbf1bd9161e4f.

137. Heyworth L, Paquin AM, Clark J, Kamenker V, Stewart M, et al. (2014) Engaging patients in medication reconciliation via a patient portal following hospital discharge. J Am Med Informatics Assoc 21: e157–e162. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84893548997&partnerID=40&md5=c99cacd80ab6b9dddd3211c7fae5abcb.

138. Schnipper JL, Gandhi TK, Wald JS, Grant RW, Poon EG, et al. (2012) Effects of an online personal health record on medication accuracy and safety: A cluster-randomized trial. J Am Med Informatics Assoc 19: 728–734. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-

84872234391&partnerID=40&md5=81dfae6b526523888ea06227753c5043.

139. Leveille SG, Mejilla R, Ngo L, Fossa A, Elmore JG, et al. (2016) Do Patients Who Access Clinical Information on Patient Internet Portals Have More Primary Care Visits? Med Care 54: 17–23. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84952872563&partnerID=40&md5=7163033934b6f810bace258f766cb36a.

140. Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, et al. (2011) Disparities in enrollment and use of an electronic patient portal. J Gen Intern Med 26: 1112–1116. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-80755175686&partnerID=40&md5=3e73a432d6ca684a8b51de60f1ce5cb3.

141. Pillemer F, Price RA, Paone S, Martich GD, Albert S, et al. (2016) Direct release of test results to patients increases patient engagement and utilization of care. PLoS One 11. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84976602594&partnerID=40&md5=70044089b3e41a866767eb7d363b1b2f.

142. Gordon NP, Hornbrook MC (2016) Differences in access to and preferences for using patient portals and other ehealth technologies based on race, ethnicity, and age: A database and survey study of seniors in a large health plan. J Med Internet Res 18. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84962052833&partnerID=40&md5=ce192bf2c432e77de5da12564c6a278c.

143. Lamsaard J, Pongthananikorn S, Theeraroungchaisri A (2016) Development of a chronic kidney disease knowledge website with electronic personal health records for patients. Thai J Pharm Sci 40: 159–162. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84975464760&partnerID=40&md5=d5ef88750510e2b16f568fda08db7e18.

144. Hulse NC, Ranade-Kharkar P, Post H, Wood GM, Williams MS, et al. (2011) Development and early usage patterns of a consumer-facing family health history tool. AMIA Annu Symp Proc 2011: 578–587. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84874203776&partnerID=40&md5=fb9940317d6bd2ee363b2969d75694d2.

145. Forchuk C, Reiss JP, O'Regan T, Ethridge P, Donelle L, et al. (2015) Client perceptions of the mental health engagement network: A qualitative analysis of an electronic personal health record. BMC Psychiatry. Available:

https://www.scopus.com/inward/record.uri?eid=2-s2.0-84944392163&partnerID=40&md5=a648ad3fc21d1283d7a8d4d1fbc494fc.

146. Emani S, Healey M, Ting DY, Lipsitz SR, Ramelson H, et al. (2016) Awareness and use of the after-Visit summary through a patient portal: Evaluation of patient characteristics and an application of the theory of planned behavior. J Med Internet Res 18. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84973281856&partnerID=40&md5=6ac1d0b5f78fa9f95f6bb6a2121a7c31.

147. Staroselsky M, Volk LA, Tsurikova R, Newmark LP, Lippincott M, et al. (2008) An effort to improve electronic health record medication list accuracy between visits: Patients' and physicians' response. Int J Med Inform 77: 153–160. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-38549111248&partnerID=40&md5=2b2955d89f0af535ec61ef1f31339bba.

148. Grant RW, Wald JS, Schnipper JL, Gandhi TK, Poon EG, et al. (2008) Practice-linked online personal health records for type 2 diabetes mellitus: A randomized controlled trial. Arch Intern Med 168: 1776–1782. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-51649116699&partnerID=40&md5=8e16b19591338274d2c620d83dafe982.

149. Lau AYS, Proudfoot J, Andrews A, Liaw S-T, Crimmins J, et al. (2013) Which Bundles of Features in a Web-Based Personally Controlled Health Management System Are Associated With Consumer Help-Seeking Behaviors for Physical and Emotional Well-Being? J Med Internet Res 15: e79. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3650927/.

150. Davis FD, Bagozzi RP, Warshaw PR (1989) User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Manag Sci 35: 982–1003. Available: http://dx.doi.org/10.1287/mnsc.35.8.982.

151. Nielson J (1993) Usability Engineering. Boston academic press.

152. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Psychology PAH and NMBT-A in, editor. Human Mental Workload. North-Holland, Vol. Volume 52. pp. 139–183. Available: http://www.sciencedirect.com/science/article/pii/S0166411508623869.

153. Mold F, Ellis B, De Lusignan S, Sheikh A, Wyatt JC, et al. (2013) The provision and impact of online patient access to their electronic health records (EHR) and

transactional services on the quality and safety of health care: systematic review protocol. Inform Prim Care 20: 271. Available: http://hijournal.bcs.org/index.php/jhi/article/view/17. Accessed 22 October 2014.

154. International Organization for Standardization (1998) Ergonomic requirements for office work with visual display terminals (VDTs) Part 11 Guidance on usability / 9241-11. Geneve.

155. Onnela J-P, Rauch SL (2016) Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. Neuropsychopharmacology 41: 1691–1696. Available: http://dx.doi.org/10.1038/npp.2016.7.

156. Bargas-Avila JA, Hornbæk K (2011) Old Wine in New Bottles or Novel Challenges: A Critical Analysis of Empirical Studies of User Experience. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '11. New York, NY, USA: ACM. pp. 2689–2698. Available: http://doi.acm.org/10.1145/1978942.1979336.

157. Danaher BG, Seeley JR (2009) Methodological Issues in Research on Web-Based Behavioral Interventions. Ann Behav Med 38: 10.1007/s12160-009-9129-0. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3846298/.

5.9 SUPPLEMENTARY MATERIAL

Supplementary Table 5.1: Syntax of searches implemented in Medline and Scopus.

Medline search	Scopus search
(User-computer interface [MeSH] OR Computers/utilization [MeSH] OR Decision making [MeSH] OR Medication Errors/[MeSH] OR Usability.tw OR (User* adj3 (experience* OR satisfaction).tw) OR UX.tw OR Interface*.tw OR User*cent*d design.tw OR Patient*cent*d design.tw OR Think*aloud.tw OR Talk*aloud.tw OR Screen*record*.tw OR Video*record*.tw OR Eye*track*.tw OR Gesture*track*.tw OR Log*.tw OR Click*through*analysis.tw OR Google*analytics.tw OR ((Risk* OR Data OR result*) adj3 interpret*).tw OR ((Out*of*range OR abnormal OR normal OR in*range) adj3 (identif* OR detect* OR spot*)).tw OR (action* OR decision*).tw OR Information*processing.tw OR Decision*making.tw OR (Medic* adj3 (error* OR discrepan* OR reconcil* OR accuracy)).tw) AND (Personal Health Record [MeSH] OR Patient Access to Records [MeSH] OR Patient* portal*.tw OR Personal* adj3 Health Record*.tw)	(TITLE-ABS-KEY("Patient Access to Record$")) OR (TITLE-ABS-KEY(personal W/3"Health Record$")) OR (TITLE-ABS-KEY("patient* portal*")) AND ((TITLE-ABS-KEY(user$ W/3 experience)) OR (TITLE-ABS-KEY(user$ W/3 satisfaction)) OR (TITLE-ABS-KEY(usability)) OR (TITLE-ABS-KEY("Decision$making")) OR (TITLE-ABS-KEY(computer$utili$ation)) OR (TITLE-ABS-KEY(talk$aloud)) OR (TITLE-ABS-KEY(think$aloud)) OR (TITLE-ABS-KEY("Patient$cent*d design")) OR (TITLE-ABS-KEY("User$cent*d design")) OR (TITLE-ABS-KEY(interface$)) OR (TITLE-ABS-KEY(ux)) OR (TITLE-ABS-KEY(google$analytics)) OR (TITLE-ABS-KEY(click$through$analysis)) OR (TITLE-ABS-KEY(log*)) OR (TITLE-ABS-KEY(eye*track*)) OR (TITLE-ABS-KEY("screen*record*")) OR (TITLE-ABS-KEY("video*record*")) OR (TITLE-ABS-KEY((risk OR data OR result) W/3 interpretation)) OR (TITLE-ABS-KEY("Information$processing")) OR (TITLE-ABS-KEY(action* OR decision*)) OR (TITLE-ABS-KEY ((Out*of*range OR abnormal OR normal OR in*range) W/3 (identif* OR detect* OR spot*))) OR (TITLE-ABS-KEY(medication*error*)))

Supplementary Table 5.2: Main items extracted from included publications.

