SIKS Dissertation Series No. 2016-02
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Cover design: Het Nieuwe Werk
ISBN 978-90-393-6472-7
© 2015 Michiel Meulendijk

Optimizing medication reviews through decision support: prescribing a better pill to swallow

Medicatiebeoordelingen optimaliseren door middel van beslissingsondersteuning: een betere pil slikken

(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op woensdag 13 januari 2016 des middags te 2.30 uur
door

Michiel Christiaan Meulendijk

geboren op 5 april 1985 te Capelle aan den IJssel

Promotor:	Prof. dr. S. Brinkkemper
	Prof. dr. M.E. Numans
Copromotor:	Dr. M.R. Spruit
	Dr. P.A.F. Jansen

This thesis was (partly) accomplished with financial support from Ephor (Expertisecentrum PHarmacotherapie bij OudeRen).

Contents

Preface

While pursuing a PhD is an individual endeavor by design, it is by no means a lonely one. During the time I worked at it, I encountered many people who – in their own various ways – contributed to the dissertation I proudly present here.

Sjaak, thank you for your confidence in me. Despite my often erratic working hours, you (usually) let me stick to my own plan and kept faith in a good outcome. I hope this dissertation does not disappoint you.

Marco, you were an approachable and enthusiastic supervisor. Especially in my first year we spent a lot of time together on the road, speaking to anyone remotely associated with medical informatics. I have fond memories of those trips.

Paul, without you the project would never have been initiated in the first place. Thank you for your unrelenting involvement with the software and my work.

Mattijs, during the project's nascence you provided valuable practical insight, which kept us firmly rooted in reality and kept this project from becoming just another academic interest. Thank you.

Clara, when I started out I didn't know a thing about the medical domain. Without your patient answering of my never-ending questions I still wouldn't. Thanks for that.

Floor, Dominique, Petra, and Jacqueline, thank you for the pleasant cooperation in our joined research effort. Putting my software in practice gave me much needed insight into its use and improved it tenfold.

Wilma, Ingeborg, Lianne, Ian and all other people involved with the OPERAM-project, thank you for continuing the work and ensuring the software we are building benefits real people!

Ravi and Amir, we spent some four years together in a room. It does, however, feel like a lifetime – and I mean that in the nicest way possible. We celebrated paper acceptances together and pulled each other through rejections. I am very happy to count you as my friends. Our other roommates who came and went, Wienand, Ivon, Marijn, Rogier, Cristobal, thanks for your company and moral support. The same goes for the guys next door, the inseparable Jaap and

Kevin, and all other PhD candidates in our group, amongst others Garm, Vincent, Erik, Sjors, Jan, and Benny.

I have experienced our research group as a welcoming, creative environment, where people are approachable and always willing to share their thoughts on your latest idea. Remko, Slinger, Sandor, Marjan, Fabiano, Floris, Jan-Martijn, Sergio, and the others who have come and gone, thank you for your collegial input and the great atmosphere.

I feel grateful to have many people in my life who force me to get away from my computer now and then. My dear friends, Joost, Tim, Liedeke, Tjing, former housemates, Amber, Ariane, Eloy, Hendrik, Ilse, Jeroen, Kathrine, Kevin, Merel, Mirte, Rydienne, Stephanie, the Utrecht crowd, Alexander, Thijs, Luuk, Nicole, Joost, my fellow PhD friends, Maya, Chris, Laurens, and countless others I've shared a beer with over the years, thanks for all the fun we've had!

Finally, I wholeheartedly thank my family for their unconditional support. Mom, dad, Pieter, and Peter, thank you for always being there, in good times and bad.

Michiel

PART I

General Introduction

Perfect nonsense goes on in the world. Sometimes there is no plausibility at all.

Nikolai Gogol, *The Nose*

1.1

Introduction

MOTIVATION

Medication Reviews for Polypharmacy Optimization

The growth of the elderly population in the past and coming decennia presents great challenges for public health. With increasing age comes a greater incidental or continuous dependency on medicine; almost half of all drug expenditures are consumed by elderly people, and of chronically ill people in the Netherlands that use five or more drugs simultaneously, half are over sixty-five years of age [110,198].

Growing numbers of elderly have more than one chronic disease or medical condition, treated with a multitude of drugs [110]. The chronic use of five or more medicinal drugs, known as polypharmacy, has been shown to have detrimental effects on patients' health [88]. In the Netherlands, 5.6% of all acute hospital admissions have medication-related causes; for admissions involving elderly people, this figure is twice as high [129]. Polypharmacy has been associated with increased risk of mortality and (co)morbidities [88].

Polypharmacy has been associated with inappropriate prescribing [88]. Suboptimal prescribing often finds its origin in secondary physicians prescribing medicine without knowledge of potentially incompatible treatments already used by patients, or in primary

care practitioners lacking the resources to review medication. Both underprescribing – the non-treatment of diseases – and overtreatment – the prescription of unnecessary medicine – have been identified as causes of adverse health events associated with polypharmacy [122,197]. Additionally, patients' adherence to their therapy decreases as the number of daily dosages increases [41]. With older patients, it has been shown that the complexity of their dosage regimens, as well as patient-specific characteristics such as health beliefs and cognitive ability, are factors influential in their drug adherence [77].

Insofar as these medication-related issues are caused by inappropriate prescribing, they can potentially be avoided by general practitioners (GPs) and pharmacists by improving the prescription process. Their large workloads, substandard IT systems, and suboptimal interdisciplinary communication, however, do not enable them to structurally improve their prescriptions. Errors resulting from this include use of incomplete patient information, insufficient communication, and mistakes because of time pressure or carelessness [111,214].

Recognizing this problem, several initiatives have been employed to reduce polypharmacy problems and improve prescriptions. Among these are the Beers criteria, the Medication Appropriateness Index (MAI), and the Geriatric Evaluation and Management (GEM) method. The effectiveness of these interventions varies; generally they appear beneficial in terms of reducing inappropriate prescribing and medication-related problems, but they have not been proven to lead to clinically significant improvement [167].

In order to significantly improve prescribing, several all-encompassing methods have been developed to address all polypharmacy problems in a systematic medication review approach. The Polypharmacy Optimization Method (POM) was designed as a "useful and rapid method for optimizing prescribing of polypharmacy in general practice" [57]; it has been shown to significantly improve GPs' prescriptions for polypharmacy patients in an experimental setting. A pharmacist-led pharmacotherapy review focusing on the use, effectiveness, indication, and safety of medicine, yielded no significant decrease in hospital admissions, but may significantly reduce medication-related hospitalizations in patients with five or more comorbidities [128]. The START- and STOPP-criteria together form a set of screening tools that warn GPs about possible underprescribing or overtreatment [75]. A randomized controlled trial investigating their impact on inappropriate prescribing showed significant improvements in prescribing appropriateness [74].

In an attempt to create a multidisciplinary all-encompassing guideline, these tools have been combined in the Systematic Tool to Reduce Inappropriate Prescribing (STRIP), which has been included in a Dutch national guideline for polypharmacy [62]. The STRIP method contains a pharmacotherapeutical analysis that checks for underprescribing, overtreatment, recommended dosage adjustments, drug effectiveness, potential adverse effects, dose frequency, clinical interactions, and medication adherence, including practical problems with medication use. To gain all stakeholders' expertise and opinions, the structured medication review is designed to be used in cooperation by GPs, pharmacists, and patients.

Medication Reviews in Primary Care

In the Netherlands and most Western-European countries, primary care serves as the principal point of continuing care for patients in a health care system. Traditionally, GPs diagnose patients and prescribe treatment, while pharmacists manage medicine dispensation and instruct on its proper use. In the last decades, however, these roles have become more intertwined, with pharmacists getting more responsibilities in the process. Especially when concerned with frail and elderly patients, pharmacists tend to play larger roles in the primary care process; older patients, practitioners, and pharmacists all value greater pharmacist participation in patient care [201].

In many medication review methods, pharmacists play pivotal roles [167]. Indeed, their involvement in performing medication reviews leads to improved pharmacotherapy for older patients, especially when they work in the context of a multidisciplinary team [194]. The Dutch multidisciplinary guideline incorporating the STRIP method explicitly states that a medication review is a "pharmacotherapeutical review conducted by patient, practitioner, and pharmacist, based on a periodical, structured, critical evaluation of medical, pharmaceutical and use information" [62].

Until now, a variety of barriers has impeded the widespread adoption of structured medication reviews in daily practice. Recently Anderson et al. conducted a systematic literature review on enablers and barriers to minimizing potentially inappropriate medications, a cornerstone of medication reviews. Barriers stopping physicians from discontinuing inappropriate medication include mostly physician-related factors. These are grouped around inertia (his or her attitudes towards discontinuation, such as fearing negative consequences), self-efficacy (his or her knowledge and available information on the topic), and awareness (his or her having poor insight or discrepant beliefs). Barriers that were not physician-related included a lack of resources, patients resisting changes to their medication, and practical and cultural factors [10]. A separate study focusing on barriers regarding pharmacist-led medication reviews reported lack of time and lack of self-confidence as the most commonly perceived barriers [157].

Among the enablers identified by Anderson et al. to improve physicians' willingness to refine patients' medication are adequate resources, integration of the process in daily practice, and access to appropriate decision support [10].

Decision Support Systems in Primary Care

Decision support systems can be found in many clinical environments. Designed to help healthcare professionals make clinical decisions, they play important roles in contemporary clinical processes. Musen, Shahar, and Shortliffe distinguish three types of clinical decision support systems, of which tools that give patient-specific recommendations form the most sophisticated category. Typically, these systems "provide custom-tailored assessments or advice based on sets of patient-specific data", using logic derived from decision theory on contextual

datasets. These systems' results may vary from assistance in diagnostics to recommendations on therapy choice [149].

The method through which clinical decision support systems arrive at their advice is an important characteristic by which to distinguish them. The simplest logical approaches have used explicit problem-specific flowcharts to generate output based on specified input. Because of their inflexibility to adapt to real-life routine, they have been mostly surpassed by more advanced approaches. Contemporary clinical decision support systems are often knowledge-based: they "symbolically encode concepts derived from experts in a field [... and] use that knowledge base to provide the kind of problem analysis and advice that experts might provide" [149].

Over time, these explicit knowledge base systems are frequently refined into recommender systems. Recommender systems are "software tools and techniques providing suggestions for items to be of use to a user" [176]. They are commonly used in consumer-focused web services to recommend products, but can be employed in clinical settings as well. Recommender systems typically work through either collaborative filtering or content-based filtering. With the former approach, systems make predictions based on similar users' historical data. In the latter approach, content-specific characteristics serve as basis for systems' recommendations. Hybrid recommender systems, combining an explicit knowledge base with content-based or collaborative filtering, have been shown to outperform their simpler counterparts [4].

In literature, there is consensus that clinical decision support has the potential to improve GPs' and pharmacists' decision-making [32]. The extent to which these systems can improve effectiveness while retaining efficiency, however, remains an open question. Insufficient studies, with varying results, have been conducted to make a conclusive statement on clinical decision support systems' efficiency.

Problems with human-computer interaction, and more specifically 'alert fatigue', are often reported as factors influential in these systems' lack of efficiency, or users' lack of satisfaction with them [116,210]. Notable system characteristics associated with these issues include whether the system assumes a consulting or a critiquing role, generates recommendations actively or passively, and to what extent human-computer interaction practices have been taken into account [149].

Utility: Effectiveness, Efficiency, and Satisfaction

Usability has long been regarded as an essential determinant for the successful application of software systems. The International Standards Organization (ISO) defines it as 'the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use' [108]. In this definition, effectiveness is understood as the 'accuracy and completeness with which users achieve specified goals', while efficiency exists of 'resources expended in relation to the accuracy and completeness with

which users achieve goals'. Satisfaction, finally, is the subjective 'degree to which user needs are satisfied when a product or system is used in a specified context of use'.

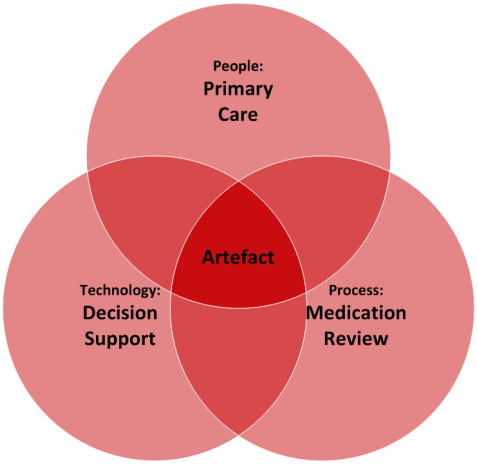
Perceived ease-of-use and perceived usefulness, concepts closely related to usability, have been included in most technology acceptance models as determinants of people's attitudes towards prolonged use of software [51,217,218]. There is a consensus in literature that in the domain of clinical decision support systems, usability has a significant impact on users' adoption behavior [47].

Effectiveness and efficiency are considered to be positively related concepts [101]. In studies measuring the usability of clinical decision support systems, however, results are mixed. While systems typically tend to improve the effectiveness of the process they focus on, there is insufficient evidence to come to a conclusion on their efficiency [32]. A study by Nilsson & Følstad, who claim that effectiveness and efficiency are conflicting requirements, may explain this discrepancy [156].

OBJECTIVE

From the domain-specific discussion described above, it is evident that there is a problem with polypharmacy resulting from inappropriate prescribing. There is ample evidence that structured medication reviews could improve polypharmacy patients' medication, but that there are major barriers impeding the direct implementation of structured medication reviews in general practice.

FIGURE 1: **The three pillars of this research positioned as a people-process-technology relationship.**



Clinical decision support systems have been shown to effectively improve prescribing processes in primary care, leading to the belief that they could be successfully implemented for the conduct of structured medication reviews. Thus, the aim of the research described in this

dissertation is to create a decision support system to facilitate the conduct of structured medication reviews by physicians and pharmacists in primary care.

In Figure 1, the three pillars of this research are positioned in terms of a people-process-technology relationship, where the primary care domain corresponds with the people concept, the medication review methods with the process concept, and the decision support system with the technology concept. Together, they form a decision support system for medication reviews in primary care, of which the usability is a definitive characteristic of the artefact's utility.

Research Perspective

The design science paradigm seeks to extend the boundaries of human and organization capabilities by creating new and innovative artefacts. Hevner et al. introduced a framework for applying design science in the information science domain. Principal in this framework is the design cycle, in which artefacts are built and evaluated in continuous iterations. During these creation iterations, there are continuous evaluative iterations with the knowledge base, containing foundations and methodologies, and the practical environment, consisting of people, organizations, and technology [95].

Peffers et al. proposed a structured process for the conduct of design science research in the information technology domain. With this framework, they aimed at creating "a conceptual model for how researchers can carry out design science research in IS and a mental model or template for readers and reviewers to recognize and evaluate it" [168]. In Figure 2 the focal points of the research described in this dissertation are positioned in their design science research process.

Peffers et al. describe how research can be initiated from any one of the first four steps of the design science research process [168]. As the problem domain in which this study was conducted was well documented, as has been described in the motivation earlier, pursuing a problem-centered approach was unsuitable. The consensus in literature on the benefits of clinical decision support systems made the choice for computerized decision support self-evident. The abundance of documentation and best practices on designing decision support systems made a design- and development-centered approach similarly unsuitable. Finally, the lack of an implemented solution made a solution-centered approach unfeasible. Thus, the research pursued in this study was positioned as an objective-centered solution.

Based on the well-documented problem domain, the intended solution's requirements were determined. Then the system was designed according to best practices and consensus in literature. A prototype was demonstrated in both experimental and real-life environments, the data of which was used to evaluate the system's usability. Insights gained from the evaluation were used to revisit the project's objectives and design decisions, and thus start a new iteration of the process. Throughout the research project, results were communicated through papers in scholarly publications.

Scientific Relevance

The relationship in design science between theory and practice is a complicated, but essential one. One of the most influential authors on design science in information science, Hevner, noted that theory and (practical) utility form a symbiont relationship: "Truth informs design and utility informs theory" and "scientific research should be evaluated in light of its practical implications" [95]. Whether or not theorizing itself is part of design science has been disputed. Some authors observe a strict division between theory and utility; March and Smith, for example, regard theories as "deep, principled explanations of phenomena" which are the domain of natural science. In their view, artefacts resulting from design science are merely evaluated for their utility in design science. They do assert, however, that the evaluated artefacts do retain knowledge in the form of models or instantiations [134].

In a paper on the role of theory in design science, Venable ponders how such tacit knowledge contained in artefacts can be evaluated by researchers, and how practitioners are guided in application of the knowledge. He proposes the use of utility theories for theorizing in design science research: "A utility theory makes an assertion that a particular type or class of technology [...] has (some level of) utility (or usefulness) in solving or improving a problematic situation (with specified characteristics)". A utility theory consists of a problem space, a solution space, and the nature of the utility that links them, all of which should be thoroughly explicated. The artefact's utility should usually be expressed in terms of its efficacy, effectiveness, or efficiency, in metrics that allow for comparison with similar solutions in the domain. Venable stresses disclosure of knowledge in a manner understandable to both researchers and practitioners [215].

In this dissertation the artefact that is created and evaluated is situated in the intersection between primary care, medication reviews, and clinical decision support systems. The accompanying utility theory regards polypharmacy as its problem space, medication reviews as its solution space, and the envisioned decision support system as the linking artefact, the utility of which is expressed in terms of its usability, i.e. its effectiveness, efficiency, and satisfaction. Usability, or its sub concepts, have been used extensively in both the decision support domain and the medication review domain, and as such makes for a dependable, comparable metric. The chapters of this dissertation contain findings relevant to the artefact and its utility, which have all been published in scholarly publications to make them available to the scientific and practical community.

Both the problem and the solution domains have been described in detail in the section on motivation earlier in this chapter. The problem of polypharmacy and its consequences are well-documented, as are the benefits medication reviews offer. While decision support systems have been developed for a wide variety of topics in primary care with varying success, no context-aware decision support system facilitating structured medication reviews has been documented or evaluated in literature.

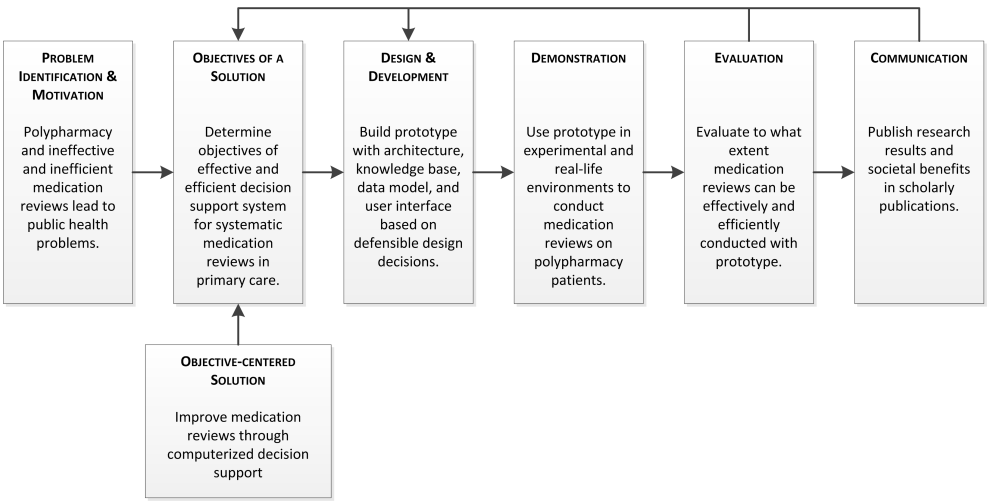
Societal Relevance

As has been extensively described in the motivation earlier, polypharmacy amongst elderly people is a growing problem leading to major challenges in public health [110]. To address this issue, various initiatives have been employed in primary care. Among these are structured methods for the conduct of medication reviews, such as the POM and the STRIP [57,62]. Decision support systems are proven solutions for effectively facilitating clinical processes in primary care [32]. A decision support system facilitating structured medication reviews in primary care has the potential to significantly improve the health of polypharmacy patients and decrease medicine costs.

The 2009 study validating the POM showed that the method for conducting structured medication reviews significantly improved GPs' prescriptions. They also prescribed fewer medicines per patient [57].

A decrease in the use of the number of medicines can be expected to lead to a decrease in medication-related hospitalizations [129]. Problems associated with polypharmacy, such as an increased risk of mortality and (co)morbidities, are likely to decrease as well [88]. Finally, a lower absolute number of dispensed drugs will lead to lower medicine costs, lightening the burden of public health on society.

FIGURE 2: **The research perspective positioned in the design science research process.**



RESEARCH APPROACH

Research Questions

In line with the objective and the research perspective described above, this dissertation will explore the utility of a decision support system facilitating the conduct of medication reviews

in primary care. The utility of this artefact will be expressed in terms of usability, leading to the following main research question:

> MRQ: How can clinical decision support systems contribute to effective and efficient medication reviews in primary care?

This main question has been divided in a number of subquestions listed below. The first three subquestions correspond with the objectives, design & development, and demonstration / validation phases of the design science research process. The fourth research question revisits the solution's objectives and design choices and can be regarded as a new iteration of the process.

Objectives

RQ1: What objectives can be identified for clinical decision support systems facilitating medication reviews in primary care from the perspectives of its stakeholders?

The problems associated with polypharmacy are well-documented, as has been described in the motivation above. In line with the design science research process, the first phase of an objective-centered solution is the identification of objectives for the artefact.

RQ1.1: How are physicians in the Dutch primary care sector, based on their experiences with decision support systems and medical formularies, likely to respond to the introduction of a clinical decision support system facilitating medication reviews?

Clinical decision support systems could play a major role in assisting physicians with performing medication reviews. In this chapter the attitudes of physicians towards the use of decision support systems are investigated. A survey was distributed amongst 500 Dutch GPs, exploring their experiences with polypharmacy, their attitudes towards current decision support systems, and their opinions towards a possible new one aiding them with medication reviews.

RQ1.2: Which non-functional requirements of medical mobile applications are deemed most important by (potential) polypharmacy patients?

The increased use of internet through smartphones and tablets enables the development of new consumer-focused mobile applications (apps) in health care, such as apps aiding users with their medication management. Concerns including these apps' safety, usability, privacy, and dependability have been raised. In this chapter the non-functional requirements of medical apps that potential users view as most important are explored through a grounded theory-approach. To this end, a document study and interviews with stakeholders are conducted, after which the discovered requirements are evaluated with potential users through a vignette study.

RQ1.3: How can the feasibility of a clinical decision support system facilitating medication reviews in primary care be investigated?

Problems with the introduction of software applications in practice may lead to the disuse of an otherwise viable product. It is common practice for researchers and entrepreneurs alike to investigate the risks that may threaten projects' ™ feasibility beforehand. However, little standardization exists regarding actual approaches. In this chapter the practice of conducting a feasibility analysis of a decision support system in the Dutch primary care sector is investigated.

Design & Development

RQ2: **What design decisions lead to the development of clinical decision support systems that effectively and efficiently facilitate medication reviews in primary care?**

With objectives having been identified, the next phase in the design science research process encompasses the design and development of the artefact. In this project, the artefact has been based on the Systematic Tool to Reduce Inappropriate Prescribing (STRIP), a structured method to perform medication reviews. To facilitate physicians' use of the STRIP method, the STRIP Assistant (STRIPA) has been developed after the objectives identified in the previous chapters. STRIPA is a stand-alone web-based decision support system that advices physicians during the pharmacotherapeutic analysis of patients' health records. In this chapter the application's architecture and rule engine, and the design decisions relating to the user interface and semantic interoperability, are described.

Demonstration & Validation

RQ3: **To what extent can clinical decision support systems contribute to effective and efficient medication reviews in primary care?**

After envisioning and developing the artefact, the design science research process emphasizes demonstrating and evaluating it in practice. The decision support system created in this project has been evaluated in both a controlled experiment and in its real-life environment.

RQ3.1: **Does a clinical decision support system significantly improve caretakers' performance when conducting medication reviews of polypharmacy patients' health records?**

In this study STRIPA's usability as a tool for physicians optimizing medical records of polypharmacy patients is validated. In an online experiment, forty-two caretakers were asked to optimize two comparable medical records of polypharmacy patients, one in their usual manner and one using STRIPA. Changes in effectiveness were measured by comparing respondents' optimized medicine prescriptions with medication prepared by an expert panel of two geriatrician-pharmacologists. Changes in efficiency were investigated by recording the time respondents took to optimize the two cases. Users' satisfaction with the software was explored with the System Usability Scale.

RQ3.2: Does the time caretakers use to optimize the medical records of polypharmacy patients with a clinical decision support system decrease over time?

The previous study's results showed improvements in effectiveness when performing medication reviews with the STRIPA decision support system, but decreases in efficiency. This lack of efficiency may be explained by the study's single-test experimental method, which does not take into account experience participants gain over time. This chapter documents a study to determine if having a group of caretakers perform decision supported structured medication reviews over a longer period of time will lead to improvements in efficiency. Four expert teams consisting of a physician and a pharmacist conducted structured medication reviews on patients in 13 general practices located in Amsterdam, the Netherlands. The time they needed to perform medication reviews was recorded over a period of thirteen months.

Refined Objectives

RQ4: What objectives can be identified for the integration and optimization of clinical decision support systems facilitating medication reviews in primary care?

The evaluation of the artefact leads to insights into its weaknesses and presents opportunities for refinement. This leads to changes of the solution's objectives and its development decisions, thus starting a new iteration of the design science research process.

RQ4.1: To what extent can patient data be meaningfully exchanged between international primary care terminologies?

The diversity of terminologies used in primary care causes significant challenges regarding semantic interoperability. Attempts to address these challenges usually focus on the creation of metaterminologies, with the peculiarities of national variations of terminologies being overlooked. In this chapter the extent to which primary care data can be meaningfully exchanged between nationally implemented terminologies is assessed. A model comprising primary care terminologies and including axioms to define their relations was developed. Generic metrics were designed to determine the completeness and accuracy of any two arbitrary vocabularies within an ontological model. These metrics were used on an implementation of the model to determine the data quality that is preserved when expressing similar data in different primary care terminologies.

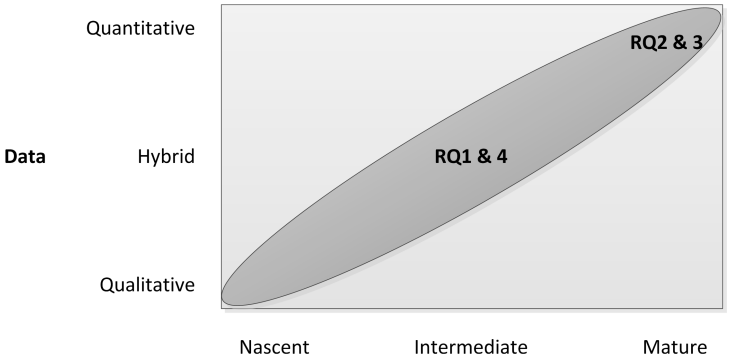
RQ4.2: Does risk incorporation in association rule mining have predictive power?

Association rule mining is one of the most prominent knowledge discovery methods in existence. The application of discovered association rules in a sensitive domain with potentially far-reaching implications, however, can be risky. In this chapter the concept of risk in association rule mining is introduced. A model for the incorporation of risk in association rules is proposed and validated in a primary

care setting. After validation, scenarios in which the combination of association rule mining and risk assessment yields useful results in practice are presented.

Research Methods

FIGURE 3: **The research questions of this dissertation positioned in Edmondson and McManus' methodological fit framework.**



The information science domain has been known to utilize a wide variety of research methods, which have their roots in different philosophies. Especially the approach of design science allows for the use of methods from both traditional positivist philosophy, and subjective-driven interpretivist philosophy. Edmondson and McManus, in their paper on methodological fit in management field research, propose a framework aiding in the decision for qualitative or quantitative methods [63]. They distinguish between nascent, mature, and intermediate theories. Using inappropriate methods leads to low methodological fit; using exploratory qualitative methods in a mature field results in reinventing the wheel, whereas using evaluative quantitative methods in a nascent field results in proposed constructs lacking validity.

Using the terminology provided by the framework of Edmondson and McManus, most research in this dissertation can be characterized as mature, as the fields composing the niche described earlier – i.e. medication reviews in primary care, decision support, and usability – are all well-established. Mature research, according to Edmondson and McManus, is exemplified by using formal hypotheses, testing existing constructs, resulting in quantitative data, ultimately leading to supported theory adding specificity to existing theories. Research questions 2 and 3 in this dissertation, and their subquestions, exemplify mature research. Positivist methods, such as experiments, were employed to test hypotheses for evaluative purposes during these phases.

The earlier steps in the design science research process, discovering problem domains and eliciting requirements, are of a more exploratory nature. This intermediate research is exemplified by relating new constructs to existing ones through the gathering and analysis of hybrid data, ultimately leading to new theoretical constructs firmly built on previous work. In

this dissertation, research questions 1 and 2, and their subquestions, investigate intermediate theory through hybrid approaches. Methods employed in these phases were mostly interpretivist, such as grounded theory or action research, though positivist methods such as surveys were used as well.

Figure 3 shows how the research questions in this dissertation are positioned in Edmondson and McManus' framework, taking into account the maturity of the relevant theory and the required type of data.

Survey Survey research is a method common in social science that has been widely adopted in the information science domain. Being a standardized measure to assess people's thoughts, opinions, and feelings, it is widely used as an exploratory or evaluative method [189]. In information science, "surveys are conducted when the use of a technique or tool already has taken place or before it is introduced" [228].

Surveys can be used in a descriptive, explanatory, or explorative manner; they are conducted to enable assertions about a population, make explanatory claims about a population, or are used as a preliminary study of a more thorough investigation [228]. Within the design science paradigm, surveys are most suitable in the objectives-gathering and evaluation phases of the design science research process.

Grounded Theory Glaser and Strauss described grounded theory as the discovery of theory from data [80]. The method was created "as a protest against what they viewed as a rather passive acceptance that all the great theories had been discovered" (Charmaz as cited in Goulding [83]). The remaining role of research was to test existing theories, not to propose new ones. Grounded theory broke with this paradigm by introducing a method to create new theory.

According to Birks, Fernandez, Levina, and Nasirin [23], grounded theory can be a powerful tool for IS scholars interested in theory development, allowing researchers to conduct pioneering research with both flexibility and rigor. Several characteristics of grounded theory make it perfectly compatible with a design science perspective. In grounded theory, data collection and analysis are interrelated processes, meaning that they can and have to happen simultaneously. Concepts are the basic unit of analysis, and have to be constantly compared to each other to form interrelationships [42]. This continuous feedback loop between data collection and analysis is compatible with the iterative design science research process.

Action Research Action research is a paradigm that has comes forth from the postmodern school of thought. It has been defined in an early 1950s paper as the "diagnosis of a social problem with a view of helping improve the situation" [25]. This definition highlights two distinctive characteristics of the paradigm that demonstrate its suitability for application in information systems research.

Firstly, following an interpretivist line of reasoning, action research is based on the assumption that "complex social interactions cannot be reduced for meaningful study" [18]. Human organizations can only be meaningfully studied within their contexts, not by isolating their parts [38]. Secondly, when the researcher then intervenes into the research setting, he or she

becomes part of the study, resulting in "the realignment of the roles of researcher and subject into more collaborative and synergistic forms" [18].

From a design science perspective, in which researchers actively intervene in the problem domain with a view of improving the situation, action research is a suitable paradigm for exploring processes in which the researchers' and actors' or developers' roles overlap.

Experiment One of the research methods essential to the scientific method is the experiment. In an experiment, the influence of one or more dependent variables on one or more independent variables is measured in a controlled environment, thereby validating or refuting a hypothesis [37]. Its main advantage over embedded methods, such as case studies, is its controlled environment, through which the unmeasured, possible influence of other variables than the dependent ones can be minimized. Precisely this aspect also highlights its main weakness: hypotheses that are verified in an experimental setting may be refuted in practice, because of the influence of variables not present in the controlled environment of the experiment.

In information science, two subdomains in which experiments are commonly used can be distinguished. Firstly, the traditional experiment involving human subjects is frequently employed to analyze human behavior when confronted with information systems. Secondly, empirical software engineering employs experiments on information systems to analyze its processes or resources [228].

In a design science perspective, traditional experiments involving human subjects fit particularly well in the demonstration and evaluation phases of the process. Even though design science emphasizes embedding in real-life processes, they can be employed to preliminarily assess the improvement of the developed artefact in a controlled environment. The empirical software engineering type of experiment can be employed in the later phases of the design science research process as well, or in the first phases of a new iteration, when a prototype can be analyzed.

Case Study Case studies are "analyses of persons, events, decisions, periods, projects, policies, institutions, or other systems that are studied holistically by one or more methods" [202]. A main advantage of conducting case studies is that phenomena can be studied in their natural environments, thereby taking into account contextual factors. The wide applicability of case studies has led to them being used in many disparate domains, including social sciences, psychology, and information science.

Yin distinguishes four types of case studies, each suited to their own kind of research questions. Holistic case studies, while ideal for the examination of phenomena in their environments, are unsuited for testing hypotheses on specific aspects of cases. Embedded case studies, in contrast, are suitable for examining isolated units of analyses embedded within a case [239].

Both the holistic and embedded case study approaches are suitable for inclusion in a research design based on a design science perspective. Understanding a problem's specifics,

determining a solution's objectives, and demonstrating and evaluating an artefact in its natural context are tasks that can be performed rigorously through case study research.

Validity

Ensuring its validity is an essential requirement for the successful conduct of any research project. Determining what constitutes validity in research, however, has been the topic of much debate for decades. Traditionally, positivist research with its quantitative methods has been assessed by a set of concepts, which have been operationalized in practical guidelines. However, since the introduction of antipositivist research approaches with their qualitative methods, the degree to which these measures can be put to use has been the topic of much debate. Rolfe summarizes the positions as follows: "those writers who wish qualitative research to be judged according to the same criteria as quantitative research; those who believe that a different set of criteria is required; and those who question the appropriateness of any predetermined criteria for judging qualitative research". He stresses that the commonly perceived dichotomy between quantitative and qualitative research be viewed as a continuum instead [177]. Recognizing that each study is unique and may thus require its own set of fitting quality constraints, below is described how the research in this dissertation addresses validity measures commonly found in the positivist approach on one hand, and the interpretivist approach on the other.

Positivist Criteria The concept of validity has been formulated by Kelley in 1927 who stated that a test is valid if "it measures what it claims to measure" [115]. Since then, the definition has been explained in terms of the commonly used internal and external validity concepts.

Internal validity refers to "the truth value that can be assigned to the conclusion that a cause-effect relationship between an independent variable and a dependent variable has been established within the context of the particular research setting" [31]. It other words, it guarantees that a certain independent variable has been influenced by the modified dependent variable, not by anything else. Threats to internal validity include homogeneity in the participant groups (e.g. because of a selection bias or the effect of repeated testing), changes in the instrument used to conduct the study, and confounding (i.e. changes in the dependent variable may be caused by or attributed to a third variable which correlates with both the dependent and the independent variables). Construct validity controls for internal validity by ensuring the constructs used in a study's instrument refer to the theoretical constructs from which the hypotheses were posited [31].

External validity refers to "the generalizability of the causal finding", or the extent to which a study's results can be generalized to other situations and other people [31]. Threats to external validity include all situational specifics (e.g. time or location) and, again, homogeneity of the participants (e.g. the sample may share common features that impede generalizability).

Interpretivist Criteria To what extent traditional validity concepts can be meaningfully applied to interpretivist research is a matter of debate [177]. Instead of evaluating their process with a strict set of validity concepts, interpretivists aim to prove that their claims to the

knowledge they have acquired through their studies are defensible. Peers should be able to arrive at similar conclusions by examining the collected evidence, copying the research process, and reproducing the context in which the research was conducted [178].

Angen reinforces the notion that, for qualitative researchers, "reaching the desired goal and meeting the requirement of trustworthiness become[s] particularly problematic due to the considerable debate about what it means to do valid research in the field of qualitative inquiry". Nonetheless, she synthesizes some criteria that can be used to evaluate research from an interpretivist perspective [11]. These include awareness and articulation of the choices made by the researcher, a written account detailing persuasive arguments, and criteria corresponding to what she calls substantive validity: documenting the evidence from which conclusions are drawn and assessing the biases inherent in the work over the lifespan of a project.

Validity in Methods This dichotomy between positivist and interpretivist notions of validity has been recognized within the communities employing the research methods described above. All communities have addressed the issue and produced guidelines for assessing validity in their particular fields.

When conducting surveys in the form of structured questionnaires, positivist notions of validity are commonly used [228]. Threats to the validity of surveys that are frequently encountered are, a.o, homogeneity of the sample, construct validity of the questions, and respondents' bias towards self-reporting.

In the classic positivist experiment, internal and external validity are concepts of paramount importance. In experiments, "results are said to have adequate validity if they are valid for the population to which we would like to generalize" [228]. Researchers are forced to make continuous trade-offs between feasibility of the conduction and validity of the results. While experiments consisting of pre- and posttests by both intervention and control groups yield the most valid results, resources may be inadequate to achieve this.

Addressing grounded theory, Corbin and Strauss have stated that "it is not appropriate [...] to use criteria ordinarily used to judge the procedures and canons of quantitative studies". They define a double set of criteria, on a study's research process and its empirical grounding of the theoretical findings, to ascertain their validity. They claim that, when a study's research process is thoroughly described, "the presented theory or theoretical formulations can be assessed in terms of degrees of plausibility" [42].

In action research there is extensive discourse on the matter of validity, as well. Using the concept of validity in action research at all has been disputed [172]; Kvale contemplates whether or not it is legitimate to "fit the qualities of action research into a traditional discourse about validity whose concerns have little to do with those of action research?" [126]. Checkland and Holwell argue for emphasizing the study's recoverability by documenting "the thought processes and models which enabled the team to make their interpretations and draw their conclusions" [38].

In case study research, validity "denotes the trustworthiness of the results, and to what extent the results are true and not biased by the researchers' subjective point of view" [228]. Yin proposes a classification scheme based on traditional positivist concepts such as internal and external validity, which are largely similar to those applied in controlled experiments [239].

Table 1 below summarizes the methods used in this dissertation per chapter, and addresses the validity issues relevant for their specific application. Validity sections in the chapters themselves contain more detail, where appropriate.

TABLE 1: **Methods and their validation measures per research question and chapter.**

Chapter	Question	Methods	Validation
2	1.1	Survey	Internal validity followed naturally from the questions' close relations to the theoretical concepts. External validity was ensured by selecting respondents through simple random sampling. Bonferroni-Holm correction was applied to the statistical analysis to correct for multiple comparison analysis.
3	1.2	Grounded Theory	Internal validity was sought by incorporating a mixed-methods design, cross-checking multiple perspectives through top-down document study and bottom-up interviews. Exploratory research was conducted until a point of saturation was reached. All interviews were transcribed and coded. The findings were evaluated by several authors to account for personal bias. The findings of the exploratory methods were evaluated through vignette study. Generalizability is limited due to the relatively small number of participants. This decision was made because of expected difficulties with performing large-scale impersonal quantitative methods with elderly respondents.
4	1.3	Action Research	Validity was sought by determining in advance the study's scope and the fields to be investigated. The authors described in as much detail as possible the thought processes and models employed to make decisions and determine consequent steps.

5	2.1	Action Research	Recoverability was ensured by documenting the thought processes that led authors to make subsequent design decisions. The choices made were explicitly articulated and defended with literature denoting applicable best practices.
6	3.1	Use Experiment	The experiment was conducted in an online setting, improving participation but threatening validity. As the participants were not observed during the experiment, external influence accounting for the changes in results between pre- and posttest cannot be ruled out. Due to the limited time between the two tests this divergence is highly unlikely, though. Likewise, a test effect making participants more adept at conducting medication reviews during the posttest cannot be ruled out. However, as most participants indicated conducting medication reviews on a regular basis, the occurrence of this effect is highly unlikely. The caretakers participated in the experiment voluntarily, which may have consequences for the study's generalizability. Additional research using different means of recruiting participants, such as reimbursement, may be useful to validate its results.
7	3.2	Case Study	Most of the cases were conducted without supervision. Users may have needed time to assess cases' characteristics, which may have influenced the stepwise analysis. The study's generalizability is limited due to the relatively small number of participants.
8	4.1	Software Experiment	Internal validity was ensured by basing the model formulations on literature. Constructs' basis in literature and practice were thoroughly described to ensure the model's recoverability. The logical formalization was reviewed by an expert to ensure its correct application.
9	4.2	Software Experiment	Internal validity was ensured by basing the model formulations on literature, and having them reviewed by expert co-authors. Generalizability of the data with which the model was validated is somewhat limited due to the relatively small number of participants.

DISSERTATION OUTLINE

Chapters 2 to 8 of this dissertation each match exactly one of the subquestions described earlier. The final chapter summarizes their results and discusses the studies' scientific contributions and practical implications. The outline of this dissertation is as follows:

1. Introduction
2. Physicians' attitudes towards decision support systems
   Published as: Meulendijk, M., Spruit, M., Drenth-van Maanen, A., Numans, M., Brinkkemper, S., & Jansen, P. (2013). General practitioners' attitudes towards decision supported prescribing: an analysis of the Dutch primary care sector. Health Informatics Journal, 19(4), 247-263.
3. Patients' attitudes towards medical apps
   Published as: Meulendijk, M., Meulendijks, E., Jansen, P., Numans, M., & Spruit, M. (2014). What concerns users of medical apps? Exploring non-functional requirements of medical mobile applications. Proceedings of the European Conference on Information Systems (ECIS) 2014. Tel Aviv, Israel.
4. Feasibility of decision support systems in primary care
   Published as: Meulendijk, M., Drenth-van Maanen, A., Jansen, P., Brinkkemper, S., Numans, M., & Spruit, M. (2013). Introducing the COrETeSt Feasibility Analysis in Medical Informatics: A Case Study of a Decision support Knowledge System in the Dutch Primary Care Sector. In M. M. Cruz-Cunha, I. M. Miranda, & P. Gonzalves, Handbook of Research on ICTs and Management Systems for Improving Efficiency in Healthcare and Social Care (pp. 1066-1087). IGI Global.
5. STRIPA: A Rule-Based Decision Support System for Medication Reviews in Primary Care
   Published as: Meulendijk, M., Spruit, M., Jansen, P., Numans, M., & Brinkkemper, S. (2015). STRIPA: A Rule-Based Decision Support System for Medication Reviews in Primary Care. ECIS 2015 Research-in-Progress Papers. Paper 29. Münster, Germany.
6. Computerized decision support improves medication review effectiveness
   Published as: Meulendijk, M., Spruit, M., Drenth-van Maanen, A., Numans, M., Brinkkemper, S., Jansen, P., et al. (2015). Computerized decision support improves medication review effectiveness: an experiment evaluating the STRIP Assistant's usability. Drugs & Aging, 32(6), 495-503.
7. Efficiency of clinical decision support systems improves over time
   Submitted as: Meulendijk, M., Spruit, M., Willeboordse, F., Numans, M., Brinkkemper, S., Knol, W., et al. (2015). Efficiency of clinical decision support systems improves over time (submitted).
8. Semantic Interoperability in Primary Care Terminologies
   Submitted as: Meulendijk, M., Spruit, M., Lefebvre, A., & Brinkkemper, S. (2015). To what extent can patient data be meaningfully exchanged between primary care

terminologies? A case study of four Western European classification systems (submitted).

9. Risk Mediation in Association Rules
   Submitted as: Meulendijk, M., Spruit, M., & Brinkkemper, S. (2015). Risk Mediation in Association Rules: The Case of Decision Support in Medication Review (submitted).
10. Conclusion

TERMINOLOGY

Below is an overview of terms and their definitions as they are used throughout this dissertation.

Advice (piece of -)
Context-dependent, non-committal recommendation generated by a decision support system to a caretaker. Syn. recommendation, suggestion.

Caretaker
Generic term indicating a GP or a pharmacist.

Clinical decision support system
Computer-based information systems aiding users with decision-making processes in clinical contexts.

Effectiveness
The accuracy and completeness with which users achieve specified goals [108].

Efficiency
Resources expended in relation to the accuracy and completeness with which users achieve goals [108].

GP (general practitioner)
Medical doctor in primary care. Syn. physician, family doctor.

Pharmacist
Chemist in primary care.

Polypharmacy
The chronic use of five or more medicinal drugs.

POM (Polypharmacy Optimization Method)
Comprehensive method for conducting structured medication reviews in primary care [57].

POMP (POM Platform)
A software platform consisting of, a.o., a decision support system aiding caretakers with conducting medication reviews according to the POM, and a mobile application aiding patients with medication management. In later publications, the POM Platform was replaced by the STRIP Assistant.

Recommender system
Software tool providing suggestions for items to be of use to a user [176]. Syn. recommendation system.

Satisfaction
Degree to which user needs are satisfied when a product or system is used in a specified context of use [108].

START (Screening Tool to Alert doctors to Right Treatment)
Set of criteria warning caretakers of potentially appropriate, indicated drugs [75].

STOPP (Screening Tool of Older Person's Prescriptions)
Set of criteria warning caretakers of potentially inappropriate drugs [75].

STRIP (Systematic Tool to Reduce Inappropriate Prescribing)
Comprehensive method for conducting structured medication reviews in primary care [62].

STRIPA (STRIP Assistant)
A rule-based, stand-alone, web-based, decision support system aiding caretakers with conducting medication reviews according to the STRIP.

SUS (System Usability Scale)
A questionnaire consisting of ten items developed for measuring usability [33].

Usability
The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use [108].

PART II

Objectives Elicitation

What a magnificent body!
How I should like to see it on the dissecting table.

Ivan Turgenev, *Fathers and Sons*

2.1

Physicians' attitudes towards decision support systems

ABSTRACT

The use of multiple drugs by patients increases the risk of medical problems. Clinical decision support could assist general practitioners (GPs) with prescribing, but is underused. This paper aims to investigate the attitudes of GPs towards using decision support systems. A survey was distributed amongst 500 Dutch GPs. Virtually all 184 respondents indicated owning a clinical information system, while only 21% indicated owning a decision support plug-in; this correlated with their use of medical formularies. Only use of one of the

Published as Meulendijk M.C., Spruit M.R., Drenth-van Maanen A.C., Numans M.E., Brinkkemper S., & Jansen P.A.F. (2013). General practitioners' attitudes towards decision-supported prescribing: an analysis of the Dutch primary care sector. *Health Informatics Journal, 19*(4), 247-263.

medical formularies correlated with the number of recognized underprescription problems. GPs' attitudes towards a newly proposed system aiding them with polypharmacy prescribing were mainly positive (57%); the perceived usefulness correlated with output quality (p=.000), time investment (p=.000), and financial stimuli (payability p=.000, reimbursement p=.015), but not with job relevance. Dutch GPs are thus likely to adopt the proposed system, under the conditions that it improves prescription quality and does not require extensive investments of time or money.

INTRODUCTION

In the past decade the development of decision support systems for the primary health care sector has greatly increased. Whether or not integrated with electronic medical records, computerized physician order entry (CPOE) systems have provided general practitioners (GPs) with appropriate tools to facilitate their drug prescriptions. By incorporating medical formularies, drug interaction databases, clinical guidelines and best practices, decision support systems can potentially optimize prescription processes[124,188,9,185].

Even though literature generally attributes prescriptions' quality improvement to decision support systems where applicable, their adoption by GPs is often lacking[56]. Pevnick, Asch, Adams, Mattke, Patel, Ettner et al.[170] report that GPs owning decision support systems use them for only one quarter of their prescriptions. Additionally, McInnes, Saltman & Kidd[137] add that, while virtually all of their adopting respondents indicate using decision support tools for drug prescriptions, only twenty percent of them actually employ them during consultations.

The primary care area of polypharmacy, occupied with the use of multiple medications by a patient, suffers from suboptimal prescribing. In the Netherlands alone, seventeen percent of the chronically ill use more than five different drugs permanently; half of these patients are over seventy years of age[198]. This polypharmacy is associated with medical problems including an increasing risk of adverse effects, under-prescribing, overtreatment, and decreased drug adherence[24,41,73,125,190,192,197,234].

Many of these problems are due to avoidable human errors GPs make during the prescribing process, including use of incomplete patient information, insufficient communication, and mistakes because of time pressure or carelessness[214,183].

To counter these problems, the Prescribing Optimization Method was designed, a step-by-step method to aid GPs in optimizing drug prescriptions. In tests, this method significantly improved their prescriptions' quality and relevance[57]. In order to fully enable the use of the method, the POM Platform (POMP) has been envisioned: a decision support knowledge system that facilitates the POM and is optimally incorporated into GPs' systems and workflows.

Recognizing the aforementioned problems regarding technology adoption by GPs, it is essential to investigate these issues in the Dutch primary care sector, in order to incorporate the results in the realization phases. Therefore the focus of this paper is to explore these issues by assessing the current use of decision support systems and by examining GPs' opinions towards potentially adopting an application assisting them with treating polypharmacy. The research question that will be investigated is the following: how are GPs in the Dutch primary care sector, based on their experiences with decision support systems and medical formularies, likely to respond to the introduction of a platform assisting them with optimizing polypharmacy?

LITERATURE ANALYSIS

Clinical Decision support Systems

Clinical decision support systems in the primary care sector have the potential to improve the decision-making of GPs and pharmacists[124,188,9,185]. In their extensive literature study, Kuperman, Bobb, Payne, Avery, Gandhi, Burns et al.[124] conclude that in order to realize the benefits of these systems, the context-specific implementation complexities have to be addressed; in their words: "there is not yet a 'one-size-fits-all' approach." Other studies that have been undertaken specifically in the Dutch primary care sector show mixed results regarding the influence implemented decision support systems have. While Martens, Van der Weijden, Severens, De Clercq, De Bruijn, Kester et al.[135] found no favorable effects for computerized messages reminding GPs to prescribe certain drugs, messages not to prescribe them "sometimes positively influence[d] the prescribing behaviour [sic] of GPs." A study by Kuilboer, Van Wijk, Mosseveld, Van der Does, De Jongste, Overbeek et al.[123] showed that the system they tested "changed the manners in which the physicians monitored their patients and, to a lesser extent, their treatment behavior."

Due to these developments, CPOE software in the primary care sector has increased in diversity in recent years. From systems that were mostly organizational in nature, they have been enhanced to facilitate consultation of electronic medical records and clinical decision support[209].

Software Market Diversity

Consequently, the Dutch software market of GPs' CPOE systems is diverse. Up to ten different systems are currently in use, all providing their own distinct features and tools in addition to the common patient record management facilities[209]. Some of these systems include their own digital prescription aids and decision support tools, while others do not. As a result, third-party developers have produced additional software that, integrated with their existing systems, assists GPs in prescribing drugs. The major plug-ins on the market are Prescriptor, which can communicate with most existing GPs' CPOE systems, and NHGDoc, which currently has limited integration capabilities[55,68].

POMP The Prescribing Optimization Method Platform (POMP) has been envisioned as a software program that, integrated with these aforementioned CPOE systems, would add to GPs' range of assistive tools. Through the decision support platform, GPs would be advised on how to prescribe in patient-specific cases; through a medication review structured by the Polypharmacy Optimization Method they would determine actual use of drugs, identify superfluous ones, and detect untreated diseases. Advice provided by the system would be based on proven clinical interactions between drugs, compatibility of medicine with patients' other diseases, and best practices extracted through knowledge management. The system would facilitate the execution of a medication review, and thus operate on-demand; it would not disturb GPs' workflows by reacting to decisions made during general consultations.

While the POM Platform will take the form of a stand-alone software program, it will be fully integrated in existing systems as far as user interaction and data exchange are concerned. Through this means the developers seek to avoid the production of a software product that is underused or performs suboptimally.

Technology Adoption

The adoption of technology by potential users has been extensively studied and has led to the development of various predictive models, including the TAM, TAM2, and UTAUT[51,217,218].

These models have not, however, performed equally well at predicting acceptance behavior for all kinds of user groups. Subsequently, the potential influence of various different aspects has been explored and the models extended accordingly. The group of physicians in general and GPs in particular has been studied extensively, and elements as diverse as financial stimuli or psychological ownership have been proposed[27,166]. Below the deviations from elements common in most models that were found in multiple studies regarding this particular group are elaborated upon.

When applied to GPs, some theoretical elements common in technology acceptance models do not seem to be supported. Notwithstanding new research by[117] suggesting otherwise, many studies show an application's ease-of-use is of minor or no importance to their acceptance by GPs, even though in literature this is usually considered a core element[39,238].

In contrast, time reduction is mentioned by various authors as a highly relevant and influential factor regarding GPs' technology adoption, while this is not an element commonly found in any of the models[27,212,238].

In their study, Chismar & Wiley-Patton[39] tested the TAM2 model on a group of pediatricians and found that, of the inherent constructs, only output quality and job relevance were determinant factors that influenced the respondents' adoption.

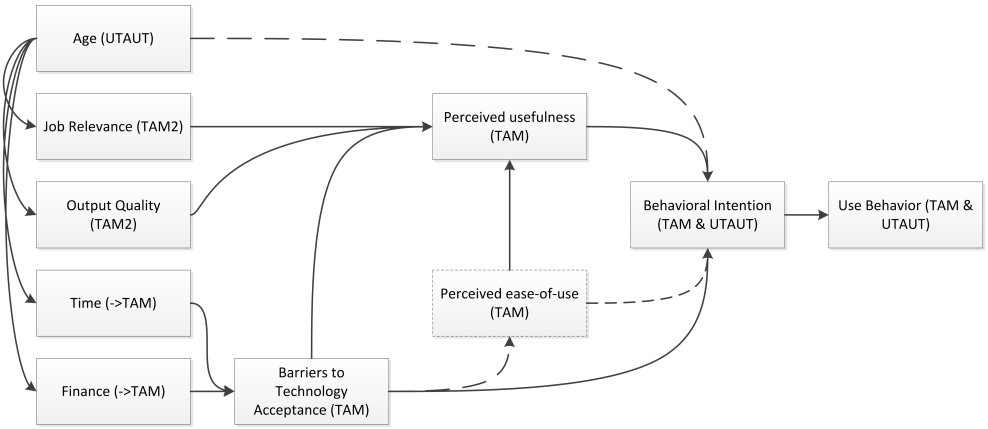
In their study on the low acceptance rates of prescription software, Boonstra, Boddy & Fischbacher[27] tested the original TAM and concluded that its factors were of hardly noticeable

influence, and that adoption was mainly affected by systems' embedding in consultation processes, financial stimuli, policy decisions and cultural boundaries instead.

THEORETICAL FRAMEWORK

Based on the main technology adoption models and the deviations common in the aforementioned literature findings, a theoretical model for testing the attitudes of GPs in the Dutch primary care sector was assembled.

FIGURE 1: **Model depicting factors influential to use behavior, and their originating theories.**



In Figure 1 this model is shown. The final use behavior is largely determined by users' behavioral intention, elements common in the TAM, TAM2 and UTAUT models. The original Technology Adoption Model is included, but in this study the perceived ease-of-use was disregarded, as it was found of no importance in a literature study by Yarbrough & Smith[238]. This leaves the factors influencing perceived usefulness, which was defined by Davis[51] as "the degree to which a person believes that using a particular system would enhance his or her job performance". The two constructs included in TAM2 that were found to be influential by Chismar & Wiley-Patton[39], job relevance and output quality, were incorporated.

Apart from those, the elements of time and finance, often mentioned as highly influential in literature, were added to the model as barriers to technology acceptance, as proposed by Yarbrough & Smith[23827,238]. Finally, the only construct considered to be influential in the UTAUT study by Venkatesh, Sykes & Zhang[219], age, was included as well. In their study, Venkatesh, Sykes & Zhang[219] related age directly to behavioral intention, but stressed that it "is not a cause but likely a symptom" and that it requires further investigation. Therefore, in this study, the authors decided to test the influence of age on the other mediating factors rather than on behavioral intention itself.

Adoption models lead to an assessment of users' final use of software. As the POM Platform under study in this research project is in its preliminary stages, its actual use cannot be assessed. However, in order to gain further insight into the acceptance of decision support systems by GPs in the Netherlands, existing products with conceptually related functionality that are currently in use were evaluated.

Hypotheses

Below the hypotheses that were tested in this study are listed. As will be explained in more detail in the section on Research Design, the hypotheses were tested for correlations, and thus they do not contain causality implications. Therefore, the survey findings are discussed in the Results section, whereas the meaning of the data analysis will be elaborated on separately in the Discussion.

Based on the consensus in literature concerning the benefits of clinical information systems, the following hypotheses were assumed and tested:

H1: Use of one decision support system correlates with use of another.
H2: Ownership of a decision support system correlates with use of a medical formulary.
H3: Ownership of a decision support system correlates with use of number of medical formularies.
H4: Use of a decision support system correlates with use of a medical formulary.
H5: Use of a decision support system correlates with use of number of medical formularies.
H6: Use of a medical formulary correlates with recognition of occurrences of a problem associated with polypharmacy.
H7: Use of a medical formulary correlates with number of recognized problems associated with polypharmacy.

Following the model in Figure 1 on the aspects influential on behavioral intention, these hypotheses were formulated and tested:

H1: Number of patients correlates with perceived usefulness.
H2: Recognition of occurrences of a problem associated with polypharmacy correlates with perceived usefulness.
H3: Number of recognized occurrences of a problem associated with polypharmacy correlates with perceived usefulness.
H4: Importance attributed to output quality correlates with perceived usefulness.
H5: Importance attributed to time investment correlates with perceived usefulness.
H6: Willingness of financial investment correlates with perceived usefulness.
H7: Importance attributed to financial reimbursement correlates with perceived usefulness.

Finally, also according to the model in Figure 1, the influence of age on the various aspects was investigated through the following hypotheses:

H1: Age correlates with number of patients.

H2: Age correlates with recognition of occurrences of a problem associated with polypharmacy.
H3: Age correlates with number of recognized occurrences of a problem associated with polypharmacy.
H4: Age correlates with importance attributed to output quality.
H5: Age correlates with importance attributed to time investment.
H6: Age correlates with willingness of financial investment.
H7: Age correlates with importance attributed to financial reimbursement.

RESEARCH DESIGN

Method

A survey was deemed an appropriate method to examine potential users´ attitudes towards a system assisting them with polypharmacy prescribing. Especially the fact that many potential users could be surveyed was a determinant factor.

First, the questionnaire investigated demographic information, such as respondents' ages and potential partnerships. The next section was dedicated to surveying GPs on their current use of information systems and plug-ins, and included questions about formularies and common problems. The last section started with a short explanation of the features the POM Platform would include, and investigated their opinions towards the newly proposed system. In this last section, only a conceptual idea of the POMP was provided, without specific details regarding its implementation or user interface.

To ensure an optimal response rate, the number of questions was kept to a minimum. Consequently, only a limited number of questions on technology adoption could be included. Consistent with the theoretical framework, questions regarding perceived ease-of-use were disregarded. The oft-mentioned time constraint was surveyed, as well as the perceived influence of financial stimuli. Specific questions on usefulness and output quality were included, but job relevance was not directly surveyed; the authors considered it unfeasible to ask about both relevance and usefulness, as the subtle difference could be lost in a questionnaire. Instead, the job relevance construct was divided over two questions; one asking about the number of polypharmacy patients each GP treated (phrased as 'patients using five or more drugs'), and one on the problems they encountered with these patients.

The questionnaire limitations restricted the explicit inclusion of less often investigated aspects, such as cultural biases mentioned by Boonstra, Boddy & Fischbacher[27] or psychological ownership proposed by Paré, Sicotte & Jacques[166]. A comment section, however, was provided where respondents could fill out any additional remarks. The survey in its final form contained fifteen questions.

The number of respondents necessary to generalize results was calculated with the 'proportion sample' and 'finite population correction' formulas provided by Israel[109], which showed that ninety-five respondents were required to fill out the questionnaire. After taking

into account an expected response rate of twenty percent, the survey was sent by postal mail to five hundred practicing GPs, who were selected through a process of simple random sampling (i.e. each individual was chosen randomly and entirely by chance). The addresses were geographically spread across the Netherlands; addresses included more practices in the country's more densely populated western provinces than elsewhere, but this can be explained by the fact that more GPs work there. A prepaid return envelope was included, giving respondents the opportunity to fill it out on paper or online. Ethical concerns described by Swanwick[199] were taken into consideration in the conduction of the survey. The responses were gathered anonymously and the respondents were assured that their results would be treated confidentially. No coercion of any kind was used to persuade potential respondents to participate, nor were any incentives provided in return for their answers. Responses were gathered during a period of three months, from mid-February to mid-May 2011. After this period, no additional responses were received. During or after this period, no attempts to increase the response rate were made.

Validity

The internal validity of most of the survey's questions on technology adoption followed naturally from their close relations to the theoretical constructs. In cases where the constructs were divided over multiple questions (such as those on job relevance and financial stimuli), results were not combined into single values to avoid distortion of the concepts.

As the respondents were selected through a process of simple random sampling, this study's findings are generalizable to the whole population of 8921 practicing GPs in the Dutch primary care sector, insofar as a confidence interval of 7.15 and a confidence level of 95% are maintained.

As it turned out, multiple questions were answered unevenly. As the frequency of answers in contingency tables was often lower than five, Fisher's exact test was favored over Pearson's chi square test to test the hypotheses[94]. The setup of the study required the authors to test for correlations only. As correlation by itself does not imply causality, the influences in Figure 1 were not directly tested. The implication this restriction has for the results interpretation will be further elaborated upon in the Discussion.

When testing a large number of hypotheses, the potential problem of multiple comparison analysis arises. This problem states that the probability of a tested result being significant by chance increases when testing multiple hypotheses. Different approaches to cope with this problem exist, but they are disputed; in their pursuit to minimize type 1 errors, they might erroneously introduce type 2 errors, i.e. falsely accepting a false hypothesis[180,182]. The techniques are most widely employed when two or more different samples are used. In this single-sample study, the Bonferroni-Holm correction technique was employed to check the results' validity whenever subgroups' correlations were tested. In order to avoid type 2 errors, however, these corrected p-values or significance levels are not incorporated into the results

reported below; applying the Bonferroni-Holm correction ensured, however, that the hypotheses from which the main conclusions were drawn remained significant after correction.

RESULTS

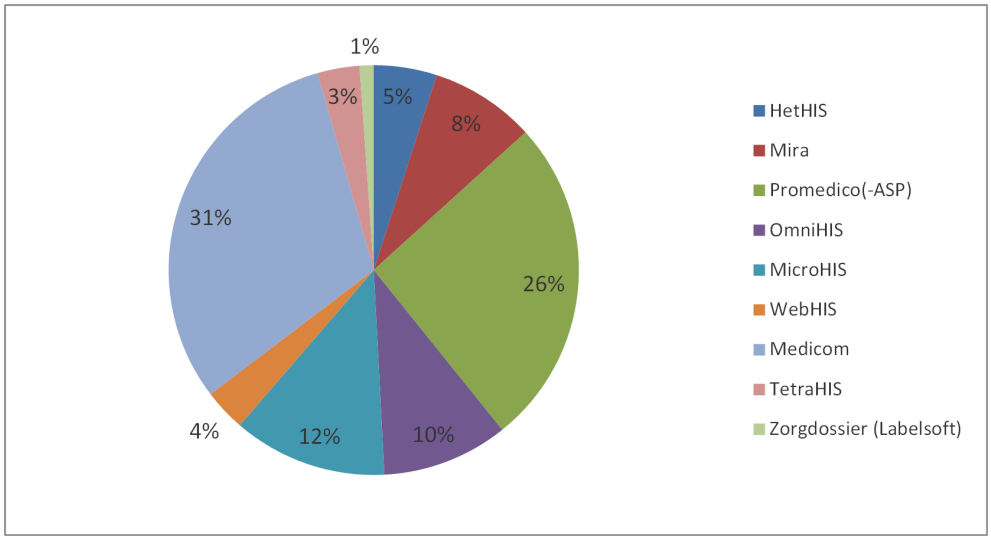
Research Group

The survey was sent by mail to five hundred practicing GPs, accompanied by a cover letter. Potential respondents were invited to either fill out the survey online or return their copy by mail. The vast majority (98%) chose for the latter option.

184 GPs returned filled-out questionnaires, which made a response rate of 37%. Of these respondents, only 15% was younger than 40; 27% were in their forties, while 47% were in their fifties. 10% were sixty years or older, with only one respondent being over 65 years of age. 21% of the respondents worked alone, 37% shared their practice with one colleague, but the majority of 42% worked with several others. All but one of the surveyed GPs reported owning a professional information system; one was unsure.

Software Market Diversity

FIGURE 2: **Model depicting the relative ownership of CPOE software by Dutch GPs.**

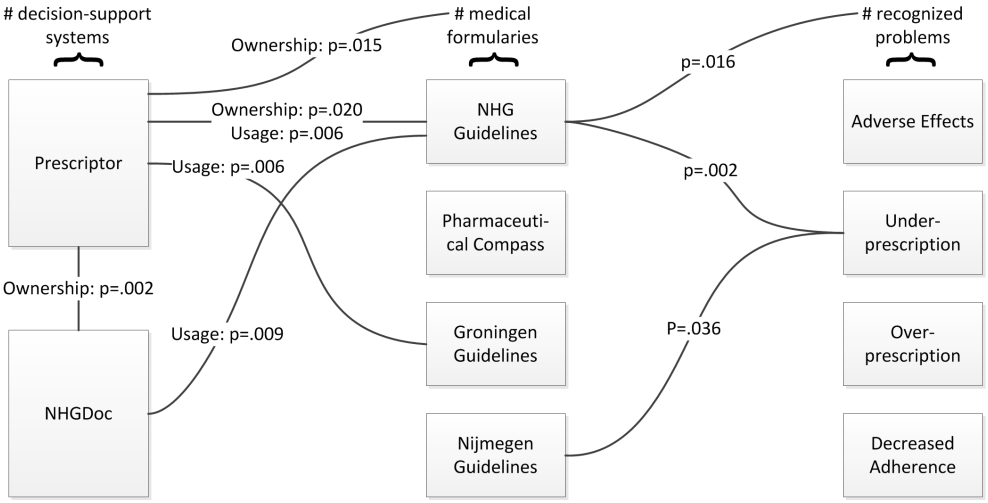


The earlier assumption regarding the diversity of the CPOE systems market was proven to be correct by the survey results. Two of the companies dominated the market, as they were used by 58% of the respondents (see Figure 2). Seven smaller manufacturers each had market shares between 1% and 12%. 29% of the respondents indicating owning Prescriptor, while 25%

was unsure of that; the remainder indicated not owning the product. Fewer respondents (13%) owned the comparable decision support system NHGDoc, with an additional 26% being unsure.