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[37]	2014 Italy	Hypertension diagnostic kit Primary care patients	Within-subject design Naturalistic	Field Longitudinal (7 days)	Target population 15	Interaction with the system: Overall usability (Questionnaire [Non-validated questionnaire]); Receiving information from the system: Data completeness (Data logs analysis [% patients with complete data recorded])
[38]	2010 US	George Washington University Patients with mental health issues	Between-subject design Controlled	Field Longitudinal (90 days)	Target population 48 (25 paper chart vs 23 patient portal)	Interaction with the system: Usage (Data logs analysis [% of patients with at least 1 day rated]); Receiving information from the system: Data completeness (Data logs analysis [% of day with complete data]); Data quality (Data logs analysis [correlation with clinicians judgement]); Data quantity (Data logs analysis [Number of recorded days])
[39]	2014 US	MyAsthma Parents of children with asthma	Within-subject design Controlled	NR (likely to be lab) Cross-sectional	Target population 5 parents and 5 clinicians	Interaction with the system: Usability issues (Think-aloud [Usability issues found]); Effectiveness (Task analysis [% of users that completed all tasks]); Overall usability (Questionnaire [Non-validated questionnaire])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[40]	2015 US	Patient centered toolkit Patients during hospitalisation and their carers	Within-subject design Naturalistic	Field Longitudinal (6-months)	Target population 239, 16 also usability questionnaire (56% respense rate)	Interaction with the system: Overall usability (Questionnaire [SUS]); Usage (Login analysis [Frequency of use]); Overall usability (Questionnaire [Non-validated questionnaire]); Receiving information from the system: Functionalities usage (Data logs analysis [% of patients inputting overall goal at least once,% of patients inputting daily goal at least once,% of patients inputting preferences at least once,% of patients sending at least one message, % of patients accessed a functionality]); Functionalities usage (System interaction logs analysis [% of patients inputting overall goal at least once,% of patients inputting daily goal at least once,% of patients inputting preferences at least once,% of patients sending at least one message, % of patients accessed a functionality])
[41]	2007 Canada	iampregnant.org Pregnant women	Between-subject design Controlled	Field Longitudinal (2 years, 11 months)	Target population 193 (97 personalised information/96 general information)	Interaction with the system: Usage (Login analysis [number of logins]); Overall usability (Questionnaire [Non-validated questionnaire, easy-of-use and perceived usefulness metioned]); Receiving information from the system: Functionalities usage (System interaction logs analysis [functionalities accessed])
[42]	2010 UK	HealthSpace Primary care patients	Within-subject design Naturalistic	Field Longitudinal (3 years)	Target population 21	Interaction with the system: Usage (System interaction logs analysis [Number of patients with at least a login,Number and % of patients that activated account to access records])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[43]	2012 US	Dietary intake monitoring application Patients in haemodialysis	Within-subject design Naturalistic	Field Longitudinal (6 weeks)	Target population 18	Interaction with the system: Usage (Login analyisis [Frequency of use]); Overall usability (Questionnaire [Validated usability questionnaire, Susan Rawl]); Receiving information from the system: Functionalitites usage (System interaction logs analysis [Number of features used per patient])
[44]	2004 US	University of Washington Patients needing orthopedic care	Within-subject design Naturalistic	Field Longitudinal (255 days)	Target population NA	Interaction with the system: Overall usability (Questionnaire [Non-validated questionnaire]); Receiving information from the system: Functionalities usage (System interaction logs analysis [Number of subject that used the referral functionality,% of requests from different users,time system used,time to complete a referral request,time to submit a request after creating account,average time spent on each category of pages]); System resource utilisation (System interaction logs analysis [Number of page requests])
[45]	2010 US	Google Health, Microsoft Healthvault General population	Within-subject design Controlled	Lab Cross-sectional	Healthy volunteers 15	Interaction with the system: Usability issues (Questionnaire [Usability issues found questionnaire]); Effectiveness (Task analysis [% of successful tasks]); Efficiency1 (UI event mining [number of clicks and keystroke behaviour]); Efficiency2 (Task analysis [time to complete tasks])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[46]	2011 US	Google Health, Microsoft HealthVault, WorldMedCard General population	Within-subject design Controlled	Lab Cross-sectional	Healthy volunteers 18	Interaction with the system: Usability issues (Video recordingand Interviews [usability issues found]); Efficiency (task analysis [Time to complete tasks])
[47]	2006 US	University of Washington Elderly	Within-subject design Naturalistic	Lab Cross-sectional	Target population 41	Interaction with the system: Effectiveness (Cognitive task analysis [% of patients that required assistance with the patient portal and task])
[48]	2015 US	MyChart Patients with mental health issues	Within-subject design Naturalistic	Remote Cross-sectional	Target population 177	Interaction with the system: Usage (Questionnaire [% of users,% of participants that used it more than once]); Overall usability (Questionnaire [Questionnaire based on previous work, easy to use]); Receiving information from the system: Functionalities usage (Questionnaire [non-validated questionnaire])
[49]	2009 US	University of Washington Elderly	Within-subject design Naturalistic	Field Longitudinal (33 months)	Target population 330 (14 replied to survey)	Interaction with the system: Usage (Login analysis [% used it at least once,% that used it more than once, % of activities that required assistance]); Overall usability (Questionnaires [Non-validated questionnaire {perceived usefulnees mention}, based on previous work]); Receiving information from the system: Functionalities usage (Questionnaire [Frequency of use])
[50]	2014 UK	myhealthlocker Patients with mental health issues	Between-subject design Naturalistic	Field Cross-sectional	Target population 8 (questionnaire) /23 (system use)	Interaction with the system: Overall usability (Questionnaire [Non-validated likert scores])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[51]	2012 Australia	Healthy.me Couples undergoing in-vitro fertilization (IVF)	Within-subject design Naturalistic	Field Longitudinal (8 weeks)	Target population 14	Receiving information from the system: Additional information accessed (Interviews [patients accessing additional resources]); Functionalities usage (Login analysis [% patients accessed each features, % patients entered data])
[52]	2008 US	Patient Gateway Primary care patients	Within-subject design Naturalistic	Field Longitudinal (18 months)	Target population 1,457	Interaction with the system: Usage (Login analysis [% of patients accessed the module]); Overall usability (Questionnaire [Non-validated questionnaire]); Overall usability (Questionnaire [Nin-validated questionnaire, easy of use]); Receiving information from the system: Data sharedness (Data logs analysis [% of information reviewed by GPs]); Functionalities Usage (Data logs analysis [% of patients accessed medication module,% of patients edited information and sent to the GP])
[53]	2010 US	University of Massachusetts Patients with non cancer chronic pain	Within-subject design Controlled	Lab Cross-sectional	Target population 4	Interaction with the system: Usability issues (Video recording [Usability issue found]); Overall usability (interview [qualitative, easy of use and perceived usefullness])
[54]	2011 US	www.MyPreventiveCare.org Primary care patients	Within-subject design Naturalistic	Field Longitudinal (6 months)	Target population 50,124	Interaction with the system: Usage (System interaction logs analysis [% of patients that establisehd and account, % of users, % of patients used it more than once]); Receiving information from the system: System resource utilisation (Google analytics [Average number of visitors per month, average visit time, average number of pages visited])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[55]	2011 US	Colorado Care Tablet Elderly	Within-subject design Controlled	Lab Cross-sectional	Target population 1	Interaction with the system: Usability issues (Think aloud [Usability issues found])
[56]	2010 US	University of Hyogo Patients with diabetes	Within-subject design Controlled	Lab Cross-sectional	Healthy volunteers 18	Interaction with the system: Overall usability (Questionnaire [Non-validated questionnaire])
[57]	2016 US/Taiwan	DiabeticLink Patients with diabetes	Within-subject design Controlled and Naturalistic	Lab/Remote/Field Cross-sectional/Longitudinal (18 months)	Target population and Healthy volunteers 24	Interaction with the system: Usage (System interaction logs analysis [Registred users, Number of users, Number of sessions]); Overall usability (Questionnaire and Video recording [non-validated questionnaire]); Receiving information from the system: System resource utilisation (System interaction logs analysis [Total page views])
[58]	2015 US	EpicCare General population	Within-subject design Naturalistic	Field Longitudinal (24 months)	Target population 534	Interaction with the system: Usage (System interaction logs analysis [% of patients that created an account, % of patients that activated the account, median number of logins per month]); Receiving information from the system: Functionalities usage (System interaction logs analysis [% of patients using each functionalities, % of patients using 0,2,3,4 functions])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[59]	2010 Canada	University of British Columbia Mothers after birth	Within-subject design Naturalistic	Field Longitudinal (1 week)	Target population 60 (27 paper /33 patient portal)	Interaction with the system: Usage (System interaction logs analysis [frequency of use]); Overall usability (Questionnaire [non validated likert-score]); Receiving information from the system: Data quantity (Data logs analysis [Differences in data recorded]); Functionalities usage (System interaction logs analysis [Frequency of use])
[60]	2015 UK	myhealthlocker Patients with mental health issues	Within-subject design Naturalistic	Field Longitudinal (12 months)	Target population 58 (10 interviews)	Interaction with the system: Effectiveness (System interaction logs analysis [Number of sessions with assistance, % of forms submitted with assistance]); Usage (System interaction logs analysis [% of users {at least two logins and one Patient Reported Outcome Form submitted},Number of logins, Mean patient portal utlisation in months]); Overall usability (Questionnaire and Interviews [Non-validated questionnaire]); Receiving information from the system: Functionalities usage (Login data [Number of Patient Reported Outcome Form submitted])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[61]	2007 US	Patient Gateway Primary care patients	Within-subject design Naturalistic	Field Longitudinal (14 months)	Target population 2,779 (437 survey)	Interaction with the system: Usage (System interaction logs analysis [% of patients that accessed the module within those that signed up]); Overall usability (Questionnaire [Non-validated questionnaire, easy of use]); Receiving information from the system: Data sharedness (Login data [% of updated records seen by clinicians]); Functionalities usage (Data logs analysis [% of patients with an upcoming visit that reviewed and updated their records])
[62]	2012 France	Sanoia Patients with idiopathic thrombocytopenic purpura	Between-subject design Controlled	Field Longitudinal (6 months)	Target population 43 (15,14,14)	Interaction with the system: Overall usability (Questionnaire [Non-validated questionnaire]); Usage (System interaction logs analysis [% adherence {minimum of 2 logins},Average frequency of use]); Receiving information from the system: Functionalities usage (System interaction logs analysis [% of patients recorded data])
[63]	2012 Canada	My Diabetes Wellness Portal, ProPortal Patients with diabetes or prostate cancer	Between-subject design Naturalistic	Field Longitudinal (3 months)	Target population 46/53	Interaction with the system: Usage (System interaction logs analysis [Average number of login per day, average session length in minutes]); Overall usability (Questionnaire and focus group [Non-validated questionnaire {based on previous work} and focus groups, Thematic analysis]); Overall usability (Questionnaire [Non-validated questionnaire {based on previous work} and focus group, based on previous work, Thematic analysis])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[64]	2014 US	MyHealthProfile,MyHealthProfile-Plus Patients with HIV	Between-subject design Naturalistic	Field/lab Cross-sectional	Target population 42 surveys/15 focus groups	Interaction with the system: Overall usability (Questionnaire [Non-validated questionnaire, based on two validated instruments, easy of use])
[65]	2014 US	Johns Hopkins University Elderly	Within-subject design Controlled	Lab Cross-sectional	Target population 14 patients/19 careers	Interaction with the system: Effectiveness (Cognite task analysis [cognitive walkthrough] [% participants asking for assistance]); Efficiency (Cognite task analysis [cognitive walkthrough] [Time to complete the task]); Overall usability (Questionnaire [non-validated questionnaire, easy of use])
[66]	2013 US	MyChart General population	Within-subject design Controlled	Lab Cross-sectional	Target population 107	Interaction with the system: Effectiveness (Cognitive task anaysis [% of patients that completed correctly each task,Overall Performance score])
[67]	2013 US	Johns Hopkins University Elderly	Within-subject design Controlled	Lab Cross-sectional	Target population 7 patients/16 caregivers	Interaction with the system: Efficiency (Task analysis [Time to complete the task]); Overall usability (Questionnaire [non-validated questionnaire, easy of use])
[68]	2002 US	PCASSO General population	Within-subject design Naturalistic	Field Longitudinal (12 months)	Target population 41	Interaction with the system: Usage (Login data [% that used It at least once, % that used it at least 5 times]); Overall usability (Questionnaire [non-validated questionnaire])
[69]	2011 US	HealthATM Underserved populations	Within-subject design Naturalistic	Field Cross-sectional	Target population 144 (115 cmpleted)	Interaction with the system: Overall usability (Questionnaire [Validated questionnaire, System usability Scale survey])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[70]	2012 US	Wellness Portal Primary care patients	Mixed Naturalistic and Controlled	Field Longitudinal (6 months + 12 months)	Target population 30/422	Interaction with the system: Usage (System interaction logs analysis [% of patients used it at least once, frequency of use]); Overall usability (Questionnaire [Non- validated questionnaire, easy of use]); Receiving information from the system: Functionalities usage (System interaction logs analysis [% of patients that complete at least one task, % of patients used every function])
[71]	2014 Canada	BCDiabetes Patients with diabetes	Within-subject design Naturalistic	Field Longitudinal (4 years)	Target population 157	Interaction with the system: Usage (Login data [% of patients that used it at east once, frequency of use])
[72]	2014 The Netherlands	www.reumacentrumtwente.nl Patients with rheumatoid arthritis	Within-subject design Naturalistic	Field Longitudinal (5 months)	Target population 194	Interaction with the system: Usage (Questionnaire [% of users, frequency of use]); Overall usability (Questionnaire [non-validated questionnaire, easy of use]); Receiving information from the system: Functionalities usage (Questionnaire [% of patients who accessed their records])
[73]	2009 US	Google Health, Microsoft HealthVault General population	Within-subject design Controlled	Lab Cross-sectional	Target population 30	Interaction with the system: Effectiveness (Task analysis [Errors qualitative]); Overall usability (Questionnaire [Non-validated questionnaire]); Overall usability (Questionnaire [Non-validated queationnaire])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[74]	2012 US	MyGroupHealth General population	Within-subject design Naturalistic	Field Cross-sectional	Target population 256	Interaction with the system: Usage (Questionnaire [% of users in the last 12-months, frequency of use]); Receiving information from the system: Functionalities usage (Questionnaire [frequency of use of each specific feature])
[75]	2016 US	myNYP Patients during hospitalisation	Within-subject design Naturalistic	Field Cross-sectional	Target population 21 phase one/12 phase two	Interaction with the system: Usability issues (Observation Interviews and Think aloud [Thematic analysis]); Usage (System interaction logs analysis [% of users {phase one}]); Overall usability (Interviews [Thematic analysis]); Receiving information from the system: Functionalities usage (System interaction logs analysis [Frequency])
[76]	2008 The Netherlands	Radboud University Nijmegen Medical Centre Couples undergoing in-vitro fertilization (IVF)	Within-subject design Naturalistic	Field Longitudinal (3 years)	Target population 1150	Receiving information from the system: Behavioural styles (Factor Analysis [% of patients pertaining to each behavioural style]); Resources utilisation (System interaction logs analysis [Number of page views for each functionality])
[77]	2011 US	MyHealtheVet Veterans	Within-subject design Controlled	Lab Cross-sectional	Target population 24	Interaction with the system: Usability issues (Video recording Face recording and interviews [Usability issues found]); Efficiency (Task analysis [Time to complete a task])
[78]	2013 Canada	PROVIDER Patients with prostate cancer	Within-subject design Naturalistic	Field Longitudinal (6 months)	Target population 22	Interaction with the system: Usage (System interaction logs analysis [Number of logins, frequency of use]); Overall usability (Questionnaire [non validated questionnaire])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[79]	2013 UK (Scotland)	MyDiabetesMyWay Patients with diabetes	Within-subject design Naturalistic	Field Longitudinal (2 years)	Target population 3,391	Interaction with the system: Usage (System interaction logs analysis [% of users,number of distinct users per months])
[80]	2010 US	MyMediHealth Parents of children with a chronic condition	Within-subject design Controlled	Remote Cross-sectional	Target population 202	Interaction with the system: Effectiveness (Task analysis [% of correct tasks,Number of attempts to complete task]); Overall usability (Questionnaire [Non-validated questionnaire])
[81]	2014 US	University of Michigan Patients with a chronic condition	Between-subject design Controlled	Remote Cross-sectional	Target population 1817 (971 with diabetes)	Interaction with the system: Effectiveness (Task analysis [% of correct tasks]); Overall usability (Questionnaire [Non-validated questionnaire]); Influence of the system on decisions-making: Patients decisions (Questionnaire [% of patients that would call their doctor immediately])
[82]	2015 US	MyAsthma Parents of children with asthma	Between-subject design Controlled	Field Longitudinal (12 months)	Target population 30/30	Interaction with the system: Usage (Questionnaire [% of submitted portal surveys, frequency]); Overall usability (Questionnaire [Non-validated questionnaire])
[83]	2004 US	Johns Hopkins University (2) Patients requiring thyroid hormone treatment	Within-subject design Controlled	Lab Cross-sectional	Target population 14	Interaction with the system: Effectiveness (Task analysis [% of correct tasks])
[84]	2004 US	MyChart General population	Within-subject design Naturalistic	Field Cross-sectional	Target population 1,421 replied to the survey/30 focus groups	Interaction with the system: Overall usability (Questionnaire and focus group [Non-validated questionnaire, easy of use])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[85]	2010 US	Boston University Parents of children with attention deficit	Within-subject design Controlled	Lab Cross-sectional	Target population NR/7	Interaction with the system: Usability issues (Think aloud [issues found]); Effectiveness (Task analysis [% of correct tasks]); Efficiency1 (Task analysis [Time to complete tasks]); Efficiency2 (Survey-Questionnaire [Validated questionnaire {Task burden, NASA Task Load Index}])
[86]	2014 UK	PatientView Patients with chronic kidney disease	Within-subject design Naturalistic	Field Longitudinal (4 years)	Target population 11,352	Interaction with the system: Usage (System interaction logs analysis [% of users, % patients used beyond one month, % of patients that used it beyond 6 months, frequency of use])
[87]	2004 UK	EMIS Primary care patients	Within-subject design Naturalistic	Lab Cross-sectional	Target population 100	Interaction with the system: Efficiency (Task analysis [Time to complete task]); Overall usability (Questionnaire [Non- validated questionnaire, easy of use])
[88]	2004 US	University of Washington Patients needing orthopedic care	Within-subject design Naturalistic	Field Longitudinal (7 months)	Target population 61	Interaction with the system: Efficiency (System interaction logs analysis [Time to complete request]); Overall usability (Questionnaire [Non-validated questionnaire]); Receiving information from the system: Functionalities usage (System interaction logs analysis [number of patients that used functionality])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[89]	2014 US	University of Iowa General population	Between-subject design Naturalistic	Field Longitudinal (1 year)	Target population 802 (portal): 273 (non portal)	Interaction with the system: Usage (System interaction logs analysis [% of patients that logged on, % of high users {defined as logging in and recording data in separate occasions}, % low users {recording data in a single occasion}]); Receiving information from the system: Functionalities usage (System interaction logs analysis [% of patients that recorded data])
[90]	2011 US	Worcester Polytechnic Institute Primary care patients	Within-subject design Naturalistic	Field Longitudinal (19 months)	Target population NA/40	Interaction with the system: Usage (System interaction logs analysis [Number of users, frequency of use]); Receiving information from the system: Behavioural styles (System interaction logs analysis and Interviews [Interaction styles found]); Functionalities usage (System interaction logs analysis [Average number of feature accessed per user])
[91]	2013 Canada	Physician PArent Decision Support System Parents of premature infants	Within-subject design Controlled	Lab Cross-sectional	Target population 8	Interaction with the system: Effectiveness (Task analysis [number of errors]); Overall usability (Questionnaire [Non-validated questionnaire {Easy of use, learnability}, number of positive comments])
[92]	2014 US	MyChart Patients with cancer	Within-subject design Naturalistic	Field Longitudinal (5 years)	Target population NA	Interaction with the system: Usage (System interaction logs analysis [Frequency of use, number of frequent infrequent users {compared to median value of total logins}]); Receiving information from the system: Functionalities usage (System interaction logs analysis [total number of times a functionalities was used, % for each functionality])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[93]	2012 US	MyHealthProfile Patients with HIV	Within-subject design Naturalistic	Field Longitudinal (2 years)	Target population NA	Interaction with the system: Usage (System interaction logs analysis [% of users {at least 2 logins on separate days}, frequency of use])
[94]	2011 US	kp.org Patients with diabetes	Within-subject design Naturalistic	Field Longitudinal (12 months)	Target population 14,102	Interaction with the system: Usage (System interaction logs analysis [% of users]); Receiving information from the system: Functionalities usage (Login data anlysis [% of patients that used a specific functionality])
[95]	2012 US	University of North Carolina General population	Mixed Controlled	Lab Cross-sectional	Target population 106	Interaction with the system: Effectiveness (Task analysis [Recall, Accuracy]); Efficiency (Task analysis [Time to complete the task]); Overall usability (Questionnaire [Non-validated questionnaire, easy of use])
[96]	2014 The Netherlands	e-Vita Patients with diabetes	Within-subject design Naturalistic	Field Longitudinal (6 weeks)	Target population 1197	Interaction with the system: Usage (System interaction logs analysis [% users, frequency of use]); Receiving information from the system: Functionalities usage (System interaction logs analysis [functionalities accessed during first login, number of functionalities accessed during the first login])
[97]	2010 Canada	InfoWell Patients with cancer	Within-subject design Naturalistic	Field Longitudinal (6 weeks)	Target population 311	Interaction with the system: Overall usability (Questionnaire [Non-validated questionnaire, based on PHWSUQ]); Receiving information from the system: Functionalities usage (Questionnaire [functionalities accessed])
[98]	2011 US	Patient Gateway Primary care patients	Within-subject design Naturalistic	Field Longitudinal (20 months)	Target population 75,056	Interaction with the system: Usage (System interaction logs analysis [% of adopters {patients who activated their account}, frequency of use])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[99]	2014 US	MyChart General population	Within-subject design Controlled	Lab Cross-sectional	Target population 51	Interaction with the system: Effectiveness (Task analysis [performance score {sum of all point got in simple and hard tasks}]); Overall usability (Questionnaire [Non-validated questionnaire])
[9]	2002 US	PatCIS General population	Within-subject design Naturalistic	Field Longitudinal (36 months)	Target population 13	Interaction with the system: Overall usability (Questionnaire [Non-validated questionnaire]); Usage (System interaction logs analysis [% of users, frequency of use]); Overall usability (Interviews [qualitative]); Receiving information from the system: Functionalities usage (System interaction logs analysis [functionalities used, number of actions per session,])
[100]	2015 US	University of Miami General population	Within-subject design Controlled	Lab Cross-sectional	Target population 51	Interaction with the system: Effectiveness (Cognitive task anaysis [overall performance score]); Overall usability (Questionnaire [Non validated questionnaire])
[101]	2015 US	MyGeisinger Primary care patients	Within-subject design Naturalistic	Field Longitudinal (12 months)	Target population 3,297	Interaction with the system: Usage (System interaction logs analysis [% of patients that used it at least twice, session length,frequency of use]); Receiving information from the system: Behavioural styles (System interaction logs analysis and PCA and hierarchical clustering [Number of patients in each cluster and characteristics of each cluster]); Functionalities usage (System interaction logs analysis [functionalities used, frequency of use, ratio between administrative and care-related functions])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[102]	2005 France	EPI-MEDICS General population	Within-subject design Naturalistic	Field Longitudinal (6 months)	Target population 50	Receiving information from the system: Data quality (Data logs analysis [% of data judged up to standard]); Data quantity (Data logs analysis [number of records]); Influence of the system on decisions-making: Clinicians decisions (Data log analysis [patients diagnosed with specific condition on the basis of the recorded data])
[103]	2010 US	I-Rounds Parents of children with cardiac disease	Within-subject design Naturalistic	Field Longitudinal (2.5 years)	Target population 270	Interaction with the system: Usage (System interaction logs analysis [% of users {did not clearly define who a user was}, frequency of use]); Receiving information from the system: Functionalities usage (System interaction logs analysis [Funcionalities accessed, frequency of use])
[104]	2012 US	MyHealthAtVanderbilt General population	Within-subject design Naturalistic	Field Longitudinal (9 weeks)	Target population NA	Interaction with the system: Usage (System interaction logs analysis [Total number of sessions])
[105]	2013 US	MyHealthAtVanderbilt Patients with diabetes	Within-subject design Naturalistic	Field Cross-sectional	Target population 75	Interaction with the system: Usage (Questionnaire [% of users, frequency of use]); Receiving information from the system: Functionalities usage (focus groups [thematic analysis])
[106]	2016 The Netherlands	e-Vita Patients with diabetes	Within-subject design Naturalistic	Field Longitudinal (1 year)	Target population 132	Interaction with the system: Usage (System interaction logs analysis [% of patients that used it at least once, % of patients that used it more tna once, frequency of use])
[107]	2014 US	MySafe-T.net Primary care patients	Within-subject design Controlled	Lab Cross-sectional	Target population 22	Interaction with the system: Usability issues (Think aloud Interviews and focus groups [Usability issues found]); Overall usability (Questionnaire [Validated instruments, preceived usefulness])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[108]	2009 US	MySafe-T.net Primary care patients	Within-subject design Controlled	Lab Cross-sectional	Target population 22	Interaction with the system: Usability issues (Interviews [Usability issues found]); Overall usability (Questionnaire [Non-validated questionnaire, preceived usefulness and ease of use])
[109]	2011 US	HealthView Patients with cardiovascular disease	Within-subject design Controlled	Lab Cross-sectional	Target population 20	Interaction with the system: Usability issues (Think aloud and Interviews [usability issues found]); Effectiveness (Task analysis [% of errors/give up, % of help requests]); Overall usability (Questionnaire [Validated questionnaire, Chin et al])
[110]	2014 US	MyChart General population	Within-subject design Controlled	Lab Cross-sectional	Target population 107	Interaction with the system: Effectiveness (Task analysis [performance score {sum of all point got in simple and hard tasks}]); Overall usability (Questionnaire and Interviews [Non-validated questionnaire])
[111]	2009 US	MyCare Connection Parents of children with diabetes, arthritis and cystic fibrosis	Within-subject design Controlled	Lab Cross-sectional	Target population 16	Interaction with the system: Usability issues (think aloud [Usability issues found]); Effectiveness (Task analysis [% of successful tasks]); Efficiency (Task analysis [time to complete tasks]); Overall usability (Questionnaire [validated questionnaire, IBM computer usability satisfaction questionnaires])
[112]	2007 US	Nanyang technological University Elderly (disabled)	Within-subject design Naturalistic	Lab Longitudinal (18 months)	Target population 46	Interaction with the system: Usage (System interaction logs analysis [frequency of use]); Receiving information from the system: Functionalities usage (System interaction logs analysis [% of patients that accessed each functionality])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[113]	2005 Denmark	akteonline.de General population	Between-subject design Controlled	Field Longitudinal (1 month)	Target population 50	Interaction with the system: Usage (Login data analys [frequency of use]); Overall usability (Questionnaire [Validated questionnaire {calculated reliability in the study}])
[114]	2008 Italy	Fondazione Bruno Kessler Primary care patients	Within-subject design Controlled	Lab Cross-sectional	Target population 16	Interaction with the system: Usability issues (Think aloud [Usability issues found]); Efficiency (Task analysis [Time to complete the tasks]); Overall usability (Questionnaire [Non-validated questionnaire])
[115]	2015 Canada	Alberta personal health record General population	Within-subject design Controlled	Lab Cross-sectional	Target population 21	Interaction with the system: Usability issues (Think aloud [Usability issues found])
[116]	2007 Portugal	P'ASMA Patients with asthma	Within-subject design Naturalistic	Field Longitudinal (8 weeks)	Target population 21	Interaction with the system: Overall usability (Questionnaire [Non-validated questionnaire]); Receiving information from the system: Data quantity (Data logs analysis [Number of records [paper-based vs patient portal]])
[117]	2001 US	PatCIS General population	Within-subject design Naturalistic	Field Longitudinal (19 months)	Target population 13	Interaction with the system: Overall usability (Questionnaire [Non-validated questionnaire]); Usage (System interaction logs analysis [% of users, frequency of use, session length]); Overall usability (Interviews [qualitative]); Receiving information from the system: Functionalities usage (System interaction logs analysis [number of actions per session])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[118]	2006 US	PatientSite General population	Within-subject design Naturalistic	Field Longitudinal (12 months)	Target population 780	Interaction with the system: Usage (System interaction logs analysis [% of patient that logged in each month]); Receiving information from the system: Functionalities usage (System interaction logs analysis [% of patients that accessed a functionality each month])
[119]	2016 Korea	My Chart in My Hand General population	Within-subject design Naturalistic	Field Longitudinal (16 months)	Target population 7,096	Interaction with the system: Usage (System interaction logs analysis [frequency of use, total number of logins, light user if below the median, heavy if above the median]); Receiving information from the system: Functionalities usage (System interaction logs analysis [frequency of use of each functionality, total number times a functionality was used])
[120]	2009 US	MyHealtheVet Veterans	Within-subject design Naturalistic	Field Cross-sectional	Target population 100,617	Interaction with the system: Overall usability (Questionnaire [validated questionnaire, American CustomerSatisfaction Index])
[121]	2013 US	PatientSite Primary care patients	Between-subject design Controlled	Field Longitudinal (3 months)	Target population 738 (375 intervntion, 363 controls)	Interaction with the system: Usage (System interaction logs analysis [% users used it at least once]); Influence of the system on decisions-making: Clinicians decisions (Data log analysis and Interviews [additional preventable/ameliorable ADEs based on patients info])
[122]	2012 US	MyHealtheVet Veterans	Within-subject design Naturalistic	Field Cross-sectional	Target population 25,898 and 18,471	Receiving information from the system: Functionalities usage (Questionnaire [validated questionnaire, American Customer])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[123]	2011 US	MiCARE Veterans	Within-subject design Naturalistic	Field Longitudinal (10 months)	Target population 250	Interaction with the system: Overall usability (Questionnaire [Non- validated questionnaire]); Receiving information from the system: System resource utilisation (Google Analytics [Visits, page count, time spent])
[124]	2016 US	PowerChart Millennium patient portal Patients during hospitalisation	Between-subject design Controlled	Field Cross-sectional	Target population 120 vs 184	Interaction with the system: Effectiveness (Task analysis [Correct information recollected]); Usage (Questionnaire [% of users, % of patients that used it more than once per day]); Overall usability (Questionnaire [Non-validated questionnaire])
[125]	2012 US	MyChart Primary care patients	Within-subject design Naturalistic	Field Longitudinal (12 months)	Target population 10,746	Interaction with the system: Usage (System interaction logs analysis [% of Patients enrolled, frequency of use]); Receiving information from the system: Functionalities usage (System interaction logs analysis [% of users using a functionality])
[126]	2013 Australia	Healthy.me General population	Within-subject design Naturalistic	Field Longitudinal (5 months)	Target population 709	Interaction with the system: Usage (Questionnaire [% of users, frequency of use]); Receiving information from the system: Functionalities usage (Questionnaire [% of patients accessed functionalities]); Influence of the system on decisions-making: Patients decisions (Questionnaire [% of users reporting higher intention to practice a healthy lifestyle])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[127]	2010 US	myHERO Patients with HIV	Within-subject design Naturalistic	Field Longitudinal (20 months)	Target population 3,760	Interaction with the system: Usage (System interaction logs analysis [% of registered patients, % of patients activated account, % of users, frequency of use]); Overall usability (Questionnaire [Non validated questionnaire]); Receiving information from the system: Functionalities usage (System interaction logs analysis [% of patients accessed functionalities and total number of visits])
[128]	2013 US	MyHealtheVet Patients with HIV	Within-subject design Naturalistic	Field Cross-sectional	Target population 1,871	Interaction with the system: Usage (Questionnaire [validated questionnaire, American CustomerSatisfaction Index])
[129]	2015 US	MyChart General population	Within-subject design Naturalistic	Field Cross-sectional	Target population 180	Interaction with the system: Usage (System interaction logs analysis [% of users, frequency of use])
[130]	2009 US	ParentLink Tiertiary care	Within-subject design Controlled	Field Cross-sectional	Target population 1,411	Receiving information from the system: Data completeness (Data logs analysis [% patients with complete data recorded]); Influence of the system on decisions-making: Clinicians decisions (Data log analysis [additional correct actions taken])
[131]	2016 US	MyChart Parents and teenager	Between-subject design Naturalistic	Field Longitudinal (30 months)	Target population 937 vs 936	Interaction with the system: Usage (System interaction logs analysis [% of users]); Receiving information from the system: Functionalities usage (System interaction logs analysis [% of users that used a functionality])
[132]	2011 US	OpenNote Primary care patients	Within-subject design Naturalistic	Field Longitudinal (12 months)	Target population 13,564	Interaction with the system: Usage (System interaction logs analysis [% of users])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[133]	2012 US	PatientSite Patients with multiple sclerosis	Between-subject design Naturalistic	Field Longitudinal (24 months)	Target population 120 and 120	Interaction with the system: Usage (System interaction logs analysis [Frequency of use]); Receiving information from the system: Functionalities usage (System interaction logs analysis [frequency of use])
[134]	2014 US	Patient Gateway Primary care patients	Within-subject design Naturalistic	Field Cross-sectional	Target population 3,389	Interaction with the system: Overall usability (Questionnaire [Non validated questionnaire]); Receiving information from the system: Data accuracy (Questionnaire [% of patients reporting an error in the list]); Data completeness (Questionnaire [% of patients with missing problems]); Influence of the system on decisions-making: Patients decisions (Questionnaire [% of patients taking actions in relation to what they saw])
[135]	2010 US	Patient Gateway Primary care patients	Between-subject design Controlled	Field Longitudinal (18 months)	Target population 3,979	Receiving information from the system: Data sharedness (Data logs [% of patients having their provider opening the eJournal]); Functionalities usage (System interaction logs analysis [% of patients who opened an eJournal and number of submitted eJournals])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[11]	2011 UK	PAERS Primary care patients	Within-subject design Naturalistic	Field Cross-sectional	Target population 213	Interaction with the system: Usage (Questionnaire [% of patients that used it, frequency of use]); Receiving information from the system: Data accuracy (Questionnaire [Patients reporting a mistake]); Functionalities usage (Questionnaire [functionalities used, qualitative]); Influence of the system on decisions-making: Patients decisions (Questionnaire [% of patients reporting a change in life style])
[136]	2013 Taiwan	Online Diabetes SelfManagement System Patients with diabetes	Between-subject design Naturalistic	Field Longitudinal (18 months)	Target population 162 (59 cases vs 103 controls)	Interaction with the system: Usage (System interaction logs analysis [% of patients that used it, frequency of use]); Receiving information from the system: Functionalities usage (Data logs analysis [% of patients using each functonality]); Influence of the system on decisions-making: Patients decisions (Questionnaire [impact on American Association of Diabetes Educators 7 Self-Care Behaviors])
[137]	2014 US	MyHealtheVet Veterans	Within-subject design Naturalistic	Field Longitudinal (6 months)	Target population 51	Receiving information from the system: Functionalities usage (Data logs analysis [% of replied messages]); Influence of the system on decisions-making: Clinicians decisions (Data log analysis [number of additional ADEs based on patients information])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[138]	2012 US	Patient Gateway Primary care patients	Between-subject design Controlled	Field Longitudinal (18 months)	Target population 267 vs 274	Receiving information from the system: Data accuracy (Data logs analysis [additional discrepancies identified]); Data sharedness (Data logs [% of patients having their provider opening the eJournal]); Functionalities usage (System interaction logs analysis [% of patients who opened an eJournal and % of patients who submitted it]); Influence of the system on decisions-making: Clinicians decisions (Data log analysis and Interviews [additional preventable/ameliorable ADEs based on patients info])
[139]	2016 US	OpenNote Primary care patients	Between-subject design Naturalistic	Field Longitudinal (24 months)	Target population 44,951	Interaction with the system: Usage (System interaction logs analysis [% of users, frequency of use]); Receiving information from the system: Functionalities usage (System interaction logs analysis [System interaction logs analysis])
[140]	2011 US	EpicCare Primary care patients	Within-subject design Naturalistic	Field Longitudinal (18 months)	Target population 7,088	Receiving information from the system: Functionalities usage (System interaction logs analysis [% used different functionalties])
[141]	2016 US	MyChart General population	Within-subject design Naturalistic	Field Longitudinal (2 months)	Target population 14,441	Receiving information from the system: Functionalities usage (System interaction logs analysis [% viewed test results])
[142]	2016 US	Kaiser Permanente Northern California General population	Within-subject design Naturalistic	Field Longitudinal (12 months)	Target population >100,000	Interaction with the system: Usage (System interaction logs analysis [% of registered patients that used it at least once]); Receiving information from the system: Functionalities usage (System interaction logs analysis [% of patients that used different functionalities])

Reference	Year Country	System name Target users	Study design Ecology	Study setting Temporal factor	Evaluators Sample size	Aspect (Method[Metric])
[143]	2016 Thailand	Chulalongkorn University Patients with chronic kidney disease	Within-subject design Controlled	Lab Cross-sectional	Target population 10	Interaction with the system: Usability issues (Think aloud [Usability issues found]); Effectiveness (task analysis [% of successful tasks]); Efficiency (task analysis [mean time spent on the tasks]); Overall usability (Questionnaire [Validated questionnaire, USE by Lund])
[144]	2011 US	OurFamilyHealth General population	Within-subject design Naturalistic	Field Longitudinal (90 days)	Target population 168	Interaction with the system: Usage (System interaction logs analysis [% of users used it only once, time spent on the patient portal])
[145]	2015 US	Lawson SMART Record Patients with mental health issues	Between-subject design Controlled	Field Longitudinal (18 months)	Target population 394	Interaction with the system: Overall usability (Questionnaire and focus groups [Non-validated questionnaire and thematic analysis])
[146]	2016 US	Patient Gateway Primary care patients	Within-subject design Naturalistic	Field Longitudinal (6-months)	Target population 4,109	Interaction with the system: Overall usability (Questionnaire [Non-validated questionnaire]); Receiving information from the system: Functionalities usage (Questionnaire [% of patients that used the functionalities under study])
[147]	2008 US	Patient Gateway Primary care patients	Within-subject design Naturalistic	Field Longitudinal (6-months)	Target population 189	Influence of the system on decisions-making: Clinicians decisions (Data log analysis [additional medication regimen changes after patient information])
[148]	2008 US	Patient Gateway Primary care patients	Between-subject design Controlled	Field Longitudinal (18 months)	Target population 126 vs 118	Interaction with the system: Usage (System interaction logs analysis [% of users])
[149]	2013 Australia	Healthy.me General population	Within-subject design Naturalistic	Field Longitudinal (5 months)	Target population 709	Interaction with the system: Usage (Questionnaire [% of users, frequency of use])

Chapter 6

PRESENTATION OF LABORATORY TEST RESULTS IN PATIENT PORTALS: EFFECT OF INTERFACE DESIGN ON RISK INTERPRETATION AND PATIENT INTERACTION

Paolo Fraccaro[1,2], Panagiotis Balatsoukas[2,3], Sabine N van der Veer[2,3], Lamiece Hassan[2,3], Richard Williams[1,2,3], Grahame Wood[4], Smeeta Sinha[4], Iain Buchan[1,2,3], Niels Peek[1,2,3]

[1] NIHR Greater Manchester Primary Care Patient Safety Translational Research Centre, Institute of Population Health, The University of Manchester, Manchester, UK.

[2] Health eResearch Centre, Farr Institute for Health Informatics Research, UK.

[3] Centre for Health Informatics, NIHR Manchester Biomedical Research Centre, The University of Manchester, Manchester, UK.

[4] Renal Clinic, Salford Royal NHS Trust, Salford, UK.

Contribution: PF, IB and NP developed the idea of the project. PF, PB, SS, and NP designed the study. PF and LH took care of all aspects regarding Patient Involvement. GW and SS created the clinical scenarios used in the experiments. PF and RW developed the prototypes used during the experiments. PF scheduled experiments with patients, collected and analysed all data. PF, PB and NP wrote the manuscript. SV and IB critically edited the manuscript. All authors approved the manuscript.

Reference: The manuscript will be submitted to the Journal of Medical Internet Research.

6.1 ABSTRACT

Background

Patient portals are considered valuable conduits for supporting patients' self-management. However, there are safety concerns over how patients might interpret and act on the information from these portals, especially laboratory test results. Contemporary human computer interaction and information visualisation research has produced methods that improve human perception and cognition in different information seeking and decision-making tasks. However, these methods have not been evaluated for presenting laboratory test results via patient portals.

Objectives

To investigate how different presentations of online laboratory test results might influence patients' interpretation of risk, perceived usefulness, information usage and processing and visual search behaviour.

Methods

We conducted a controlled study with 20 patients who underwent a kidney transplant and had quarterly blood tests. Participants visited our human computer interaction lab and interacted with different clinical scenarios, designed by nephrologists, to reflect high, medium and low health risks. These were shown using three different web-based presentations. Each presentation was tile-based, with a baseline presentation (based on PatientView, a system currently available to patients) and two more advanced presentations providing different visual cues, colours and interaction techniques to show normal and abnormal values. After viewing each clinical scenario, patients were asked how they would have acted in real life: 1) call their doctor immediately (high perceived risk); 2) ask for an appointment within four weeks (medium perceived risk); or 3) wait for their next scheduled appointment (low perceived risk). We tested each presentation in terms of accuracy of risk interpretation, perceived usefulness, level of understanding, information processing, and visual search behaviour.

Results

We found no statistically significant differences between the three presentations in terms of the accuracy of risk interpretation. Misinterpretation of risk information was high, with 65% of patients underestimating the severity of risk across all presentations at least once. Particularly, patients decided to wait for their next appointment in 50% of the medium and high risk cases. Patients found it particularly difficult to interpret medium risk. No statistically

significant effect of patient's health literacy and numeracy was found on risk interpretation. The two advanced presentations were perceived as more useful (P=0.023). Differences in information usage and level of information processing were associated with personal characteristics, such frequency of PatientView and internet usage, education and graph literacy. In terms of visual search behaviour, patients followed similar visual search behaviours across the three presentations. Specifically, we observed greater interest for areas showing the latest laboratory test results, and lower interest for the comparison of longitudinal information for two laboratory tests, which was rarely used.

Conclusions

Although limited by a small sample size, our study is the first to bridge interface design and patients' interpretation of risk, and it provides unique data on how patients interact with and make sense of laboratory results in patient portals. This study confirms patients' difficulties in interpreting laboratory results, with many patients underestimating risk across different presentations, even when abnormal values were highlighted or grouped together.

Keywords: user-computer interface; computers/utilization; decision making; personal health record; patient access to records; patient portals; decision support systems, clinical.

6.2 INTRODUCTION

Patient portals are seen as important instruments to motivate and involve patients in having an active role in their health [1–6]. These systems allow patients to book appointments online, view laboratory test results, or communicate with their physicians. Currently, many patient portals are for patients living with chronic conditions [7,8], who undertake complex longitudinal follow-ups, where self-management is a key component [9].

Patients mainly use patient portals to check their laboratory test results [10–14]. Yet, laboratory test results are among the most difficult features of a patient portal for patients to understand [15–18], with concerns on the effectiveness of current patient portals in supporting patients in this task [7,23]. This could lead to detrimental self-interpretation [14], especially when patients have access to laboratory test results outside clinical consultation [17]. Misinterpretation of laboratory test results can have adverse effects on patient safety, including increased patient anxiety, or inability of self-management [7,14,19]. With the increasing availability of patient portals [20,21], it is therefore important to understand how patients interact with and process laboratory test results to inform the evidence-based and empirically-sound presentation of this information online for accurate risk interpretation and improved user interaction.

It is known that several contextual factors, like numeracy and health literacy, can influence risk interpretation in the context of online laboratory test results [17], however little is known about the effect of presentation on patients risk interpretation and interaction with this type of information. Previous studies have reported that patients found it difficult to understand laboratory test results shown in tables [15–17] and graphs [15,18], even when patients were familiar with the clinical scenarios at hand (i.e. glucose level monitoring for diabetes patients) [17]. A possible explanation for this phenomenon is the way numerical data were presented to patients [22]. However, given the web-based nature of patient portals, user interface design elements like colour and luminance can play an important role in the perception of risk and decision making process of lay people [23]. Yet, there is no evidence of their effect in the case of web-based laboratory test results. Furthermore, new interface design and human computer interaction techniques have been proven to enhance human decision making and information seeking behaviours in other web-based contexts [24–26]. However, to date it is not clear which of these techniques are the most effective in the context of online laboratory test results.

We conducted a controlled study to investigate the effect of different web-based presentations of laboratory test results on perceived risk interpretation and user interaction. In this study, we used the term presentation to refer to effect of different user interface design principles and

techniques on patient's behaviour [27], without investigating the effect of numerical information presentation (e.g. as frequencies vs. percentages). In particular the objectives were:

- To measure the effect of presentation of web-based laboratory test results, across different clinical scenarios, on: the accuracy of risk interpretation; perceived usefulness and level of understanding; level of information usage and processing; and visual search behaviour;
- To investigate the effect of the interaction between individual patient characteristics, types of presentation and clinical scenarios on the aforementioned measured outcomes.

6.3 METHODS

6.3.1 Study population

We focused on patients with Chronic Kidney Disease (CKD), for whom there is an online platform, named PatientView [28] (formerly known as Renal Patient View), that provides access to laboratory results and is available in almost 90% of renal units within the UK. We included patients who underwent a renal transplantation at least 12 months before recruitment. These patients undergo a longitudinal follow-up after the kidney transplant, which consists of quarterly visits to monitor their laboratory test values. This allowed us to obtain a more homogeneous group of participants in terms of their experience with and knowledge of the disease. We excluded patients with any visual impairment, to avoid ineffective eye tracking data collection, and patients that did not use the internet in their everyday life, as not being potential users of online patient portals.

We recruited 20 patients from the Renal Transplant Clinic at Salford Royal NHS Foundation Trust (SRFT), which has one of the largest communities of PatientView users in the UK. The study received ethical approval from NHS and local R&D ethical committees (IRAS ID: 183845). Research nurses from the NIHR Clinical Research Network Portfolio (Study CPMS ID: 20645) were responsible for approaching eligible patients and collecting signed informed consents of patients willing to participate.

6.3.2 Controlled study design

The study followed a "3x3" repeated measures, within-subjects design according to which

each participant used three different presentations of web-based laboratory test results to complete the same simulated task in three different clinical scenarios. These were designed by nephrologists at SRFT to reflect:

- *High risk clinical scenarios:* characterised by life threatening situations, where creatinine, estimated Glomerular Filtration Rate (eGFR) (i.e. the main indicators of kidney disease [29]) and potassium (i.e. associated to higher mortality in kidney patients in presence of hyperkalemia and hypokalemia [30]) had high deviance from the standard range;

- *Medium risk clinical scenarios:* identified by abnormal creatinine and eGFR, but normal potassium and stable conditions, which would require further tests;

- *Low risk clinical scenarios:* characterised by normal creatinine, eGFR and potassium, and not requiring any action until the next scheduled appointment.

In addition to creatinine, eGFR and potassium, each scenario included 25 more laboratory test results, with different deviance from the standard range in relation to the reflected risk (i.e. more concomitant abnormal values for high risk scenarios). Each laboratory test had longitudinal information for up-to two years follow-up.

The task, previously used by Zikmund-Fisher et al [17] to evaluate patients ability of interpreting laboratory test results in a static tabular format, consisted in exploring the laboratory test results as if the participant was at home and received new results from the clinic. After exploring the results, participants were asked to respond to a set of questions about:

- *Perceived risk*. Patients were asked what they would do in real life if the results they had just explored were their own. They could choose between: 1) calling their doctor immediately (high perceived risk); 2) trying to arrange an appointment within the next four weeks (medium perceived risk); 3) waiting for the next appointment in three months (low perceived risk);

- *Perceived usefulness and level of understanding*. We used two questions adapted from Zikmund-Fisher et al [17], with responses arranged on a 5-point Likert scale. The first question asked the participants how well they understood what the laboratory results showed, while the second question asked them to rate how useful the presentation of laboratory test results was;

- *Information usage and processing.* We evaluated different aspects of how patients processed information:

    - *Focused Immersion*. It is defined as the "state of deep involvement with a software" [31]. Focused immersion was measured via a self-reported 7-point Likert scale questionnaire, previously used by Harle et al [31].

    - *Heuristic and systematic processing*. Heuristic processing is the quick exploration of a message with little cognitive effort in judging the message validity (i.e. skimming a web page), while systematic processing involves a more detailed consideration of a message with a consequence of a greater cognitive effort [31,33]. An increase in systematic processing and decrease in heuristic processing mean a greater engagement of the user, which in this case actively tries to make sense of the information provided. We measured these features using a 7-point Likert scale from Harle et al [31].

While performing a task, there was no time limit and participants could decide to terminate exploration of the laboratory test results whenever they felt ready to reply to the follow up questions.

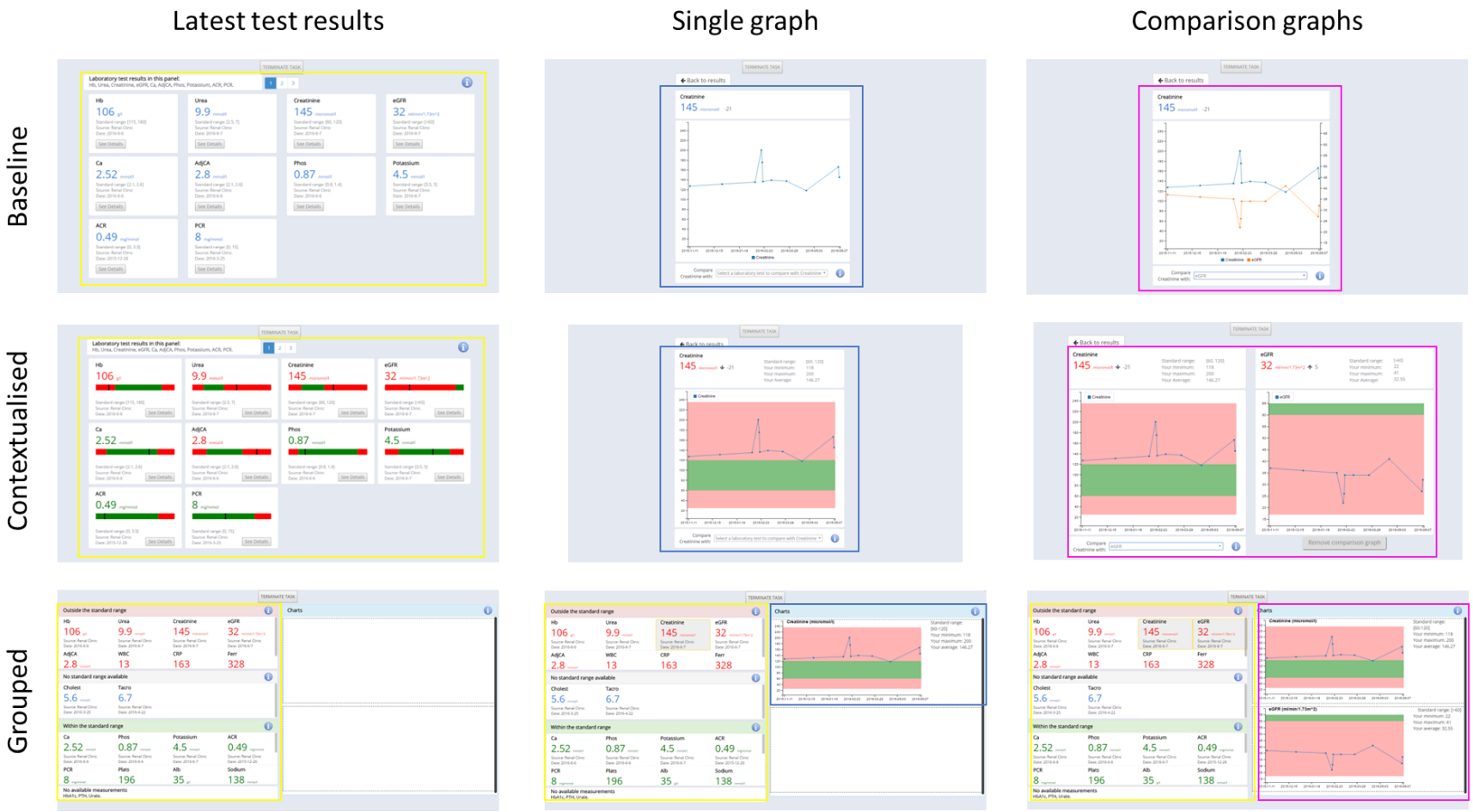
At the beginning of the experiment each participant filled four electronic background questionnaires that measured individual characteristics in terms of: demographics (age, gender, education, years since transplant, frequency of internet usage, and frequency of PatientView use); Subjective Numeracy Scale (SNC) [34] calculated on a 1.0-6.0 range; self-reported health literacy on a 0.0-4.0 range based on Chew *et al.* [35], with values closer to 0 indicating better self-reported health literacy; graph literacy, calculated as % of correct answers on the questionnaire from Galesic *et al.* [36].