Of the respondents indicating owning one of these systems, the actual use differed. Of the owners of Prescriptor, the majority (55%) indicated using it never or rarely (less than once a week). 39% used it once or multiple times per consultation, with only 6% using it once per week. NHGDoc numbers were comparable; 54% used the software package rarely or never, 38% used it intensively, and 8% just once per week.

FIGURE 3: **Model depicting the probability values of the statistically significant relationships between decision support systems, medical formularies, and recognized problems.**



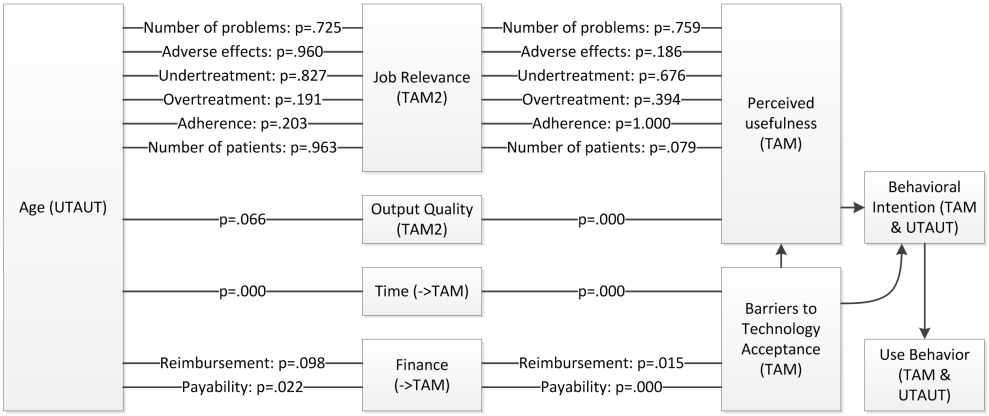
Decision Support Systems & Medical Formularies

Statistical analysis showed that owners using one decision support product (either Prescriptor or NHGDoc) are more likely to use the other one as well than GPs who do not (p=.002, Fisher's exact test); this led to acceptance of H1. The medical formularies, often integrated with GPs' CPOE systems and facilitated by Prescriptor and NHGDoc, were widely used. 78% specified using the Pharmaceutical Compass, 38% used the guidelines by the Dutch College of General Practitioners (NHG), and respectively 13% and 18% also adhered to regional rules as established in Groningen and Nijmegen. There appeared to be some evidence that ownership and use of plug-ins was related to GPs' use of medical formularies. Owning Prescriptor correlated with using the NHG-guidelines (p=.020, FET) and the regional guidelines from Groningen (p=.029, FET). The use of Prescriptor was correlated with the use of NHG-guidelines only (p=.006, FET). Additionally, the number of formularies used correlated with Prescriptor ownership (p=.015, FET).

Relations including ownership or use of NHGDoc did not occur in the same number; use of this plug-in correlated with use of the NHG-guidelines (p=.009, FET), but none of the other relations.

These results validated H4 in the above-mentioned cases. H2 and H3 were accepted for Prescriptor but not for NHGDoc, while H5 was rejected altogether.

FIGURE 4: **Model depicting the probability values of all tested relationships between the factors relating to use behavior, according to Figure 1.**



Problem Recognition

The problems commonly associated with polypharmacy were widely recognized. 95% of respondents recognized a higher risk of adverse effects with polypharmacy patients. Patients of 83% of the surveyed GPs showed signs of decreased adherence to their drug prescriptions. Respectively 49% and 40% of the respondents recognized increased risks of under- and overmedication. Overall, after correcting for double counting due to the possible inclusion of GPs sharing the same clientele, GPs each treated 159 polypharmacy patients on average.

The sixth and seventh hypotheses were tested by checking whether the use of any of the formularies was related to the problems encountered. Especially relations including the NHG-formulary showed significant results; underprescription (p=.002, FET) and decreased adherence (p=.044, FET) were more often reported by GPs using the NHG-guidelines. Overall, GPs using the NHG-formulary reported more problems than those who did not (p=.016, FET). None of the other formularies showed any relations, except for the Nijmegen-guidelines, which showed that use of those clinical rules correlated with the recognized occurrences of underprescription (p=.036, FET). This partly validated H6 for the cases involving the NHG- and the Nijmegen-guidelines. H7 was valid only for the NHG-guidelines.

With the testing of these hypotheses the current situation of the use and impact of decision support systems in the Dutch primary care sector was investigated. The results are visually represented in Figure 3.

Perceived Usefulness

When informed about the possible new system aiding them with treating polypharmacy, the majority (57%) of GPs responded positively; only 5% responded outwards negatively. Subsequently, the constructs presented in the model on technology acceptance in Figure 1 were tested. The two questions on job relevance showed different results; the number of problems encountered with polypharmacy patients did not appear to have any relation with the POM Platform's perceived usefulness (p=.759, FET); neither did any of the specific problems themselves appear to correlate with GPs' attitudes (adverse effects: p=.186, FET; undertreatment: p=.676, FET; overtreatment: p=.394, FET; adherence: p=1.000, FET). The total number of patients using five or more different drugs that each GP reported did not significantly correlate either (p=.079, FET). The aspect of output quality, however, appeared to correlate strongly (p=.000, FET) with the POM Platform's usefulness as perceived by the GPs. A similar result was found for the relation between time investment (p=.000, FET) and the system's usefulness. The two questions used to measure the financial phenomenon, on payability and reimbursement, showed significant correlations with respectively p=.000 (FET) and p=.015 (FET). This led to the acceptance of H11, H12, H13 and H14. H8, H9 and H10 were rejected.

Age

Finally, the hypotheses regarding the impact of respondents' ages on the constructs were tested. As it appeared, age correlated significantly with the constructs of time (p=.016, FET) and payability (p=.022, FET). The aspects of output quality (p=.066, FET) and reimbursement (p=.098) were not related. Age appeared to significantly correlate neither with the number of encountered problems (p=.725, FET) nor with the number of patients (p=.963, FET); no evidence was found to assume any relations between age and the recognition of specific problems (adverse effects: p=.960, FET; undertreatment: p=.827, FET; overtreatment: p=.191, FET; adherence: p=.203, FET). This validated H19 and H21, but led to the rejection of H15, H16 and H17, H18 and H20.

All probabilities found with Fisher's exact test that are part of the model in Figure 1 are depicted in Figure 4.

DISCUSSION

With this paper the authors aimed at exploring the motivations of GPs to adopt decision support systems. This study being part of a greater project involving polypharmacy complications, the data collection focused on GPs' experiences with polypharmacy prescribing, treatment and assistive software.

The questionnaire's results showed that the Dutch market of GPs' CPOE systems is very diverse, even though the two largest providers control half of the market. Assistive plug-ins for these products are roughly distributed amongst a fourth of the GPs; their functionality, however,

is seriously underused. The majority of the owners of this assistive software indicate never or rarely ever using it, with just under forty percent employing it regularly during consultations.

Simultaneously, the results indicate that – in several cases – use of decision support systems is related to use of medical formularies. Furthermore, in the case of the NHG-guidelines the adherence to these clinical rules is related to GPs' recognition of polypharmacy problems.

As mentioned earlier, the data analysis only revealed correlations between hypotheses; causality could not be inferred from the statistical results. However, since the hypotheses were based on the influential relations between the variables in Figure 1, which in turn were based on consensus in literature, causality may be assumed when interpreting these results. Thus, in the discussion below, these aspects are interpreted in the light of this justifiable causality.

When exploring GPs' attitudes towards using an assistive software program aiding them with polypharmacy treatment, subjects generally responded positively to the idea of such a decision support system. Some, however, uttered doubts as to its added value; several respondents questioned the proposed application's form as a separate product, suggesting to "improve existing systems rather than introducing new methods." The generally positive response to clinical decision support systems corresponds with the findings of a study among Irish physicians by[100].

The theoretical framework in Figure 1 includes two constructs that are commonly included in technology adoption models; job relevance and output quality. Of these, the only one that appears to be strongly valued by GPs is that of output quality. The other construct, job relevance, seems to be of lesser importance; neither the number of patients suffering from polypharmacy a GP treats seems to positively influence his or her opinion towards the proposed system, nor does the (number of) recognized problems. This finding does not concur with GPs' answers regarding the system's usefulness, nor does it reflect views in literature that specifically mention job relevance as vital to GPs' attitudes[39]. Reasons for this discrepancy may be found in the methodical approach to this aspect, which will be further elaborated upon in the section on limitations.

An aspect not included in any of the common technology adoption models by default, time, appears to have a very strong influence on GPs' motivations. This finding concurs with several other studies investigating this aspect[27,212,238]. Due to their high workload GPs may be unwilling to adhere to new guidelines requiring time investment, but welcome aids reducing consultation time per patient[208].

Another aspect explored in literature as possibly influential is that of financial stimuli. Just as time, it is supposedly a barrier against adoption of new technology[238]. The survey results show that payability and, to a lesser extent, reimbursement influence GPs' attitudes towards the newly proposed system. One of the respondents, for example, indicated that he would be willing to consider purchasing the finished product "only if reimbursement is guaranteed in writing."

The importance of financial stimuli found in this study mirrors the findings of Boonstra, Boddy & Fischbacher[27] who mentioned it as one of the four influential factors in their research.

Age, finally, has a moderate influence on the importance GPs attribute to time and financial stimuli. No effect on output quality or job relevance was discovered. As an aspect commonly investigated in technology adoption literature, its discovered influence corresponds with the scientific consensus, most notably with the work of Venkatesh, Sykes & Zhang[219], who found that age was the only construct common in adoption models that appeared to be influential to GPs' attitudes.

CONCLUSION

While GPs' CPOE systems are widely in use, the use of decision support systems during patient consultations is lacking. When being proposed a conceptual decision support system aimed at assisting polypharmacy treatment, however, GPs indicate expecting such a system to be valuable. Specifically, they indicate output quality, time investment and financial stimuli as important determinants in their attitudes towards acceptance. Job relevance seems not to have as strong an influence as is commonly assumed in literature[39].

The original research question which the authors attempted to answer through this study, reads: how are GPs in the Dutch primary care sector, based on their experiences with decision support systems and medical formularies, likely to respond to the introduction of a platform assisting them with treating polypharmacy? Following the result that the vast majority of respondents judged the perceived usefulness of the proposed system as high, GPs working in the Dutch primary care sector are likely to adopt such a system, under the conditions that it improves prescription quality and does not require extensive investments of time or money.

Limitations

Even though utmost care was taken in the conduction of this study in order to ensure its reliability and validity, some considerations should be taken into account when applying its findings.

The survey employed in the study contained some questions that did not directly represent theoretical constructs, most notably the one on job relevance. This choice was made because of the possible ambiguity respondents may perceive between that concept and usefulness, but ultimately it hinders generalizability. The large differences between the rough estimates given by the respondents may explain the lack of significance in the relations measured. Considering that this study's outcomes on job relevance do not concur with scientific consensus, the authors suggest that these specific findings should be generalized outside of this study with reservation.

Additionally, given that respondents voluntarily participated in the study, they may be more interested in, and thus hold more positive attitudes towards, clinical decision support systems than non-participants.

Further Research

This study has investigated factors of influence on GPs' attitudes towards adopting decision support systems. Factors commonly not included in technology adoption models, such as time or financial investments, have been proven to be influential in shaping GPs' attitudes towards such systems.

Further research should focus on investigating the wider generalizability of these claims. Additionally, the salient beliefs held by health care professionals should be explored, as proposed by Holden & Karsh[99], in order to find out if indeed some fundamental concepts of technology adoption, most notably perceived ease-of-use, do not influence this group's attitudes, while other aspects generally not included in such models, do.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

APPENDIX: QUESTIONNAIRE (TRANSLATED FROM DUTCH)

1. Do you work in a partnership with others?
    - No, I work alone
    - Yes, with one other colleague
    - Yes, with several other colleagues
2. What is your age?
    - Younger than 30
    - 30 – 34
    - 35 – 39
    - 40 – 44
    - 45 – 49
    - 50 – 54
    - 55 – 59
    - 60 – 64
    - 65 or older
3. What GPs' CPOE do you use?
    - HetHIS
    - Mira
    - Promedico
    - OmniHIS
    - MicroHIS
    - WebHIS
    - Medicom
    - None

- I don't know
- None of the above (please specify)

4. Which formularies do you use during drug prescriptions?
    - (ETAS-)NHG
    - Groningen
    - Nijmegen
    - Pharmacotherapeutical Compass
5. Do you own the electronic prescription system Digitalis Prescriptor?
    - Yes
    - No
    - I don't know
6. How often do you use Digitalis Prescriptor?
    - Once per consultation
    - Several times per consultation
    - Once per week
    - Less than once per week
    - Never
7. Do you own the electronic prescription system NHGDoc?
    - Yes
    - No
    - I don't know
8. How often do you use NHGDoc?
    - Once per consultation
    - Several times per consultation
    - Once per week
    - Less than once per week
    - Never
9. How many patients using five or more drugs do you treat approximately?
    - (Please specify)
10. Do you recognize one or more of these problems with your polypharmacy patients?
    - An increased risk of adverse effects
    - An increased risk of under-prescription
    - An increased risk of over-prescription
    - A decreased medication adherence
    - Another (please specify)

The Prescribing Optimization Method (POM) is a method meant to support GPs with prescribing drugs for elderly experiencing polypharmacy. Advice is given by checking, among others, if the used drugs conflict with each other or to what extent patients adhere to their prescriptions. This can cause drugs to be replaced by alternatives or to have their dosages modified. Because

of this patients experiencing polypharmacy can be treated quicker and more effective; GPs' prescription behavior is significantly improved through this method.

Through integration with CPOEs and in combination with formularies the POM can be easily used digitally.

1. Would you consider the POM valuable for prescribing drugs for polypharmacy patients?
    - Yes
    - No
    - I don't know
2. Would you consider time reduction important in your decision whether or not to use the POM?
    - Yes
    - No
    - I don't know
3. Would you consider quality improvement important in your decision whether or not to use the POM?
    - Yes
    - No
    - I don't know
4. Would you be willing to pay to use the POM?
    - Yes
    - No
    - I don't know
5. Would your intention to use the POM be influenced if its use were reimbursable?
    - Yes
    - No
    - I don't know

2.2

Patients' attitudes towards medical apps

ABSTRACT

The increased use of internet through smartphones and tablets enables the development of new consumer-focused mobile applications (apps) in health care. Concerns including these apps' safety, usability, privacy, and dependability have been raised. In this paper the authors present the results of a grounded theory-approach to finding what non-functional requirements of medical apps potential users view as most important. A document study and interviews with stakeholders yielded nine non-functional requirements for medical

Published as Meulendijk, M., Meulendijks, E., Jansen, P., Numans, M., & Spruit, M. (2014). What concerns users of medical apps? Exploring non-functional requirements of medical mobile applications. *Proceedings of the European Conference on Information Systems (ECIS) 2014*. Tel Aviv, Israel.

apps: accessibility, certifiability, portability, privacy, safety, security, stability, trustability, and usability. Six of these were evaluated with two groups (differing by age) of potential users through a vignette study. This revealed differences between the age groups regarding the importance each attributed to apps' usability and certifiability. Furthermore, and contrary to consensus in literature, privacy was considered one of the least important attributes for medical apps by both groups. Trustability, security, and, for the younger group, certifiability, were considered the most important non-functional requirements for medical apps. The implications of these results for developing medical mobile applications are briefly visited.

INTRODUCTION

The increased use of internet through smartphones and tablets has enabled the development of new consumer-focused mobile applications in areas such as education, personal finances, transportation, and health care[49,236,71]. In this last area, health care, various initiatives have been employed to improve patients' well-being through mobile technology. These include mobile applications that aim at aiding users directly through advice or information, as well as software that collects data for research purposes or planned operations[1,163,181].

While these new applications have the possibility to improve both patients' empowerment and their health, they are not without risks. Concerns include applications' safety, usability, privacy, and dependability[181,35,147,118]. Further exploration into these concerns is necessary in order to advance the development of patient-beneficial applications. Involvement of potential users in the assessment of these concerns has been argued for in literature[181].

The use of multiple different medicinal drugs by people suffering from chronic conditions is known as polypharmacy. It has been associated with problems including decreased adherence and misinformation, leading to patients having greater risks of adverse effects or hospitalization[24,41,73,125,190,192,197,234].

In this paper the authors present the results of a grounded theory-approach to finding what non-functional requirements of consumer-focused medical mobile applications (potential) polypharmacy patients view as most important. They aim to answer the following research question: which non-functional requirements of medical mobile applications are deemed most important by (potential) polypharmacy patients?

The research approach revolves around eliciting and subsequently evaluating these non-functional requirements by interviewing stakeholders, including developers, experts and potential users. Section 2 describes the theoretical background that provides the foundation for this study. The subsequent sections, 3 and 4, detail the research approach and discovered results. Finally, the discussion relates the results to existing theory, while the conclusion relates the findings to practical implications.

BACKGROUND

Medical Mobile Applications (Medical "Apps")

The increase in the use of smartphones and tablets has been widely recognized as having the potential to be of great use in the medical domain[163,181,147,118]. Mobile applications, or "apps", could be applied in both professional and personal settings, by patients-consumers and professional caretakers alike.

As is evident from the research question, this study is aimed at researching what non-functional requirements consumer-focused medical apps should meet. Thus, applications aimed at professionals, or to be used during consultations, are outside of the scope of this study. A systematic literature review by Ozdalga, Ozdalga, & Ahuja yielded a list of consumer-focused medical apps containing, among other features, monitoring, educating, and communicating[163]. Adding to this, in the course of this study the authors investigated the current state-of-the-art of medical apps in the Dutch market, as the research was conducted in the Netherlands.

The search was performed on the common platforms for mobile app distribution, Google Play and Apple Store. Sixteen relevant apps were found and examined. Several of the apps were mainly aimed at advising users about medical issues (e.g. advise them whether or not to visit their GP, based on analysis of a medical problem). Others' main functionalities revolved around acting as a personal assistant for patient-specific health issues (e.g. keeping track of food intake for patients suffering from diabetes, and generating alerts when necessary). Popular features of apps included maintaining users' medical records, generating alerts and reminders, and generating advice or information based on user input.

Requirements

In their book, Hull, Jackson, & Dick define requirements engineering as "the subset of systems engineering concerned with discovering, developing, tracing, analyzing, qualifying, communicating, and managing requirements that define the system at successive levels of abstraction"[103]. The authors see a requirement as the basis for every project, "defining what the stakeholders in a potential new system need from it, and also what the system must do in order to satisfy that need". Stakeholders, in their view, can be any person or entity that uses, benefits from, is disadvantaged by, or is responsible for a system. The importance of continuous engagement of stakeholders throughout the development process has been recognized[96].

Requirements are generally divided into functional and non-functional requirements. Functional requirements state 'what the system should do', while the non-functional requirements are 'attributes of or constraints on a system'[40] or 'how the system (should) behave'[72]. In an attempt to minimize the ambiguity surrounding the definitions of non-functional requirements, Glinz created a taxonomy in which they have been divided into performance requirements, specific qualities, and constraints[81]. Examples of performance

requirements are timing and throughput efficiency, quality requirements include usability, reliability, and portability, and constraints can be physical or legal barriers[81]. Additional attempts to classify and standardize often identified quality indicators for software have resulted in a variety of industry standards[106,108,84].

According to Paech & Kerkow non-functional requirements are often poorly understood, and neglecting them is one of the top ten risks in requirements engineering[165].

Grounded Theory

Grounded theory was first described in 1967 by Glaser & Strauss as the discovery of theory from data[80]. According to Charmaz the method was created "as a protest against what they viewed as a rather passive acceptance that all the great theories had been discovered" (as cited in Goulding[83]). The remaining role of research was to test existing theories, not to propose new ones. Grounded theory broke with this paradigm by introducing a method to create new theory.

According to Corbin & Strauss there are a number of procedures to follow when adopting grounded theory as a research method[42]. A first one is that data collection and analysis are interrelated processes, which means that they can and have to happen simultaneously. Concepts are furthermore the basic units of analysis, and subsequently, categories must be developed from these concepts and related to one another. Another important procedure is that the analysis makes use of constant comparison. This means that everything that is formed into concepts or categories is constantly compared to all the other elements and aspects of the study[42].

According to Birks, Fernandez, Levina, & Nasirin, grounded theory can be a powerful tool for IS scholars interested in theory development, allowing researchers to conduct pioneering research with both flexibility and rigor[23]. An important aspect of grounded theory is that 'all is data'[79]. This implies that not just methodical interviews, surveys and observations are data, but that anything the researcher comes into contact with, including respondents' behavior or attitudes, is data[83].

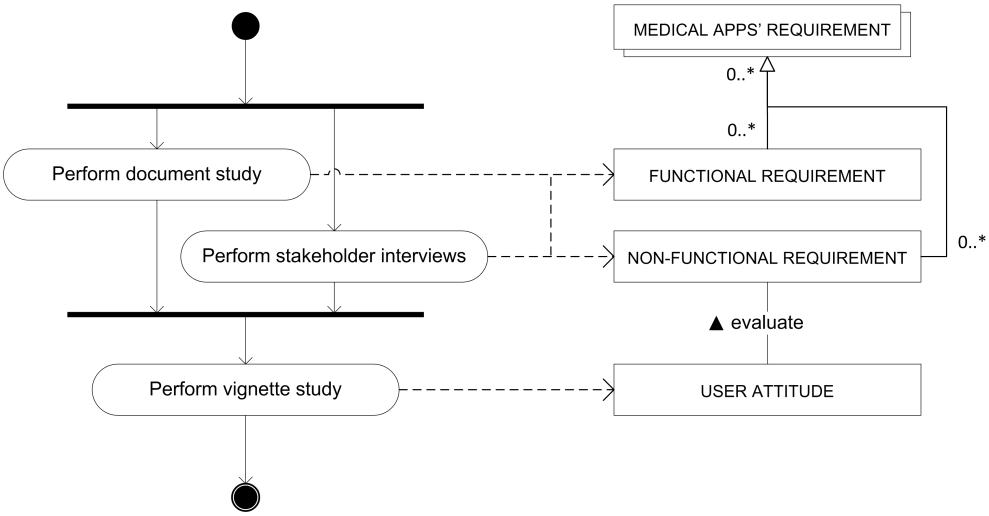
RESEARCH DESIGN

In order to elicit and evaluate the non-functional requirements of consumer-focused medical apps, a mixed-methods research design was adopted. This choice was made because of the broader perspectives a mixed-methods design offers, as well as the ability to evaluate findings after initial exploration. More specifically, the research design was an exploratory sequential design, as described by Creswell & Plano Clark[44]. However, due to expected difficulties with performing large-scale impersonal quantitative methods with elderly respondents, the second evaluative step was performed through vignette-guided interviews. While limiting the results' generalizability, this approach did allow for intensive interaction with respondents, in line with grounded theory.

Two methods were used to perform the first step of eliciting the requirements: a document study and interviews with appropriate stakeholders. The document study was performed in order to explore the consensus in public discourse on relevant non-functional requirements. Next to this top-down approach, the stakeholder interviews were used as a bottom-up approach to discover what attitudes both potential users and experts had.

For evaluation, vignette-based interviews with potential users were performed. The vignettes were based on the non-functional requirements discovered in the previous elicitation phase. This method was employed to find out which requirements potential users viewed as most important in their decisions whether or not to use a medical app.

FIGURE 1: **Process-deliverable diagram detailing the research process and its outcomes[209].**



Document Study

The document study was performed to discover both practical and generic sources pertaining to non-functional requirements of medical apps. Included in the study were news articles, directives, laws, roadmaps, working instructions, and guides on polypharmacy and the development of medical applications.

The initial documents were found by searching a variety of general and domain-specific search engines: the generic Google search engine, the mobile app stores Google Play and Apple App Store, and the Dutch primary care web portal Artsennet. Initial queries consisted of combinations of 'polypharmacy' and 'medical applications' and their derivatives; Dutch equivalents of these terms were used as well.

Besides the documents found through this process, additional sources were found through the so-called ancestry approach, i.e. through reference lists and related works of included documents. All documents were judged on source quality and topical relevance; those found

to be relevant and reliable were included in the study. They were processed by recording the source, summarizing the document information, and extracting the relevant data.

Stakeholder Interviews

Semi-structured interviews with a variety of stakeholders were performed to discover what requirements interviewees mentioned as essential for consumer-focused medical apps. Among the people interviewed were potential users, medical professionals and information systems developers. In total, two polypharmacy patients, two family caregivers, two professional caregivers, two information system experts, and two pharmacists were interviewed. The group of interviewees was purposefully diverse, in order to accommodate many different perspectives. The results were coded during analysis and collected in concepts. In line with grounded theory the interview phase was ended when a point of saturation was reached, i.e. when the analysed results no longer yielded new concepts.

The interviews were performed in person at respondents' homes, and were recorded and later transcribed. Respondents were asked about their opinions and attitudes towards a potential mobile app that would help them manage and monitor their drug use in a variety of ways. Questions included whether or not respondents valued the idea of such an app, what functionalities they thought it should have, and if they would consider using it if it were available.

Vignette-Guided Interviews

In order to evaluate the identified non-functional requirements, potential users were approached. Two groups were interviewed. As this study focused specifically on polypharmacy patients' concerns, members of the first group were 65 years or older and used five or more medicinal drugs chronically. A second group, consisting of people over 50 years of age and using at least one drug chronically, was included to examine differences in attitude between polypharmacy patients and people whose health is starting to decline and who may become reliant on multiple drugs in the future.

A vignette-study is an assessment study in which information is gathered through vignettes: "a short description of a person or situation that contains relevant information which is presented to respondents to obtain a value judgment about that described person or situation"[213]. A principal characteristic of vignettes is that the presented scenario allows for researchers to include both advantages and disadvantages of a respondent's choice. Through this design, respondents are made to carefully consider the (hypothetical) implications of their decisions when answering. Although primarily a quantitative technique, vignettes have been used to guide interview sessions as well, as they have in this research[17,146].

Vignette-guided interviews with the two groups of respondents were performed. The vignettes were created after the non-functional requirements discovered in the document study and stakeholder interviews. Even though a total of nine non-functional requirements were

found, only six of these were evaluated with potential users. It proved impossible for potential users of medical apps to judge technical aspects such as stability, portability and accessibility during vignette-guided interviews.

Each respondent was read all six vignettes presenting a scenario revolving around a non-functional requirement. Then they were presented with three options highlighting advantages and disadvantages of their decision, asked to pick one, and motivate their choice. Below is an example of a vignette. It contrasts the non-functional requirement usability with the app's number of functionalities; if respondents opt for a highly functional app, this will negatively impact its usability, whereas, if they decide for an app with limited functionalities, its usability will be positively impacted.

> The medical app comes in multiple versions. Please pick the option that most closely resembles your attitude.
>
> 1. The app is very easy to use, but only contains a medication overview.
> 2. Besides the medication overview, the app contains some other helpful functions and is still rather easy to use.
> 3. The app contains, among others, a medication overview, a drug reminder, information about your diseases, and an option to immediately contact your GP. Because of all these functionalities, it is somewhat harder to use.

RESULTS

Document Study

The document study yielded fourteen documents reliable and relevant to the authors' research question. Example sources of documents that were included are European Union directives, the Dutch College of General Practitioners, and the Dutch Ministry of Health, Sport, & Welfare. The most important issues that were identified as non-functional requirements were certifiability and privacy.

The first requirement, the necessity of certifiability, is based on a European Union directive, which in turn has been converted into laws in its member states. The directive defines that a medical device or aid has to be CE-certified to be used legally, and proposes a definition for what a medical device is[43]. Software that assists patients in monitoring or preventing diseases and medications is in fact classed as a medical device.

The second requirement revolves around privacy regulations and how an application that deals with personal or medical information should incorporate those. The processing of any personal information is limited by the Dutch Data Protection Act; both the transfer and processing of information by third parties, entered by patients, will need to be approved by those patients. Apart from getting approval, other conditions that have to be met include the

aspect that the patients' privacy is not disproportionately harmed, and that the process favors a 'general cause'[2].

Stakeholder Interviews

The interviews were performed with ten stakeholders, two of each of the following groups: polypharmacy patients, family caregivers, professional caregivers, information system experts, and pharmacists. All sessions were recorded and transcribed. The length of the interviews ranged from approximately twenty to sixty minutes, which an average of thirty-seven minutes. Results from the document study were incorporated in the interview questions.

Using the coding techniques common in grounded theory, the transcribed interviews were analyzed for themes. The non-functional requirements that were found using this method are accessibility, portability, privacy, safety, security, stability, trustability and usability.

Accessibility, portability and stability are application-specific characteristics that are related. Accessibility is commonly understood as the degree to which the application is available to users, usually taking into account disability measures. Portability refers to the number of operating systems and devices the application supports, and stability to the technical robustness and dependability of the application. Information system expert #2 stressed the importance of technically catering to the right audience, saying "Family caregivers visit patients at home, and so do nurses. So smartphones, and likewise tablets, would be obvious choices.", while information system expert #1 commented on the variety in functionality that different user roles bring with them: "[We previously talked] about the two modes: whether or not having access to [a patient's] health record. You should adjust those to the different stakeholders."

Privacy, safety, and security all relate to safeguarding users' well-being when using the application. Privacy was for example discussed with pharmacist #1, who was able to tell that "if you can decide for yourself who can have access, it doesn't have to be a problem". These concepts are related to the requirement of trust, which entails the users' perceptions of these characteristics. Making sure the application is private, safe, and secure is essential for having people 'trust' the application. Pharmacist #1 warned for providing laypersons with incomplete or ambiguous advices, implying that "if you tell a patient of a [clinical] interaction [between his drugs], he will panic, risk that he stops using something, while nothing may be wrong."

Finally, the last non-functional requirement mentioned by the stakeholders is the usability of the application, which is its ease of use and learnability. Family caregiver #1 wondered if the mobile application could be made simple enough for her to understand at her old age: "Then I think, once I would get to know the application, I would certainly be willing to use it".

Vignette-Guided Interviews

To evaluate the gathered non-functional requirements with potential users, vignettes were created which forced respondents to judge the pros and cons of certain functions.

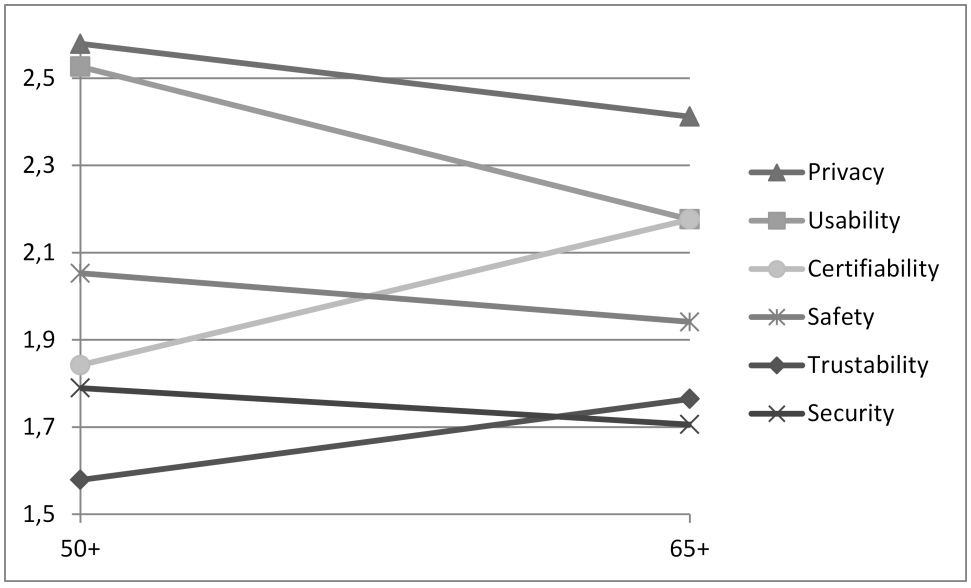
Three of the non-functional requirements had to be excluded from the study, as they were too technical to be judged by non-professional respondents; these are accessibility, portability, and stability. This leaves the requirements certifiability, privacy, safety, security, stability, and trustability to be tested in the vignette-guided interviews. The first group that was interviewed included seventeen respondents, who on average were aged 77 and used 9 drugs. The second group consisted of nineteen persons, who had an average age of 60 and used 3 drugs.

Firstly, most respondents indicated putting at least some degree of trust in an application like the one proposed. Only four people, all of the first group, chose the most conservative trust option, indicating they had no faith in automatically generated advices from a medical application whatsoever.

Secondly, the majority of respondents, especially of the second group, indicated favoring functionality over usability. The first group held more conservative attitudes toward this requirement.

The respondents showed progressive results for the requirement concerning privacy as well. Only one interviewee refused to have her personal data used at all. The others were willing to let their data be used if it would benefit the advices they would receive; about half the respondents demanded their data be used anonymously.

FIGURE 2: **Importance respondents attributed to non-functional requirements per age group. The horizontal axis represents the average answer respondents gave to each vignette; as lower-numbered answers were more conservative, a lower score means a requirement was considered more important.**



Regarding the security requirement, most respondents indicated wanting some form of protection for their personal data in the proposed application. Of the first group, most people favored password protection in the app itself, while the majority of the second group settled with their smartphone's default mode of security.