The controlled study was conducted at the Interaction Analysis and Modelling (IAM) laboratory of the University of Manchester, and each patient participated individually. All participants performed the tasks using a desktop computer with a 17-inch screen with an embedded eye tracker (Tobii T60), which permits a 60-Hz sampling rate, 0.5 degrees gaze point accuracy, and free head motion. Patients' eye movements were recorded throughout the experiment, with recalibration at the beginning of each task.

6.3.3 Presentations

We implemented three different presentations of laboratory test results (see Figure 6.1).

Figure 6.1: Latest test results overview and longitudinal detailed information for one or two tests at the same time across the three presentations. Coloured rectangles represents the different areas of interest (AoIs) we defined across the three presentations (yellow: AoI 1; blue: AoI 2; purple: AoI 3).



The *Baseline presentation* was directly based on, and very similar to the current PatientView [28] *overview-preview* interface, which uses tiles to show the available latest laboratory test results (overview interface). Tiles are scattered across multiple pages of the overview interface in a similar way that a search engine displays search results across several pages. Each tile reports information about the laboratory test result value and units, the date of the test, data source, and standard range. By clicking on a tile the user can access detailed longitudinal information (i.e. univariate time series) about the selected laboratory test in a new screen (preview interface). This information is displayed in a 2D point-type line graph, which allows to compare the selected laboratory test result with another test within the same graph (i.e. bivariate time series).

The Baseline presentation did not make use of mainstream information seeking and data visualisation techniques that are known to improve human cognition and perception across various types of information seeking and decision making tasks. For example, this presentation did not use colour (to denote within-range and out-of-range values), following contrast and analogy principles [37]. Furthermore, no graphical elements to contextualise the presentation of laboratory test results for each patient [16], or categorical and clustered views to filter and organise the display of information on the screen [38] were used.

We developed the comparison presentations by using these techniques to display normal and abnormal values with the aim of improving patients' interpretation of risk and user interaction. Particularly, the second presentation (*Contextualised presentation*) was built on the Baseline presentation and used coloured horizontal bars that contextualised the latest value in relation to the standard range in each tile of the overview interface [16,39]. This approach was previously shown to outperform tables in terms of ease of use and perceived usefulness when reporting values about laboratory test results [16]. The third presentation (*Grouped presentation*) made use of *categorised overviews* [40–42] to dynamically group the tiles in "Outside the standard range", "No standard range available", and "Inside the standard range" categories. Categorised overviews support exploratory search of relevant information through the effective organisation of information in meaningful categories, which supports spatial grouping of information and reduces the amount of data stored in short term memory during information processing. This technique has previously been proven to be more effective than tables or list-based presentations when searching and evaluating information for various types of decision-making tasks [40–42]. However, it was never tested before in the context of online laboratory test results. As opposed to the other two presentations (Contextualised and Baseline), the use of categorised overviews allowed to have all information about laboratory test results displayed in a single page, which is an important interface design feature when

presenting health data [43].

Graphs showing the detailed longitudinal information in both the Contextualised and Grouped presentations reported personalised statistics for a selected test [44–46] and used colours to show the area inside the standard range in green and the remaining area, outside the normal range, in red [47]. We applied the same approach to colour the latest laboratory test results in the tiles in both presentations, as colour positively influences human cognition for risk interpretation tasks [48]. This happens because colour is a powerful pre-attentive property, when it comes to the design of visual presentations, that facilitates fast identification of information without the need for sequential search or conscious attention [49,50]. In addition to red and green colour, we used blue to present values in laboratory test results for which no range information was available. Finally, we did not control for colour blindness, since only patients with normal vision were selected for participation in the study.

Throughout the project, we involved three patients from the local CKD patient community (http://gmkin.org.uk/), of which two were experienced in using PatientView. They gave us feedback on the different presentation techniques and research protocol, with one of them participating in a pilot study.

6.3.4 Data analysis

6.3.4.1 Presentation's effect on risk interpretation

To assess the effect of the presentations (Baseline, Contextualised and Grouped) on the accuracy of risk interpretation, we created a 3x3 confusion matrix for each presentation that reported the judgments made by patients versus our gold standard (i.e. nephrologists' clinical judgement). From the confusion matrices, we calculated precision (i.e. proportion of correct interpretations of all interpretations as risk X), recall (i.e. proportion of correct interpretations on clinical scenarios with risk X) and accuracy (i.e. proportion of correct interpretations of all interpretations) for each presentation, and compared these using chi-squared tests. We assessed the influence of patient characteristics (age, education, years since transplant, frequency of internet usage, and frequency of PatientView use, SNC, health and graph literacy) on accuracy of risk interpretation using mixed-effects logistic regression analysis, adjusting for the presentations (i.e. a categorical variable with values Baseline, Contextualised and Grouped presentation) as fixed effect, and adjusting for correlations between repeated observations by including random intercepts for patient and clinical scenario.

We repeated the analysis with a secondary definition of the outcome, which aimed at investigating a situation in which, from a safety perspective, a misjudgement could have

serious consequences. Particularly, we evaluated the presentation's performance in allowing patients to identify when an action was needed (i.e. at least medium risk versus low risk). To evaluate whether performance was driven by single patients (i.e. there were some patients that misinterpreted most of the information), we counted the mistakes that each patient made. We distinguished between: *safety mistake*, defined as the patient under-estimating risk and not acting when needed (i.e. confusing medium or high risk with low risk); and *requiring unnecessary care*, defined as asking for help when not needed (i.e. over-estimating low risk clinical scenarios). We assessed the influence of patient specific characteristics on making one or the other type of mistake with t-tests for continuous variables (i.e. age, years since kidney transplant, SNC score, graph literacy score, and health literacy score) and Wilcoxon-sum-rank test for categorical ones (i.e. level of education, internet usage and level of experience with PatientView).

6.3.4.2 Presentation's effect on self-reported measures

For perceived usefulness, level of understanding, and information usage and processing measures, we performed a three-way within subject analysis of variance (ANOVA) to compare the presentations and a Wilcoxon sum rank test for head-to-head comparisons. We assessed the influence of patient characteristics on the self-reported measures with mixed-effects linear regression analyses, adjusting for the presentations as fixed effect, and for correlations between repeated observations by including random intercepts for patient and clinical scenario.

6.3.4.3 Presentation's effect on visual search behaviour

We analysed participants' eye movements to investigate the effect, if any, of presentation on the visual search behaviour of participants. First, we measured participants' *interest* as the number of fixations (each fixation was defined as a stable gaze lasting at least 180ms [51]) and total dwell time spent on specific areas of interest (AoIs). There were three AoIs in each presentation (see Figure 6.1): 1) tiles showing the latest values for all laboratory tests; 2) the graph showing detailed longitudinal information for a single laboratory test; 3) the graphs comparing detailed longitudinal information for two laboratory tests. We used a three-way within subject ANOVA to evaluate the effect of presentation on participant's interest. Also, we used paired t-tests for head-to-head comparisons in order to measure differences in the level of interest between AoIs across presentations and clinical scenarios, applying Bonferroni correction to account for multiple comparisons. Furthermore, mixed-effects linear regression analysis was performed to assess the influence of patient characteristics on visual interest. Again, we adjusted for the presentations as fixed effect, and included random intercepts for

patient and clinical scenario to adjust for correlations between repeated observations. Finally, we calculated heatmaps, which show a qualitative overall representation of visual behaviours for all participants, with red indicating more, and green indicating less, eye fixations.

Data analysis was performed in "R". All visual search behaviour data were extracted using the Tobii Studio (version 3.4.0).

6.4 RESULTS

6.4.1 Patient characteristics

Table 6.1 reports the characteristics of the 20 patients who participated in the study. The majority were male, had at least a college education and used the internet for more than five hours per week. Frequency of PatientView use was balanced within our study population, with eleven participants who were regular users (i.e. quarterly use) and the remaining nine using it less than twice per year. Mean age was 51.8 (Standard deviation [SD] = 10.3) and mean years since kidney transplant was 10.7 (SD = 8.7). Mean scores for SNC, health literacy and graph literacy were 4 (SD = 0.8), 0.5 (SD = 0.6) and 73.5 (SD = 11.3), respectively.

Table 6.1: Patient's characteristics. Abbreviations: SD, Standard deviation; SNC: Subjective Numeracy Scale; GCSE: General Certificate of Secondary Education.

Parameters		**Values**
Number of patients		20
Gender	Female (%)	4 (20)
	Male (%)	16 (80)
Age (years)		51.8 (10.3)
Years since kidney transplant (mean, SD)		10.7 (8.7)
SNC score (mean, SD)		4 (0.8)
Health literacy (mean, SD)		0.5 (0.6)
Graph literacy score (mean, SD)		73.5 (11.3)
Education	Lower than GCSE (%)	1 (5)
	GCSE (%)	7 (35)
	A-level/College (%)	5 (25)
	Higher education/University degree (%)	7 (35)
Internet use	Less than one hour per week (%)	1 (5)
	One to five hours per week (%)	5 (25)
	Five to 10 hours per week (%)	5 (25)
	More than 10 hours per week (%)	9 (45)
PatientView use	Never used (%)	3 (15)
	Less or equal than once per year (%)	5 (25)
	Twice per year (%)	1 (5)
	Quarterly (%)	11 (55)

6.4.2 Data analysis

6.4.2.1 Effect of presentation on risk interpretation

Table 6.2 shows the confusion matrixes and performance measures (precision and recall) for all presentations across the different clinical scenarios (i.e. low, medium and high risk). The majority of patients correctly interpreted low and high risk clinical scenarios across all the presentations, while medium risk clinical scenarios were often confused with low risk ones. For each presentation, at least two patients misinterpreted high risk clinical scenarios for low risk ones. The precision scores of the interpreted risks were similar for the Baseline and Contextualised presentations, and ranged from 0.33 to 0.69 and from 0.38 to 0.70, respectively. For the Grouped presentation performance was lower, and ranged between 0.28 and 0.58. Likewise, recall scores for the Baseline and Contextualised presentations were similar, and ranged from 0.30 to 0.70, with a drop in performance from the Grouped presentation. Overall accuracy was 0.55 for the Contextualised presentation, 0.52 for the Baseline presentation and 0.45 for the Grouped presentation. These differences were not statistically significant according to the chi squared test.

Table 6.2: Confusion matrix of the interpreted risk by patients versus the gold standard (nephrologists judgement), and performance (i.e. precision and recall) for each presentation.

Presentation	**Interpreted risk**	**Gold standard (Nephrologists)**			**Performance**	
		Low risk	**Medium risk**	**High risk**	**Precision**	**Recall**
Baseline	Low	14	10	2	0.54	0.70
	Medium	5	6	7	0.33	0.30
	High	1	4	11	0.69	0.55
Contextualised	Low	13	9	2	0.54	0.65
	Medium	6	6	4	0.38	0.30
	High	1	5	14	0.70	0.70
Grouped	Low	11	9	3	0.48	0.55
	Medium	7	5	6	0.28	0.25
	High	2	6	11	0.58	0.55

The mixed effects logistic regression showed two statistically significant coefficients in our data (see Table 6.3). First, the frequency of PatientView use was associated to better risk interpretation accuracy. Second, as already shown in the confusion matrixes, medium risk clinical scenarios were associated to a lower level of risk interpretation accuracy.

Table 6.3: Adjusted odds ratios of interpreting correctly the risk. Abbreviations: CI: confidence interval; SNC: Subjective Numeracy Scale.

Covariate	Adjusted odds ratios [95% CI]
(intercept)	0.78 [0.01-47.67]
Age	1.01 [0.96-1.07]
Years since transplant	1.04 [0.98-1.1]
Contextualised presentation vs Baseline	1.15 [0.52-2.55]
Grouped presentation vs Baseline	0.73 [0.33-1.6]
Medium risk clinical scenario vs Low risk	**0.2 [0.08-0.54]**[b)]
High risk clinical scenario vs Low risk	0.86 [0.33-2.23]
Education[a)]	0.98 [0.68-1.43]
Frequency of internet use[a)]	1.08 [0.7-1.66]
Frequency of PatientView use[a)]	**1.38 [1.04-1.85]**[b)]
SNC	0.77 [0.47-1.26]
Health literacy	1.38 [0.73-2.61]
Graph literacy	0.99 [0.95-1.02]

a) Ordered categorical factors considered as numeric covariates in the analysis.
b) Statistically significant at 0.05 level.

In addition, we calculated confusion matrixes to compare situations in which some sort of action was needed (i.e. at least medium risk scenarios) versus those that did not require any action (i.e. low risk clinical scenarios) (see Table 6.4). For scenarios where an action was needed (i.e. call the doctor to arrange an appointment immediately or in the next four weeks), the three presentations performed similarly with a precision around 0.8 and a recall around 0.7. However, the precision for interpretations where patients did not choose to take any action was around 0.5 for all presentations. Nevertheless, we did not find any statistically significant differences between the different presentations in the chi-squared test.

Table 6.4: Confusion matrix of the interpreted risk by patients versus the gold standard (nephrologists judgement) and performance (i.e. precision and recall) for the analysis by comparing interpretation of clinical scenarios where an action was needed (i.e. at least medium risk) to those where no action was needed (i.e. low risk).

		Gold standard (Nephrologists)		Performance	
Presentation	**Interpreted risk**	**Action needed**	**No action needed**	**Precision**	**Recall**
Baseline	Action needed	28	6	0.82	0.70
	No action needed	12	14	0.54	0.70
Contextualised	Action needed	29	7	0.81	0.72
	No action needed	11	13	0.54	0.65
Grouped	Action needed	28	9	0.76	0.70
	No action needed	12	11	0.48	0.55

Finally, Table 6.5 shows the number of times patients made safety mistakes (i.e.

underestimated the severity of risk) or required unnecessary care (i.e. overestimated the severity of risk) across all presentations and clinical scenarios. Thirteen patients (65%) made at least one safety mistake, of whom seven (35%) misinterpreted at least half of the scenarios requiring an action. Fourteen patients (70%) required unnecessary care at least once. In both cases, individual patient characteristics (see Supplementary Table 6.1 and Supplementary Table 6.2) were not associated with the level of risk interpretation accuracy according to both the t-tests and Wilcoxon-sum-rank tests.

Table 6.5: Contingency table for the number of patients who made safety mistakes and requiring unnecessary care. Safety mistakes were defined as confusing medium or high risk tasks with low risk ones. Requiring unnecessary care was defined as confusing low risk with medium or high.

Number of mistakes	**Number of patients (n=20)**	
	Safety mistake (%)	**Requiring unnecessary care (%)**
0	7 (35)	6 (30)
1	3 (15)	8 (40)
2	3 (15)	4 (20)
3	4 (20)	2 (10)
4	2 (10)	/
5	0 (0)	/
6	1 (5)	/

6.4.2.2 Effect of presentation on self-reported measures

As shown in Table 6.6, results for the self-reported measures were similar across presentations and clinical scenarios. We found a significant result in the ANOVA tests only for perceived usefulness, with a statistically significant difference between the presentations at a 0.05 level (p-value = 0.023). However, head-to-head comparisons with Wilcoxon-sum-rank test did not show any statistically significant difference between the Contextualised and Baseline presentations (p-value = 0.057) or Grouped and Baseline presentations (p-value = 0.380).

Table 6.6: Performance for all the self-reported measures. Perceived usefulness and level of understanding are measured on a 5-point Likert scale, while the others on 7-point Likert scale. Abbreviations: SD. Standard deviation.

Presentation	**Risk**	**Perceived usefulness (SD)**	**Understanding level (SD)**	**Focused immersion (SD)**	**Systematic processing (SD)**	**Heuristic processing (SD)**
Baseline	Low	4.0 (1.1)	4.1 (1.0)	5.5 (1.6)	5.5 (1.3)	4.4 (2.0)
	Medium	3.8 (1.3)	3.7 (1.0)	5.6 (2.0)	5.5 (1.8)	3.8 (2.0)
	High	4.0 (1.0)	4.1 (1.0)	5.5 (1.7)	5.6 (1.6)	3.4 (1.7)
Contextualised	Low	4.4 (0.9)	4.1 (0.9)	5.7 (1.3)	6.0 (1.4)	3.9 (2.1)
	Medium	4.0 (1.2)	4.0 (1.1)	5.3 (1.8)	5.6 (1.3)	3.6 (1.8)
	High	4.2 (1.0)	4.0 (1.0)	5.7 (1.5)	5.7 (1.2)	3.6 (2.1)
Grouped	Low	4.3 (0.9)	4.0 (0.8)	5.7 (1.5)	5.9 (0.9)	3.8 (1.7)
	Medium	4.0 (1.00)	4.1 (0.7)	5.5 (1.6)	5.5 (1.4)	3.5 (1.8)
	High	4.1 (1.1)	4.1 (1.0)	5.2 (1.7)	5.4 (1.5)	3.8 (1.9)

Table 6.7 reports the coefficient estimates from the mixed-effects linear regressions on the self-reported measures. Frequency of internet usage was negatively associated with all the measures with the exception of the heuristic processing. The frequency of PatientView use was positively associated with level of understanding, focused immersion and systematic processing. Finally, medium risk clinical scenarios were associated with lower perceived usefulness, whereas level of education and graph literacy were linked to greater focused immersion.

Table 6.7: Coefficient estimates in the mixed-effects linear regression for the self-reported measures. Abbreviations: CI. Confidence interval.

Covariate	**Perceived usefulness coefficient estimate [95% CI]**	**Understanding level coefficient estimate [95% CI]**	**Focused immersion coefficient estimate [95% CI]**	**Systematic processing coefficient estimate [95% CI]**	**Heuristic processing coefficient estimate [95% CI]**
(intercept)	3.47 [0.47,6.48]	2.44 [0.12,4.76]	3.56 [-1.43,8.54]	3.51 [-0.26,7.29]	7.50 [-1.30,16.29]
Age	-0.02 [-0.06,0.02]	-0.01 [-0.04,0.02]	-0.04 [-0.11,0.03]	-0.04 [-0.09,0.01]	-0.11 [-0.22,0.01]
Years since transplant	0.02 [-0.03,0.06]	0 [-0.03,0.03]	0.03 [-0.04,0.11]	0.03 [-0.02,0.09]	0.08 [-0.04,0.21]
Contextualised presentation vs Baseline	0.27 [-0.02,0.56]	0.05 [-0.21,0.31]	0.04 [-0.27,0.36]	0.25 [-0.08,0.57]	-0.19 [-0.60,0.21]
Grouped presentation vs Baseline	0.2 [-0.09,0.49]	0.07 [-0.19,0.32]	-0.05 [-0.36,0.27]	0.08 [-0.24,0.41]	-0.16 [-0.57,0.24]
Medium risk clinical scenario vs Low risk	**-0.32 [-0.61,-0.03]**[b)]	-0.15 [-0.5,0.2]	-0.17 [-0.53,0.2]	-0.25 [-0.6,0.1]	-0.38 [-1.02,0.25]
High risk clinical scenario vs Low risk	-0.12 [-0.41,0.17]	0.02 [-0.34,0.37]	-0.15 [-0.51,0.21]	-0.22 [-0.57,0.13]	-0.45 [-1.08,0.18]
Education[a)]	0.06 [-0.22,0.34]	0.08 [-0.13,0.3]	**0.49 [0.02,0.95]** [b)]	0.23 [-0.12,0.58]	0.55 [-0.27,1.36]
Frequency of internet use[a)]	**-0.54 [-0.86,-0.21]**[b)]	**-0.38 [-0.63,-0.13]**[b)]	**-1.03 [-1.57,-0.49]**[b)]	**-0.66 [-1.07,-0.25]**[b)]	-0.71 [-1.67,0.24]
Frequency of PatientView use[a)]	0.09 [-0.12,0.3]	**0.21 [0.04,0.37]**[b)]	**0.39 [0.04,0.74]**[b)]	**0.42 [0.15,0.69]**[b)]	0.20 [-0.42,0.82]
SNC	0.26 [-0.11,0.64]	0.2 [-0.09,0.49]	0.07 [-0.56,0.69]	0.28 [-0.2,0.75]	0.90 [-0.20,2.00]
Health literacy	-0.02 [-0.5,0.46]	-0.02 [-0.38,0.35]	0.01 [-0.79,0.8]	0.11 [-0.49,0.71]	0.54 [-0.86,1.94]
Graph literacy	0.02 [-0.01,0.05]	0.02 [-0.01,0.04]	**0.05 [0.00,0.09]**[b)]	0.03 [0,0.06]	-0.04 [-0.12,0.04]

a) Ordered categorical factors considered as numeric covariates in the analysis.
b) Statistically significant at 0.05 level.

6.4.2.3 The effect of presentation on visual search behaviour

Out of the 20 participants we analysed eye movements for 15, for whom we were able to collect good quality eye movement data for all nine tasks. Encountering difficult eye-tracking circumstances in some participants is not uncommon in eye tracking studies [52]. This can be related to participant's characteristics like glasses, particular shapes of eye-lids or very small pupils [52].

Figure 6.2 shows the mean number of fixations across the different presentations for each AoI and clinical scenario. Values were similar across the different presentations and risks, with the AoI with the latest test results receiving the highest number of fixations. Conversely, the AoI with the comparison graphs received the lowest number of fixations. The ANOVA tests showed that the difference in the number of fixations between the three AoI was statistically significant at the 0.05 level (p-value = 0.0003). However, no statistically significant differences were found in head-to-head comparisons with Bonferroni-corrected paired t-tests across the different presentations, showing similar visual search behaviour across the three presentations. In terms of the effect of the individual patient characteristics on visual search behaviour, the results of the mixed-effects linear regression analysis (see Supplementary Table 6.3) showed an effect only for the frequency of PatientView use variable. Specifically, the frequency of PatientView use was associated to lower eye fixation counts at 0.05 level. Similar figures were found for dwell time across the three presentations, clinical scenarios and AoI (see Figure 6.3). Again, the frequency of PatientView use variable had a statistically significant association at a 0.05 level, with lower dwell time values for patients who used PatientView more often.

Figure 6.2: Participants' eye fixation count on the different areas of interest (i.e. latest laboratory tests, single graph, comparison graph) across clinical scenarios (i.e. reflecting low, medium, and high risk) and presentations (i.e. Baseline, Contextualised, Grouped).

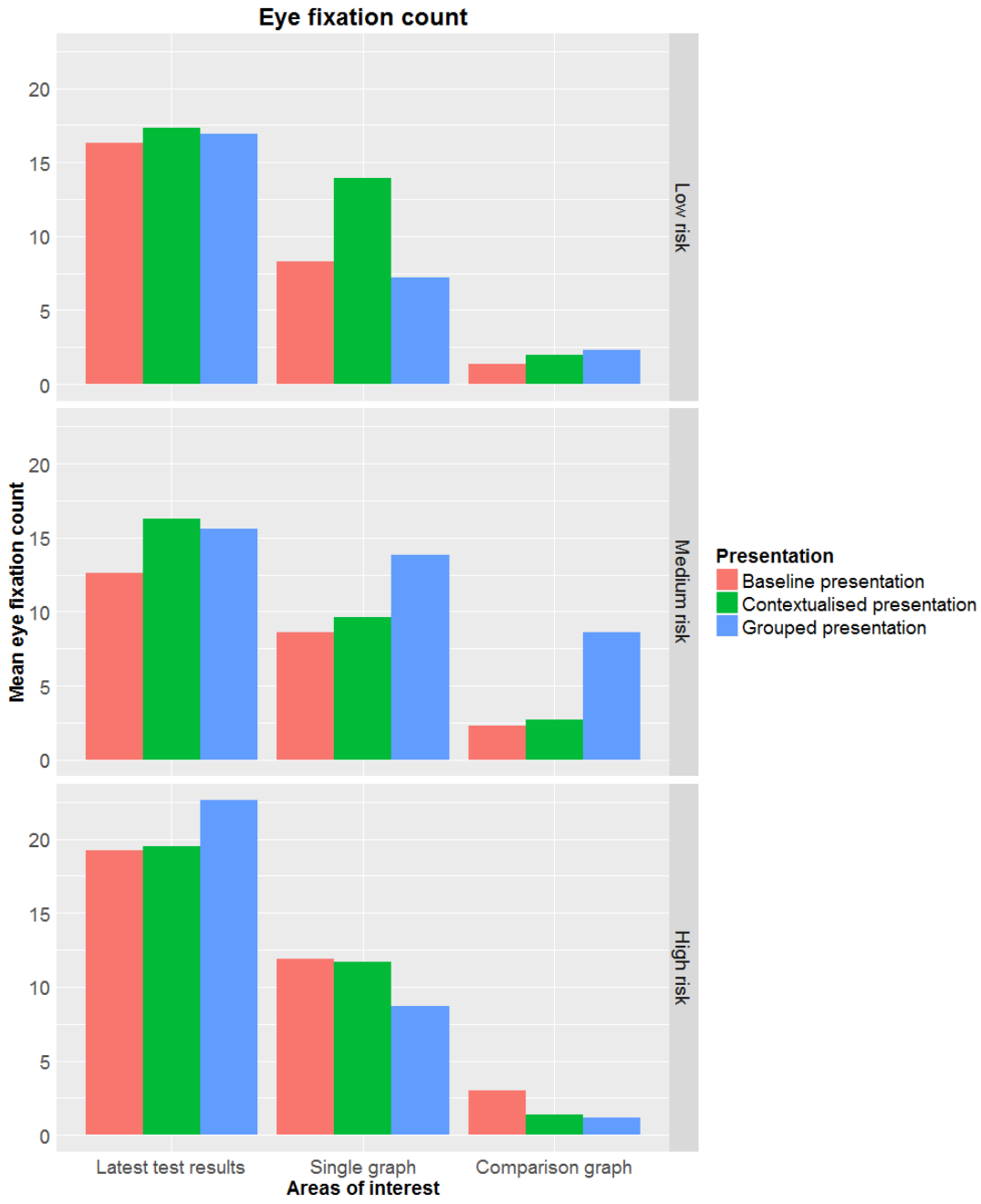
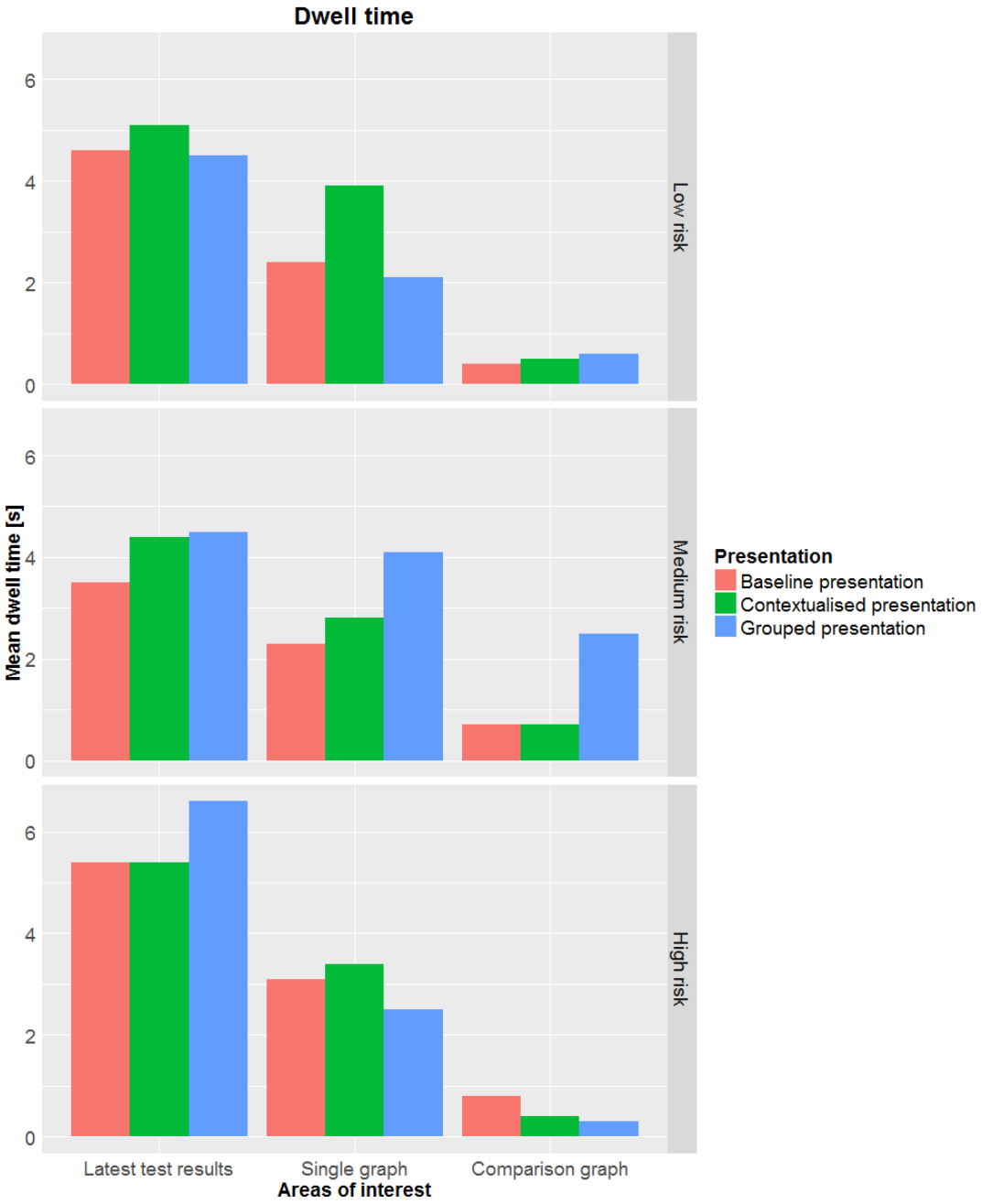


Figure 6.3: Participants' dwell time on the different areas of interest (i.e. latest laboratory tests, single graph, comparison graph) across clinical scenarios (i.e. reflecting low, medium, and high risk) and presentations (i.e. Baseline, Contextualised, Grouped).



Despite the lack of statistically significant differences between the three presentations in terms participants' interest (i.e. fixation count and dwell time), a few patterns of user interaction arose from the heatmaps in Figure 6.4. Specifically, patients focused less on the AoI displaying the comparison graphs, with an increased interest only for the Grouped presentation. This was especially observed for medium risk clinical scenarios (see Figure 6.2 and Figure 6.3). Another interesting result we can extract from the heatmaps is that, when viewing the AoI with the single graph of the Contextualised presentation, patients mainly looked at the area of the screen reporting the personalised statistics (increased level of attention is indicated by the

red spots in that area of the heatmap). Despite the absence of significant differences with the other presentations, this was noted for low and high risk clinical scenario, for which the observed higher levels of interest Contextualised presentation (see Figure 6.2 and Figure 6.3).

Figure 6.4: Heatmaps for all AOIs across the three different presentations showing a qualitative overall representation of visual behaviours for all participants. The figure shows higher presence of fixations in the Single graph screen for the Contextualised presentation and in the Comparison graphs screen for the Grouped presentation.

6.5 DISCUSSION

We found no effect of presentation on the accuracy of risk interpretation, with misinterpretation of risk being consistently high across the three presentations. Particularly, 65% of patients made safety mistakes at least once, while the precision of risk interpretation when patients decided not to act on the information (i.e. high and medium risk scenarios) was low (precision score = 0.5). This happened despite the use of visual properties (i.e. colour to denote normal and abnormal values), graphic elements (i.e. horizontal contextual bars), categorised overviews (i.e. automatically grouping of the latest laboratory test result values into normal and abnormal categories) and personalised descriptive statistics. These findings raise patient safety concerns and highlight that the presentation of laboratory test results, in terms of interface design alone, is not enough to improve the accuracy of risk interpretation [53]. To help patients translate test results into actionable information additional features about their specific clinical context could be necessary [53]. In this regard, going beyond population-based reference ranges, which are often not helpful for chronic patients, should be a priority [53,54].

To the best of our knowledge this is the first study to assess the effect of different web-based presentations of laboratory test results on risk interpretation. Previous studies have examined how patients interpreted risk in the context of laboratory test results [15–17,55]. While the authors of these studies also found that patients had difficulties in interpreting risk correctly, they did not make comparisons between different presentation formats. Instead, they evaluated patients' behaviour using a single presentation format (i.e. tabular or graphical). The only exception was the study by Brewer et al [16] who compared a tabular format to personalised horizontal coloured bars. In this study the authors did not find significant improvement in recall and understanding of laboratory test results.

Participants in our study found more difficult to interpret information in the case of medium risk clinical scenarios, rather than low and high risk ones. This finding was confirmed by the mixed effects regression analyses that showed a statistically significant association between medium risk clinical scenarios and incorrect interpretation, in addition to lower perceived usefulness, across all three interfaces. This is a unique finding of our study, as previous studies that looked at patients' interpretation of laboratory test results limited the task to the identification of abnormal values [15,16,56]. In our study, instead of asking patients to make dichotomous judgments, by identifying whether a specific laboratory test result value was inside or outside the normal range, we employed a more naturalistic task, as prescribed in [17], in order to assess what they would have done in real life if the results presented were theirs.

This approach allowed us to examine patients' interpretation using clinical scenarios of different levels of risk.

In addition to the medium risk clinical scenarios, another factor that influenced significantly risk interpretation accuracy was the frequency of PatientView use. In particular, frequency of PatientView use was positively associated with correct risk interpretation and higher self-reported measures (i.e level of understanding, increased focused immersion and systematic processing). Despite the lack of similar studies in this context, there is rich evidence in human computer interaction research that the level of experience or training with a specific application can have a positive effect on user performance and satisfaction [43,57,58].

Conversely to previous research [17,59], we found no association between individual characteristics (i.e. SNC, health or graph literacy) and risk interpretation accuracy. Particularly, Zikmund-Fisher et al. [17] found that numeracy, as measured by SNC, and health literacy were the main patient characteristics associated to a correct interpretation of laboratory test results. In our study, we tested the same variables but did not find the same association. A possible explanation of these results are the differences between our study population and the one studied by Zikmund-Fisher et al [17]. On the one side, Zikmund-Fisher et al [17] employed a much larger sample and included a broader group of people, with and without the condition under examination (i.e. diabetes). On the other side, differences in personal characteristics might have been less important in our controlled study population, as we selected a much smaller and more homogeneous group of patients who had an advanced level of knowledge about their condition. Furthermore, the majority of our participants were also familiar with the task at hand, having previous experience in checking laboratory test results in online portals.

Finally, nevertheless some patterns emerged from the analysis of the visual search behaviours (i.e. greater interest in the Grouped presentation when viewing the longitudinal values in the comparison graphs, and the Contextualised presentation when inspecting the AoI displaying the longitudinal values in a single graph), there was no statistically significant effect of presentation on eye movement data. Furthermore, there was no association between presentations and any of the self-reported measures, especially those measures related to conscious processing of information, like focused immersion and systematic processing. These results are not in accordance with existing evidence in information seeking and human computer interaction research. For example, earlier studies showed the presence of more systematic visual search behaviour in the context of interfaces that make use of categorized overviews [40]. In our case, categorised overviews in the Grouped interface did not evoke a significantly more in-depth immersion in the task as it would be expected. Moreover, the use

of contextual personalised horizontal bars in the Contextualised interface did not produce significant differences in the way participants inspected the AoI that displayed all latest laboratory test values. A possible explanation for this behaviour is that patients in our study, given their good level of knowledge about their disease and previous experience with interpreting their laboratory test results, did not need any additional eye fixation on the additional features introduced by the Contextualised and Grouped presentation. Ultimately, patients might have had a specific search strategy that was used in each task. In this case, the changes in the interface design alone were not enough to alter this visual search behaviour as shown in Figure 6.4.

The findings of this study should be interpreted with caution due to the presence of several limitations. First, we used a relatively small sample size. A larger sample size would have strengthen the statistical power of the results. Also, a larger and more diverse sample would have allowed us to better evaluate the influence of personal characteristics on risk interpretation. Second, we tested only two alternative web-based presentations of laboratory test results. Using more comparison interfaces would have enabled us to record more data about the effect of interface design on risk interpretation and user interaction. Finally, patients were given fictitious/simulated clinical scenarios to assess risk. Observing patients' interpretation of data from their own laboratory test results could simulate a more naturalistic research environment and strengthen the applicability of our results in clinical practice.

6.6 CONCLUSIONS

The presentations we tested were equivalent in supporting patients in interpreting risk correctly. This study confirmed patients' difficulties in interpreting laboratory test results, with many patients underestimating risk across all presentations, even when abnormal values were highlighted using visual elements or grouped together. Further research is needed to generalise the findings of this study in the context of other types of web-based presentations or interaction design styles and inform the evidence-based design of interfaces for the presentation of laboratory test results in patient portals.