Respondents of both groups indicated accepting a limited risk in the generated advices, provided the advices were customized to their situations.

Finally, the majority of the second group indicated valuing a form of certification for the proposed app. Most interviewees of the first group answered attaching more importance to recommendations by their general practitioners.

In Figure 2 the importance the respondents attributed to each of the queried non-functional requirements is graphically displayed.

Recapitulation

As mentioned earlier in this paper, the field of requirements engineering is well-established, and has brought forth a multitude of theories and taxonomies attempting to list quality indicators of software[106,108,84]. These industry standards have different perspectives on concepts regarding non-functional requirements; in Table 1 all non-functional requirements found in this study are shown, and an attempt has been made to map the existing concepts of the main standards to the newly found ones in this study[40]. As can be observed, some of these concepts translate well to those in existing taxonomies, while others are new to them.

Among the concepts that are common in most frameworks is usability. This user-focused requirement has long been a part of software requirements models. While the sub concepts belonging to the concept of usability greatly differ between theories, it is principally understood as "the effort needed for use, and on the individual assessment of such use"[106]. Sub concepts may include ease-of-use, learnability, aesthetical attractiveness, and, in some cases, accessibility[108]. Accessibility, or software's ability to be "use[d] by people with a wide range of characteristics", is found explicitly in only one of the references theories, but is often understood as an implicit aspect of usability[108].

Concepts aimed at more technical aspects of software are rather common in most frameworks as well. Portability, defined in the ISO 9126 standard as "the ability of software to be transferred from one environment to another" is present in most theories, although sometimes referred to as adaptability, installability or compatibility, or included as sub sets of these[106]. Likewise, the concept of security, interpreted as the degree "to which a system prevents unauthorized access to data", is common throughout theory[108]. Finally, stability is found in most frameworks, and is understood as systems being "operational and accessible when required" to "perform specified functions under specified conditions for a specified period of time"[108]. It includes aspects such as systems' robustness, availability or uptime, and recoverability from errors.

While the concept of safety is present in various forms in most frameworks, it is usually implied through combinations of broader concepts such as accuracy, correctness, or reliability. The concept as found in this study is defined as the "degree to which a product or system provides the correct results with the needed degree of precision"[108].

The remaining concepts discovered in this study, privacy, trustability and certifiability, are not or hardly reflected in existing taxonomies of requirements engineering. Although privacy is included in the ISO 25010 standard as the assurance that "data are accessible only to those authorized to have access", it is non-existent in others[108]. Likewise, the concept of trust is included in the ISO 25010 standard as a subjective user criterion, but the concept of trustability found in this study, i.e. the ability of systems to convince users of their dependability and responsibility, is not[108]. Finally, the related concept of certifiability, the degree to which software's behavior is approved by authorities, was not found in any form in any of the referenced frameworks.

TABLE 1: **The non-functional requirements identified in this study and their equivalents in leading industry standards.**

This study's NFR's	ISO/IEC 25010[106]	ISO/IEC 9126[108]	FURPS(+)[84]
Accessibility	Accessibility		
Certifiability			
Portability	Adaptability, installability	Adaptability, installability	Adaptability, installability
Privacy	Confidentiality		
Safety	Functional correctness	Accuracy	Accuracy
Security	Security	Security	Security
Stability	Availability, fault tolerance	Stability, fault tolerance	Availability, stability
Trustability	Trust (user criterion)		
Usability	Usability	Usability	Usability

DISCUSSION

The grounded theory approach described in this paper resulted in a set of non-functional requirements that are applicable to the development of consumer-focused mobile software applications in a medical domain.

Uncommon Non-Functional Requirements

Some of the concepts discovered in this study can easily be mapped to those prominent in existing taxonomies, as shown in Table 1; this is not true for all of them, though. It is notable

to see how concepts revolving around trustworthiness of both the output of the software and the intentions of its developers are found important in this study, although they are generally underrepresented in existing theory.

Indeed, issues revolving around the safety of medical apps' output have been recognized. Buijink, Visser, & Marshall argue that users should be made aware that "some apps contain unreliable, non-peer-reviewed content"[35]. In studies where the quality of medical apps' advices was assessed, they varied greatly from product to product[229,221,87,70]. As incorrect advices of medical apps can negatively impact users' health by delaying correct diagnoses or advocating incorrect self-medication, measures to ensure their reliability have been proposed, including testing apps before publication and certification[35,229,221,70]. While in most taxonomies the concept of safety is present in an implicit form or through combinations of other requirements, in the medical domain it seems to take a more prominent and explicit role.

It has been proposed that the scientific evidence for medical apps' quality should be reflected in certifications by recognized authorities[35,221,70]. Buijink, Visser, & Marshall call for "government health authorities to provide official certification marks guaranteeing the quality of apps"[35]. While the changeable nature of apps conflicts with the process of certification, this measure is advocated for by several authors in the field, and is reflected by findings in this study[35,229,221,70].

The issue of privacy has long been at odds with information technology in the medical domain, with research showing caretakers and patients alike being concerned about using health data for multiple purposes[169]. In the words of Sarasohn-Kahn: "keeping personally identifiable health information secure is a long-standing challenge"[181]. In the domain of apps, studies have shown apps often do not provide users with adequate control over and visibility into how applications use and share their personal data[193,65]. Marceglia et al. mention privacy breaches, through technical, organizational, or human factors, as one of the foremost concerns regarding apps in the medical domain[133].

Trust is a cornerstone of the medical practice and is essential in patient-caretaker relationships[46], in the literature on medical apps, this aspect is underexposed. Marceglia, Bonacina, Zaccaria, Pagliari, & Pinciroli mention it as an aspect pertaining to their concern of privacy, but as a requirement in itself it is usually considered implicit[133].

Finally, the concept of accessibility, or ensuring users with diverse physical and psychological abilities are able to use software, has long been included as a success factor for software[108]. It has, however, often been considered an implicit requirement, or categorized as a sub-concept of usability. From the results of this study it appears to have a more explicit role in the medical domain. An explanation for this finding can be sought in the likeliness that consumer-focused medical apps are to be used by people with medical histories, and thus form a diverse group regarding abilities. While apps show great promise for empowering frailer users[181], users differing in age or ability have different usability needs from one another[48,150,152,226].

Relative Importance Attributed To Non-Functional Requirements

After eliciting the non-functional requirements through interviewing stakeholders and investigating documentation, potential users were presented with vignettes to evaluate six relevant ones. The two groups of interviewees differed on age and amount of medication. In Figure 2 the results of this evaluation are depicted in a graph, in order to highlight the differences between the groups. While not a quantitative study, these averages do reveal some unexpected findings that are in line with the researchers' understanding of the interviews.

One of the outstanding insights is the difference in judgment of usability between the two groups; the group of older respondents indicated assigning greater importance to applications' ease of use over extended functionality. As is evident from the requirement frameworks shown in Table 1, usability has long been recognized as a crucial factor for software's success. This holds true for all demographics, but research has shown elderly have different usability needs from younger users[48,226]. Reading small-sized text or clicking small areas for several seconds may pose problems for some elderly[48]. Renaud & Van Biljon prove that in their senior-focused extension of Davis' technology acceptance model, software's limited ease of use may make elderly users reject it[175]. Thus, while the rise of touch interfaces on smartphones and tablets is often seen as presenting new opportunities for catering to an aging user base, the importance of usability for this demographic is evident. Family caregiver #1 expressed her concerns regarding this: "I cannot use those [smartphones or tablets]. I even used to find phones difficult back in the day. [...] I do not think I would learn to use new devices quickly at my age."

Another relative difference between the younger and older response groups can be observed with the concept of certifiability. The younger respondents indicated valuing certified apps more than older respondents, the majority of whom put more confidence in their GPs' suggestions. A sense of certifiability, thus, may be related to the authority people assign to their physicians. A recent study by[46] showed that people's trust in their GPs increases with their age. This finding is reflected in the results of this study. As polypharmacy patient #2 remarked, "I would rather leave [my drugs management] to my GP; he knows what he is doing."

A final insightful observation is the fact that most the respondents rated privacy as one of the least important non-functional requirements. Only one of them indicated not wanting to share their information with other patients at all; the others were willing to share their data, either anonymously or personally, if it would benefit their own experiences with the software. This finding does not reflect the importance that is generally attributed to matters of privacy[169,133]. It does reflect the findings of a recent study on laymen's perspectives on health technology, which found that "especially for the chronically and acutely ill, privacy is of far less concern to patients than to health professionals"[222]. The benefits software brings to polypharmacy patients managing taxing daily routines may outweigh the disadvantages of privacy infringement.

CONCLUSION

In this study, the authors sought to explore non-functional requirements that apply to mobile medical apps. Through methods of stakeholder interviews, document analyses, and user evaluations, the following nine non-functional requirements were found: accessibility, certifiability, portability, privacy, safety, security, stability, trustability, and usability.

Six of these were evaluated with potential users of medical apps, which revealed that the importance placed in apps' certification decreases with age and destitution, in favor of the GPs' judgment. In contrast, the value placed in usability increases with age and destitution. The concept of privacy was the concept least valued by respondents of any age group. Trustability, security, and, for the younger group, certifiability, were considered the most important non-functional requirements for medical apps.

Implications for Design

If these non-functional requirements lead to augmentations or limitations of software depends on the context and implementation of the application. Glinz argues that "the notion of non-functional requirements is representation-dependent", meaning that requirements can take the form of constraints, performance requirements, quality attributes, or even functional requirements, depending on how they are modeled[81].

Depending on which age group the app is focused on, more or less attention should be placed on certifiability or usability. Older, frailer users would adopt an application quicker if it were recommended by their GPs, while younger, healthier persons would put more faith in certifications. Moreover, and in contrast to older people, younger persons would favor more extended functionality over ease of use.

During the evaluation interviews, respondents were asked what features they would like to see included in a mobile medical application (out of a shortlist of ten preselected items). Out of these, respondents rated the more comprehensive features (overview of their health records, ability to log side effects or complaints, automatic medicine checks) higher than the more conventional ones (medicine reminders, sharing experiences with other patients, directly contacting their GPs). These more comprehensive features come into conflict with the non-functional requirements quicker than the other ones, which implies that a careful trade-off between user-requested functionalities and the infringements these make on non-functional requirements should be made.

Limitations & Further Research

Even though the utmost care was taken in the conduction of this study, some reservations should be made when interpreting its results. Even though a total of nine non-functional requirements were discovered in the elicitation process, only six of these were evaluated with potential users. This decision was made because it proved impossible for potential users of medical apps to judge technical aspects such as stability, portability and accessibility during

vignette-guided interviews. In future studies, other methods, such as semi-structured interviews or card sorting, may be applied to further discover the importance potential users attribute to these non-functional requirements.

A wide variety of stakeholders and potential users were included in this study to gain insights into non-functional requirements for medical mobile applications. However, the qualitative nature of the research and the limited number of interviewees make generalizing its results difficult. In future studies, the results of this study should be validated by quantitatively testing them with a representative sample of potential users.

2.3

Feasibility of decision support systems in primary care

ABSTRACT

Although the conduction of feasibility analyses before executing comprehensive projects is often urged for in literature, little standardization exists regarding actual approaches. In this chapter the authors describe the practice of conducting a feasibility analysis of

Published as: Meulendijk, M., Drenth-van Maanen, A., Jansen, P., Brinkkemper, S., Numans, M., & Spruit, M. (2013). Introducing the COrETeSt Feasibility Analysis in Medical Informatics: A Case Study of a Decision support Knowledge System in the Dutch Primary Care Sector. In M. M. Cruz-Cunha, I. M. Miranda, & P. Gonçalves, Handbook of Research on ICTs and Management Systems for Improving Efficiency in Healthcare and Social Care (pp. 1066-1087). IGI Global.

a decision-supportive knowledge platform in the Dutch primary care sector, and present recommendations for researchers and entrepreneurs performing similar projects accordingly. The research question – how can the feasibility of a decision-supportive knowledge platform in the primary care domain be investigated? – is answered by describing into detail the issues encountered during the process and included as CORETEST-recommendations: the investigation of conceptual, organizational, economic, technological and societal aspects comprises an extensive feasibility analysis in the primary care domain.

INTRODUCTION & BACKGROUND

Whenever conducting comprehensive projects, it is common practice for researchers and entrepreneurs alike to investigate the risks that may threaten their feasibility beforehand. Although definitions vary, the aim of feasibility studies is generally understood as objectively and rationally determining projects' risks and ultimate viability. Various studies define feasibility analyses in similar terms, from a tool "to determine if a business opportunity is possible, practical and viable" to "a controlled process for identifying problems and opportunities, determining objectives, describing situations, defining successful outcomes and assessing the range of costs and benefits" [98,203].

Feasibility studies are designed to show how ventures would operate under a controlled "set of assumptions", testing "such factors as the technology used, financing, marketing, and so on" [136]. Based on these studies entrepreneurs can identify potential pitfalls and adjust their plans accordingly. Aside from the results' practical implications, benefits may also include proved economic viability and thus influence prospective investors.

While the initial costs of conducting feasibility analyses may discourage entrepreneurs from doing so, their application is widely propagated in literature: start-up ventures' limited success rates are often quoted to stress the benefits of testing their feasibility in advance [136,203]. As [136] state: "The study is usually the first time in a project development process that many key pieces and information about the project are assembled into one overall analysis. The study must show how well all of these pieces fit and perform together." Feasibility studies should be carried out before starting work on the actual project; estimations regarding the time their conduction should take vary, but "a good rule of thumb for the feasibility analysis step for most development projects is 3 to 6 months" [136].

Feasibility studies are common in the fields of medicine and IT [204,216]. And even though much literature urges for the conduction of feasibility analyses, little standardization exists regarding the actual approach that should be taken. No proven methods have been proposed as step-by-step procedures for performing feasibility research. Even though this apparent lack of formalization remains unattended, various authors have proposed frameworks that

aid entrepreneurs and researchers in determining which aspects to investigate. The most structured one of these is the feasibility analysis that [196] describe as part of the Systems Development Life Cycle. The TELOS, an acronym for its factors of interest, comprises the areas of technical, economic, legal, operational, and schedule feasibility. Other studies on feasibility include similar components, often based on the authors' areas of expertise. [98] greatly extend the economic and operational areas, including factors such as organizational and marketing aspects. [136] describe various practical considerations regarding studies' approaches (especially applicable in their field of agriculture), including consultants' independency and stakeholders' interests.

Research Approach

Research Question This paper documents the process of conducting a feasibility analysis of an IT system in the field of medical informatics, thereby attempting to add to the apparently limited body of knowledge available in this field. The research question that will be investigated is the following: how can the feasibility of a decision-supportive knowledge platform in the primary care domain be investigated?

As will be described in more detail below, the approach consisted of action research employing a variety of methods. Following the components mentioned earlier, the chapter will mainly encompass external aspects of feasibility analyses, such as economic, organizational and, to a lesser extent, technological ones. The creation of a conceptual model will be covered as well, as its compatibility with the ones created for the different aspects is essential to the conduction of a successful feasibility study.

Validity & Reliability The results presented in this chapter were reached through a process conventionally called action research. In an early paper on this topic, action research was defined as "diagnosis of a social problem with a view of helping improve the situation" [25]. Action research is closely related with the postmodern school of thought and is based on the assumption that "complex social interactions cannot be reduced for meaningful study" [18]. Following a holistic approach, human organizations should be studied within their contexts, not by isolating their parts [38]. When the researcher then intervenes into the research setting, he becomes part of the study, resulting in "the realignment of the roles of researcher and subject into more collaborative and synergistic forms" [18]. Consequently, action research is well-suited to reflectively exploring processes in which the researchers' and actors' roles overlap, such as feasibility analyses.

Having originated from the positivist school of thought, the usage of the concept of validity in action research has been disputed [172]; as [126] contemplates, is it legitimate to "fit the qualities of action research into a traditional discourse about validity whose concerns have little to do with those of action research?" Although replicability, a traditionally important criterion of validity in positivist thought, is infeasible in action research, [38] argue for the importance of the study's recoverability. Following that construct, researchers should "make clear to interested observers the thought processes and models which enabled the team to

make their interpretations and draw their conclusions." This construct will provide the research a certain validity that does not match the concepts of the positivist school, but is nonetheless stronger than mere plausibility. In this project, this construct was operationalized by determining in advance the study's scope and the fields to be investigated. Furthermore, the authors strived to describe in as much detail as possible the thought processes employed to make decisions and determine consequent steps.

Research Methods When researching any type of phenomenon, selecting the appropriate method for the intended purpose is essential. While qualitative and quantitative methods are often stated as being clear opposites of each other, they both come with their specific advantages and shortcomings. Qualitative approaches, grounded in understanding phenomena in their context-specific settings without employing means of quantification, are ideal for exploring new fields or stakeholders' opinions. Quantitative research, based on employing experimental methods and statistical analysis to test hypothetical generalizations, can be employed to test the applicability of claims among wider populations [111].

While action research traditionally employs mainly qualitative methods, in this project quantitative survey results were used as well. In a feasibility study, where both non-quantifiable details of a field need to be investigated and sheer numbers are required, such a combination can prove fruitful.

In the POMP project the initial approach was to explore the field of decision-supportive systems and to examine to what extent they had been employed in the medical area; a literature study was conducted for this purpose. In an early phase, stakeholders were interviewed to determine their views regarding the facilitation of medical methods through IT. Results of these sessions were used to create and distribute a survey among five hundred physicians, regarding their opinions on the usage of specific IT systems in medicine. The process itself was described by retrospectively collecting and analyzing documents, including research journals, reports, papers, and interview results.

Case: Prescribing Optimization Method Platform (POMP)

The different aspects of the IT system of which the feasibility has been analyzed in the research phase prior to the writing of this paper will be explored into detail throughout the text. An introduction to the system, the field in which it operates, and the problems it aims to solve is described below.

In the Netherlands, seventeen percent of the chronically ill use more than five different drugs permanently; half of these patients are over seventy years of age [198]. Although this polypharmacy is often unavoidable, it significantly increases their chances of hospitalization. Problems often associated with polypharmacy are, among others, an increasing risk of adverse effects, under-prescribing, decreased adherence to daily doses, and overtreatment due to increased medication usage [73,24,190,192,234,41,125,197].

Recognizing these problems, the Prescribing Optimization Method (POM) was devised, a six-step routine through which General Practitioners (GPs) can optimize the prescriptions of patients experiencing polypharmacy. During tests usage of this method significantly increased the relevance of their prescriptions [57].

Currently, the development of a POM Platform (POMP) is being researched, within a vision of optimizing and capitalizing the POM through the use of information and communication technology. Through the decision-supportive platform, GPs are to be advised on how to prescribe in patient-specific cases; advice will be based on proven clinical interactions between drugs, compatibility of medicine with patients' other diseases, and best practices extracted through knowledge management.

CONCEPTUAL FEASIBILITY

Problem Exploration

An early step in any feasibility study is the exploration of the problems that the intended venture is meant to solve. Even if the original foundation is based on sound scientific analyses, such as in the case of the POMP project, the change in scope when investigating a technological application may require additional study to accommodate the different perspective. Both literature studies and expert interviews are well-equipped methods for this step, as they can be used to identify determinants relevant to the field.

As mentioned above, in the POMP project the literature analysis on which the original medical method was based showed that a major problem lies in patients' usage of multiple drugs. The claims on hospitalization made earlier were investigated into more detail to determine what steps in the prescription process specifically caused them; this revealed that a significant amount of them were caused by various, especially procedural, factors. With regard to polypharmacy, seven percent of errors were generally attributed to procedural factors; these included incomplete information on patients' medical histories, inadequate communication between health professionals, and errors due to time restraints or carelessness.

After identifying into detail the problems and their causes and possible solutions, the extent to which they are being recognized in the applicable field can be investigated. Through additional literature analysis experiences regarding the practical occurrences of the problems can be learned; in the case of the POMP, papers containing interview results from physicians showed that they recognized the problems of polypharmacy; for example, they indicated often feeling obliged to prescribe medicine even if they knew the patient's situation did not demand it.

In order to identify the magnitude of the awareness of the problems, quantitative data can be beneficial. If this sort of data is not readily available from earlier published papers or statistics agencies, surveys are appropriate methods to gather it. People likely to experience the investigated problems should be asked to what extent they recognize them and, if applicable,

what they do to avoid or limit them. In the POMP project, a survey was distributed amongst general practitioners that included questions on this. The 184 respondents indicated recognizing many of the problems with their patients and consequently answered positively to the concept of the POMP. They also indicated using various (software) tools to contain the problems.

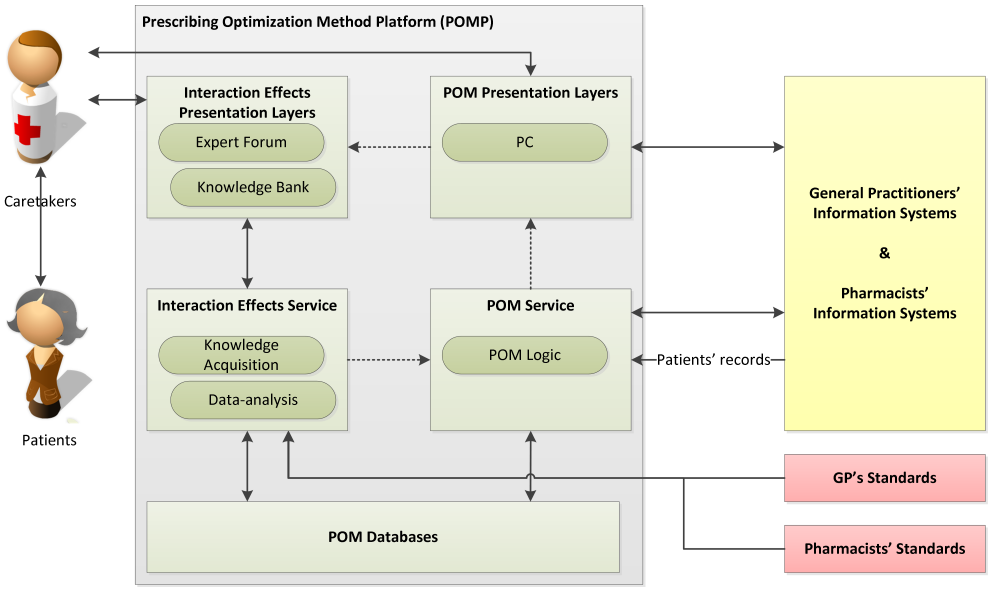
Having explored and quantitatively established the medical aspects of the problem, the informatics approach was researched. The uniqueness of the proposed project can be determined by exploring what similar systems have already been developed regarding this problem. In the primary care field the introduction of decision-supportive systems is often advocated, as they can, if correctly implemented, significantly improve GPs' prescriptions [12]; literature analysis showed that under more exaggerated expectations of those systems by users can cause disappointment and disuse [148]. In a more general sense, authors did encourage developers to create systems aiding physicians with the prescription process, while focusing on patient-specific characteristics [171].

Concept Modeling

When the original field of the problem has been explored, as well as the relevant information scientific and computer scientific approaches, a first attempt at creating a conceptual model of the proposed system can be made. Such a concept can, optionally, be modeled through established techniques, or, given the early phase of development, in an informal way based on practical considerations. In Figure 1 the early concept of the POMP drawn up after the literature analysis is shown. During the modeling no formal technique was applied, allowing the researchers to freely imagine the concept, unrestrained by modeling languages' limiting rules. The concept was based on best practices in knowledge system design, modified for usage in the medical field. Created by the information scientists, the concept was commented on by the POM's medical instigators and modified accordingly.

In the model, the POMP concept consists of five sub-concepts. The main one of these is the POM logic module, where the medical basis for the decision-supportive system is used through logical rules. Through physicians' information systems the POM interface is displayed to the platform's users. They interact with the POMP through their own native systems, and share the results with their patients. Additionally, physicians interact directly with the POMP's knowledge-based interfaces, such as expert forums. The feedback provided by them via these channels is interpreted and judged by the interaction effects-module, which creates new logical rules. These are, combined with rules based on best practices, stored in the POMP's databases. Information provided by third parties, such as physicians' and pharmacists' industry standards, is stored and used similarly.

FIGURE 1: **Conceptual model of the POMP.**



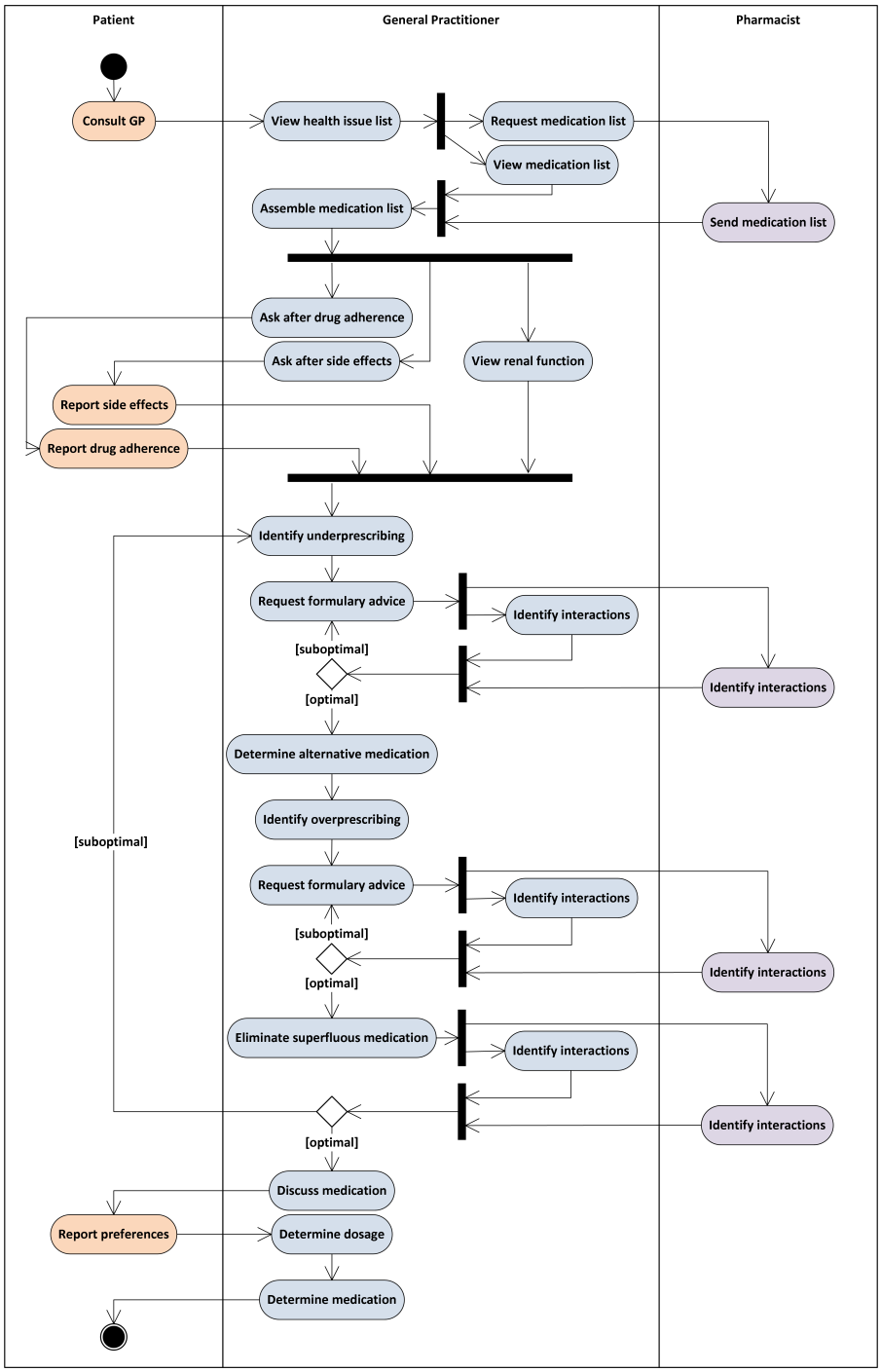
Process Modeling

Not only high level models can be insightful at this point. Although diagrams that show intermediary steps are difficult to create in this phase, as the required level of detail for these models has not yet been decided upon, lower level usability diagrams can show users' interactions with systems.

One of the formal techniques commonly used in these approaches is UML; especially following the effective approach described by Ambler[7], this method is applicable for drawing up early concepts. Through this vision of Agile Modeling, only diagrams required for the current phase of development have to be designed, unlike UML's formal rules that demand all diagrams up to a certain point be modeled. By assuming simplicity while designing and embracing change in subsequent phases, diagrams can be designed efficiently.

Apart from the high level conceptual model shown above, additional diagrams can be employed to explore the processes through which the eventual system would be put to use. UML provides multiple types of diagrams that are capable of graphically depicting this, mainly state diagrams and activity diagrams. In deciding which type of diagram should be employed, the type of process being modeled should be taken into account. Whereas activity diagrams are meant to display sequential processes, state diagrams are better suited for displaying componential changes. In most usability processes, using activity diagrams will be advantageous [8]; in the POMP project, these were employed to display user processes.

FIGURE 2: **Activity diagram of the POM.**



Special attention was paid to the procedure through which the activity diagrams were created. First, the POM method's researchers were interviewed to describe the overall process they envisioned; during these sessions, the concepts they mentioned were abstracted and simplified to step-by-step approaches. Then these results, along with the concepts proposed in the POM paper [57], were incorporated into basic workflow diagrams. These were checked by the researchers, whose comments were taken into account while refining the diagrams. After several of these expert exchanges, consensus was reached on how to abstract the method into an activity diagram. The Prescribing Optimization Method, as it was depicted in a workflow diagram, is shown in Figure 2.

ORGANIZATION FEASIBILITY

Market Modeling

After having modeled the system's concept and its interaction with potential users, the organizational feasibility can be assessed by looking into the market in which the system is intended to participate. Based on this, key players, competitors and potential partners can be identified after which an organizational framework can be built.

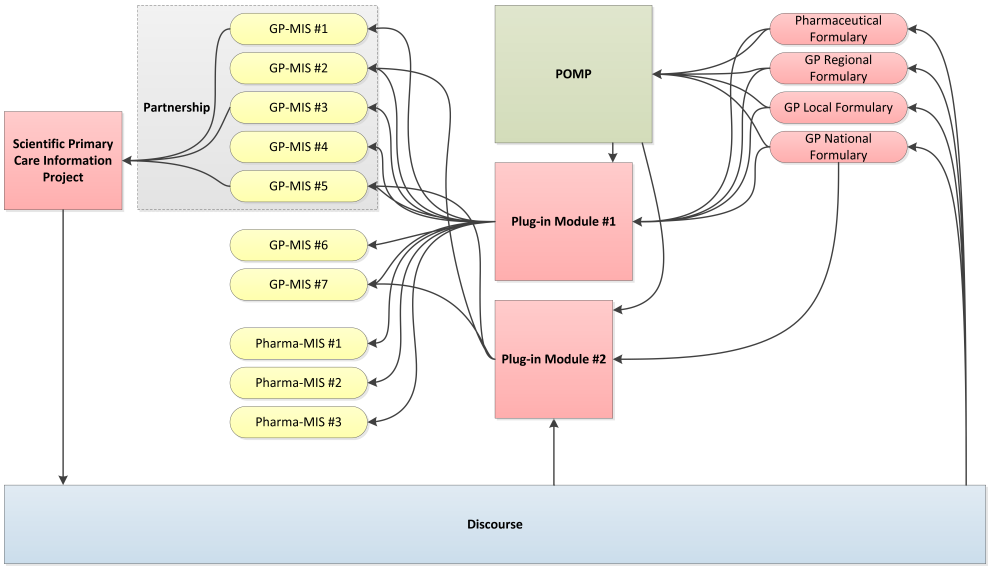
This step commences with the creation of diagrams depicting stakeholders in the field. Through these, overviews of how the system behaves in relation to third-party software and which dependencies logically follow from different approaches can be made insightful early on. While several common formal methods are available for this sort of modeling (including UML collaboration diagrams, ArchiMate models and software ecosystems), the POMP project's early diagrams were not modeled through strict techniques. The aforementioned modeling techniques were used in the feasibility study's later phases, but in the exploratory stage the liberty and simplicity of free modeling were preferred.

In Figure 3 the (anonymous) diagram drawn to depict the POMP's interactions with third-party software and methods is shown. The information required to visualize the data was gathered by market research and literature analysis. The many rounded yellow boxes on the left represent the third-party software that the POMP would have to be integrated with, in order to function properly. Although the market shares of these medical information systems vary, the diversity makes direct integration difficult. Two alternative methods of integration are shown as red squares; other software companies have already developed integration modules with several of the third-party systems. As such, this diagram clearly depicts that a potential partnership with either of these companies would prove beneficial, but would simultaneously greatly increase the project's dependency on these possible partners. The red rounded rectangles on the right depict the various medical rule sets that have to be incorporated into the POMP and that, up to different extents, are already part of the third-party systems.

Also shown is the feedback on the usage of the third-party systems, assembled by a primary care information-project. In this case, the arrows show exactly on which information these results are based. For a scientific project that requires validation in its later phases, such as POMP, this is valuable information; the one or two medical information systems required for a preliminary pilot study could be selected based on this information, as additional data on those systems' usage would be available through the information-project.

Finally, the primary care information-project's data adds to the scientific discourse of the decision-supportive medical field, displayed as a blue rectangle. Ideas and developments discussed in the field are incorporated into the POMP, the third-party software, and the medical rules.

FIGURE 3: **POMP's stakeholders relationships.**



Organization Modeling

After identifying the key players in the relevant market and still prior to developing economic strategies, potential partners can be approached through interviews. The goal of this activity is both to deepen the researchers' understanding of players' roles, and to explore the latter's attitudes towards the newly proposed system.

The survey distributed amongst general practitioners included questions on their system usage. This delivered insight into the market share each of these systems has, which in turn proved useful by determining which players were essential to the field. It appeared that two of the medical information systems together provide software for more than half of the general practitioners. The rest of the market is divided over the remaining suppliers. The adoption of the

POM Platform by the users of these two largest systems can be expected to greatly influence it becoming a common plug-in, and as such needs attention in the economic analysis.

When, through interviews and survey analysis, the various players' roles and attitudes have been determined, they can be modeled in a more formal manner. This provides an overview of all stakeholders in the project and their specific functions regarding it. Various formal modeling techniques exist that can handle tasks like this, but in the POMP project ArchiMate was chosen for its flexibility [127]. This modeling language is able to display relationships between concepts at various levels, so it can be expanded in later phases if necessary. In line with the principles of Agile Modeling, however, only the organizational perspective desired was modeled in the POMP project [7].

FIGURE 4: **ArchiMate organizational model of the POMP.**

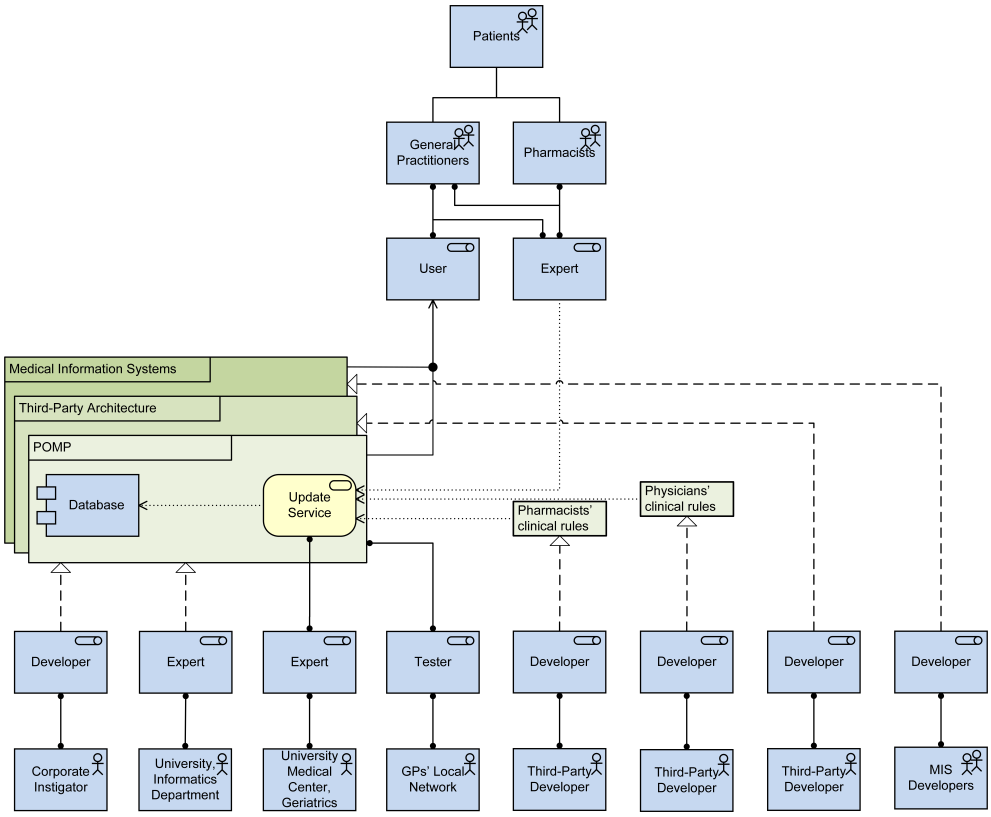


Figure 4 depicts the POMP organizational model as created through the ArchiMate standard. The big green squares represent the various architectural layers on top of which the POMP was designed. The blue rectangles show all stakeholders and their roles regarding the project. In this way, the different roles some players have can be precisely modeled. General practitioners, for example, act both as users and as experts towards the system; the same is true for pharmacists.

The relationships towards the POMP and third-party software can also be specifically depicted; developers usually have creation-relationships towards their products, whereas others are merely assigned to them.

ECONOMIC FEASIBILITY

Development Strategy

At this point the market of medical information systems and the specific stakeholders directly involved with the POMP project have been investigated. The specific roles of third-party developers and their willingness to cooperate and invest have been explored. With this information, development strategies for the ultimate product can be designed.

In a market as diverse and difficult to enter as the Dutch primary care one, no single option should solely be relied upon. With as many stakeholders as the POMP project evidently has, consequently the project's dependency on these partners increases as well.

It should be noted that all partners have different stakes regarding the POMP project. Some have clear commercial interests, such as the third-party software developers. Others, such as the university and the academic medical center, have scientific concerns. As software users, GPs' and pharmacists' interests lie in the improvement of their medical advice, and the reduction of their time spent on patients. Patients' interests lie, obviously, in the improvement of their health, but possibly in the reduction of medicine costs as well.

Concerned with all these partners and their different interests, the POMP's instigators have had to ensure that the project's dependency on any of these stakeholders is limited. Consequently, they defined a number of different development strategies, each pursuing different partnerships. This ensured the project's longevity, in the event that any of the partners would withdraw their support.

Three different development strategies were created, each sharing only the 'core' partners (i.e. the instigators and the academic geriatric and information science departments). The strategies are aimed to be executed at different points in time. An explanatory overview of them is given below.

1. Short-term strategy; in an early phase, the instigators aim at developing an alpha-version of the platform that will only contain the essential elements. Through this implementation the method's effectiveness in a digital form can be assessed. Additionally, a start with exploring other aspects, such as technology acceptance and usability, can then be made. Technologically, this implementation is envisioned as an add-on to existing systems, either using the host's built-in capacities or as a parasitical application, residing in the host without it being aware. No additional partnerships are required for this implementation, so no dependencies are created.

2. Mid-term strategy; in a later phase, a wider distribution of the beta-version of the platform is pursued. Hereto, partnerships with the third-party developers of plug-ins are required. Through these joined efforts, availability of the platform in many different medical information systems can quickly be realized. This is helped by the technological integration of the POMP with the third-party plug-ins, thereby enabling joined distribution of the software. Notwithstanding the relative speed and spread of this approach, in this phase the platform's development would become dependent on partnerships.
3. Long-term strategy; the final goal of the POMP's development is a seamless integration with all primary care medical information systems. In order to realize this, collaboration with all manufacturers of these software packages is necessary. As the market is very diverse, this objective requires careful negotiation with all players and as such can only be seen as a long-term goal. Reaching this objective, however, would eliminate the dependency of the platform on the third-party plug-in developers of the mid-term strategy.

Even though the different development strategies greatly reduce the project's dependencies, partnerships are still essential to successfully achieving its objectives. As a result, ensuring the prolonged interest of the partners is important. In the POMP project, this was done by ensuring that all stakeholders would benefit from a fully-developed POMP. Commercial, scientific and practical benefits were all negotiated with stakeholders and secured on paper. In return, immediate partners (up to the second phase) signed memoranda of understanding, declaring their intentions to enable fully realizing the POMP.

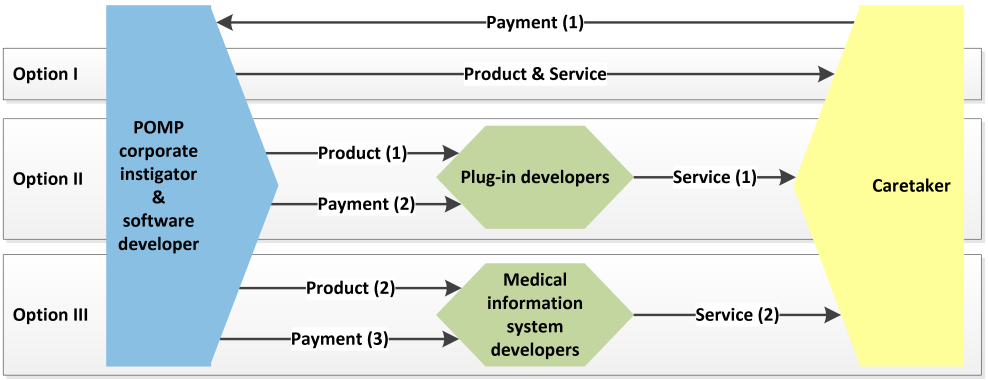
Software Ecosystem Modeling

With the three development strategies created, economic perspectives can be modeled. Especially insight into the products and services delivered by the various players, and how they are compensated, is important in an early stage of the project. An adequate method of visualizing this is through the use of software ecosystems. A software ecosystem is defined as "a set of actors functioning as a unit and interacting with a shared market for software and services, together with the relationships among them" [112]. These models are meant to display software supply networks, exploring the relationships between producers, suppliers, intermediaries and customers. [29] attempted to standardize the approach to designing software ecosystems by formalizing rules.

Based on their standardization, a software ecosystem for the POMP project was created. It is shown in Figure 5. Note that only the economically relevant partners are displayed, while the scientific and practical partners are not. The blue shape represents the corporate instigator and developer of the POMP. The yellow shape depicts the GPs and pharmacists as end users of the software product. Finally, the green hexagons represent the third-party plug-in developers and the medical information system developers.

The relations between the economically viable partners for all three different development strategies are depicted. In all options, health professionals pay the corporate instigator for the usage of the POM Platform, but only in the first option are the product and services directly supplied to them by POMP's developer. In the second option, the POMP is supplied to plug-in developers, along with their share of the revenue. Then, they deliver the final service to the end users. In the final option, the arrangement is equal to that of the second option, but in this scenario the medical information system developers function as intermediaries, instead of the plug-in developers.

FIGURE 5: Partial software ecosystem of the POMP.



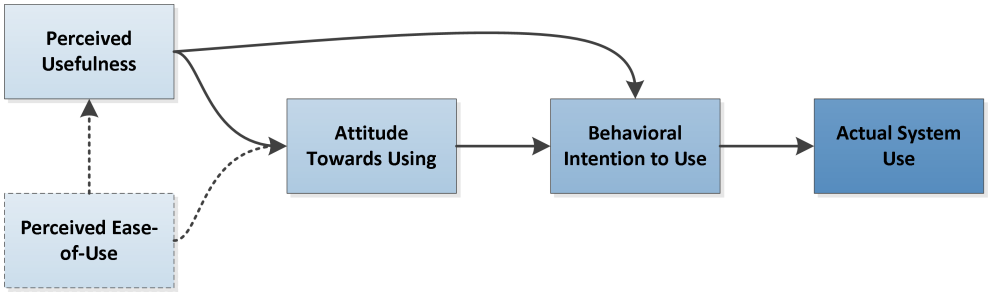
Technology Acceptance

While a firm understanding of the market and cooperation with the key players on it is essential to successfully exploiting a software product, its actual acquisition and usage are the final determinants of success. The field of technology acceptance is dedicated to researching people's motivations to adopt new technological innovations. One of the foremost theories in this field is the Technology Acceptance Model (TAM) [51]. Originally designed to function in business environments, it has been expanded numerous times to take into account aspects inherent to other fields.

As such, the TAM has been used to study people's motivations to (dis)use software in medical environments as well. While employing the theory to explore general practitioners' attitudes towards software adoption, varying studies have found similar aspects to be of greater or lesser importance. Interestingly, software products' ease-of-use, generally considered to be one of the main determinants in technology acceptance, appears to have a diminished impact on GPs. The perceived usefulness, however, appears to be of prevailing importance, especially considering job relevance and quality improvement [39,99]. Figure 6 visualizes this modification in the TAM using dotted lines for the less influential aspects.

Apart from that, studies show that physicians may have objections towards technology acceptance because of the individuality of their jobs [22,238]. This is supposedly aggravated by the fact that new systems imply a disturbance of their routines and introduce a learning curve [238].

FIGURE 6: **Technology Acceptance Model applied to the Dutch primary care sector.**



In the early phases of any project, predicting potential users' acceptance is difficult. In the POMP project, an attempt was made by asking GPs about their software usage and their experiences with polypharmacy in the survey. It turned out that the majority of respondents recognized the problems and had a positive attitude towards an assistive method. Over ninety percent of them indicated finding quality improvement a major factor in their acceptance of it.

Apart from their experiences with polypharmacy, GPs were also asked about their usage of third-party software. Especially the degree to which they owned and used plug-ins in their medical information systems was investigated. The results showed that at least twenty percent and at most fifty percent owned one or both of the relevant plug-ins. However, only three quarters of them actually use the software on any regular basis.

These newly learned factors were taken into consideration in subsequent steps of the feasibility analysis. When predicting societal gains or creating a business model, this information and these numbers prove valuable in calculating expected benefits and revenues.

TECHNOLOGICAL FEASIBILITY

Information Exchange Modeling

When the conceptual, organizational and economic perspectives have been modeled, researchers can focus on the technological aspects. In an early stage it is difficult to provide detailed technological approaches, as many of the steps required for that have not yet been decided upon. However, certain high level technical concepts can be particularized. The goal of this step is to make actual technologically realistic decisions visible following the conceptual framework.

The POM Platform had been envisioned in the early stages as a web service integrated with GPs' and, optionally, pharmacists' information systems. As many different systems exist,

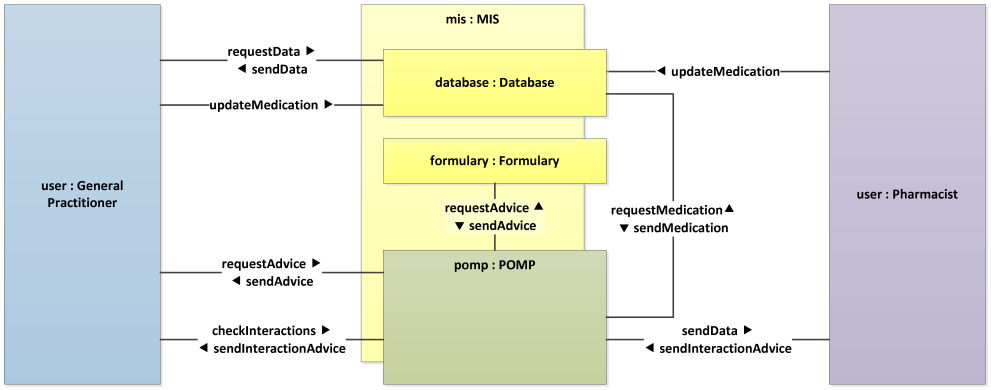
implementation strategies for various time scales have been formed, and specified in the development strategies mentioned earlier. Attempts were made to visualize the exchange of information through these systems in the various approaches through UML collaboration diagrams.

As the principles of Agile Modeling were pursued during the creation of these diagrams as well, only aspects deemed relevant were modeled [7]. Following this concept, the first and last development strategies were considered too similar, and consequently just a single collaboration diagram for these scenarios was created.

In Figure 7 the collaboration diagram for the directly implemented POMP is displayed, as the information exchange would function in the first and third development strategies. The POM Platform is directly integrated into the medical information system (MIS), without being dependent on any intermediary architecture. In the first stand-alone development strategy, the platform would have a parasitic relationship with the medical information system, with the latter being unaware of the former's presence. Through screen-scraping techniques [211], relevant information would be read from the host into the POMP. In the third and optimal strategy, the POM Platform would operate as a perfect plug-in to the default system, updating information by requesting it from the host's corresponding interfaces.

As the diagram shows, both the general practitioner and the pharmacist act as users to the POMP and their own systems. Many of the tasks required for the POMP are still directly executed in the existing interfaces of the original medical information systems, such as updating patients' drug usage. Only the specific functionality added by the POMP is performed directly through the platform's interface, such as the requesting and sending of drug interaction effects. Furthermore, the POMP exchanges information with the components built into the original systems, as its functionality demands.

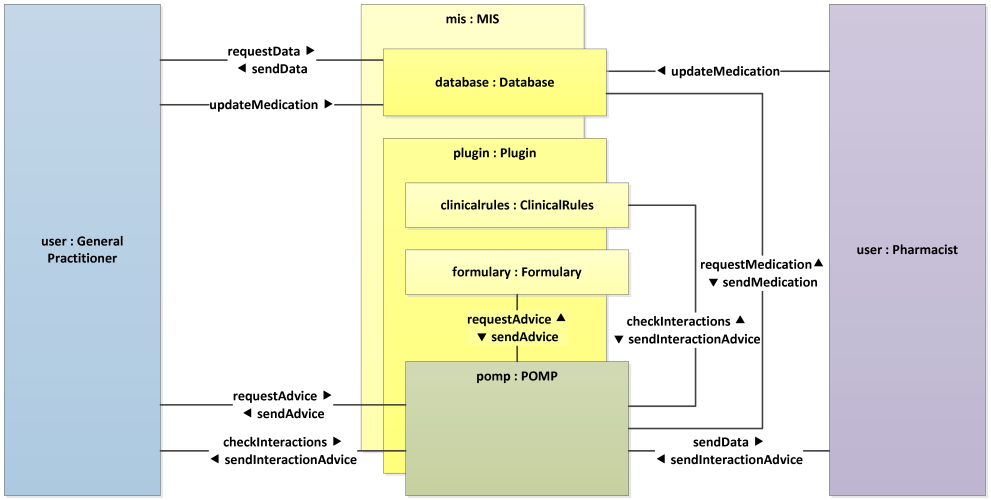
FIGURE 7: **UML collaboration diagram of the first and third development strategies of the POMP.**



In Figure 8, a collaboration diagram for the second development strategy is displayed. In this strategy, the POMP does not communicate directly with the medical information systems but

uses intermediary software. Technically, it would be a component of the third-party plug-ins and use their architecture directly. As can be derived from the diagram, users still use their own systems and the POMP's interfaces in a similar way. However, the POMP relies more heavily on the components built into the plug-in, such as formularies and clinical rule-based systems. In this scenario, patient and drugs data are still collected as directly as possible through the medical information system.

FIGURE 8: **UML collaboration diagram of the second development strategy of the POMP.**



Technological Process Modeling

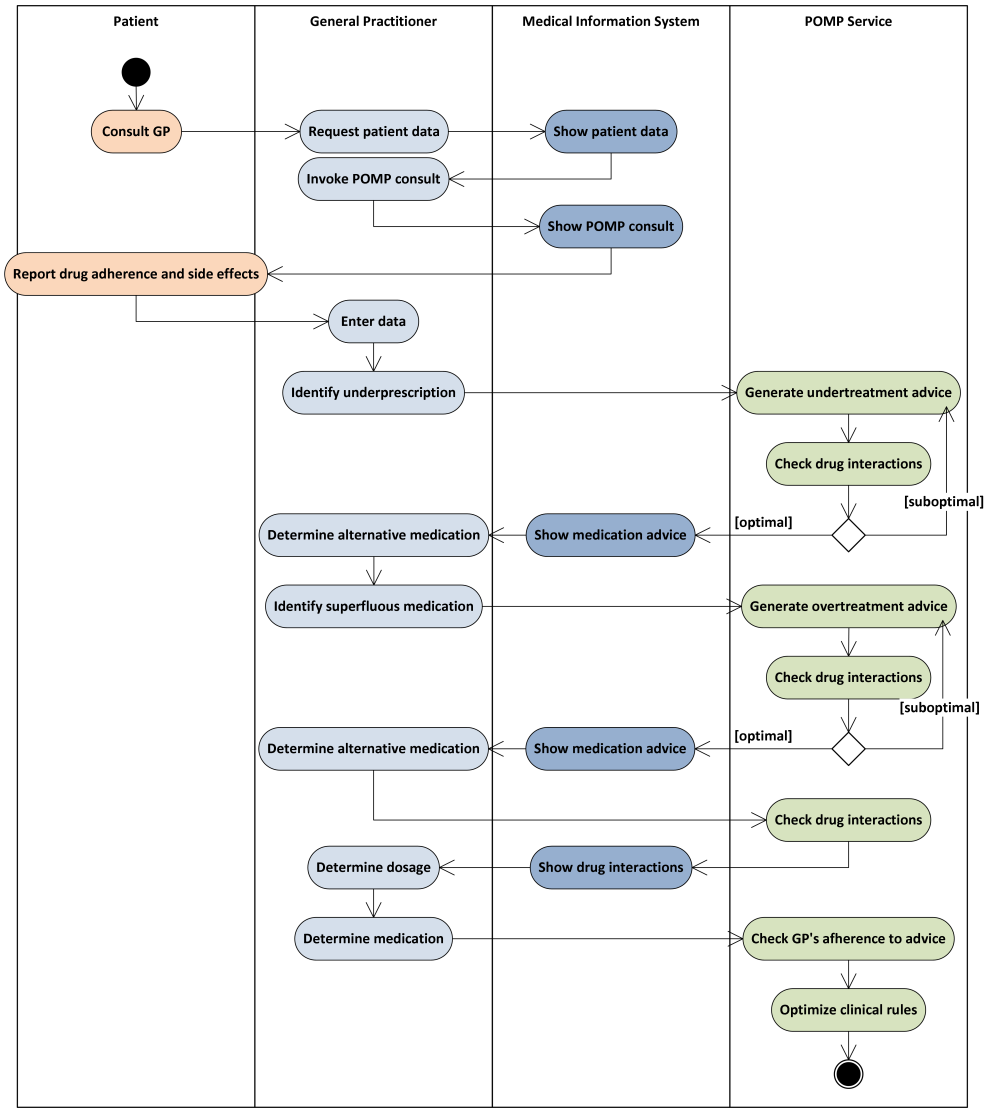
Having determined the high level technological architecture the system is to be built on, the procedural diagram created before can be adapted to accommodate the technological perspective. Through such a diagram, the user perspective on performing the method within the system can be visualized, including the responses executed by the system itself.

In Figure 9 a UML activity diagram depicting the POM process as performed within the POMP system is showed. It was created by fitting the original procedural activity diagram in Figure 2 into the conceptual and technological perspectives, displayed in Figure 7 and Figure 8. The four columns display the actors performing their specific tasks; the leftmost actor is the patient consulting his GP, the second one is the physician performing the POM tasks and operating his own software, the third is the medical information system answering the GP's requests, and the last one is the POMP service responding to the GP's requests through his own system.

Considering the integration of the POMP within the GP's original systems, the service only responds when absolutely necessary. Early steps of the method, such as updating patients' medication lists are still executed within the medical information systems. The actual checking of prescription errors and drug interactions is performed by the POMP service, but its results are displayed in the original software. After determining the medication the user's role comes to an

end, but the system's self-learning capabilities check to what extent the given advice has been followed by the GP and it updates its clinical rules accordingly.

FIGURE 9: **Activity diagram of the POMP.**



SOCIETAL FEASIBILITY

As a final step in any feasibility study that strives to contribute to the solution of socially relevant issues, an assessment of societal gains should be made. Although not all software products aim at solving societal problems, for many this is a vital step. As feasibility studies

tend to be carried out not only in entrepreneurs' own interests but especially in those of funding agencies, investigating communal gains is relevant. In the POMP project a simple yet defendable cost-benefit analysis was employed in this respect. Although it is out of scope to go into too much detail regarding its specific outcome in this text, some generally applicable approaches and factors inherent to the medical technological field are worth describing.

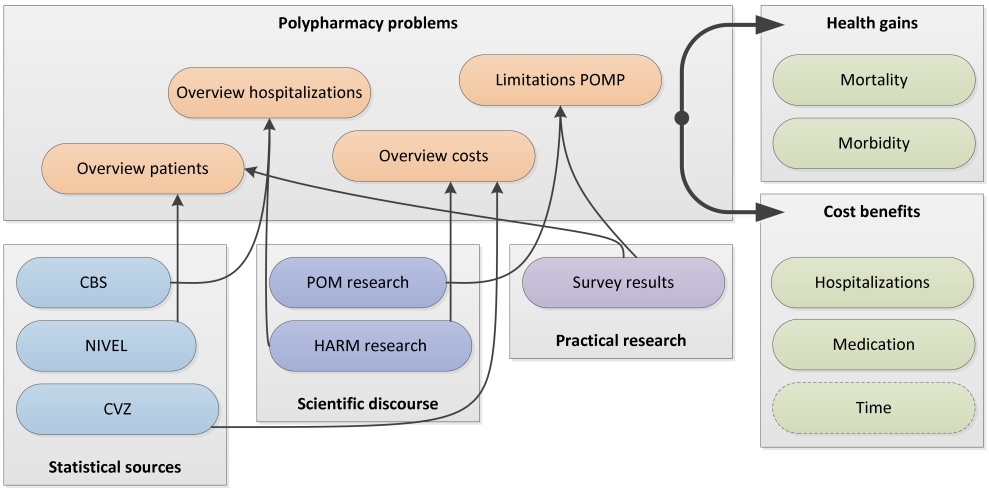
Specific to societal gains in the medical field is that benefits can not only be described in monetary terms but in health itself; health gains can be quantified in reduced mortality and morbidity rates, the latter of which implicates prolonged healthy living. Monetization of these factors is possible by estimating costs associated with provided medical care on the one hand, and persisting productivity gains on the other. The benefits' consequences can be both direct, such as reduced medicine costs, and indirect, such as improved productivity.

In the POMP project the researchers attempted to make a calculated guess as to the direct implications the successfully implemented POM Platform would have. Indirect consequences were disregarded as it proved impossible to make valid calculations with the limited test results. In Figure 10 the different components used for the societal analysis are shown. In the grey square on the right the direct implications that were measured are depicted. The health gains (mortality and morbidity) were assessed independently from the monetary ones; of the direct consequences only hospitalization and medicine costs were calculated. GPs' time spent on patients proved impossible to take into account as the POM's test results did not decisively promise reduction.

In order to determine the factors necessary for the calculations (the ones depicted in the orange ovals) the limited sources available were used to full effect. Where possible, available statistical sources were consulted; for example, data from various Dutch statistics agencies were used to determine the number of hospitalizations per year in the Netherlands, their average cost, and the national number of patients and GPs. The results from scientific papers were used to quantitatively determine polypharmacy issues, such as the number of hospitalizations related to medicinal errors. Finally, quantitative data on the number of polypharmacy patients and the POMP's possible adoption were extracted from the survey results.

Using these numbers in simple equations the POMP's potential societal gains could be calculated. It appeared that the number of lives the implemented system could save ranged from five to twenty. The annual costs that could be saved numbered in the tens of millions of Euros.

FIGURE 10: **Diagram of the POMP's expected societal gains.**



RECOMMENDATIONS

In this chapter, a case study of the conduction of a feasibility analysis of a decision-supportive knowledge system in the Dutch primary care sector was described. The usefulness of feasibility analyses is undisputed in literature and often called for in practice [136]. Standardized methods for the conduction of feasibility studies, however, do not exist. Considering this lack of formalization, various authors have proposed structured approaches to assess projects' feasibility beforehand [196,98,136], which do not offer step-by-step methods but rather provide areas that should be investigated.

The description of feasibility analyses in this chapter was meant to shed some light into their conduction in the field of medical informatics. By no means claiming to be all-encompassing, a summary of aspects worthy of exploring is included here as recommendations for researchers and entrepreneurs pursuing similar projects. These aspects, abbreviated as CORETEST, are:

- Conceptual feasibility; one of the earliest steps in any feasibility study should be the development of a conceptual model of the intended venture. Though it may seem too straight-forward to be included as a separate step, it should be investigated unequivocally. If multiple partners are included, it is essential that consensus is reached on the conceptual design before any other analyses are pursued. Diagrams suited for these tasks can either be designed freely, or through high-level UML templates.
- Organizational feasibility; exploring the market and its key players is an important early step, so the intended venture's final design can be accommodated to take any irregularities into account. Simultaneously, not only competitors but potential partners as well can be identified in this step, as such acting as precursor for the

economic analysis. Diagrams can still be modeled freely in this step, although more formal techniques such as ArchiMate can be employed to map the specific relations of players to the market and the project.

- Economic feasibility; based on the market exploration, optionally partnerships with key players can be formed. Then development strategies for the intended venture can be drawn up, each one preferably involving different partners and thereby limiting the risks of cooperation. Predictions regarding actual software purchase and usage can be made through literature and survey results. Software ecosystems can then be used to formalize the financial and distributive details.
- Technological feasibility; although detailed technological approaches cannot yet be modeled in preliminary stages, some aspects of the cooperation between the newly envisioned product and already existing ones can be visualized. Both the information exchange occurring between those systems and the usability approach from a technological perspective can be modeled. UML collaboration and activity diagrams are well suited for representing these perspectives.
- Societal feasibility; finally, predictive calculations can be made to determine the societal gains resulting from implementing the proposed project. Based on the field, these gains can be monetized in different ways; in the medical field, societal gains would usually comprise reduced mortality and morbidity rates, and as such improved health. Direct consequences and, if available, indirect ones can be calculated and lead to a defendable calculation of cost benefits.

CONCLUSION & FURTHER RESEARCH

In this chapter the authors described the practice of conducting a feasibility analysis of a decision-supportive knowledge platform in the Dutch primary care sector, and presented recommendations accordingly. The research question, "how can the feasibility of a decision-supportive knowledge platform in the primary care domain be investigated?" was answered by describing the aspects mentioned earlier as CORETEST-recommendations: by investigating the project's conceptual, organizational, economic, technological and societal feasibility.

Not within the scope of this chapter are the practical validity checks of the various feasibility components that would occur in later phases. The implementation of early versions of the software product in practical testing environments, for example, would yield results indicating to what extent the conceptual and technological predictions were adequate. In a similar fashion, the approximation of societal gains, albeit admittedly calculated with limited data, would have to be proven correctly in practical application. Finally, the organizational and economic aspects researched could be used in a business model, the worth of which would be found after implementation. In conclusion, the long-term applicability of the practices advocated in this text is impossible to assess, and forms a limitation.

Other limitations include the ones following the action research approach. As the roles of researchers and participants in the feasibility analysis process intertwined, biases regarding observations and descriptions may not have been avoided completely. These were minimized, however, by the usage of various different methods to study the intended phenomena.

The authors aimed at contributing to the limited body of knowledge available on the conduction of feasibility analyses in general, and especially in the medical informatics sector. While various authors have attempted to introduce structured approaches to these studies, in this text some factors not mentioned in any of their works were deemed relevant. Therefore the authors advocate the development of a generally applicable standardized framework comprising aspects universally relevant to feasibility studies, based on which applicable methods could be generated.

PART III

Design & Development

What's important is that twice two is four; all the rest is nonsense!

Ivan Turgenev, *Fathers and Sons*

3.1

STRIPA: A Rule-Based Decision Support System for Medication Reviews in Primary Care

ABSTRACT

The chronic use of multiple medicinal drugs is growing, partly because individual patients' drugs have not been adequately prescribed by primary care physicians. In order to reduce these polypharmacy problems, the Systematic Tool to Reduce Inappropriate Prescribing (STRIP) has been created. To facilitate physicians' use of the STRIP method, the STRIP Assistant (STRIPA) has been developed. STRIPA is a stand-alone web-based decision

Meulendijk, M., Spruit, M., Jansen, P., Numans, M., & Brinkkemper, S. (2015). STRIPA: A Rule-Based Decision Support System for Medication Reviews in Primary Care. *ECIS 2015 Research-in-Progress Papers. Paper 29.* Münster, Germany.

support system that advices physicians during the pharmacotherapeutic analysis of patients' health records. In this paper the application's architecture and rule engine, and the design decisions relating to the user interface and semantic interoperability, are described. An experimental validation of the prototype by general practitioners and pharmacists showed that users perform significantly better when optimizing medication with STRIPA than without. This leads the authors to believe that one process-oriented decision support system, built around a context-aware rule engine, operated through an intuitive user interface, is able to contribute to improving drug prescription practices.

INTRODUCTION

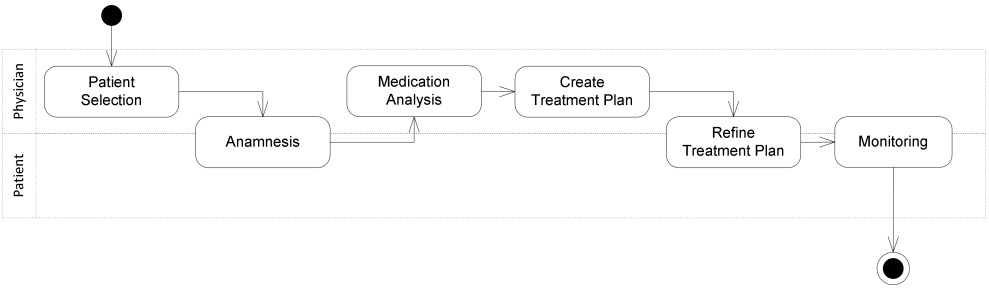
The chronic use of multiple medicinal drugs is growing. In the Netherlands alone, seventeen percent of the chronically ill use more than five different drugs permanently; half of these patients are over seventy years of age [198]. Over ten percent of Dutch hospital admissions are related to medication use [129]. This development, known as polypharmacy, has been demonstrated to lead to a variety of clinical problems for its users, including an increasing risk of adverse effects, under-prescribing, overtreatment, and decreased drug adherence [24,41,73,125,190,192,197,234].