6.7 REFERENCES

1. Pearce C, Bainbridge M. A personally controlled electronic health record for Australia. J Am Med Inform Assoc. 2014;21: 707–13. doi:10.1136/amiajnl-2013-002068

2. Institute of Medicine (USA). Achieve Meaningful use - Stage 2 [Internet]. Available: http://www.webcitation.org/6qBAzdht7

3. Urowitz S, Wiljer D, Apatu E, Eysenbach G, Delenardo C, Harth T, et al. Is Canada ready for patient accessible electronic health records? A national scan. BMC Med Inform Decis Mak. 2008;8: 33. doi:10.1186/1472-6947-8-33

4. Pagliari C, Shand T, Fisher B. Embedding online patient record access in UK primary care: a survey of stakeholder experiences. JRSM Short Rep. Royal Society of Medicine Press; 2012;3: 34. doi:10.1258/shorts.2012.012009

5. Fisher B. Patients' access to their electronic record: offer patients access as soon as you can. Br J Gen Pract. British Journal of General Practice; 2013;63: e423-5. doi:10.3399/bjgp13X668384

6. Adler-Milstein J, Sarma N, Woskie LR, Jha AK. A Comparison Of How Four Countries Use Health IT To Support Care For People With Chronic Conditions. Health Aff (Millwood). 2014;33: 1559–66. doi:10.1377/hlthaff.2014.0424

7. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and Provider Attitudes Toward the Use of Patient Portals for the Management of Chronic Disease: A Systematic Review. Brixey J, editor. J Med Internet Res. Toronto, Canada: JMIR Publications Inc.; 2015;17: e40. doi:10.2196/jmir.3703

8. Price M, Bellwood P, Kitson N, Davies I, Weber J, Lau F. Conditions potentially sensitive to a Personal Health Record (PHR) intervention, a systematic review. BMC Med Inform Decis Mak. 2015;15: 1–12. doi:10.1186/s12911-015-0159-1

9. Bayliss EA, Ellis JL, Steiner JF. Barriers to self-management and quality-of-life outcomes in seniors with multimorbidities. Ann Fam Med. 2007;5: 395–402. doi:10.1370/afm.722

10. Phelps GR, Taylor J, Simpson K, Samuel J, Turner NA. Patients' Continuing Use of an Online Health Record: A Quantitative Evaluation of 14,000 Patient Years of Access Data. J Med Internet Res. 2014;16: e241. doi:10.2196/jmir.3371

11. Cimino JJ, Patel VL, Kushniruk AW. The patient clinical information system (PatCIS): technical solutions for and experience with giving patients access to their electronic medical records. Int J Med Inform. 2002;68: 113–127. doi:10.1016/S1386-5056(02)00070-9

12. Shah SGS, Fittonb R, Hannanc A, Fisherd B, Younga T, Barnette J. Accessing personal medical records online: A means to what ends? Int J Med Inform. 2015;84: 111–8. doi:10.1016/j.ijmedinf.2014.10.005

13. Bhavnani V, Fisher B, Winfield M, Seed P. How patients use access to their electronic GP record--a quantitative study. Fam Pract. 2011;28: 188–94. doi:10.1093/fampra/cmq092

14. Young MJ, Scheinberg E, Bursztajn H. Direct-to-patient laboratory test reporting: balancing access with effective clinical communication. JAMA. American Medical Association; 2014;312: 127–8. doi:10.1001/jama.2014.5823

15. Taha J, Sharit J, Czaja SJ. Usability of an electronic personal health record (PHR) among a diverse group of adults. Proceedings of the Human Factors and Ergonomics Society. University of Miami, Miller School of Medicine, United States; 2014. pp. 619–623. doi:10.1177/1541931214581132

16. Brewer NT, Gilkey MB, Lillie SE, Hesse BW, Sheridan SL. Tables or bar graphs? Presenting test results in electronic medical records. Med Decis Making. United States; 2012;32: 545–553. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=22472914

17. Zikmund-Fisher BJ, Exe NL, Witteman HO. Numeracy and Literacy Independently Predict Patients' Ability to Identify Out-of-Range Test Results. J Med Internet Res. Journal of Medical Internet Research; 2014;16: e187. doi:10.2196/jmir.3241

18. Britto MT, Jimison HB, Munafo JK, Wissman J, Rogers ML, Hersh W. Usability testing finds problems for novice users of pediatric portals. J Am Med Inform Assoc. United States; 2009;16: 660–669. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=19567793

19. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. Helping Patients Decide: Ten Steps to Better Risk Communication. JNCI J Natl Cancer Inst. Oxford University Press; 2011;103:

1436–1443. doi:10.1093/jnci/djr318

20. Simborg DW, Detmer DE, Berner ES. The wave has finally broken: now what? J Am Med Inform Assoc. BMA House, Tavistock Square, London, WC1H 9JR: BMJ Publishing Group; 2013;20: e21–e25. doi:10.1136/amiajnl-2012-001508

21. Iacobucci G. Patients promised online access to their medical records by 2018. BMJ. 2015;351. Available: http://www.bmj.com/content/351/bmj.h4740.abstract

22. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How Numeracy Influences Risk Comprehension and Medical Decision Making. Psychol Bull. 2009;135: 943–973. doi:10.1037/a0017327

23. Hill S, Spink J, Cadilhac D, Edwards A, Kaufman C, Rogers S, et al. Absolute risk representation in cardiovascular disease prevention: comprehension and preferences of health care consumers and general practitioners involved in a focus group study. BMC Public Health. 2010;10: 1–13. doi:10.1186/1471-2458-10-108

24. Klimov D, Shahar Y. A Framework for Intelligent Visualization of Multiple Time-Oriented Medical Records. AMIA Annu Symp Proc. American Medical Informatics Association; 2005;2005: 405–409. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1560450/

25. Gotz D, Sun J, Cao N, Ebadollahi S. Visual Cluster Analysis in Support of Clinical Decision Intelligence. AMIA Annu Symp Proc. American Medical Informatics Association; 2011;2011: 481–490. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3243281/

26. Wang TD, Wongsuphasawat K, Plaisant C, Shneiderman B. Visual Information Seeking in Multiple Electronic Health Records: Design Recommendations and a Process Model. Proceedings of the 1st ACM International Health Informatics Symposium. New York, NY, USA: ACM; 2010. pp. 46–55. doi:10.1145/1882992.1883001

27. Shneiderman B, Plaisant C, Cohen M, Jacobs S. Designing the User Interface: Strategies for Effective Human-Computer Interaction. Pearson Prentice Hall/Pearson education international; 2009.

28. Bartlett C, Simpson K, Turner AN. Patient access to complex chronic disease records on the Internet. BMC Med Inform Decis Mak. 2012;12: 87. doi:10.1186/1472-6947-

12-87

29. Radhakrishnan J, Remuzzi G, Saran R, Williams DE, Rios-Burrows N, Powe N, et al. Taming the chronic kidney disease epidemic: a global view of surveillance efforts. Kidney Int. 2014;86: 246–250. doi:10.1038/ki.2014.190

30. Luo J, Brunelli SM, Jensen DE, Yang A. Association between Serum Potassium and Outcomes in Patients with Reduced Kidney Function. Clin J Am Soc Nephrol . 2016;11: 90–100. Available: http://cjasn.asnjournals.org/content/11/1/90.abstract

31. Harle CA, Padman R, Downs JS. Design, Implementation, and Preliminary Evaluation of a Web-Based Health Risk Calculator. System Sciences, 2009 HICSS '09 42nd Hawaii International Conference on. 2009. pp. 1–10. doi:10.1109/HICSS.2009.148

32. Evans J. Heuristic and analytic processes in reasoning. Br J Psychol. 1984;75: 451–468.

33. Chaiken S. Heuristic versus systematic information processing and the use if source versus message cues in persuasion. J Pers Soc Psychol. 1980;39: 752–766.

34. Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM. Measuring Numeracy without a Math Test: Development of the Subjective Numeracy Scale. Med Decis Mak. 2007;27: 672–680. Available: http://mdm.sagepub.com/content/27/5/672.abstract

35. Chew L, Griffin J, Partin M, Noorbaloochi S, Grill J, Snyder A, et al. Validation of Screening Questions for Limited Health Literacy in a Large VA Outpatient Population. J Gen Intern Med. Springer-Verlag; 2008;23: 561–566. doi:10.1007/s11606-008-0520-5

36. Galesic M, Garcia-Retamero R. Graph Literacy: A Cross-Cultural Comparison. Med Decis Mak. 2011;31: 444–457. Available: http://mdm.sagepub.com/content/31/3/444.abstract

37. Tufte ER. The Visual Display of Quantitative Information. Cheshire, Connecticut: Graphics Press USA; 2001.

38. White RW. Interaction with Search Systems. New York, NY, USA: Cambridge University Press; 2016.

39. Elder NC, Barney K. "But what does it mean for me?" Primary care patients'

communication preferences for test results notification. Jt Comm J Qual Patient Saf. 2012;38: 168–76. Available: http://www.ncbi.nlm.nih.gov/pubmed/22533129

40. Kules B, Shneiderman B. Users can change their web search tactics: Design guidelines for categorized overviews. Inf Process Manag. 2008;44: 463–484. doi:http://dx.doi.org/10.1016/j.ipm.2007.07.014

41. Dumais S, Cutrell E, Chen H. Optimizing Search by Showing Results in Context. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM; 2001. pp. 277–284. doi:10.1145/365024.365116

42. Greene S, Marchionini G, Plaisant C, Shneiderman B. Previews and overviews in digital libraries: Designing surrogates to support visual information seeking. J Am Soc Inf Sci. John Wiley & Sons, Inc.; 2000;51: 380–393. doi:10.1002/(SICI)1097-4571(2000)51:4<380::AID-ASI7>3.0.CO;2-5

43. West VL, Borland D, Hammond WE. Innovative information visualization of electronic health record data: a systematic review. J Am Med Informatics Assoc. 2015;22: 330 LP-339. Available: http://jamia.oxfordjournals.org/content/22/2/330.abstract

44. Johnson CM, Shaw RJ. A Usability Problem: Conveying Health Risks to Consumers On the Internet. AMIA Annu Symp Proc. American Medical Informatics Association; 2012;2012: 427–435. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540473/

45. Rothman AJ, Kiviniemi MT. Treating People With Information: an Analysis and Review of Approaches to Communicating Health Risk Information. JNCI Monogr. 1999;1999: 44–51. Available: http://jncimono.oxfordjournals.org/content/1999/25/44.abstract

46. Ahmed H, Naik G, Willoughby H, Edwards AGK. Communicating risk. BMJ. 2012;344. Available: http://www.bmj.com/content/344/bmj.e3996.abstract

47. Andry F, Naval G, Nicholson D, Lee M, Kosoy I, Puzankov L. Data visualization in a personal health record using rich internet application graphic components. HEALTHINF 2009 - Proceedings of the 2nd International Conference on Health Informatics. InterComponentWare Inc., One Waters Park Drive- Suite 280, San Mateo, CA 94403, United States; 2009. pp. 111–116. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-

77954660690&partnerID=40&md5=8eb9214249dcd66275f4ad5af5d9645a

48. Hill S, Spink J, Cadilhac D, Edwards A, Kaufman C, Rogers S, et al. Absolute risk representation in cardiovascular disease prevention: comprehension and preferences of health care consumers and general practitioners involved in a focus group study. BMC Public Health. BioMed Central; 2010;10: 108. doi:10.1186/1471-2458-10-108

49. Ware C. Information Visualization, 2nd Edition. Morgan Kaufmann; 2004.

50. Mazza R. Introduction to Information Visualization. London, UK: Springer London; 2009. doi:10.1007/978-1-84800-219-7

51. Rayner K. Eye movements and attention in reading, scene perception, and visual search. Q J Exp Psychol. Routledge; 2009;62: 1457–1506. doi:10.1080/17470210902816461

52. Matos R. Designing eye tracking experiments to measure human behavior [Internet]. 2010. Available: http://www.webcitation.org/6qB95Hc4m

53. O'Kane M, Freedman D, Zikmund-Fisher BJ. Can patients use test results effectively if they have direct access? BMJ Br Med J. 2015;350. Available: http://www.bmj.com/content/350/bmj.h673.abstract

54. Fraccaro P, Brown B, Prosperi M, Sperrin M, Buchan I, Peek N. Development and preliminary validation of a dynamic, patient-tailored method to detect abnormal laboratory test results. Stud Health Technol Inform. 2015;216: 701–705. doi:10.3233/978-1-61499-564-7-701

55. Segall N, Saville JG, L'Engle P, Carlson B, Wright MC, Schulman K, et al. Usability evaluation of a personal health record. AMIA Annu Symp Proc. 2011;2011: 1233–42. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3243224&tool=pmcentrez&rendertype=abstract

56. Segall N, Saville JG, L'Engle P, Carlson B, Wright MC, Schulman K, et al. Usability evaluation of a personal health record. AMIA Annu Symp Proc. United States; 2011;2011: 1233–1242. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=22195184

57. Gist ME, Schwoerer C, Rosen B. Effects of alternative training methods on self-efficacy and performance in computer software training. J Appl Psychol. US: American Psychological Association; 1989;74: 884–891. doi:10.1037/0021-9010.74.6.884

58. Hasan B. The influence of specific computer experiences on computer self-efficacy beliefs. Comput Human Behav. 2003;19: 443–450. doi:http://dx.doi.org/10.1016/S0747-5632(02)00079-1

59. Schwartz LM, Woloshin S, Welch HG. Can Patients Interpret Health Information? An Assessment of the Medical Data Interpretation Test. Med Decis Mak. 2005;25: 290–300. Available: http://mdm.sagepub.com/content/25/3/290.abstract

6.8 SUPPLEMENTARY MATERIAL

Supplementary Table 6.1: Characteristics of patients who did not make safety mistakes and those who did. Abbreviations: SD, Standard deviation; SNC: Subjective Numeracy Scale; GCSE: General Certificate of Secondary Education.

Parameters		**Patients who did not make safety mistakes**	**Patients who made safety mistakes**
Number of patients		7	13
Gender	Female (%)	1 (14)	3 (23)
	Male (%)	6 (86)	10 (77)
Age (years)		51.1 (12.9)	52.2 (9.1)
Years since kidney transplant (mean, SD)		11.9 (9.6)	10 (8.5)
SNC score (mean, SD)		3.7 (0.9)	4.1 (0.9)
Health literacy (mean, SD)		0.3 (0.3)	0.6 (0.7)
Graph literacy score (mean, SD)		72.5 (11.6)	74.5 (11.1)
Education	Lower than GCSE (%)	1 (14)	0 (0)
	GCSE (%)	1 (14)	6 (46)
	A-level/College (%)	2 (29)	3 (23)
	Higher education/University degree (%)	3 (43)	4 (31)
Internet use	Less than one hour per week (%)	0 (0)	1 (8)
	One to five hours per week (%)	1 (14)	4 (31)
	Five to 10 hours per week (%)	2 (29)	3 (23)
	More than 10 hours per week (%)	4 (57)	5 (38)
PatientView use	Never used (%)	1 (14)	2 (15)
	Less or equal than once per year (%)	1 (14)	4 (31)
	Twice per year (%)	0	1 (8)
	Quarterly (%)	5 (71)	6 (46)

Supplementary Table 6.2: Characteristics of patients who did not require unnecessary care and those who did. Abbreviations: SD, Standard deviation; SNC: Subjective Numeracy Scale; GCSE: General Certificate of Secondary Education.

Parameters		**Patients who did not require unnecessary care**	**Patients who required unnecessary care**
Number of patients		6	14
Gender	Female (%)	3 (50)	1 (7)
	Male (%)	3 (50)	13 (93)
Age (years)		52.8 (5.8)	51.4 (11.8)
Years since kidney transplant (mean, SD)		15.5 (9.9)	8.6 (7.5)
SNC score (mean, SD)		3.8 (0.8)	4.1 (0.8)
Health literacy (mean, SD)		0.7 (0.4)	0.4 (0.6)
Graph literacy score (mean, SD)		71.8 (10.5)	74.7 (11.5)
Education	Lower than GCSE (%)	1 (17)	0 (0)
	GCSE (%)	3 (50)	4 (29)
	A-level/College (%)	1 (17)	4 (29)
	Higher education/University degree (%)	1 (17)	6 (43)
Internet use	Less than one hour per week (%)	0 (0)	1 (7)
	One to five hours per week (%)	3 (50)	2 (14)
	Five to 10 hours per week (%)	1 (17)	4 (29)
	More than 10 hours per week (%)	2 (33)	7 (50)
PatientView use	Never used (%)	1 (17)	2 (14)
	Less or equal than once per year (%)	0 (0)	5 (36)
	Twice per year (%)	0 (0)	1 (7)
	Quarterly (%)	5 (83)	6 (43)

Supplementary Table 6.3: Coefficient estimates in the mixed-effects linear regression for the eye fixation count and dwell time. Abbreviations: CI. Confidence interval.

Covariate	**Eye fixation count coefficient estimate [95% CI]**	**Dwell time coefficient estimate [95% CI]**
(intercept)	24.99 [-25.31,75.3]	7.24 [-6.7,21.17]
Age	0.01 [-0.66,0.69]	0 [-0.19,0.19]
Years since transplant	-0.13 [-0.89,0.63]	-0.02 [-0.23,0.19]
Contextualised presentation vs Baseline	1.19 [-2.82,5.21]	0.39 [-0.75,1.54]
Grouped presentation vs Baseline	1.46 [-2.55,5.47]	0.52 [-0.63,1.66]
Medium risk clinical scenario vs Low risk	0.5 [-3.52,4.51]	0.17 [-0.98,1.31]
High risk clinical scenario vs Low risk	1.51 [-2.5,5.52]	0.42 [-0.73,1.56]
Education[a)]	-1.69 [-6.45,3.07]	-0.4 [-1.72,0.92]
Frequency of internet use[a)]	-5.28 [-13.88,3.32]	-1.67 [-4.06,0.71]
Frequency of PatientView use[a)]	**-5.22 [-9.81,-0.63][b)]**	**-1.47 [-2.74,-0.2][b)]**
SNC	4.13 [-3.89,12.16]	1.25 [-0.97,3.47]
Health literacy	4.6 [-5.14,14.34]	1.39 [-1.31,4.09]
Graph literacy	0.21 [-0.41,0.84]	0.06 [-0.11,0.23]

a) Ordered categorical factors considered as numeric covariates in the analysis.
b) Statistically significant at 0.05 level.

Chapter 7

DEVELOPMENT AND PRELIMINARY VALIDATION OF A DYNAMIC, PATIENT-TAILORED METHOD TO DETECT ABNORMAL LABORATORY TEST RESULTS

Paolo Fraccaro [1,2], Benjamin Brown [1,2], Mattia Prosperi [1], Matthew Sperrin [1], Iain E. Buchan [1,2], Niels Peek [1,2]

[1] Health eResearch Centre, Farr Institute for Health Informatics Research, University of Manchester, UK.

[2] NIHR Greater Manchester Primary Care Patient Safety Translational Research Centre, Institute of Population Health, The University of Manchester, Manchester, United Kingdom.

Contribution: PF (corresponding author), BB and MS developed the tested method. PF and NP designed the validation study. PF collected the data and performed all analyses. PF and NP drafted the manuscript. BB, MP, MS and IB critically edited the manuscript. PF replied to the reviewers' comments during the review process.

Based on: Fraccaro P, Brown B, Prosperi M, Sperrin M, Buchan I, et al. (2015) Development and preliminary validation of a dynamic, patient-tailored method to detect abnormal laboratory test results. Stud Health Technol Inform 216: 701–705. Available: http://ebooks.iospress.nl/publication/40300.

7.1 ABSTRACT

Laboratory test results in primary care are flagged as 'abnormal' when they fall outside a population-based Reference Interval (RI), typically generating many alerts with a low specificity. In order to decrease alert frequency while retaining clinical relevance, we developed a method to assess dynamic, patient-tailored RIs based on mixed-effects linear regression models. Potassium test results from primary care were used as proof-of-concept test bed. Clinical relevance was assessed via a survey administered to general practitioners (GPs). Overall, the dynamic, patient-tailored method and the combination of both methods flagged 20% and 36% fewer values as abnormal than the population-based method. Nineteen out of 43 invited GPs (44%) completed the survey. The population-based method yielded a better sensitivity than the patient-tailored and the combined methods (0.51 vs 0.41 and 0.38, respectively) but a lower PPV (0.66 vs 0.67 and 0.76, respectively). We conclude that a combination of population-based and patient-tailored RIs can improve the detection of abnormal laboratory results. We suggest that lab values outside both RIs be flagged with high priority in clinical practice.

Keywords: Clinical Decision Support Systems; Point-of-Care Systems; Biochemistry.

7.2 INTRODUCTION

Failure to follow up laboratory test results is a major concern in primary care [1]. Missed test results can lead to delayed interventions or to decisions made on the basis of incomplete information with potential compromise of patient safety [2]. For patients, poor tests handling may generate increased number of visits, repeated laboratory examinations, and unnecessary stress or harm [3]. Callen et al. [1] showed that the extent of failure to follow up test results in primary care ranges from 7% to 62% of results; and it can cause delayed diagnosis, preventable hospitalisation and adverse drug reactions.

Electronic Health Records (EHRs) can improve test result follow-up and management [4]. However, general practicioners' (GPs) satisfaction of informatics systems in place to manage tests is low [5]. One of the main barriers for timely follow-up is the high number of often unecessary alerts that physicians are presented with on a daily basis [6,7]. It is estimated that each general practitioner spends almost an hour per day processing alerts generated by primary care information systems, the majority of which is composed of test result alerts [5,6]. As a consequence, physicians may not have the necessary time to focus on the most important alerts [8]. This information overload contributes to alert fatigue [9] and it can potentially generate patient safety issues [1].

The high number of alerts produced is directly related to the lack of specificity of current threshold-based methods for detecting abnormal values. The majority of EHRs use population-based Reference Intervals (RIs), which are defined as "intervals that, when applied to the population serviced by the laboratory correctly include most of the subjects with characteristics similar to the reference group and exclude the others." [10]. RIs are usually calculated by assuming a Gaussian distribution of test results for the given physiological measure (e.g. potassium, creatinine, heamoglobin), and estimating population mean and variance to calculate the RI as the 95% reference range. By definition a value outside the RI is flagged as abnormal [11]. As this approach is based on population estimates, it may flag values as abnormal that may be considered normal in the context of a specific patient's medical history. For instance, some patients have persistently high or low levels of certain physiological parameters, and this is dealth with via the clinician's knowledge of the patient. Conversely, some patients may experience sudden changes in critical parameters that need prompt medical action, but these changes are not flagged as abnormal because measured values are still within the population-based RI. In this regard, the adoption of more personalised methods that adapt to the patient's history could reveal key patterns and insights the interpretation based on population thresholds would not [12].

The aim of the paper is to develop a dynamic patient-tailored method for detecting abnormal laboratory test results in primary care, and assess the incidence of abnormal test results when using this method alone and in combination with the population-based method. Furthermore, we aimed to assess the potential clinical relevance through a survey administered to experienced practicioners in UK primary care.