Many of the problems persist because individual patients' drugs have not been adequately prescribed by primary care physicians; their dosages may be too high, or they may be incompatible with each other altogether. The causes for these problems are in many cases avoidable; errors on the part of the general practitioner or pharmacist, such as time pressure, carelessness, and the use of incomplete health records have a major impact [214,183].

In order to reduce these polypharmacy problems, a structured method has been developed: the Systematic Tool to Reduce Inappropriate Prescribing (STRIP) is a drug optimization process, consisting of an anamnesis of patients' actual perceived problems and drug use, and a pharmacotherapeutic analysis. It has been included in the Netherlands in a multidisciplinary guideline on polypharmacy and elderly patients [62]. The method leads to a newly refined individual treatment plan, omitting superfluous drugs and adjusting dosages where applicable. The STRIP process is shown in Figure 1.

To ensure that the STRIP method is incorporated into GPs' and pharmacists' daily practice, a software tool has been developed to facilitate the most time-consuming and complex step in the process: the pharmacotherapeutic analysis.

FIGURE 1: **UML Activity Diagram depicting the STRIP process.**



Status Quo of Clinical Decision Support

In the last decade, computerized physician order entry (CPOE) systems have changed from being mainly organizational aids to supporting the GP with the prescription process. Apart from the ability to register consultations, primary care information systems now also contain medical formularies, electronic medical records, and decision support systems [111].

Nonetheless, the adoption of these systems by general practitioners is suboptimal. A recent study shows that GPs owning decision support systems use them for only one quarter of their prescriptions [170]. Another study showed that only twenty percent of GPs owning decision support systems actually use them during consultations [137]. A survey among Dutch physicians found that more than half of the owners of decision support systems rarely or never use them [142].

As to why potentially beneficial systems are underused, studies give different answers. Some show that the traditional indicators of technology adoption, perceived usefulness and perceived ease-of-use, are at least somewhat of influence to this group's attitudes as well [117,39,238]. Studies exploring influential factors particular to physicians have shown that a.o. output quality, embedding in practice, and especially time efficiency are important indicators of GPs' attitudes towards software [27,212,238,39].

Design Objectives

In an attempt to overcome these problems and create a useful and easy-to-use decision support tool, we developed the STRIP Assistant (STRIPA). The design objectives when developing STRIPA were to create a stand-alone web application that effectively and efficiently facilitates physicians' use of the STRIP method, by optimally advising them during the pharmacotherapeutic analysis. In order to avoid the 'alert fatigue' that users complain about in comparable information systems, the decision was made not to disrupt their everyday routines, but instead design the application as an independently invokeable process [116]. This decision also enabled the creation of an autonomous user interface, free from restrictions that regularly limit plug-in applications and lead to suboptimal workflows.

STRIPA was designed to directly react to users' actions, generating its advices in response to users adding or removing drugs. In this approach it differs from most decision support systems integrated in CPOE systems, which base their advices on input available at specified moments in time (e.g. upon opening a patient's health record). To ensure that only relevant advices are displayed, the rule engine was designed to incorporate context-specific characteristics, and to work with both complex and simple rules.

While taking advantage of the independence from existing information systems, industry standards were incorporated into STRIPA to ensure successful communication with existing CPOE systems. Complete health records can be successfully transferred between STRIPA and third-party applications, leaving the classification systems of the underlying objects (e.g. drugs and diagnosed diseases) intact.

STRIPA PROTOTYPE

Architecture

The STRIP Assistant has been developed as a stand-alone web service, relying on Java and MySQL in the back-end and on JavaScript in the front-end. The communication between the front- and back-ends is facilitated through AJAX, using JSON as data format for its brevity. The expert system in the back-end is powered by the Drools rule engine [173].

Figure 2 depicts STRIPA's sub systems, their most important components, and the interfaces they communicate through. The MySQL database holds all patient records and clinical data required to execute the decision rules. A database management system (DBMS) provides the query capabilities necessary to use it.

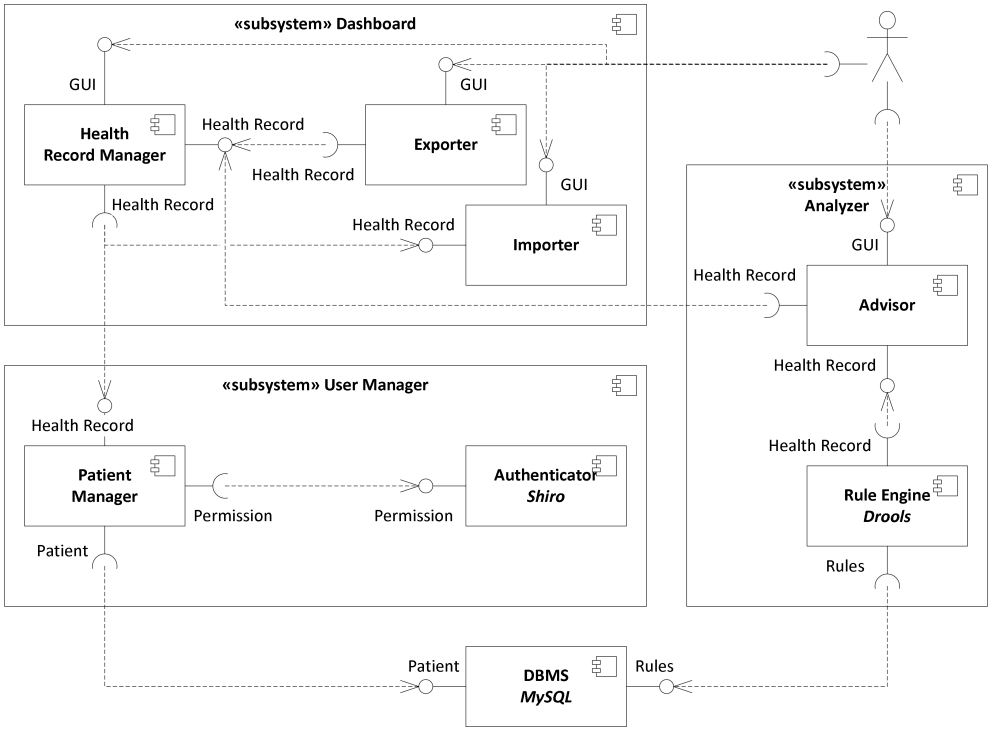
The User Manager sub system has two primary functions; it manages the current user's session and his or her specific permissions, and it supplies the rest of the system with requested patient data. Acting as a gatekeeper, every request is authenticated before it is executed and its results are returned to the calling function. The open-source Java security framework Shiro is used to manage user sessions and authenticate requests [13].

The Dashboard sub system provides a user interface for recording the results of patients' anamneses. An anamnesis typically provides an individual's diagnosed diseases, complaints, prescribed drugs, self-medication, and recorded measurements. These values are either obtained or validated through communication with the patient, but GPs' and pharmacists' information systems can serve as initial sources of patient data. The Importer component provides users the ability to upload health records from third-party sources. These can then be edited through the user interface provided by the Health Record Manager component.

Finally, the Analyzer sub system provides a user interface for performing the pharmacotherapeutic analysis. The changes made to the health record during the process are sent to the Rule Engine component, which holds them in working memory. The appropriate

rules are then executed and its results returned to the Advisor component, which shows them to the user. The user is free to heed or reject the advices provided by the rule engine. After completing an analysis, a patient's health record is updated by the Health Record Manager component, which in turn updates the patient data in the database. When required, the analysis' results or complete health record can be exported to a third-party application through the Exporter component.

FIGURE 2: **UML Component Diagram depicting STRIPA's architecture.**



Decision Rules

The rules included in STRIPA's decision support system come from a variety of sources. Some, such as detections of drugs having clinical interactions or having the same active components, are provided as datasets by (inter)national organizations. Others, especially the rules incorporating more contextual variables, have been established by expert panels [52]. All rules' parameters are given unique ids to ensure they can be reliably detected by the expert system. Classification systems that have been implemented are a.o. ATC for drugs and ICPC for diseases [232,230].

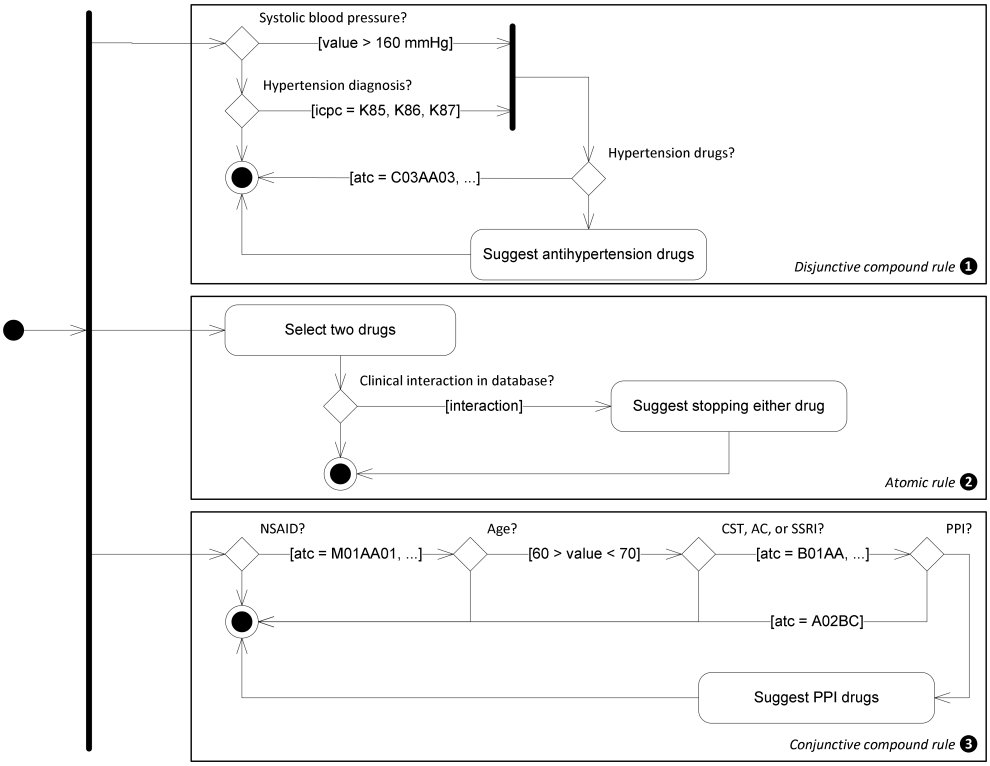
The expert system in STRIPA relies on rules with a varying degree of complexity. All rules, however, result in advices that can either be heeded or rejected by the user. Advices can propose

to start new medication, stop specific drugs, or change the dosage or frequency of a drug already used; any or all of these actions may be combined in any given advice. Providing multiple options (stop Drug A or Drug B) or even combinations of options (either start Drug A and Drug B, or start Drug C) is possible. The conditions that triggered the rule are included in the advice as well.

Some of the rules in STRIPA are as simple as 'Drug A and Drug B may not be used simultaneously', some require a multitude of variables. Because of this diversity, an implementation choice had to be made that was flexible, easily extendable, and did not require double declarations for conceptually similar rules.

Because of its flexibility, platform-independency, and wide-ranging capabilities, Drools was chosen as an adequate rule engine for STRIPA [173]. Both simple and complex rules can be modelled easily in Drools, without redundancy. Its ability for forward-chaining inference enables explaining to end-users how given advices were produced (i.e. which specific causes triggered the rules).

FIGURE 3: **UML Activity Diagram depicting rules suggesting addition or discontinuation of drugs, if necessary conditions are satisfied.**



The medication review domain could be modelled using three distinguishable types of rules: those based on atomic formulae, conjunctive compound formulae, and disjunctive compound formulae. Atomic formulae depend on a single condition only, containing no deeper propositional structure. Compound formulae, in contrast, do contain logical connectives to incorporate multiple conditions. While conjunctive operators require several conditions to be satisfied before a consequence is implied, disjunctive operators require only one of several conditions to be met. These three types of rules have been implemented in STRIPA's rule engine; the ability to incorporate multiple types of rules is one of the characteristics which sets STRIPA apart from other decision support systems in primary care. Examples of each type of rule are shown in Figure 3.

The atomic rule, of which hundreds of thousands of possible variations exist, was modelled by creating a single rule that could be triggered by any combination of objects validating the condition. Rule (2) in Figure 3 illustrates how this principle works. Any two drugs are checked for potentially dangerous clinical interactions in the database. If a match is found, an advice is created recommending the user to remove either one of the conflicting drugs.

The more complicated conjunctive compound rules, which depend on several conditions, were modelled independently. A table format was used to allow for more convenient correcting and editing by members of the expert panel, which was later automatically converted into code. Rule (3) in Figure 3 illustrates the working of a complex rule: the suggestion of adding a proton pump inhibitor, dependent on patients' age and their use of NSAIDs (non-steroidal anti-inflammatory drugs) and SSRIs (selective serotonin reuptake inhibitors), corticosteroids, or anticoagulants. Only if all conditions are satisfied (including, obviously, the current lack of a proton pump inhibitor) a suggestion is made.

Finally, some rules can be triggered by several (combinations of) conditions. Often, these rules have a preference for one (combination of) conditions over another; a more precise measurement is usually preferred over a more generic diagnosis. Rule (2) in Figure 3 presents an example of such a disjunctive compound rule which can be activated through several ways: recommending the addition of antihypertension drugs in case a patient suffers from untreated hypertension. Ideally, hypertension is determined by one's systolic blood pressure being higher than 160 mmHg. However, if this value is unavailable, diagnoses indicating hypertension are considered. In either case, a suggestion to add antihypertension drugs is given.

User Interface

Taking into account the literature on user interface mistakes in CPOE systems, much attention was paid to designing the user interface for STRIPA. The aim was to create an application that had the potential to be accessed through different mediums, i.e. both PCs and mobile devices. A study of non-functional requirements of medical apps was conducted, and revealed that aging users attribute more importance to user-friendliness than younger ones [141]. After designing an initial wireframe version based on interviews with medical experts and potential users,

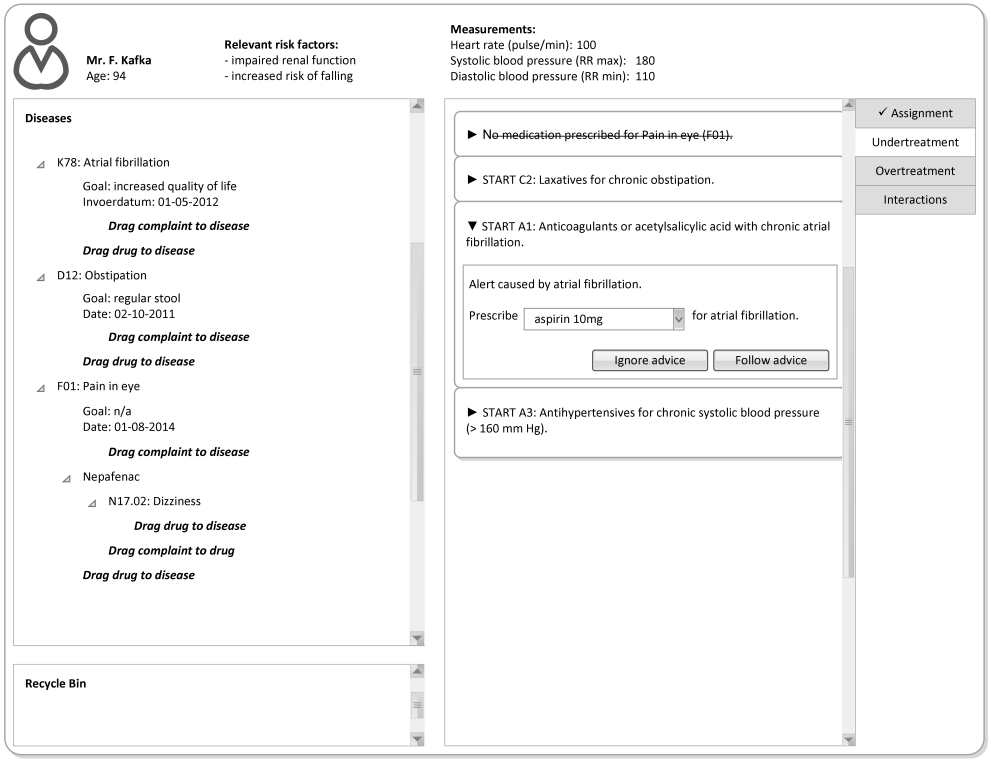
a prototyping process using prototypes of increasing fidelity was used to refine it [223]. Early prototypes of the application were used in test sessions, where users were invited to comment on its usefulness and user-friendliness. Their remarks were used to further improve the user interface. In Figure 4 the preliminary design of the user interface is displayed, before it was refined through the prototyping method. The appendix contains a screenshot of the final version of the user interface of the analyser component.

In designing, Nielsen's broadly applied usability heuristics served as guidelines [154]. In essence, the aim was to create an aesthetically pleasing, uncluttered interface that allowed users to effectively and efficiently perform the pharmacotherapeutic analysis according to the STRIP method (aesthetic and minimalist design). The purpose of the method is to create a newly refined list of a patient's diseases and diagnoses. Since this is a core aspect of the application, the main panel containing the new list is always in view. This also holds for the current patient's personalia and relevant measurements, which are displayed in the top row. The steps of the STRIP method (e.g. checking for undertreatment, over-prescribing, clinical interactions etc.) are shown on the far right, with the currently selected step coloured differently from the others (visibility of system status). The generated advices are shown together with their causes and textual explanations. All possible actions users can take (e.g. adding or removing drugs) are shown only when they can actually be performed; consequently, all available options are immediately visible (recognition rather than recall). Throughout the interface, concepts and terms that are either common medical knowledge (e.g. episodes for diagnosed diseases) or have a solid basis as computer concepts (e.g. recycle bin to store discontinued drugs) are used (match between system and the real world, consistency and standards).

The STRIP workflow consists of six steps, which were translated into six advice-supported panels in the interface. While users are guided from one step to the next with a Next-button, they do not have to follow this designated order. They can review earlier steps in later phases, or skip unnecessary checks for uncomplicated cases (flexibility and efficiency of use). Assigning drugs to diagnosed diseases is a core principle of the STRIP method; in the application this feature was extended by allowing users to further assign side effects to drugs that cause them. This approach results in a structured health record showing the relations between diseases, complaints, and drugs. The assignment of drugs to diseases has been implemented through a drag 'n drop mechanism.

The system is designed in such a way that technical errors are hidden from the user. Users can always recover from their own mistakes through a single action, such as dragging back a drug that was accidentally placed in the recycle bin to its original position (error prevention, help users recognize, diagnose, and recover from errors). Each panel contains a context-aware help-button which shortly tells the user what is expected of him or her in the current step. An instructional video explaining the use of the application is available on-demand (help and documentation). This demonstration video can be viewed here: http://videodemo.stripa.eu/english/ [145].

FIGURE 4: **Initial wireframe user interface design.**



Semantic Interoperability

Essential to the usefulness and time efficiency of any application relying on external data, is its ability to communicate with third-party applications. In the primary care sector sharing data is notoriously difficult because of its sensitive nature and multitude of (often incompatible) information systems. While a more direct and safe approach is being researched, STRIPA currently uses an old but proven health exchange format known as MEDOVD [153]; it is accessed by users through the Importer and Exporter classes in Figure 2. MEDOVD is a de facto standard based on the EDIFACT format that can be locally exported and imported by physicians, and which is supported by all Dutch CPOE systems [140]. It can contain complete health records and is flexible enough to include different classification systems.

While international classification standards for drugs and diseases exist, most countries have implemented their own modified versions, or even completely disconnected ontologies; the Netherlands is no exception. A Dutch extension of the ICPC ontology for diseases is widely used in primary care, as well as custom national standards for dosage and measurements. In STRIPA, objects (such as drugs and diseases) have been designed in such a way that they can contain values from multiple classification systems. As cardinalities between these ontologies

are not necessarily one-to-one, objects can contain several values of some classification systems. For example, diseases may have a single ICPC value, but multiple entries of the more specific ICD10 classification [231]. STRIPA's rule engine relies on coded objects, and rules incorporate either the most common or the most specific system of classification. To enable the use of different ontologies as input for the application, conversion rules have been implemented to convert values from one classification system to their equivalents in another.

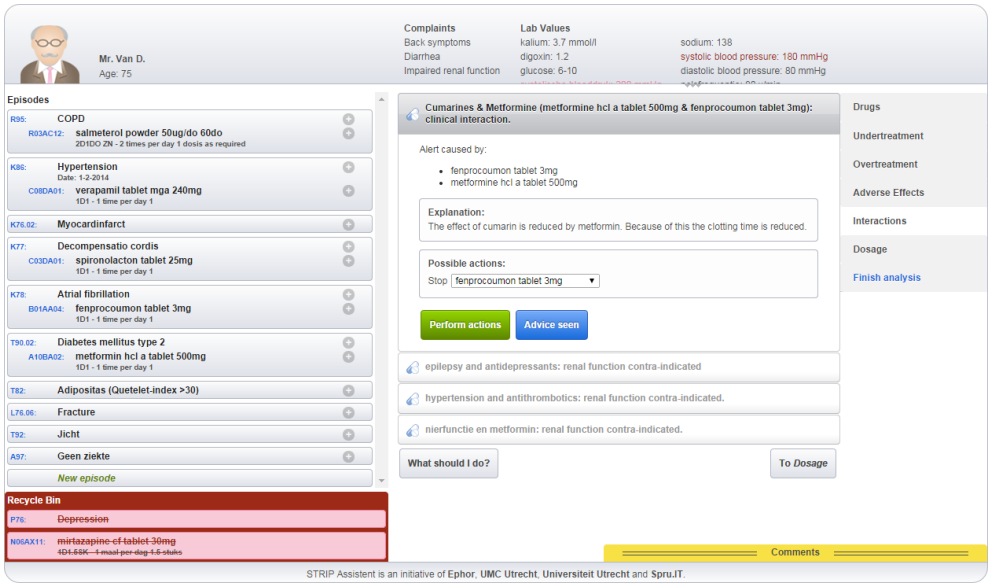
EVALUATION & CONTRIBUTION

At the moment of writing, multiple studies investigating STRIPA's effectiveness and efficiency are being conducted. The results of these studies have not yet been published, but early analyses show that caretakers perform significantly better when optimizing medical records with help by STRIPA than through their usual care practice: 55% of their unsupported decisions are correct, versus 76% of their choices made with the decision support system. Subjective comments of users in this experimental study included unfamiliarity with the interface ("I wanted to bisect [the dosage for] digoxin, but couldn't.") and concerns about the output quality ("The system failed to recognize the impaired renal function [...] resulting in a suboptimal drug balance."). The current prototype does increase the time practitioners require to perform medication reviews. Among the primary concerns of the developers are ways to decrease the time needed to complete analyses, by improving the relevance of the generated advices and through a more complete integration with existing information systems.

In this paper the authors strived to demonstrate how an elegantly implemented decision support system could contribute to more effective and efficient medication reviews in primary care. The preliminary results of an experimental study give reason for optimism. This leads the authors to believe that one process-oriented decision support system, built around a context-aware rule engine, operated through an intuitive user interface, is able to contribute to improving drug prescription practices. A product-ready version of the STRIP Assistant is planned for release in 2015, making the tool available for primary care practitioners in the Netherlands. Further research will focus on performing more extensive analyses of the studies into STRIPA's effectiveness and efficiency.

APPENDIX: STRIPA SCREENSHOT

FIGURE 5: Screenshot of the STRIP Assistant generating a recommendation to stop fenprocoumon. Heeding this advice would result in fenprocoumon tablet 3mg in the left overview being removed to the Recycle Bin.



PART IV

Demonstration & Evaluation

Logic may indeed be unshakable, but it cannot withstand a man determined to live.

Franz Kafka, *The Trial*

4.1

Computerized decision support improves medication review effectiveness

ABSTRACT

Background: Polypharmacy poses threats to patients' health. The Systematic Tool to Reduce Inappropriate Prescribing (STRIP) is a drug optimization process in primary care for conducting medication reviews. To effectively and efficiently incorporate this method into daily practice, the STRIP Assistant has been developed, a decision support system that

Published as: Meulendijk, M., Spruit, M., Drenth-van Maanen, A., Numans, M., Brinkkemper, S., Jansen, P., et al. (2015). Computerized decision support improves medication review effectiveness: an experiment evaluating the STRIP Assistant's usability. *Drugs & Aging, 32*(6), 495-503.

aims at assisting physicians with the pharmacotherapeutic analysis of patients' medical records. It generates context-specific advice based on clinical guidelines.
Objective: The aim of this study was to validate the STRIP Assistant's usability as a tool for physicians optimizing medical records of polypharmacy patients.
Methods: In an online experiment, 42 physicians were asked to optimize two comparable medical records of polypharmacy patients, one in their usual manner and one using the STRIP Assistant. Changes in effectiveness were measured by comparing respondents' optimized medicine prescriptions with medication prepared by an expert panel of two geriatrician-pharmacologists. Efficiency was operationalized by recording the time respondents took to optimize the two cases. Satisfaction was measured with the System Usability Scale (SUS). Independent and paired t-tests were used for analysis.
Results: Medication optimization significantly improved with the STRIP Assistant. Appropriate decisions increased from 58% without the STRIP Assistant to 76% with it; p<0.0001. Inappropriate decisions decreased from 42% without the STRIP Assistant to 24% with it; p<0.0001. Participants spent significantly more time optimizing medication with the STRIP Assistant (24 minutes) than without it (13 minutes); p<0.0001. They filled out a below-average SUS-score of 63.25.
Conclusion: The STRIP Assistant improves the effectiveness of medication reviews with polypharmacy patients.

BACKGROUND

Polypharmacy and Inappropriate Prescribing

Polypharmacy, or the chronic use of multiple medicines, poses significant threats to patients' health. A consensual definition of polypharmacy is lacking, but it is often described as the concurrent use of five or more different chronically used drugs [110]. Polypharmacy has been associated with negative health consequences. Drugs may cause clinical interactions or adverse effects that may aggravate patients' symptoms instead of relieving them. Medicine issues including underprescribing, overtreatment, and decreased drug adherence have been associated with polypharmacy [24,41,73,125,190,192,197,234]. A 2008 study showed that in the Netherlands, 5.6% of all acute hospital admissions have medication-related causes [129]. For elderly patients, who constitute half of all chronically ill polypharmacy patients, this figure was twice as high [198].

The concurrent use of multiple medications is not entirely undesirable, as in many patient cases polypharmacy is indicated or even unavoidable. However, inappropriate prescribing of medications is prevalent amongst elderly patients [74]. An incidence-focused study found that inappropriate medication use increased elderly persons' risks of hospitalization and mortality

[54]. Geriatric assessment and medication review have been shown to be effective methods in aiding prescribers with optimizing polypharmacy [57,187].

A multitude of initiatives has been developed to assess the appropriateness of drugs prescribed for individual patients. These approaches can be divided in implicit and explicit methods. The former implicit methods use patient-specific information, combined with medical knowledge, to determine medication appropriateness, while the latter explicit methods provide screening tools, containing lists of clinical interactions or contraindications [110]. Amongst the explicit methods are the Beers criteria and the Screening Tool to Alert to Right Treatment (START) and Screening Tool of Older People's Prescriptions (STOPP) criteria, while the implicit methods include the Medication Appropriateness Index and the pharmacotherapy review focused on drugs' use, indication, safety, and effectiveness (GIVE) [36,75,90,130]. The effectiveness of these interventions varies; generally they appear beneficial in terms of reducing inappropriate prescribing and medication-related problems, but they have not been proven to lead to clinically significant improvement [167].

In order to improve medication prescribing in primary care, several implicit and explicit methods have been combined into an all-encompassing systematic medication review approach, the Polypharmacy Optimization Method (POM). It has been shown to significantly improve GPs' prescriptions for polypharmacy patients in an experimental setting [57].

A variety of barriers is impeding the widespread adoption of structured medication reviews in daily practice. Recently Anderson et al. conducted a systematic literature review on enablers and barriers to minimizing potentially inappropriate medications by GPs. Most factors revolved around physicians, and included inertia (his or her attitudes towards discontinuation, such as fearing negative consequences), self-efficacy (his or her knowledge and available information on the topic), and awareness (his or her having poor insight or discrepant beliefs). Barriers that were not physician-related included a lack of resources, patients resisting changes to their medication, and practical and cultural factors [10]. A separate study focusing on barriers regarding pharmacist-led medication reviews reported lack of time and lack of self-confidence as the most commonly perceived barriers [157].

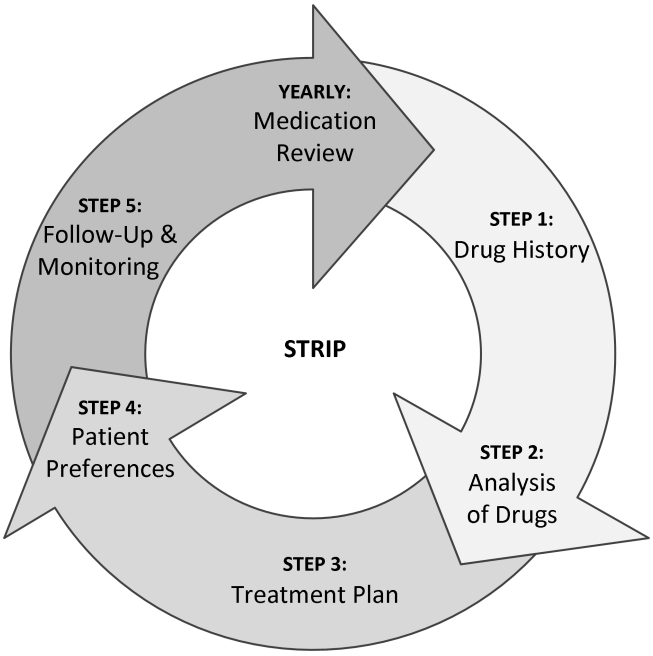
STRIP Recently, the POM, GIVE, and START and STOPP criteria have been combined into the Systematic Tool to Reduce Inappropriate Prescribing (STRIP), which has consequently been included as part of a Dutch multidisciplinary guideline on polypharmacy in elderly patients [62]. The STRIP has been designed to be an all-encompassing drug optimization process in primary care, focusing not just on the pharmacotherapeutic analysis but also on patients' medication histories and preferences; Figure 1 shows the STRIP method's different steps.

The STRIP analysis is more extensive than its predecessors [57,130,75]. It combines both the implicit approaches of the POM and the GIVE, and the explicit lists of the first version of the START and STOPP criteria. The pharmacotherapeutic analysis in the STRIP includes checks on underprescribing, overtreatment, recommended dosage adjustments, drug effectiveness, potential adverse effects, dose frequency, clinical interactions, and medication adherence,

including practical problems with medication use. The START and STOPP criteria are implemented in the pharmacotherapeutic analysis. This extensive medication review results in a patient-specific treatment plan in which new drugs are gradually added and superfluous ones discontinued. This approach to conducting structured medication reviews is based on consensus rather than evidence, synthesizing the results of the earlier optimization methods mentioned above. Currently, solid evidence for choosing specific strategies for the optimization of pharmacotherapy in the elderly over others is lacking [205].

Involvement of patients in the medication review is emphasized to guarantee their therapy adherence; patients' preferences are taken into account as much as possible. The pursuit of the treatment plan is monitored through regular communication between practitioner, pharmacist, and patient. The involvement of pharmacists in medication reviews, as part of multidisciplinary teams, has been shown to lead to improved pharmacotherapy for older patients [194]. Educating patients on their medication use and treatment goals, simplifying their drugs regimens, and preventing adverse drug reactions have all been identified as factors influencing patients' adherence to their treatments [205].

FIGURE 1: **The five steps of the STRIP method depicted as a yearly repeating cycle (STRIP = Systematic Tool to Reduce Inappropriate Prescribing).**



Clinical Decision Support Systems

In recent years, computerized physician order entry (CPOE) systems have gradually changed in terms of functionality. From systems that were traditionally organizational in nature, they have

been enhanced to facilitate management of electronic medical records and clinical decision support [209]. There is consensus in literature that clinical decision support has the potential to improve GPs' and pharmacists' decision-making [32]: "Both commercially and locally developed CDSSs are effective at improving health care process measures across diverse settings". Evidence for concurrent improvement in efficiency, cost-effectiveness, or clinical effectiveness is inadequate or ambiguous. A study investigating the attitudes of Dutch GPs to the introduction of a decision support system specifically aiding them with conducting medication reviews revealed that the majority were positively inclined to using such a system [142].

STRIP Assistant In order to enable GPs and pharmacists to effectively and efficiently incorporate the STRIP method into their daily practice, the STRIP Assistant has been developed. The STRIP Assistant has been designed as a stand-alone web application that aims at assisting GPs and pharmacists with the pharmacotherapeutic analysis of patients' medical records. Based on patients' records and the decisions that GPs and pharmacists make during the medication review, the application generates context-specific advice. The STRIP Assistant's design decisions adhere to best practices in information science research; user interface conception and decision rule implementation have been designed to balance efficiency and information completeness, aiming to minimize previously mentioned barriers such as users' lack of confidence and lack of time [144].

The knowledge used to generate the STRIP Assistant's advice consists of well-established guidelines on clinical interactions, double-medication, contra-indications, dosage strength and frequency, and specific implementations of version 1 of the START- and STOPP-criteria [240,52]. The rules incorporate not only patients' diseases and drugs, but also their contra-indications, complaints, and relevant physical properties (such as renal functioning, and weight). This results in items of advice that recommend users to add new drugs or to remove superfluous ones, or to change dosages of existing medicines.

In the future, the STRIP Assistant has been planned to integrate with existing CPOE systems, thereby increasing the efficiency with which the method can be performed. Additionally, use of datamining techniques on historical data should reveal patterns in users' behavior towards generated advice, which could be used to improve recommendations [164].

Usability

Usability has since long been regarded as an essential factor for the success of software applications. In the widely used definition by the International Organization for Standardization (ISO), usability is defined as 'the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use' [108]. In this context, effectiveness is understood as the 'accuracy and completeness with which users achieve specified goals', while efficiency exists of 'resources expended in relation to the accuracy and completeness with which users achieve goals'. Satisfaction, finally, is the

subjective 'degree to which user needs are satisfied when a product or system is used in a specified context of use'.

A recent systematic literature review on clinical decision support systems showed that there is ample evidence that these systems can improve effectiveness; not enough research on efficiency and satisfaction is available to make generalizations regarding these aspects [32]. In technology adoption literature, it has been shown that systems' perceived usefulness and ease-of-use, aspects closely related to usability, are the major determinants of people's attitudes towards using technology [51,217,218].

Hornbaek described the current practices in evaluating usability [102]. A multitude of metrics and instruments has been used to measure the three main factors of usability identified in the ISO-definition. Measurements of effectiveness usually involve the degree to which a task has been successfully completed, leading to metrics such as accuracy, recall and completeness. Efficiency metrics mostly revolve around the time spent completing a task, but can also involve mental efforts. The subjective satisfaction criterion is often measured through standardized questionnaires or interface ranking.

OBJECTIVES

The above-mentioned considerations lead us to believe clinical decision support has the potential to successfully aid GPs and pharmacists in incorporating structured medication reviews in daily practice. Therefore, we aim to validate the STRIP Assistant instrument's usability as a tool for physicians optimizing medical records of polypharmacy patients in an experimental setting. This main research question has been divided in the following sub questions:

RQ1: Do GPs and pharmacists make significantly more appropriate decisions when optimizing the medical records of polypharmacy patients with the STRIP Assistant than without it?

RQ2: Do GPs and pharmacists make significantly fewer inappropriate decisions when optimizing the medical records of polypharmacy patients with the STRIP Assistant than without it?

RQ3: Do GPs and pharmacists take significantly less time to optimize the medical records of polypharmacy patients with the STRIP Assistant than without it?

RQ4: Do GPs and pharmacists perceive using the STRIP Assistant for the optimization of medical records of polypharmacy patients as satisfactory?

In this context appropriate decisions means decisions that correspond to those agreed upon by an expert panel.

METHODS

In order to explore to what degree the STRIP Assistant is usable for aiding GPs and pharmacists with performing medication reviews, an experiment was conducted.

Participants

The experiment was aimed at GPs and pharmacists. 52 respondents were selected through opportunity sampling, as the researchers lacked the resources to guarantee participants' cooperation through reimbursement. All participants were required to be either general practitioners or pharmacists in Dutch primary care, and had to fully complete both parts of the experiment to warrant inclusion. Of the 52 respondents, 9 had to be discarded because of corruptions in the data: 3 participants did not fill out the unassisted first part of the experiment, 5 did not assign drugs to diseases or did not respond to advice during the assisted part, and 1 record was a duplicate. Finally, 43 participants' results were eligible for inclusion in the data analysis.

Respondents were recruited through the researchers' personal networks (i.e. symposia, conferences, and (training) conventions). They were briefly informed about the experiment's goal and ensured that their anonymity would be guaranteed. As an incentive, respondents were offered three months of use of the software application with their own patients free of charge.

Study Design

The experiment took the form of a pre-experiment with a one-group pre-test post-test design, as described by 't Hart et al. [91]. Respondents were placed in a single research group; an initial test was performed, after which a stimulus was applied and the test was repeated.

In the test, two medical records of polypharmacy patients, having been earlier selected from the geriatric ward of an academic medical center for the study by Drenth-van Maanen et al., were used [57]; they were actualized (i.e. drugs that were no longer available were replaced by their contemporary counterparts) and confirmed to be of comparable difficulty by an expert panel of geriatricians specialized in clinical pharmacology (P.J. and W.K.) During the experiment, respondents were asked to optimize the first case in their usual manner and the second one using the STRIP Assistant.

The three usability aspects of effectiveness, efficiency and satisfaction were operationalized in the experiment as follows: effectiveness was measured by recording the respondents' medicine prescriptions, after their optimization. The decisions made by the respondents were then compared with the medication list that the aforementioned expert panel of two geriatrician-pharmacologists prepared. They reached consensus on the pharmacotherapeutic changes that should be made in the records optimized by the respondents, and classified the decisions as correct, neutral or potentially harmful. Efficiency was operationalized by recording the time respondents took to optimize the two cases. Finally, satisfaction was measured through a standardized questionnaire, the System Usability Scale, consisting of ten statements with which respondents had to indicate their (dis)agreement on a Likert scale.

All data was gathered between November 2013 and June 2014. During this time, no changes of any kind were made to the software.

Outcome Measures

The main outcome measure was the difference in the percentage of appropriate decisions made by the participants without and with use of the STRIP Assistant. Secondary outcome measures were the difference in the number of inappropriate decisions taken by participants without and with use of the STRIP Assistant, the difference in time needed to perform the medication review without and with use of the STRIP Assistant, and the extent to which participants experienced their use of the STRIP Assistant as satisfactory.

Instrument

The STRIP Assistant has been designed as a stand-alone web application that aims at assisting GPs and pharmacists with the pharmacotherapeutic analysis of patients' medical records. The user interface accommodates the six phases of the STRIP medication review (i.e. drugs-disease assignment, undertreatment, overtreatment, side effects-drugs assignment, clinical interactions, and dosage frequency). In most phases, users are shown advice on missing, superfluous, or incompatible drugs. The items of advice are patient-specific, incorporating their diseases, drugs, side effects, and users' actions up to that point. The STRIP Assistant's rule base consists of a combination of well-established clinical rule databases and specific implementations of the START and STOPP criteria.

For the experiment, the user interface was enhanced to first display one of the patient cases in a bulleted list, summing up his/her diseases, drugs, side effects, complaints, measurements, and lab values.

Procedure

Respondents were asked to optimize the first case in their usual manner, specifying in an adjacent text field which drugs should be added or removed for optimal treatment. They were then shown a 1.5 minute video explaining the use of the STRIP Assistant, after which they were presented with the second patient case in the STRIP Assistant user interface. Respondents were asked to optimize this case through the STRIP process, reacting to the advice generated by the application. Each screen contained a help button explaining what was expected of the respondents.

After optimizing the second case, respondents were presented with the SUS, consisting of ten statements with which they had to indicate their (dis)agreement on a Likert scale. Finally, respondents' demographics (age and sex) were collected, alongside their experience with medication reviews and CPOE systems. In a text field, respondents could optionally leave their comments.

Statistical Analysis

In all cases, an expert panel determined the correctness of the decisions made by the participants. Slight corrections to the data had to be made to account for the differences in the

potential number of appropriate decisions respondents could make in each case: respectively 17 in the unassisted case and 20 in the assisted one. Similar corrections were applied to account for differences in the possible number of inappropriate decisions: respectively 30 in the unassisted case and 40 in the assisted one. Paired t-tests were used to analyze the data pertaining to appropriateness and inappropriateness of decisions, and the differences in time spent.

The results of the SUS were formatted in the manner described by Brooke [33]: the values of the odd questions were subtracted by 1 to get the corrected score; the values of the even questions were in turn subtracted from 5 to get their corrected score. The sum of all questions was multiplied by 2.5 to calculate the final score ranging from 0 to 100.

RESULTS

Descriptive Statistics

TABLE 1: Overview of participants' characteristics.

		Frequency	Percentage
Sex	Male	12	27.9
	Female	27	62.8
	No data	4	9.3
Age (y)	<= 30	5	11.6
	31-40	7	16.3
	41-50	8	18.7
	51-60	14	32.5
	=> 61	5	11.6
	No data	4	9.3
Participant's role	GP	31	72.1
	Pharmacist	4	9.3
	Dispensing GP	2	4.7
	GP in training	2	4.7
	No data	4	9.3
Experience with medication reviews	STRIP	12	27.9
	Other medication review method	18	41.9
	None	9	20.9
	No data	4	9.3

Of the 43 respondents whose answers were valid, all but 4 filled out the questions pertaining to their personal characteristics. The majority of these were female (62.8%). Most respondents were in their fifties (32.5%) or forties (18.7%). 7 participants (16.3%) were aged between 31 and 40, and 5 (11.6%) were in their twenties. 5 (11.6%) were over 60 years of age. Most were either general practitioners (72.1%) or pharmacists (9.3%). 2 were dispensing GPs (4.7%), and 2 were GPs in training (4.7%). Most were experienced with performing medication reviews: 18 participants (41.9%) did not use STRIP for their reviews, while 12 (27.9%) did. 9 (20.9%) had no experience performing medication reviews at all.

Usability Hypotheses

TABLE 2: Overview of the tested hypotheses and their statistical outcomes.

	Usual care	STRIP Assistant	Statistics
The STRIP Assistant positively influences the number of *appropriate decisions* made in a medication review: accepted.	418 (58%; M=11.44; SD=2.63)	656 (76%; M=15.26; SD=2.05)	Paired t-test: t(42)=8.80, p<0.0001
The STRIP Assistant negatively influences the number of *inappropriate decisions* made in a medication review: accepted.	302 (42%; M=9.36; SD=2.53)	210 (24%; M=4.88; SD=2.23)	Paired t-test: t(42)=8.93, p<0.0001
The STRIP Assistant negatively influences the *time* taken to perform a medication review: rejected.	13 minutes (M=0.94; SD=0.40)	24 minutes (M=1.34; SD=0.20)	Paired t-test: t(42)=7.07, p<0.0001
Users perceive using the STRIP Assistant as *satisfactory*: rejected.		63.25 (SUS-score)	Quality consensus test: 63.25 < 70

In total, 86 medication reviews were performed by the participants; in half of these cases they used their usual care methods to perform the optimization, in the other half they were aided by the software application. An overview of all tested hypotheses is shown in Table 2. On average, the participants prescribed 8 drugs for the unassisted case, and 14 for the assisted one.

A paired t-test showed a statistical difference in the appropriateness of the decisions made without the STRIP Assistant (M = 11.44, SD = 2.63) and with the STRIP Assistant (mean (M) = 15.26, standard deviation (SD) = 2.05); t(42) = 8.80, p < 0.0001. A Wilcoxon signed-rank test showed similar results; Z = -5.40, p < 0.0001. From a total of 418 unassisted correct decisions

and 656 aided ones, on a decision total of respectively 720 and 866, follows that the proportion of appropriate decisions increased from 58% without help to 76% with the STRIP Assistant.

A paired t-test showed a statistical difference in the inappropriateness of the decisions made without the STRIP Assistant (M = 9.36, SD = 2.53) and with the STRIP Assistant (M = 4.88, SD = 2.23); $t(42) = 8.93$, $p < 0.0001$. The percentage of inappropriate decisions decreased from 42% in the unassisted case to 24% in the assisted one.

On average, participants took 13 minutes to complete the unassisted part of the experiment, and 24 minutes to complete the assisted medication review. A paired t-test of the base 10 logarithm of these values showed a statistical difference in the time taken without the STRIP Assistant (M = 0.94, SD = 0.40) and with the STRIP Assistant (M = 1.34, SD = 0.20); $t(42) = 7.07$, $p < 0.0001$. This indicates that participants spent significantly more time optimizing medication with the STRIP Assistant.

On average, the respondents filled out a SUS-score of 63.25, out of a possible maximum of 100. This value is lower than the quality threshold of 70 arrived at by Bangor et al. and corresponds with a marginal acceptance rate in a later paper by the same authors [15,16].

DISCUSSION

Effectiveness

This study has shown that a decision support system can make general practitioners and pharmacists perform better medication reviews, albeit in an experimental setting with preselected patient cases. This is in line with the consensus on the effectiveness of health recommendation systems in literature [32]. More specifically, the results indicate that the choice for a recommender based on a pre-determined explicit knowledge base yields viable results in a complex domain with potentially far-reaching implications. Rather than relying solely on collaborative or content-based filtering, a knowledge base guarantees a minimal quality level when generating recommendations [4].

Even though the medication reviews performed with the STRIP Assistant are significantly better than those without, a non-negligible number of mistakes respondents made (15%) can be attributed to software suggestions. In this experiment, each START advice was presented as an alphabetically ordered list of medicines that users had the possibility to prescribe. In practice, many users picked the first item in the list, resulting in an overabundance of suboptimal choices; when adding a vitamin D supplement, for example, many users picked alfacalcidol instead of cholecalciferol, even though the former has fewer and more specific indications. Few publications have touched upon the subject of decision support systems generating incorrect recommendations, and consequently strategies to prevent them are lacking [66,121,14]. Hybrid recommendation systems, combining an explicit knowledge base with content-based or collaborative filtering, have been shown to outperform their simpler counterparts [4]. As long as the risk associated with automatic learning systems in a precarious domain such as health

care is accounted for, a hybrid approach may prove beneficial to improving recommenders' effectiveness.

Efficiency

Contrary to our assumption, performing medication reviews with the STRIP Assistant is less efficient (i.e. takes more time) than optimizing drugs manually. Traditionally, the three aspects of usability are assumed to be positively correlated [101]. However, a different perspective viewing effectiveness and efficiency as conflicting requirements in a project has been proposed by Nilsson & Følstad [156]. In an experiment such as the one in this study, where respondents either use their habitual approach, or have to learn a new structured method, a drop in efficiency can be reasonably attributed to effectiveness and efficiency conflicting. Due to the experimental setting, unfamiliarity with the method and the user interface is likely to play a role as well.

Conducting experiments in which more gradual changes to the method are applied may result in improvements in both effectiveness and efficiency; in a study related to this one, a paper version of the earlier polypharmacy optimization method was tested in an experiment [57]. It, too, proved to be less efficient than performing a medication review manually. However, the software-aided reviews performed in this study took less time than the paper-based ones in the previous study. This lends credibility to the assumption that gradual changes may improve all aspects of usability simultaneously.

Satisfaction

Respondents perceived using the STRIP Assistant as only marginally acceptable. The SUS-average of 63.25 was lower than the commonly accepted quality indicator of 70 [15,16]. This aspect, too, can be understood by viewing the usability aspects as conflicting requirements [156]. The suboptimal prototypical design of the software's user interface, and the unfamiliarity of the respondents with the application may explain this inconsistency with consensus in literature.

Clinical Relevance

Methods for medication review have been proven to be valuable in improving prescriptions of polypharmacy patients. The Polypharmacy Optimization Method, which served as a foundation for the STRIP method, has led to improvements in appropriate decisions in medication reviews [57]. The START and STOPP criteria, which constitute a major part of the STRIP Assistant's knowledge base, have been shown to be associated with improvements in medication appropriateness, reductions in adverse drug reactions, and decreases in drug use and costs [75,158].

The two patient cases used in this experiment were comparable in complexity and number of medicines, but there could, for reasons of validity, not be a complete overlap of diseases and drugs. This makes determining the clinical relevance of the intervention difficult. Nevertheless, the most noticeable improvements in the adequate prescribing of drugs were the treatment of

osteoporosis with bisphosphonates, calcium and vitamin D, and the treatment of systolic heart failure with ACE inhibitors. The most important improvement relating to stopping medicine use is the discontinuation of digoxin when atrial fibrillation is adequately treated with beta blockers. These interventions correspond with guidelines of the START and STOPP criteria [75].

Thus, the results in this study confirm the results of previous studies that structured methods for medication review significantly improve medication appropriateness of prescriptions.

Limitations

When interpreting these results, the experimental nature of the method should be taken into account. The STRIP Assistant's usability has been tested and validated with real patient cases in a controlled environment, but has not been validated in practice with users reviewing their own patients. While the results lend credibility to the STRIP method being useful in practice, this study does not prove its clinical relevance.

When generalizing the results of this study, the limited number of participants should be considered, as well as the sampling method. 42 GPs and pharmacists participated voluntarily, enabling the possibility that they were positively biased towards use of a clinical decision support system aiding them with medication reviews.

Further Research

A randomized controlled trial incorporating a large representative sample should be conducted to conclude the STRIP Assistant's effectiveness, efficiency, and satisfaction. Further research should focus on its usability through evaluation in a real life setting over a longer period of time, exploring to what extent experience influences users' effectiveness and efficiency in working with the software. Furthermore, longitudinal research could show if the STRIP Assistant is clinically relevant in practice and evaluate its impact on adverse effects and medicine costs.

CONCLUSION

In this study, a clinical decision support system (the STRIP Assistant) designed to aid GPs and pharmacists with conducting medication reviews was validated in an experimental setting. The results show that using the STRIP Assistant positively influences the number of appropriate decisions made in a medication review of elderly polypharmacy patients and decreases the number of inappropriate choices. Contrary to our assumptions, users spent more time optimizing health records with the STRIP Assistant than without it. Users perceive using the software only as marginally acceptable. Further research is needed to determine whether optimization of polypharmacy with the help of the STRIP Assistant is clinically beneficial.

4.2

Efficiency of clinical decision support systems improves over time

ABSTRACT

Polypharmacy endangers patients' health. Several structured medication review methods have been developed to aid general practitioners (GPs) and pharmacists in optimizing their prescriptions for polypharmacy patients. These methods have been shown to improve medication reviews' effectiveness, but are less efficient than usual care approaches. This lack of efficiency may be explained by the studies' single-test experimental methods, which

Submitted as: Meulendijk, M., Spruit, M., Willeboordse, F., Numans, M., Brinkkemper, S., Knol, W., et al. (2015). Efficiency of clinical decision support systems improves over time (submitted).

do not take into account experience participants gain over time. The aim of this study was to determine if having a group of physicians perform decision supported structured medication reviews over a longer period of time will lead to improvements in efficiency. Four expert teams consisting of a GP and a pharmacist conducted structured medication reviews on patients in 13 general practices located in Amsterdam, the Netherlands. During thirteen months, the expert teams performed 261 medication reviews. An independent t-test showed a statistical difference in the time users spent during the first half of the medication reviews (M = 15.70, SD = 8.81) and the second half (M = 10.67, SD = 5.21); t(259) = 5.625, p = .000. This leads the authors to conclude that the amount of time users needed to perform similar tasks decreased significantly as they gained experience over time.

INTRODUCTION

The chronic use of multiple medicines, while often being indicated in older patients, has been shown to lead to adverse drug reactions. A comprehensive Dutch study showed that 5.6 percent of Dutch hospital admissions are due to medication-related problems [129]. Appropriate prescribing in older people is challenging, because of age-related changes in pharmacokinetics and pharmacodynamics, multimorbidity and polypharmacy [132,195].

Several structured medication review methods have been developed to aid general practitioners (GPs) and pharmacists in optimizing their prescriptions for these polypharmacy patients [75,57]. Implementations of these methods as decision support systems have been proven to improve the quality of prescriptions in research settings [57,143].

A major disadvantage of these methods, however, is that physicians need more time to perform them than they do to perform their usual care methods. In a 2009 study in which the Polypharmacy Optimization Method (POM) was tested, Drenth-van Maanen et al. found that performing a medication review with the structured method took more time (16.7 minutes) than performing one without (8 minutes)[57]. A more recent study, in which a decision support system facilitating the Systematic Tool to Reduce Inappropriate Prescribing (STRIP) was validated, Meulendijk et al. found that the average time physicians need to optimize a patient's health record with the STRIP method was 24 minutes, while the time needed without any structured approach was 13 minutes [143].

Both studies testing structured medication review methods used experimental approaches with a pre-test posttest design. A single group of participants was asked to optimize a patient record employing their usual care approach, after which a stimulus was applied: in the POM study, participants were educated in the use of the method through a 2-hour lecture, while in the STRIP study, respondents were shown a 1.5-minute video explaining the use of the software. Afterwards, participants were asked to optimize a comparable patient record using the structured method.

Literature treats efficiency, or the resources spent while performing a specific task, as one the determinants of usability [108]. Users' experience with software has also been shown to influence their attitudes towards prolonged use of it [217,218]. However, only limited longitudinal research measuring the effects of prolonged software use on efficiency has been conducted; some studies, in this regard, have indicated that determinants common in conventional usability research differ for experienced users [139,113].

Following this line of thought, the lack of efficiency of structured medication reviews may be explained by the studies' single-test experimental methods. The authors of this paper hypothesize that having a group of physicians perform decision supported structured medication reviews over a longer period of time will lead to improvements in efficiency.

BACKGROUND

Efficiency in Usability

Efficiency is being regarded as one of the main aspects determining software applications' usability; it has been included in all major definitions of the concept [108,155]. Experience users gain through prolonged exposure to a system has been shown to be a major determinant on their attitudes towards accepting technology [217,218,51,142].

The commonly used definition proposed by the International Standards Organization (ISO) defines usability as 'the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use' [108]. Effectiveness is understood as the 'accuracy and completeness with which users achieve specified goals', while efficiency exists of 'resources expended in relation to the accuracy and completeness with which users achieve goals'. Satisfaction is the subjective 'degree to which user needs are satisfied when a product or system is used in a specified context of use'. Theoretically the notion of efficiency can encompass any kind of resource, including money or knowledge, but in practice the concept is usually limited to indicate the amount of time spent on a certain task.

The consensus in literature is that effectiveness and efficiency are positively correlated concepts [101]. In the studies in which systematic medication reviews were validated, however, this was not the case; participants needed more time to conduct medication reviews in a methodical manner [57,143]. In the latter study, the authors attempted to explain this discrepancy by adhering to the view proposed by Nilsson & Følstad, who claim that effectiveness and efficiency are conflicting requirements [156].

Most usability studies use experimental methods to determine systems' effectiveness, efficiency, and satisfaction. The effect of temporality, or the degree to which these factors change over time, is largely overlooked [113]. At the same time, however, aspects dependent on temporality are accepted as major determinants: a common definition proposed by Nielsen

emphasizes the importance of learnability and memorability of user interfaces [155]. Learnability encompasses the ease with which users can accomplish basic tasks when they first encounter a design, while memorability concerns the degree to which they can reestablish proficiency when reusing it. Thus, testing of usability through experiments often entails nothing but testing of learnability, measuring factors of novice rather than experienced users [139]. In experimental studies, in which participants are unfamiliar with the method or the user interface, a drop in efficiency can reasonably be expected [143].

Efficiency in Clinical Decision Support Systems

In literature, there is consensus that usability has a significant impact on users' adoption of clinical decision support systems [47]. However, many clinical decision support systems have not been shown to improve efficiency. While systems' improvements in effectiveness are well-documented, studies on efficiency are lacking. A recent systematic literature review found that "contamination of clinicians in the control group, [...] evaluation periods that were too brief to demonstrate an effect on efficiency, and small clinician sample sizes" made it impossible to generalize on efficiency consequences [32].

A study synthesizing features of proven effective clinical decision support systems found mixed results regarding efficiency: while embedding systems in physicians' workflows was associated with improved effectiveness, advice presented within computerized physician order entry (CPOE) systems had negative correlations regarding success [179]. The authors explained this apparent anomaly by subscribing to the oft-mentioned 'alert fatigue' found in clinical decision support systems; in many systems, users are presented with a multitude of warnings and suggestions, forcing them to ignore the majority [116,210].

RESEARCH DESIGN

Rationale

Following the above-mentioned considerations, the authors hypothesize that users grow more proficient performing decision supported medication reviews over time. The effects of experience on applying the application effectively and memorizing its functionality are assumed to lead to gradual improvements in efficiency.

This objective is investigated in the following research question: does the time physicians use to systematically optimize the medical records of polypharmacy patients decrease over time?

A secondary research interest focuses on users' perceived satisfaction regarding their use of a clinical decision support system. Furthermore, the practical implications an efficiency investigation may have on a clinical decision support system are explored.

Clinical Decision Support System: STRIP Assistant

The clinical decision support system that will be used in this study to explore the research question is the STRIP Assistant (STRIPA), a web application that facilitates the structured pharmacotherapeutical analysis of the STRIP method [144]. STRIPA displays a single patient's health record at a time, showing his or her diseases, medicines, complaints, and relevant measurements. Physicians are guided through the analysis steps to optimize a patient's medication; first, they manually assign drugs to the diseases for which they have been prescribed; second, they add any missing medicine for which there is an indication; third, they eliminate superfluous ineffective drugs or drugs for which there is no appropriate indication; fourth, they check for any relevant clinical interactions between drugs; finally, they readjust dosages if necessary. In all these steps, except the first one, STRIPA generates advice based on (inter)national guidelines, most notably the START/STOPP criteria and guidelines for medication safety recommended by the Royal Dutch Pharmacists Association [158,240]. The pieces of advice arrived at through these guidelines are used to alert users of anomalies and suggest appropriate actions.

Method

The research question was investigated within the intervention arm of a larger, multidisciplinary study [227]. A randomized controlled trial was performed in 25 general practices located in Amsterdam, the Netherlands, including 500 patients. The patients in the 13 intervention practices received systematic medication reviews, which yielded 261 analyses relevant for this study.

Four expert teams consisting of a GP (never the patient's GP) and a pharmacist (never the patient's pharmacist) conducted structured medication reviews according to the STRIP method. The pharmacotherapeutical analysis was carried out using the STRIP Assistant, and served to identify undertreatment, overtreatment, clinical interactions, contra-indications, and inappropriate dosages in patients' health records. The software application enabled physicians to adjust medicines where necessary, advising them where possible. Prior to the study, the expert teams were trained in the use of the STRIP Assistant, using it for test cases under supervision.

The data was gathered during a period of thirteen months, from November 2013 to November 2014. To explore the applications' efficiency, users' actions were recorded in logs. Not only what they did was recorded, but also how long they hesitated in between actions. Logs consisted of timestamps, action descriptions, interface phase descriptions, and an array of associated objects. Clickstream analysis was used to gain insight into the actions and the time users needed in between them. A clickstream is an "electronic record of a user's activity on the Internet" and is an often used method for gaining insight into people's behavior with web applications [34,164]. An example of one of these logs is shown below:

TABLE 1: **A sample log showing the addition of a medicine (lactulose) to a disease (obstipation) as part of an undertreatment intervention.**

Log ID	Value
Action description:	addedObj[ect]
User interface phase:	Undertreatment
Timestamp:	November 1st, 2013, 15:00:01
Associated objects:	Medicine: lactulose 667mg/ml, apply once daily. Disease: chronic obstipation
Disease:	Medicine: lactulose 667mg/ml, apply once daily.

The satisfaction aspect was measured with the System Usability Scale (SUS), a commonly used method for determining users' perceived satisfaction [33]. Participants were asked to fill out a survey consisting of ten statements with which they had to indicate their (dis)agreement on a Likert scale.

RESULTS

Descriptive Statistics

The expert teams conducted a total of 261 analyses for all patients in the intervention group. All analyses were performed in the period of November 2013 to November 2014. On average, it took them 13.2 minutes (SD = 7.65) to complete a medication review, with outliers ranging from 2.1 to 45.7 minutes.

On average, the cases consisted of 13.5 (SD = 5.76) episodes (either symptoms or diagnosed diseases) to which the users assigned 7.5 (SD = 4.38) medicines; these medicines comprised both ones patients already used, and new ones that the users prescribed. A one-way ANOVA showed that there were no significant differences between the numbers of diseases ($F(12, 248) = 1.230$, $p = 0.263$) and medicines ($(F(12, 248) = 1.012$, $p = 0.438$) over time, indicating that the cases were of comparable complexity. The number of medicines was moderately positively correlated with the time spent on medication reviews ($r = 0.415$, $n = 261$, $p = 0.000$), showing that users took longer to perform reviews for more difficult cases.

Overall Efficiency

FIGURE 1: Decreasing trend in average time spent per medication review over time.

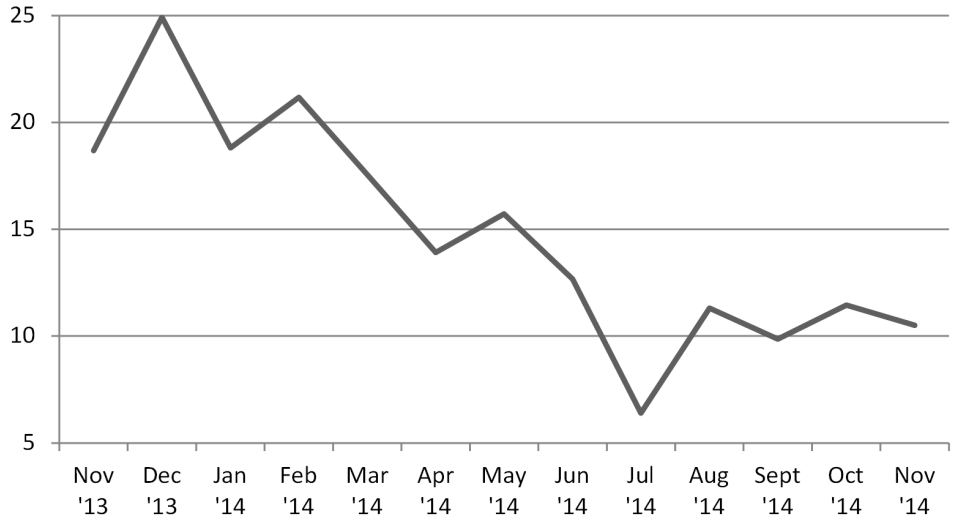


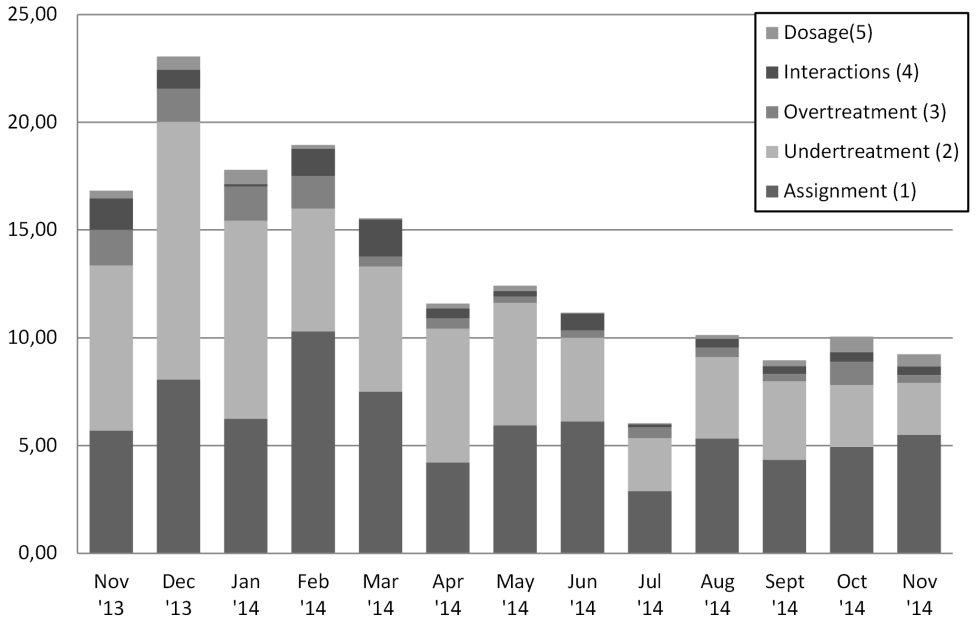
Figure 1 shows a discernable trend in decreasing time spent on medication reviews over time. To determine statistical differences, the number of analyses was bisected, resulting in two groups consisting of 130 and 131 analyses respectively. An independent t-test showed a statistical difference in the time users spent during the first half of the medication reviews (M = 15.70, SD = 8.81) and the second half (M = 10.67, SD = 5.21); t(259) = 5.625, p = .000.

Stepwise Efficiency

Further analysis of the clickstream showed that the time users spent on the different phases of the web application differed greatly. The first step, in which users were tasked with manually assigning medicines to relevant diseases, was responsible for 50% (5.6 minutes) of the time they spent on a medication review. 37% (4.2 minutes) of the time users spent on the second step, concerning undertreatment. The other phases were performed much quicker, with the final step being accountable for 2.4% (0.3 minutes) of the time. Figure 2 shows the time spent on different steps within medication reviews over time.

Independent t-tests were used to determine statistical differences in the time users spent per step. Three of the steps showed significant discernable trends in time allocation; the steps concerning medicine assignment (t(259) = 2.526, p = .012), undertreatment (t(259) = 5.028, p = .000), and interactions (t(259) = 2.792, p = .006) were performed faster in the second half of the study than the first one. The other steps, concerning overtreatment and dosage adjustments, showed no significant differences in time spent during the first and second halves.

FIGURE 2: Average time spent on each step in the decision support system per month.



Activity-Based Efficiency

In other to explore users' behavior towards distinct actions, descriptive statistics were used to determine which advices took users longer to follow than others. Of advice that was heeded at least 5 times, users on average took 22.9 seconds to decide, ranging from a minimum of 0.08 seconds to 2.4 minutes. The three recommendations that took users longest to respond to were all part of the undertreatment step.

TABLE 2: Overview of least efficient recommendations that were heeded or overridden.