7.3 METHODS

7.3.1 Model

Mixed-effects modelling allows the analysis of data with complex patterns of variability and hierarchical structure [13]. Specifically, when estimating RI for laboratory tests, mixed-effects modelling can take into account both the population variability and the intra-patient variability. Consider a patient i and its corresponding test result set $Y_i=y_{i1}$, by applying the introduced hierarchical structure and calculating the maximum likelihood estimate:

$$\textit{If } y_{ij} \sim N(\alpha_i, \sigma^2),\ \alpha_i \sim N(\mu,\omega^2),\ \textit{then } \alpha_i \mid \bar{y}_{ij}, \sigma^2, \mu,\omega^2 \sim N(\tilde{\mu}_{ij}, V_{ij}) \textit{ where}$$

$$\tilde{\mu}_{ij}= \frac{\mu\omega^{-2}+\bar{y}_{ij}\frac{n_{ij}}{\sigma^2}}{\omega^{-2}+\frac{n_{ij}}{\sigma^2}} \qquad \text{and} \qquad V_{ij}= (\omega^{-2} + \frac{n_{ij}}{\sigma^2})^{-1}$$

Equivalently,

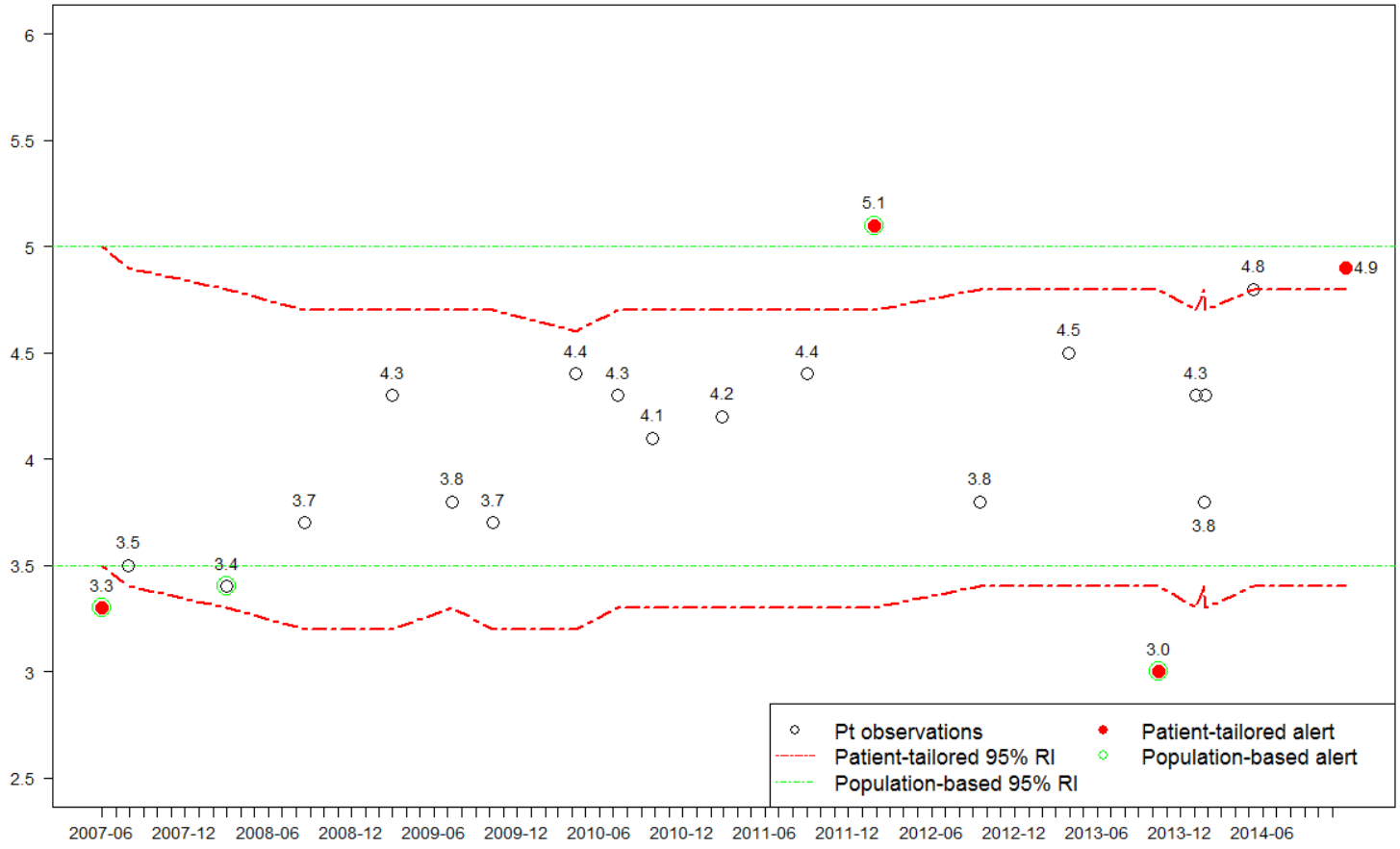
$$\tilde{\mu}_{ij}= \mu\lambda_{ij} + (1 - \lambda_{ij})\bar{y}_{ij} \tag{1}$$

$$\lambda_{ij}= \frac{\omega^{-2}}{\omega^{-2} + \frac{n_{ij}}{\sigma^2}} = \frac{\frac{\sigma^2}{n_{ij}}}{\omega^2 + \frac{\sigma^2}{n_{ij}}} = \frac{V_{ij}}{V_{ij} + \omega^2} \tag{2}$$

Here, μ and ω^2 are population mean and variance, y_{ij} is the jth observation of patient i, α_i is the mean of patient i, σ^2 is the intra-patient variance. Moreover, $\bar{y}_{ij}$ and n_{ij} are the sample mean and number of observations for patient i after j observations. Finally, $\tilde{\mu}_{ij}$ and V_{ij} are the maximum likelihood estimates of α_i and σ^2, and λ_{ij} is a shrinkage factor, that as soon as there

are more observations for patient i, increases the weight of the sample mean $\bar{y}_{ij}$ compared to the population mean μ. Accordingly, by using $\tilde{\mu}_{ij}$ and V_{ij} it is possible calculate an adaptive patient-tailored RI for patient i (defined as in the standard method as the 95% reference range), which is dynamically updated at each new test result and tells us what is normal/abnormal in a specific patient's context. Figure 7.1 shows an example of the application of the population-based and patient-tailored RIs to a patient's potassium results time series to detect abnormal values. At the first observation (on the left), with no previous information about the spefic patient, the population-based and patient tailored RIs coincide. Conversely, with new observations coming, the patient-tailored RI is updated and adapts to the the patient's context. For example, the third value, that is similar to the previous ones but slighlty outside the population-based RI, is not flagged as abnormal according the patient-tailored method but it is according to the population-based ones. Finally, the last value, that is inside the RI but "unusual" for the patient, is flagged as abnormal by following the patient-tailored RI and it would be missed with the population-based one.

Figure 7.1: Example of application of the patient-tailored and population-based RIs to a patient's potassium results time series to detect abnormal values.



7.3.2 Data

The Salford Integrated Record (SIR) database was used to test our method. SIR is an anonymized electronic health record database from the City of Salford (population 234k, UK), which collects data from 49 primary care providers and one secondary care provider. Records are stored using Read codes v2 and Clinical terms Version 3 standards [14].

We focused our analysis on potassium results and we extracted data from SIR for patients aged 18 to 85 in the period 1/1/1990–31/12/2012. Potassium results are influenced by medications and chronic conditions, and in case of hypokalemia (<2.5 mmol/l) or hyperkalemia (>6 mmol/l) it can result in significant cardiac dysrhythmias. In addition, some patients can have levels persistently outside the RI without any effect on heart ryhthm. Accordingly, a patient-tailored alerting system would be particularly helpful for this laboratory parameter. As a consensus RI across UK laboratories we implemented in our study a potassium RI of 3.5 to 5 mmol/l [15].

7.3.3 Parameter estimation

From SIR data extraction, after outliers exclusion in order to reproduce the implemented potassium RI, we derived two datasets: 1) a test set made by 500 patients randomly extracted, with all their potassium results; 2) a training set composed by all the remaining data extract used for fitting the mixed-effects model and estimate its parameters to implement the abovementioned adaptive patient-tailored RI.

Potassium results in the test set were flagged as abnormal if out of the RI range according either to the standard threshold-based method using our UK consensus RI (referred from now on as *standard method*) or the adaptive patient-tailored RI *(patient-tailored method)*.

7.3.4 Evaluation

7.3.4.1 Survey design

In order to test the clinical relevance of flagged values by the patient-tailored method as well as the standard method, we used clinical judgement by GPs as gold standard. This judgment was obtained through a survey, as follows. From the analysed testing dataset, we randomly selected 15 values for each possible combination of the two methods, i.e. values alerted by: 1) only the standard method; 2) only the patient-tailored method; 3) the standard AND patient-tailored methods at the same time (referred in the text as *combined method*); 4) none. From these values, we randomly included two per combination (for a total of eight) in a survey. This

choice of numbers (i.e. 15 randomly selected values for each combination and two per combination in each survey) was derived via simulation, with the aim of increasing the chances of each value to be evaluated by multiple GPs. Particualry, we assumed to have 40 respondents and derived the combination of numbers that would allow us to have each value assessed by a median of three different GPs.

In the survey, GPs were asked to rate on a colour-based scale how abnormal each value was in the context of the specific patient:

- *Green* (normal value; i.e. no actions required);
- *Yellow* (probably abnormal; i.e. repeat in more than a week, do further test, change medication);
- *Red* (definitely abnormal; i.e. repeat urgently / hospital admission).

For each value GPs were provided with patient characteristics (i.e. age and gender), a graph showing the patient's previous potassium results, a brief summary of past medical history (i.e. comorbidities and time since diagnosis), and all medications that were prescribed during the last four months (see Supplementary Figure 7.1). In order to avoid priming no information about the standard and patient-tailored RIs were provided. Since values were randomly selected each time from the 15 values per combination included, it is noteworthy that all surveys administered to participants were different from each other, while maintaining the proportion between the different combination of the two methods; each value could be assessed by more than one general practitioner.

The survey also contained three questions about respondent's working days (1-3 days, 4-5 days), years of experience (1-10 years, 10-20 years, >20 years), and opinion about abnormal test results alerts ("not enough", "about right", "too much").

7.3.4.2 Participants and survey setting

We administered the survey to a group of 43 GPs taking part in a 5 days Continuing Professional Development course for leadership development held in the city of Manchester (UK) in Octorber 2014.

7.3.4.3 Data analysis

We considered as clinically relevant a value that was judged by GPs as at least probably abnormal (i.e. yellow and red in the coloured scale in the survey). For each type of value we

calculated the percentage of agreement between GPs as the mean over all values evaluated by more than one GP. Furthermore, in order to test performance, we calculated sensitivity and PPV for the standard, patient-tailored and combined methods by taking the prevalence of the different type of flagged values (i.e. flagged only by the standard method, only by the patient-tailored or by both methods at the same time) in the original dataset into account.

In order to assess variable importance, possible intra-assessor and intra-value correlation, a mixed-effects logistic regression was employed, using the flags by standard and patient-tailored methods as well as respondent characteristics as independent variables and the values clinical relevance as binary outcome.

All analyses were performed using the R software (http://www.r-project.org/).

7.4 RESULTS

7.4.1 Data analysis

We extracted 1,411,757 unique potassium results from the SIR database, for a total of 151,681 patients. Mean age at first potassium record was 45.4 (17), female accounted for 49.8% and the mean follow-up time was 4.7 (4). The test dataset was composed of 500 patients and 4,144 potassium results. Of these, 470 (11.3%) values were flagged as abnormal by the standard method, 372 (9%) by the patient-tailored method and 301 (7.3%) by the combined one. The patient-tailored and combined methods registered a 20% (98 values) and 36% (169 values) reduction of the number of abnormal flagged values compared to the standard method.

7.4.2 Evaluation

Of the 43 GPs that received the survey, 19 completed it (response rate of 44%). Each value was assessed by a median of 3 GPs. Table 7.1 reports respondents' characteristics. The majority of GPs had more than 20 years of experience in general practice (63.2%) and worked 1-3 days in general practice. Furthermore, 42.1% thought that there are too many test results alerts in general practice.

Table 7.1: Baseline characteristics of general practitioners.

Question	**Reply**	**N (%)**
Days per week in practice	1-3 days	10 (52.6%)
	4-5 days	9 (47.4%)
Years of experience	<10 years	2 (10.5%)
	10-20 years	5 (26.3%)
	>20 years	12 (63.2%)
Opinion about tests alerts in general practice	Not enough	4 (21.1%)
	About right	7 (36.8%)
	Too much	8 (42.1%)

Out of the 152 values assessed by GPs, 92 values were cosidered normal (green), 54 as probably abnormal (yellow), and 6 values as definitely abnormal (red). On average, each general practitioner identified 4.8 normal values (minimum: 1, maximum: 8), 2.8 probably abnormal values (minimum: 0; maximum: 7), and 0.3 definitely abnormal values (minimum: 0; maximum: 1).

Table 7.2 reports cross tabulation of values flagged as abnormal by all methods and whatever values were considered clinically relevant by GPs.

Table 7.2: Cross tabulation between standard, patient-tailored and combined methods and general practitioners judgement as clinically relevant.

		Clinically relevant		
		Neg	*Pos*	*Total*
Standard method	*Neg*	63	13	76
	Pos	29	47	76
	Total	92	60	152
		Clinically relevant		
		Neg	*Pos*	*Total*
Patient-tailored method	*Neg*	55	21	76
	Pos	37	39	76
	Total	92	60	152
		Clinically relevant		
		Neg	*Pos*	*Total*
Combined method	*Neg*	83	31	114
	Pos	9	29	38
	Total	92	60	152

The mean percentage of agreement between GPs was 94% for values not alerted by any method, 76% for values alerted only by the standard method, 87% for values alerted only by the patient tailored method, and 84% for values alerted by both methods.

Table 7.3 reports sensitivity and PPV of all three methods, calculated by taking the prevalence in the original dataset into account. The standard and combined methods had the best performance in terms of sensitiviy and PPV respectively.

Table 7.3: PPVs and sensitivities based on general practitioners judgements as clinically relevant.

Parameter	Standard method	Patient- tailored method	Combined method
Prevalence in testing dataset (n=4,144)	11.3%	9%	7.3%
Prevalence in values assessed by GPs	50%	50%	25%
Sensitivity	0.51	0.41	0.38
PPV	0.66	0.67	0.76

Table 7.4 shows the adjusted ORs for the fixed effects in the mixed-effects logistic regression of values identified as clinically relevant by GPs. The estimated variance of the random effects for assessor and value were 1.5 (SD:1.2) and 0.4 (SD: 0.6) respectively. The only two significant variables were the flags by the standard and patient-tailored methods.

Table 7.4: Adjusted ORs for values identified as clinically relevant by general practitioners.(GP: General Practice).

Parameter	Adjusted OR [95% CI]
Standard method pos. vs neg.	24.5 [5.3,113.7]
Patient tailored method pos. vs. neg.	6.2 [2.0,19.1]
Weekly working days in GP: 4-5 days vs 1-3 days	2.2 [0.4,11.3]
Years of experience in GP: 10-20 years vs <10 years	3.5 [0.4,11.3]
Years of experience in GP: >20 years vs <10 years	6.0 [0.3,103.1]
Opinion about tests alerts in GP: not enough vs about right	0.5 [0.7,3.7]
Opinion about tests alerts in GP: too much vs about right	0.2 [0,1.3]

7.5 DISCUSSION

This paper describes the development and preliminary evaluation of a method to produce dynamic, patient-tailored alerts for abnormal test results. The evaluation was focused on

potassium results and clinical relevance was assessed via a survey administred to a group of experienced GPs in Manchester (UK).

Looking at performance of the standard and patient-tailored method, overall the standard method yielded a better trade-off between sensitivity and PPV. Although the dynamic, patient-tailored method achieved a reduction in the number of flagged abnormal values and a slightly better PPV, the standard method was significantly more sensitive.

The combined method showed poor sensitivity (similar in absolute numbers to the patient-tailored method). However, PPV was the best by far. This result suggests that combining population RI with patient-specific contextual information, improves the clinical relevance of flagged values. This finding is not unexpected: values flagged by the combined method are potassium results that are abnormal for the healthy population as well as in the context of a particular patient. The importance of the additional information provided by the patient-tailored method is also confirmed by the mixed-effects logistic regression modelling.

To decrease the number of not relevant test results alerts, some investigators have suggested alerting physicians only about those values that have a high deviance from the RI [16,17]. We performed a sensitivity analysis by adopting a threshold-based approach with potassium RI of 2.5-6 mmol/l (previously proposed in [17]) that confirmed our main analysis. In detail, weighted sensitivity and PPV were 0.1 and 0.89 respectively. Although values outside the adopted RI had a very high deviance from the UK potassium RI not all flagged values were considered clinically relevant and just a small proportion was considered definitely abnormal. This confirms the importance of the context and the bluntness of using fixed-thresholds when looking at individual patients.

Our main focus was on PPV, which is key to avoiding alert fatigue as high PPV values reflect more relevant alerts likely to be well received by clinicians [18,19]. However, we note that low sensitivities obtained by the patient-tailored and combined methods would lead to clinically relevant abnormal values being missed, with possible consequences for patient safety. It is noteworthy that in absolute numbers even the sensitivity by the standard method (0.51) cannot be considered satisfactory and would lead to miss many values that are abnormal in the context of a specific patient but lay within the population RI. We have shown in this study that modelling patient's values over time in isolation is not enough to identify these clinically relevant values laying within the population RI. Future studies should focus on incorporating more covariates in the modelling (e.g. medications and comorbidities). Furthermore, future research should focus on trying to go beyond the abstract concept of normality/abnormality towards a more concrete definition of clinical relevance in terms of

laboratory tests that triggered further clinical actions (e.g. further testing, change in medications or adverse event).

In addition to the lack of sensitivity, the patient-tailored and stardard methods have further drawbacks directly related to how they are defined. On one side, because of his adaptative nature, the patient-tailored method in presence of patients that are not stable but worsen over time (i.e. the value for a specific parameter keeps increasing or decreasing) would slowly adapt to the abnormal values and produce a biased patient-tailored RI. On the other side, the standard method, which uses fixed thresholds, would keep flagging tests results as abnormal when patients' values are stable but consistently outside the population RI. Improvements are needed in these regards for all methods.

Our analysis has several limitations. First, we focused on only one biochemical factor (potassium). Second, we carried out our evaluation within a relatively small panel of GPs in one geographical area, with a low response rate to our survey. Although this might limit the generalisability of our findings, mean percentage of agreement between GPs was high. Therefore, we expect this not to have an impact on our interpretation of the results. Third, the UK potassium RI adopted might have been slightly different than the one used in other laboratories that the clinicians may be used to; furthermore there might have been GPs that slavishly applied the abovementioned RI when replying to the survey.

7.6 CONCLUSION

This study demonstrates that the combined adoption of a patient-tailored method and a the standard threshold-based method for assessing potassium levels can improve the PPV of results flagged as abnormal. This could be particularly important to prioritise alerts by making the values flagged by both methods more prominent.

We plan to extend our experiments to a wider panel of laboratory tests (e.g. creatinine, eGFR, calcium, heamoglobin) and to a larger number of GPs, as well as investigating alternative statistical approaches (including Bayesian inference). We also plan to relate the alerting performance to adverse health outcomes.

This study represents a first step towards a next generation of context-aware alerting systems that in future may enhance patient safety.

7.7 REFERENCES

1. Callen JL, Westbrook JI, Georgiou A, Li J. Failure to follow-up test results for ambulatory patients: a systematic review. J Gen Intern Med. 2012;27: 1334–48. doi:10.1007/s11606-011-1949-5

2. Elder NC, Dovey SM. Classification of medical errors and preventable adverse events in primary care: a synthesis of the literature. J Fam Pract. 2002;51: 927–32. Available: http://www.ncbi.nlm.nih.gov/pubmed/12485545

3. McKay J, Bradley N, Lough M, Bowie P. A review of significant events analysed in general practice: implications for the quality and safety of patient care. BMC Fam Pract. 2009;10: 61. doi:10.1186/1471-2296-10-61

4. Sittig DF, Singh H. Electronic Health Records and National Patient-Safety Goals. N Engl J Med. Massachusetts Medical Society; 2012;367: 1854–1860. doi:10.1056/NEJMsb1205420

5. Poon EG, Gandhi TK, Sequist TD, Murff HJ, Karson AS, Bates DW. "I wish I had seen this test result earlier!": Dissatisfaction with test result management systems in primary care. Arch Intern Med. American Medical Association; 2004;164: 2223–8. doi:10.1001/archinte.164.20.2223

6. Murphy DR, Reis B, Sittig DF, Singh H. Notifications received by primary care practitioners in electronic health records: a taxonomy and time analysis. Am J Med. 2012;125: 209.e1-7. doi:10.1016/j.amjmed.2011.07.029

7. Singh H, Spitzmueller C, Petersen NJ, Sawhney MK, Sittig DF. Information overload and missed test results in electronic health record-based settings. JAMA Intern Med. 2013;173: 702–4. doi:10.1001/2013.jamainternmed.61

8. Singh H, Spitzmueller C, Petersen NJ, Sawhney MK, Smith MW, Murphy DR, et al. Primary care practitioners' views on test result management in EHR-enabled health systems: a national survey. J Am Med Inform Assoc. 2013;20: 727–35. doi:10.1136/amiajnl-2012-001267

9. Sijs IH van der (Heleen). Drug safety alerting in computerized physician order entry: Unraveling and counteracting alert fatigue. 2009; Available: http://www.webcitation.org/6NYq64xRe

10. Ceriotti F. Prerequisites for use of common reference intervals. Clin Biochem Rev. 2007;28: 115–21. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1994109&tool=pmcentrez&rendertype=abstract

11. Katayev A, Balciza C, Seccombe DW. Establishing reference intervals for clinical laboratory test results: is there a better way? Am J Clin Pathol. 2010;133: 180–6. doi:10.1309/AJCPN5BMTSF1CDYP

12. Baron JM, Dighe AS, Arnaout R, Balis UJ, Black-Schaffer WS, Carter AB, et al. The 2013 symposium on pathology data integration and clinical decision support and the current state of field. J Pathol Inform. 2014;5: 2. doi:10.4103/2153-3539.126145

13. Galecki A, Burzykowski T. Linear Mixed-Effects Models Using R A Step-by-Step Approach. Casella G, Fienberg SE, Olkin I, editors. New York, New York, USA: Springer New York; 2012.

14. NHS England. Read Codes [Internet]. [cited 16 Jun 2014]. Available: http://www.connectingforhealth.nhs.uk/systemsandservices/data/uktc/readcodes

15. reference range (potassium) - General Practice Notebook [Internet]. [cited 4 Dec 2014]. Available: http://www.gpnotebook.co.uk/simplepage.cfm?ID=584056899

16. Salinas M, López-Garrigós M, Gutiérrez M, Lugo J, Flors L, Leiva-Salinas C. Should we customise critical value procedure according to patient origin and laboratory turnaround time? J Clin Pathol. 2013;66: 269–72. doi:10.1136/jclinpath-2012-201030

17. Mathew G, Kho A, Dexter P, Bloodworth N, Fantz C, Spell N, et al. Concept and development of a discharge alert filter for abnormal laboratory values coupled with computerized provider order entry: a tool for quality improvement and hospital risk management. J Patient Saf. 2012;8: 69–75. doi:10.1097/PTS.0b013e31824aba75

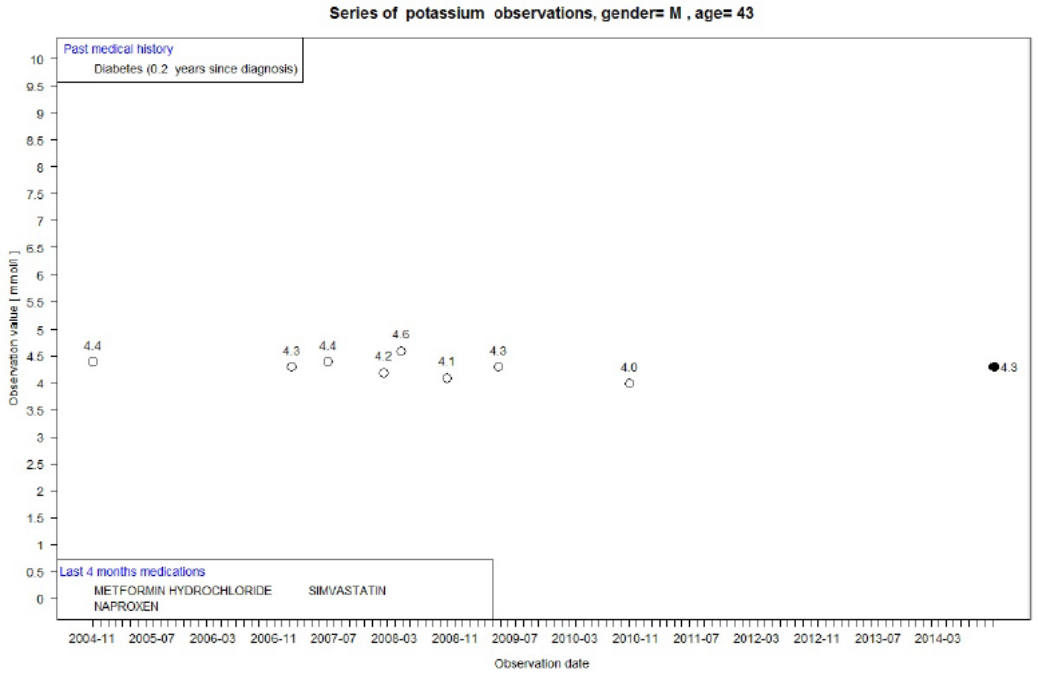
18. Bitan Y, O'Connor MF. Correlating data from different sensors to increase the positive predictive value of alarms: an empiric assessment. F1000Research. London, UK: F1000Research; 2012;1: 45. doi:10.12688/f1000research.1-45.v1

19. Eppenga WL, Derijks HJ, Conemans JMH, Hermens WAJJ, Wensing M, De Smet PAGM. Comparison of a basic and an advanced pharmacotherapy-related clinical decision support system in a hospital care setting in the Netherlands. J Am Med Inform Assoc. BMA House, Tavistock Square, London, WC1H 9JR: BMJ Group; 2012;19:

66–71. doi:10.1136/amiajnl-2011-000360

7.8 SUPPLEMENTARY MATERIAL

Supplementary Figure 7.1: Example of chart provided to the general practitioners (GPs) in the evaluation survey.



What colour this value (black dot) should be flagged?

- ☑ Green (normal value; i.e. no actions required)
- ☐ Yellow (probably abnormal; i.e. repeat in more than a week, do further test, change medication)
- ☐ Red (definitely abnormal; i.e. repeat urgently, hospital admission)

Chapter 8

DISCUSSION AND CONCLUSIONS

Although extensive discussions and specific conclusions were reported at the end of each chapter, here we aim at drawing together the key findings for each research question and highlighting their significance. For each research question, we also describe to what extent the knowledge produced as part of this thesis is actionable in the healthcare contexts in Table 1.1 (i.e. clinical practice, technology development, population health management and research). Finally, we provide some conclusions and future directions.

8.1 SUMMARY OF FINDINGS AND SIGNIFICANCE OF WORK

This thesis made secondary use of EHR data with the aim of producing actionable information to overcome some of the current challenges in predictive modelling using EHR data and patients access to their EHR data in patient portals. Below, we report how we addressed the three research questions we explored in this thesis.

8.1.1 RQ1: How can we use longitudinal information in EHRs more effectively to investigate multimorbidity?

The main context in which this research question provided actionable information is research. Particularly, the systematic review of the literature in Chapter 2 showed how multimorbidity is an under-investigated area in the informatics community. This is an important finding, as multimorbidity is an area of public health and clinical importance [1–3]. Therefore, research in this area should have high priority. We also suggest that the health informatics community can potentially have a key role in plugging evidence gaps around multimorbidity through research using EHR data. In this regard, Chapter 3 explores how to more effectively use the vast amount of longitudinal information available in EHRs, which is often disregarded by studies in the literature [4–6]. Particularly, we used the Charlson Comorbidity Index (CCI) [7] as a proxy of comorbidity burden, and showed that multimorbidity is indeed a dynamic process, with 10% of patients in our study population experiencing a comorbidity change in 5 years follow-up. We also demonstrated that the importance of such longitudinal changes in survival models is two-fold. First, accounting for this changes with a time-dependent variable produces survival models that better capture the effect of comorbidity burden on mortality. Second, we showed how the

comorbidity changes have an important prognostic impact, with the patients experiencing them who had an increased mortality risk. We suggest that this type of survival model should be considered when using EHR data for research.

The knowledge produced to address this research question cannot be considered actionable in any of the other healthcare contexts in Table 1.1. As a matter of fact, the predictions from the survival models tested in Chapter 3 would not support decision making by clinicians or policy makers in the UK primary care context.

8.1.2 RQ2: How can we use EHR data to externally validate existing prediction models?

Primary care EHR data, with their large size and broad population base, offer ample opportunities to perform external validation studies of predictive models. It is essential that such studies are conducted before models are used in clinical practice [8–11]. In Chapter 4 we presented an external validation study of existing Chronic Kidney Disease (CKD) predictive models in the UK primary care context. The study did not develop any new methodology, but followed international guidelines to perform external validation analyses [12,13]. Furthermore, it is one of the first studies to do a head-to-head comparison of multiple predictive models (both logistic and Cox proportional hazard regressions) on a large EHR dataset.

In our external validation study, we compared seven CKD onset predictive models that we selected from the literature. We tested their performance on a five-year horizon of disease onset in terms of calibration and discrimination by using multiple metrics. All predictive models discriminated well between patients who developed CKD and those who did not, with a c-statistics around 0.90. However, most of included predictive models were not calibrated to the UK primary care context and substantially over-predicted the risk of developing CKD. Only two models did not need any recalibration. Overall, QKidney [14], the only model originally developed in the UK, outperformed the other predictive models. Particularly, QKidney demonstrated to support a high risk approach to CKD prevention.

This finding is actionable at a population health management level. On the basis of our results, policy makers should consider to update clinical practice guidelines by including QKidney among the CKD screening criteria. For example, policy makers could decide to recommend monitoring of estimated Glomerular Filtration Rate (eGFR) and albumin-creatinine ratio (ACR) for all patients with a prediction greater than 0.0692 (i.e. the threshold probability associated with the top 10% of predicted risk that we found in our study). We showed that with this approach it would be possible to identify 64.5% of those developing CKD in five years by targeting 10% of the study population.

The evidence produced by our study is not currently actionable in the technology development and clinical practice contexts. However, the eventual decision from population health management to include QKidney [14] in clinical practice guidelines might influence decisions in the other two contexts. For example, since all information needed to calculate predictions is in EHRs, EHR providers could consider to integrate QKidney in EHR systems. At this point, with clinical practice guidelines prescribing what to do in case of a QKidney prediction above threshold and the model in form of alert in EHR systems, GPs could use it in clinical practice for a timely identification of individual at a high risk of developing CKD.

8.1.3 RQ3: How can we use interface design and predictive modelling to enhance interpretation of clinical laboratory test results?

The research question aimed at using different methods (i.e. interface design and predictive modelling) to produce actionable information in context of technology development, in order to enhance interpretation of laboratory results on patient portals and EHR systems. Although interesting results came out as part of our studies, none of our findings can be considered actionable at this stage. First, in Chapter 6, where we described a controlled study assessing whether interpretation and decision-making is influenced by the way the laboratory test results are presented in patient portals, did not find any statistically significant difference between the three presentations we tested. Particularly, misinterpretation of risk was high for all three presentations, with patients often not recognising scenarios where they should have acted on the information (e.g., requiring an appointment earlier than scheduled or calling their doctor immediately). The information that is relevant in a technology development context produced by Chapter 6 is that participants preferred the alternative presentations to the Baseline system, which was based on an established patient portal for kidney patients in the UK (called PatientView [15]). However, this would not be enough to inform significant changes of the system currently in use. Second, the method to calculate dynamic, patient-tailored alerts developed in Chapter 7 was only preliminarily tested on one type of laboratory test value (i.e. potassium) among a small group of GPs. Therefore, although promising results were obtained, this is not yet ready to be implemented in EHR systems.

Although not directly actionable, the fact that many patients in the study in Chapter 6 often did not recognise that an action was needed may be a relevant finding for policy makers. In fact, this raises patient safety concerns and reinforces the idea that the presentation of laboratory test results, in terms of interface design alone (even when abnormal values are highlighted with colours or grouped together), is not enough to improve the accuracy of patients' risk interpretation [16]. This might also limit the potential of patient portals to actively involve patients in their own healthcare in clinical practice.

So, the main context in which we produced actionable information is research. On one hand, in Chapter 5 we reviewed the literature and found that the influence of patient portals on decision-making is rarely investigated. This aspect is particularly relevant when patients access information outside clinical consultations because misinterpretation and erroneous decisions can have adverse effects on patient safety and self-management [17–19]. Therefore, influence on decision making should be one of the priorities in patient portal research, despite this aspect being particularly difficult to investigate. On the other hand, the method developed as part of Chapter 7 represents a methodological improvement in prediction modelling, as it takes advantage of the vast amount of longitudinal information available in EHRs (often disregarded in prediction modelling [8]).

8.2 FUTURE DIRECTIONS

There are several ways how the work we presented could be taken forward. Below, we report some future directions on the topics considered in this PhD thesis.

- **RQ1:** The survival model proposed in Chapter 3 represents an improvement to the current approaches to investigate multimorbidity with EHR data. However, this is only a first step towards fully taking advantage of longitudinal information available for patients in EHRs. In addition to testing our models on different outcomes (e.g. disability, function loss, hospitalisation and resource utilisation), future research should investigate new scores that account for the interactions between different comorbidities. Particularly, such scores could be developed by using large EHR databases like the Clinical Practice Research Datalink to find clusters of patients with common comorbidities and evaluate how different groups of clinical conditions might lead to different health outcomes. Potentially, this could be integrated with life style information (e.g. smoking, body mass index and alcohol consumption) to obtain an estimate of biological age, as opposed to demographic age, with a similar approach to Framingham Hearth Age estimator [20].

- **RQ2:** There are two main directions for future research building on our findings. First, after having demonstrated that, from a statistical perspective, the QKidney model could be used in clinical practice, future studies should focus on evaluating the influence that the model would have on GPs' decision making. Second, as we showed that all predictive models we tested had a substantial decrease in performance with patients with established CKD risk factors, future research should evaluate whether more complex models might be more suitable for these patients. For example, it would be interesting to explore the inclusion of longitudinal information about important CKD markers (i.e. serum creatinine or eGFR) to identify different trajectories towards CKD onset.

- **RQ3:** Although our work provided important findings on how patients interpret and act on laboratory test results, the main limitation of the study presented in Chapter 6 is that patients did not access their own information but rather fictitious scenarios. This might have been one of the reasons why we did not observe any effect of presentation on patients' interpretation. Future studies should therefore focus on evaluating this aspect in a more realistic scenario, with patients accessing their own laboratory test results. This could be done by randomising them to different presentations of the laboratory test results and monitoring their actions remotely. Furthermore, the method to develop patient-tailored reference ranges described in Chapter 7 needs further development and testing before it can be made available to patients. Particularly, after evaluating whether the current method is generalisable to other laboratory parameters (e.g. serum creatinine or haemoglobin), it would be interesting to test its performance in clinical practice by integrating patient-tailored alerts in EHRs and assessing their impact on GPs' behaviours (i.e. alert override versus clinical action taken).

8.3 CONCLUSIONS

In this thesis, we used routinely collected EHR data to investigate decision-making in different contexts and involving different stakeholders. These included patients, clinicians and policy makers, as well as EHR developers and researchers. Ultimately, we produced actionable information that spans across health research and population health management. For health research, this thesis provided important methodological advances in predictive modelling using EHR data, with the development of innovative ways to investigate multimorbidity (Chapter 3) and to contextualise reference ranges for laboratory test results (Chapter 7). For population health management, we demonstrated that QKidney is fit for use in clinical practice. Policy makers should consider to integrate this in model in clinical practice guidelines to support timely identification of patients at a higher risk of developing CKD.

8.4 REFERENCES

1. Taylor AW, Price K, Gill TK, Adams R, Pilkington R, et al. (2010) Multimorbidity - not just an older person's issue. Results from an Australian biomedical study. BMC Public Health 10: 718. Available: http://www.webcitation.org/6NYkN8vmX. Accessed 17 January 2014.

2. Anderson G (2010) Chronic Care: Making the Case for Ongoing Care. Robert Wood Johnson Foundation. Available: http://www.webcitation.org/6NYkTSQ5Z.

3. Salisbury C, Johnson L, Purdy S, Valderas JM, Montgomery AA (2011) Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. Br J Gen Pract 61: e12-21. Available: http://www.webcitation.org/6NYkZNkgB. Accessed 17 January 2014.

4. Wang CY, Baldwin L-M, Saver BG, Dobie SA, Green PK, et al. (2009) The Contribution of Longitudinal Comorbidity Measurements to Survival Analysis. Med Care 47: 813–821. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2701975/.

5. Zeng C, Ellis JL, Steiner JF, Shoup JA, McQuillan DB, et al. (2014) Assessment of Morbidity Over Time in Predicting Health Outcomes. Med Care 52. Available: http://journals.lww.com/lww-medicalcare/Fulltext/2014/03001/Assessment_of_Morbidity_Over_Time_in_Predicting.11.aspx.

6. Strauss VY, Jones PW, Kadam UT, Jordan KP (2014) Distinct trajectories of multimorbidity in primary care were identified using latent class growth analysis. J Clin Epidemiol 67: 1163–1171. Available: http://www.sciencedirect.com/science/article/pii/S0895435614002182.

7. Charlson ME, Pompei P, Ales K, Mackenzie C (1987) A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. J Chronic Dis 40: 373–383. doi:10.1016/0021-9681(87)90171-8.

8. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA (2016) Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Informatics Assoc. Available: http://jamia.oxfordjournals.org/content/early/2016/05/17/jamia.ocw042.abstract.

9. Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA (2015) External validation of new

risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol 68: 25–34. Available: http://www.sciencedirect.com/science/article/pii/S0895435614003539.

10. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, et al. (2014) External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol 14: 40. Available: http://www.biomedcentral.com/1471-2288/14/40. Accessed 10 May 2014.

11. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, et al. (2016) External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ 353. Available: http://www.bmj.com/content/353/bmj.i3140.abstract.

12. Collins GS, Reitsma JB, Altman DG, Moons KGM (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med 13: 1–10. Available: http://dx.doi.org/10.1186/s12916-014-0241-z.

13. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, et al. (2015) Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and ElaborationThe TRIPOD Statement: Explanation and Elaboration. Ann Intern Med 162: W1–W73. Available: +.

14. Hippisley-Cox J, Coupland C (2010) Predicting the risk of chronic Kidney Disease in men and women in England and Wales: prospective derivation and external validation of the QKidney Scores. BMC Fam Pract 11: 49. Available: http://www.biomedcentral.com/1471-2296/11/49. Accessed 8 May 2014.

15. Bartlett C, Simpson K, Turner AN (2012) Patient access to complex chronic disease records on the Internet. BMC Med Inform Decis Mak 12: 87. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3438097&tool=pmcentrez&rendertype=abstract. Accessed 9 October 2014.

16. O'Kane M, Freedman D, Zikmund-Fisher BJ (2015) Can patients use test results effectively if they have direct access? BMJ Br Med J 350. Available: http://www.bmj.com/content/350/bmj.h673.abstract.

17. Kruse CS, Argueta DA, Lopez L, Nair A (2015) Patient and Provider Attitudes Toward the Use of Patient Portals for the Management of Chronic Disease: A Systematic Review.

J Med Internet Res 17: e40. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4376181/.

18. Fagerlin A, Zikmund-Fisher BJ, Ubel PA (2011) Helping Patients Decide: Ten Steps to Better Risk Communication. JNCI J Natl Cancer Inst 103: 1436–1443. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3218625/.

19. Young MJ, Scheinberg E, Bursztajn H (2014) Direct-to-patient laboratory test reporting: balancing access with effective clinical communication. JAMA 312: 127–128. Available: http://jama.jamanetwork.com/article.aspx?articleid=1882585. Accessed 21 October 2014.

20. Framingham Heart Study (n.d.) Framigham Heart Age Calculator. Available: https://www.framinghamheartstudy.org/risk-functions/cardiovascular-disease/general-cvd-risk-prediction-using-bmi.php#.