	Advice	Average time (seconds)
Followed	Beta-2-sympaticomimetica or anticholinergics with asthma or COPD	37.4
	Antihypertensives with chronic systolic blood pressure	37.8
	ACE inhibitor with acute myocardinfarct	33.6
Ignored	Antihypertensives with chronic systolic blood pressure	35.7
	Bisfosfonates with corticosteroids	30.8
	Antihypertensives with chronic systolic blood pressure or untreated hypertension episode	21.2

Analysis of the advice users had overridden showed that on average they took 11.4 seconds to decide, with ranges from 0.05 seconds to 4.6 minutes. When ignoring advice, the top three recommendations that took users longest belonged to the undertreatment step. Table 2 summarizes these least efficient recommendations.

Satisfaction

The answers the participants gave to the SUS-statements were formatted in the manner described by Brooke to calculate their final scores [33]. They filled out an average SUS-score of 78. This value is above the quality threshold of 70 arrived at by Bangor et al. and corresponds with a 'very good' adjective rating in their later paper [15,16].

DISCUSSION

Efficiency Improvement

In this paper, the results of a longitudinal study on the effects of temporality on physicians' efficiency of performing decision supported medication reviews were reported. In line with the authors' hypothesis, the amount of time users needed to perform similar tasks decreased significantly as they gained experience.

Earlier validations of the STRIP method showed significant increases in effectiveness over usual care approaches, but decreases in efficiency [57,143]. In the latest validation of the method embedded in a decision support system, the authors hypothesized that "experiments in which more gradual changes to the method are applied may result in improvements in both effectiveness and efficiency" [143]. This assumption was in line with the view proposed by Nilsson & Følstad of effectiveness and efficiency as conflicting requirements [156]. The results of this study confirm the hypothesis that users' familiarity with an application and their experience with performing similar tasks leads to increases in efficiency.

Limited research has been done into the effects of temporality on usability in general, and efficiency in particular. Mendoza and Novick reported, though, that frustration levels of participants using unfamiliar software decreased over time, leading them to remark that "factors such as features being hard to find and operators committing slips and mistakes really are the principal causes of severe frustration" [139]. They report that with increased familiarity these frustrations decrease; it can be reasonably expected that, in a similar fashion, time spent on repetitive tasks decreases.

In fields related to usability, temporality has been recognized as being influential in shaping people's motivations towards using software. In a user experience study, temporality appeared to be a determinant in the changing motivations of people who use software for a prolonged period of time: "prolonged experiences became increasingly more tied to aspects reflecting how the product becomes meaningful in one's life" [113]. Experience has been shown to be a determinant in users' attitudes towards acceptance of technology [51,217,218].

Practical Implications in STRIPA

The clickstream analysis pinpointed which steps of the STRIP process were least efficient and could be improved in the software application. In an attempt to decrease the workload of the first, medicine assignment, step, knowledge discovery methods were employed. Based on historical data, the application was improved to automatically assign medicines to appropriate diseases. All recommendations that took physicians the longest to respond to belonged to the undertreatment step. By redesigning the user interface, the developers combined pieces of advice and limited the number of medicines initially shown in dropdown lists; this resulted in users having to perform fewer actions to accomplish the most common actions, without limiting the application's functionality.

A limited evaluation of these practical implications showed that they succeeded in further improving efficiency. A total of 32 medication reviews were conducted by the same expert teams with the functional improvements in place. With the new functionality, the teams on average spent 9.4 minutes (SD = 4.82), which is a minute faster than they did in the last month of the study.

CONCLUSION

This paper reports the results of a longitudinal study on the effects of temporality on physicians' efficiency of performing decision supported medication reviews. Corresponding with the authors' hypothesis, the amount of time users needed to perform similar tasks decreased significantly as they gained experience.

Limitations of this study include the limited number of participants and the fact that the analyses were performed unsupervised. The time required to assess patients' characteristics and review their information in the STRIP Assistant's interface may partially account for the relatively long amount of time spent during the analysis' first step.

PART V

Refined Objectives Elicitation

An intellectual is someone whose mind watches itself.
[...]
"I despise intelligence" really means: "I cannot bear my doubts."

Albert Camus

5.1

Semantic Interoperability in Primary Care Terminologies

ABSTRACT

Objective: The diversity of terminologies used in primary care causes significant challenges regarding semantic interoperability. Attempts to address these challenges usually focus on the creation of metaterminologies, with the peculiarities of national variations of terminologies being overlooked. In this paper the extent to which primary care data can be meaningfully exchanged between nationally implemented terminologies is assessed.

Submitted as: Meulendijk, M., Spruit, M., Lefebvre, A., & Brinkkemper, S. (2015). To what extent can patient data be meaningfully exchanged between primary care terminologies? A case study of four Western European classification systems (submitted)

Method: In order to determine this, a model comprising primary care terminologies and including axioms to define their relations was developed. Generic metrics were designed to determine the completeness and accuracy of any two arbitrary vocabularies within an ontological model. These metrics were used on an implementation of the model to determine the data quality that is preserved when expressing similar data in different primary care terminologies.

Results: The results show that values of terminologies which are closely related can express each other's concepts relatively well. While less extensive terminologies' concepts often have equivalents in larger classification systems, concepts of more comprehensive terminologies can only be expressed in simpler terminologies to a very limited degree.

Conclusion: The authors conclude that the current state of accuracy and completeness between primary care terminologies does not allow for sufficiently meaningful semantic interoperability. They argue for the standardization and unaltered use of existing national terminologies and the mapping of their concepts to metaterminologies, thus creating metaterminology sub sets which would be semantically interoperable.

INTRODUCTION

There is a great need for a more sophisticated level of semantic interoperability in primary care. Errors made during the prescription process often lead back to physicians' use of incomplete health records and miscommunication with caretakers in secondary care [214,183]. This lack of semantic interoperability causes information exchange problems, which in turn can lead to medical errors; a loss of information when a drug prescription is communicated from a prescriber to a pharmacist could result in harmful changes to a patient's drug regimen (e.g. an incorrect change in dosage).

In the last two decades, this issue has been addressed by approaches to standardize document structures, communication protocols, and classification systems in health care, leading to a wide variety of international and national de facto standards [69,53]. In primary care, a large number of classification systems are in use, including taxonomies to systematically classify diseases, drugs, and laboratory tests, among many others [231,230,174].

This proliferation of medical classification systems has in turn led to further difficulties regarding semantic interoperability; concept ambiguity and differences in perspective make information exchange between systems highly problematic. Attempts to address these problems usually focus on the creation of metaterminologies, to which lower-level terminologies are related [105,206]. The many national variations on standards and custom-made terminologies are usually overlooked in these studies.

In order to improve semantic interoperability in primary care, the extent to which different national classification systems are able to express each other's concepts is worth exploring. In this paper, the results of an attempt to match and map common primary care terminologies to each other in four Western European countries are described.

BACKGROUND

Semantic Interoperability in Primary Care

The benefits in prescribing quality offered by systems facilitating electronic prescribing and medical decision support have been demonstrated [32,143]. The advent of these systems makes the need for semantic interoperability in health care evident [92]. In primary care, large amounts of information are stored and communicated between patients, practitioners, pharmacists, and secondary caretakers. Objective concepts including physical measurements and prescribed drugs, but also subjective ones such as patients' complaints and living situations are recorded. As will be described more extensively in following sections, a multitude of classification systems aimed at systematically storing this information has been developed. Various approaches to achieve semantic interoperability between these ontologies have been explored.

The various classification systems in health care differ in comprehensiveness and intended use. All of them are based on a shared nomenclature in their subdomain of choice, from which a controlled vocabulary, or terminology, is created. The taxonomical classification consists of categorizing the concepts in hierarchies and defining generalization relations between parents and children. Additionally, most classification systems contain some ontological features. An ontology can be defined as a formal, explicit specification of a shared conceptualization; it formally represents knowledge as a hierarchy of concepts within a domain, defining a terminology to describe those concepts and their relationships [85]. While most classification systems are not that exhaustive, some include ontological rules, restrictions, and axioms that determine how concepts are related.

Attempts to map values from one terminology to their closest relatives in another are all around. After the second revision of the popular primary care terminology International Classification of Primary Care (ICPC), a mapping to the more extensive International Classification of Diseases version 10 (ICD10) terminology was published [159]. In several instances, mappings between international standards and their national adaptations were created to facilitate information exchange [78,50]. While these approaches have led to feasible information exchange between various standards, they are not without their deficiencies. The continuous changes made to both national and international terminologies require constant maintenance of all mapping systems. As many of these mappings have been published by national government-dependent organizations, this threatens their continuity. The lack of alignment between these organizations may lead to different interpretations of concepts and thus to inaccuracy in their mappings.

A more recent development has been to utilize Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT) for semantic interoperability in primary care. SNOMED-CT is a "comprehensive, multilingual clinical healthcare terminology", consisting of coded concepts with descriptions [105]. It allows for the representation of relationships between concepts, complete with cardinalities and synonyms. When applying it for purposes of semantic interoperability, it can either completely replace existing terminologies or serve as an intermediate taxonomy to which concepts of other terminologies can be mapped. SNOMED-CT's comprehensiveness and unique identification guarantee correctness when transferring data between terminologies. Exploration into the mapping of values from clinical terminologies to their equivalents in SNOMED-CT is ongoing [3,114,26].

A comparable approach to achieve semantic interoperability in the health care domain is the Unified Medical Language System (UMLS). Developed by the American government, it aims to "bring together many health and biomedical vocabularies and standards to enable interoperability between computer systems" [206]. It consists of a metathesaurus containing uniquely identifiable concepts with their equivalents in commonly used terminologies, such as ICD10 and ICPC. Like SNOMED-CT, it includes relationships between its concepts, providing synonyms and shared definitions.

Data Quality

The reliability of ontology-based systems is commonly recognized to be partially dependent on the data quality of the ontologies used as their input. As "the validity and quality of the ontology data directly affects the validity and quality of the system using the ontology", assessing the data quality through metrics is a logical and essential step in creating effective and efficient systems [161].

Over the years, a wide variety of variables, and metrics to measure them, have been developed. In their study on quality dimensions in ontological foundations, Wand and Wang (1996) discerned between accuracy and precision, reliability, timeliness and currency, completeness, and consistency [224]. In their more recent approach, Batini and Scannapieco (2006) simplified these to accuracy, completeness, time-related dimensions, consistency, and minor parameters of lesser importance (i.e. accessibility and source quality) [19].

In the medical domain, completeness, correctness, concordance, plausibility, and currency are commonly used to assess data quality [225]. In their literature study on data quality in electronic health records, Weiskopf and Wang (2013) found that especially completeness and correctness (or accuracy) are regarded as insightful factors. Metrics most frequently employed to assess these dimensions are element presence (in the case of completeness) and comparison to a gold standard (for both correctness and completeness).

Considering the research question, the information exchange model validated in this study will be tested for its accuracy and completeness. Even though some of the other parameters mentioned are essential to the quality of data, they can only be ensured through successful

implementation. Consistency of data throughout systems depends heavily on information architectures and communication protocols. The same goes for time-related dimensions and accessibility. Source quality in the primary care domain is partly guaranteed by the non-profit organizations maintaining the ontologies, and partly by the authority of whoever implements and uses them.

Defining accuracy is usually done by defining its antonym: inaccuracy implies that a "real-world state different from the one that should have been represented" is displayed [224]. From this follows that accuracy is the correct representation of a real-life phenomenon. Flaws in data, for example due to input errors, are occurrences of inaccuracy; specifying a drug dosage as 100mg instead of 10mg is an accuracy error with potentially far-reaching implications. Lack of precision or ambiguity of concepts can cause data to be inaccurate as well; a recording of a bone fracture without specifying which bone is broken may be too inaccurate to be meaningful.

An ontology is generally considered to be complete if it "represents every fact of the real world" within its domain [19]. Meeting this criterion requires an ontology to have a narrowly described domain, specifying which attributes are expected to be included. Missing data does not automatically imply that this criterion is not met, however, since values may be expected to be missing if their attributes are optional; drug prescriptions, for example, may include specific brands if alternatives have been proven to be ineffective for a specific patient, but usually do not.

RESEARCH DESIGN

Rationale

The diversity of clinical terminologies causes significant challenges regarding semantic interoperability in primary care. Their different perspectives, levels of detail, national customizations, and use in various phases of the prescribing process make information exchange difficult. While implementation of a comprehensive metaterminology such as SNOMED-CT would facilitate meaningful interoperability of data, this would be problematic in practice. Terminologies in primary care have long histories and are often custom-made to suit specific requirements of daily practice. Their different perspectives and levels of detail justify their continued use in health care. It is not expectable, nor desirable, that these terminologies are replaced by more comprehensive but generic alternatives. This means that, since not all concepts from current terminologies can be expressed in every other terminology, data quality suffers. Doubts regarding concepts' completeness and accuracy when transferring data appear.

This leads to the following research question, which will be explored in this study: to what extent can primary care data be meaningfully exchanged between nationally implemented terminologies?

Approach

In order to investigate this issue, the researchers strive to create a model of primary care concepts. This model will consist of the concepts present in a selection of nationally implemented terminologies. The relations between these concepts will be determined, specifying how concepts' attributes are interconnected and how relations' cardinalities exist. Based on this model, a set of rules will be proposed through which concepts from any one classification system can be expressed in those of another. Finally, testing an implementation of this rule-based model with sample data will address the issues put forward in the research question.

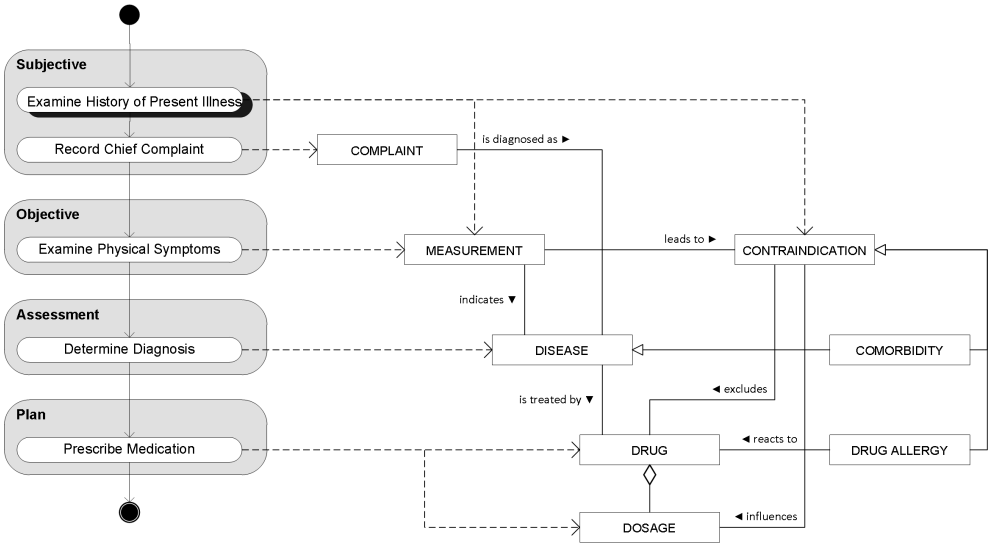
First the current status quo of primary care will be explored. In his 2013 paper, Kierkegaard describes the current situation in Europe regarding the adoption of e-prescriptions [119]. He distinguishes between leading, trailing, and passive countries. To compare health terminologies to one another, the countries included in this study should have reached a certain maturity with regard to their health care technology. Consequently, two leaders and two trailers were selected as case studies. The countries whose primary care situations will be explored are technology leaders the Netherlands and Denmark, and trailing countries Germany and the United Kingdom.

After the preliminary investigation, a process involving terminology matching and mapping is performed to map values from one terminology to its equivalents in another, insofar as this is possible [191]. Bellahsene, Bonifati and Rahm (2011) define matching as "the task of finding semantic correspondences between elements of two schemas" and mapping as the task that "derives the relationships between elements and structures in heterogeneous schemas" [20]; in order for mapping to be successful, matching should be completed beforehand. This process results in a comprehensive model detailing to what extent information exchange between these terminologies can be facilitated.

To evaluate the model, the terminology mappings are stored in a database. Then, a rule engine with rules describing the transfer conditions between the terminologies is implemented. Using this implementation, the model is finally tested against the data quality parameters of completeness and accuracy.

PRESCRIBING IN PRIMARY CARE

FIGURE 1: **Process-data diagram depicting concepts relevant during a consultation.**



Process

The central process in primary care is the patient's consultation of his or her physician. In its traditional arrangement, this process leads to the practitioner determining a diagnosis and prescribing medicine. In order to determine what information in this process should be interoperable, the steps and resulting concepts pertaining to it should be evident. For this purpose, the process-data diagram shown in Figure 1 was employed.

One of the most influential approaches to structuring patients' records is the Problem-Oriented Medical Record. The second part of this structure consists of notes describing patients' problems, including their diagnoses and treatment plans. A widely accepted documentation form for these notes is SOAP, an acronym for the four steps of the prescription process [21]:

1. Subjective: recording the patient's chief complaint, and checking his or her history of present illness;
2. Objective: performing physical examinations, and checking the results of earlier measurements;
3. Assessment: determining a diagnosis for the patient's chief complaint;
4. Plan: determining a treatment plan which, in the case of a prescribing process, consists of a drug prescription.

The activity diagram on the left side of Figure 1 displays the four steps of a prescribing process structured by SOAP.

Concepts

The SOAP process results in a set of concepts, such as a diagnosed disease and a prescribed drug. These concepts correspond with those identified by the Dutch College of General Practitioners [28]. The right side of Figure 1 displays the concepts resulting from a prescribing process structured by SOAP.

The main action a GP performs in the first, subjective, step is recording the patient's grievances, which leads to a complaint. After, optionally, performing one or more measurements, this is diagnosed as a disease. In the prescription process, this disease is treated by a drug, with a specified dosage. If a patient suffers drug allergies, the choice of medications available for treatment is reduced. Likewise, other diseases a patient suffers from, or comorbidities, may lead to exclusion of one or more drugs. These contra-indications may be indicated by performed measurements, and result in excluding drugs or adjusting their dosages.

Terminologies

Within primary care, a multitude of terminologies exist to classify the concepts described in the prescribing process. Both national and international terminologies are in use to structure diagnoses, drug prescriptions, and measurements. Their use differs based on their purpose, geographical region, caretakers' preference, and history. While some terminologies overlap, most offer a unique perspective on the data they attempt to classify. For example, both ICPC and ICD provide categorizations for diseases; the latter, however, strives to offer an exhaustive set of disorders, while the former aims at providing a manageable list of both subjective complaints and objective diagnoses.

Terminologies classifying the concepts mentioned in the previous section are described below. Since contra-indications can be expressed as diseases or (allergies to) drugs, and complaints are subjective and impossible to categorize, this results in four classifiable concepts: diseases, drugs, dosages, and measurements. The terminologies in use in the four countries explored in this study are briefly described, including their national adaptations.

Diseases **ICPC** The International Classification of Primary Care (ICPC) is a method of classification for encounters in primary care [232]. It mainly allows for the patients' complaints and the GPs' diagnoses to be recorded in a structured way. It has been developed to be used in combination with the SOAP process.

The ICPC classification contains around 1300 concepts, each with both a unique id and a description (e.g. T90: Diabetes mellitus). It has been criticized for its lack of specificity, but nonetheless is widely used throughout the world. The latest version recommended for use is version 2.

Denmark has implemented a directly translated version of ICPC2, known as ICPC2-DK [50]. Its concepts and codes fully correspond with those in the original WHO-publication. The

Netherlands still use a variation of the first ICPC-version, which has been adapted to conform to requirements in Dutch daily practice [60].

ICD The International Statistical Classification of Diseases and Related Health Problems (ICD) is a classification system aimed at recording medical diagnoses [231]. Depending on which version is used, the coding system contains over 12.000 concepts, existing of combinations of codes and descriptions.

Having originally been developed as a statistical tool, ICD has been updated and expanded many times to accommodate different uses. While not commonly used in primary care, version 10 is a popular disease classification system in secondary care throughout Europe.

Nonetheless, Germany has adapted ICD10 for use by primary care practitioners. This variation, known as ICD10-GM, omits some diagnoses rarely ever used in Western Europe and simplifies the coding of frequently used ones. It is the predominant classification scheme for diseases in German primary care [78].

READ Having been developed separately from its international counterparts, READ is currently the predominant diagnosis classification system in the United Kingdom. Its scope is much broader, however, than any comparable terminologies. It supports not only the recording of diseases, but also that of symptoms, observations, laboratory tests, surgical procedures, and personal characteristics [93].

READ was revised during the 1990s to form its current third version, also known as Clinical Terms Version 3 (CTV3). The terminology contains around 300.000 concepts with accompanying unique identifiers. Even though its use is restricted to the United Kingdom, sections of READ were directly incorporated into SNOMED-CT.

Drugs **ATC** The Anatomical Therapeutic Chemical (ATC) classification system is a terminology for the categorization of medicinal drugs [230]. It categorizes substances at five different levels according to the organ or system on which they act and their chemical, pharmacological and therapeutic properties. Each bottom-level ATC code denotes a single pharmaceutically used substance (optionally including adjuvants), for a specific indication. This means that the same substance can have multiple ATC codes, based on its intended effect. Likewise, final products can share an ATC code, if they have the same active substance and indication.

Up to 2000 ATC codes are in use, dependent on which drugs are authorized for distribution in a specific country. It is the most widely used pharmaceutical classification system. All countries included in this study use ATC to systematically classify drugs.

Dosages Contrary to most other concepts common in primary care, few standardized approaches exist to structurally denote drug dosages. Instead, most countries use notations augmented with abbreviations which are deeply rooted in tradition, and are called SIG codes. After studying the situations in four Western European countries, however, there appears to be consensus on what a drug dosage should consist of: a frequency, prescribing how often medication should be used (e.g. twice daily, once a week); an amount, preferably expressed in

a specified form (e.g. one tablet, two drops); and, optionally, instructions advising patients on how to use drugs (e.g. add water, in between meals).

The Netherlands is the only one of the four included countries having implemented a national dosage terminology. It supports all dosage elements described above, and provides a structured way of recording them. Lists of abbreviations for frequencies (e.g. D for day, W for week), forms (T for tablet, DR for drop), and instructions (e.g. MW for add water) are included [58]. Germany and the United Kingdom use abbreviations of dosage elements in both their native languages and in Latin (e.g. q.s., quantum satis = as much as required). Denmark, finally, has no de facto tradition of using descriptive SIG codes; the Danish national health protocol does, however, support the systematic recording of amount (without explicit reference to the form) per day [138].

Measurements **LOINC** The Logical Observation Identifiers Names and Codes (LOINC) is a classification system for laboratory data, whose purpose is to "facilitate the exchange and pooling of results for clinical care, outcomes management, and research" [174]. It provides unique identifiers, descriptions, and units of measurement for use in primary care. It has its origins in the United States, but is used in some European countries, including Germany, as well.

IFCC-IUPAC NPU The Nomenclature, Properties and Units (NPU) coding system is the combination of a collaboration between the International Federation of Clinical Chemistry (IFCC) and the International Union of Pure and Applied Chemistry (IUPAC) [104]. Like LOINC, it strives to systematically classify laboratory requests and results of patient examinations. It contains not only codes and descriptions for each entry, but also scale types and units of measurement. It is widely used in Europe, including in Denmark.

READ The above-mentioned British READ nomenclature also contains unique values to identify laboratory concepts [93]. It is exclusively used in the United Kingdom and parts of its commonwealth.

NHG Table 25 Of the countries included in this study, the Netherlands is the only one where a custom laboratory terminology is nationally implemented. Developed by the Dutch College of General Practitioners, this 'table 25' contains identifiers, descriptions, units of measurement, and minimal and maximal values for diagnostics [59].

Summary The above-mentioned terminologies are internationally the most widely recognized classification systems in primary care. As noted, however, few countries use them in practice without any alternations. The reasons for these modifications differ: lacking concepts (i.e. adding missing diseases), refinement of concepts (i.e. adding sub concepts), and reduction (i.e. superfluous concepts are removed). Table 1 shows how countries have implemented the international terminologies or, if they haven't, which national system they use instead.

TABLE 1: **Overview of terminologies implemented in four European countries.**

	United Kingdom (UK)	Netherlands (NL)	Germany (GM)	Denmark (DK)
Disease	READ-UK	ICPC1-NL	ICD10-GM	ICPC2-DK
Drug	ATC	ATC	ATC	ATC
Dosage	SIG-UK	SIG-NL	SIG-GM	SIG-DK
Measurement	READ-UK	NHG Table 45	LOINC	IFCC-IUPAC NPU

RELATIONSHIPS BETWEEN TERMINOLOGIES

Matching

The process of matching encompasses the finding of values with similar meanings in two terminologies. It is a precursor for the mapping process, in which the terminologies are related in a schema, with regard to structures and cardinalities. In this study, many terminology mappings were already available, and the matching process was limited to the few classification systems which had not been matched yet. The process of matching, in these cases, was performed through manual assignment. The authors would list all values from a certain terminology that needed matching, and would iterate over every other value from a different terminology looking for an equivalent. As the terminologies that needed matching were relatively small, this approach was feasible.

For most disease terminologies, mappings (and thus matchings) were readily available. The only exception to this was the Dutch ICPC1-NL variation; the sub values were manually related to equivalents in ICPC2, where possible. Table 2 summarizes the routes taken to exchange terms of any one disease terminology to another one.

As most dosage terminologies were not standardized, no mappings existed between them. By dividing the lists of common British and German SIG codes into the Dutch format for frequency, form, and instructions, they could be matched with Dutch equivalents. Use of the Dutch T25 table as an intermediate terminology in all cases makes presentation of a table for dosage routes similar to Table 2 superfluous.

TABLE 2: **Overview of the routes when matching one disease terminology's terms to those of another. Values between brackets refer to the matching available in references, [M] implies manual assignment by the researchers, lacking available resources.**

		To:				
		ICPC1-NL	ICPC2(-DK)	ICD10	ICD10-GM	READ-UK
From:	ICPC1-NL	-	->ICPC2 [M]	->ICPC2 [M] -> ICD10 [159]	->ICPC2 [M] -> ICD10 [159] -> ICD10GM [78]	->ICPC2 [M] -> ICD10 [159] -> READ-UK [93]
	ICPC2(-DK)	->ICPC1-NL [M]	-	->ICD10 [159]	->ICD10 [159] -> ICPC2 [78]	->ICD10 [159] -> READ-UK [93]
	ICD10	->ICPC2 [159] -> ICPC1-NL [M]	->ICPC2 [159]	-	->ICD10-GM [78]	->READ-UK [93]
	ICD10-GM	->ICD10 [78] -> ICPC2 [159] -> ICPC1-NL [M]	->ICD10 [78] -> ICPC2 [159]	->ICD10 [78]	-	->ICD10 [78] -> READ-UK [93]
	READ-UK	->ICD10 [93] -> ICPC2 [159] -> ICPC1-NL [M]	->ICD10 [93] -> ICPC2 [159]	->ICD10 [93]	->ICD10 [93] -> ICD10-GM [78]	-

Mapping

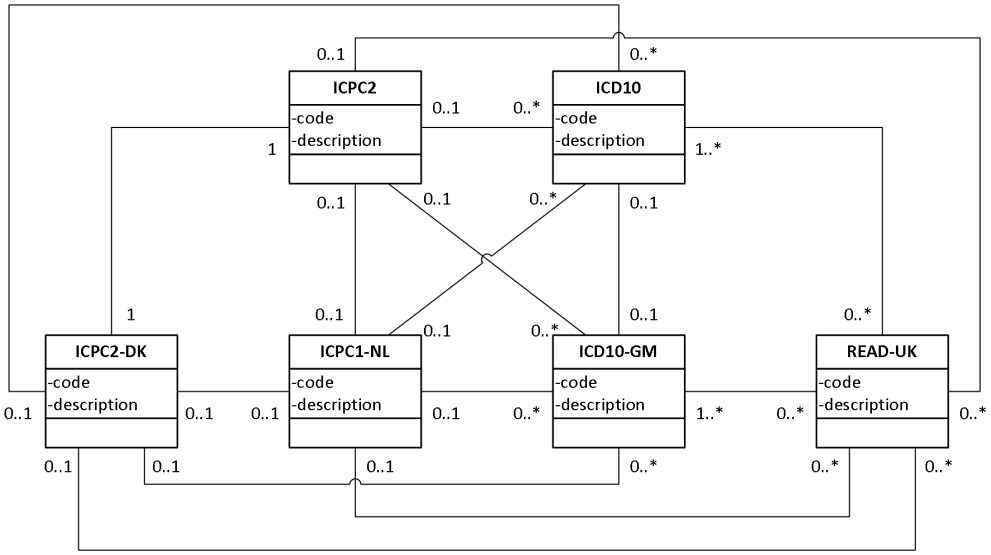
In order to create a model for information exchange between existing medical terminologies, the relations between them should be defined. As shown in the previous section, all concepts shown in Figure 1 can be modeled by at least one terminology (with the exception of complaint). The complaint is a (very) short statement by the patient about his or her illness; as it is subjective and uninterpreted, it cannot be linked to predefined terminologies. On basis of Figure 1, it can be concluded that the remaining concepts can be expressed as one of these four: disease, drug, dosage and measurement. Contraindications, being either drug allergies or comorbidities, can be expressed as (restrictions on) drugs or as diseases, respectively. As explained above, complaints cannot be expressed through any terminology.

In the model, all terminologies containing similar concepts are related to one another. Where possible, existing mappings were used to determine the degree to which concepts of one terminology could be meaningfully expressed in another. Where existing mappings were not available, they were created specifically for this study. With the exception of ICPC1-NL, all terminologies depicting diseases had mappings available. Between the dosage terminologies no mappings existed at all. Between the international measurement terminologies, no mappings existed; only a limited mapping of the Dutch system to LOINC has been published.

The results of this modeling approach can be seen in Figure 2 and Figure 3. Both international standards (e.g. ICPC2), regardless of whether they have been implemented, and actual national adaptations (e.g. ICPC1-NL) have been included. The relations' cardinalities depict to what degree the terminologies' concepts can be expressed in those of another. A 1-1 relationship would infer that all concepts of either terminology have exactly one match in the other. A 1-0..1 relationship defines that any concept in the first terminology has either one or no equivalents in the second terminology. Similarly, a 1-0..* relationship means that any concept of the first terminology has any number (including zero) of equivalent concepts in the other. The cardinalities are a comprehensive result of the matching process, which linked concepts from different terminologies to one another. This means that two terminologies between which a 0..*-0..* relationship exist may contain concepts for which 1-1 relations exist. These specific matching details will be addressed in the implementation phase and the creation of the inference rules, which will be described later.

Diseases

FIGURE 2: **UML class diagram depicting terminologies of diseases and their interrelations.**



A total of six terminologies able to classify diseases have been included in the model. All of them describe concepts as code-description combinations, whereby the code acts as a unique identifier.

The international standards ICPC2 and ICD10 were originally mapped when the former was released. As ICD10 holds much more specific diseases, only a small number of its concepts have equivalents in ICPC2. On the other hand, most ICPC2 concepts can meaningfully be expressed in ICD10 codes; for some unspecified concepts, such as L76 – Fracture (other), manual intervention

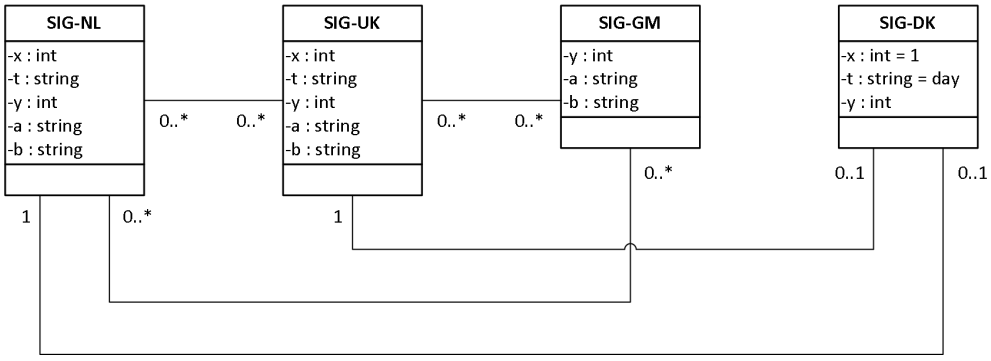
would be required to successfully select a corresponding ICD10 code out of a list of possibilities. Some subjective complaints, such as P05 – Feeling old, cannot be mapped to ICD10 concepts.

Three of the national terminologies studied, the Dutch, German, and Danish, are based on international standards. The Dutch ICPC1-NL is an amended version of the original ICPC1 terminology, enhanced for primary care practice with sub concepts (such as T90.01 – Diabetes mellitus type 1). Some of these sub concepts were added to the international version when ICPC2 was released, enabling the successful mapping of a small number of ICPC1-NL sub concepts to ICPC2 main concepts. The Danish version of ICPC2 is merely a translation of the original terminology, meaning that all concepts can be meaningfully transferred between the two. The German version of ICD10 was altered to better correspond to primary care needs. It excludes some diseases that are very rare in Western Europe (such as bubonic plague) and chapters detailing morbidity and mortality causes. This results in most, but not all, concepts being mapped between the two terminologies.

The READ terminology, finally, is a separate classification having been developed and widely used in the United Kingdom. The official issue of READ codes by the British Health and Social Care Information Centre contains a mapping to ICD10. As it contains a very comprehensive terminology, most ICD10 codes can be successfully mapped to READ concepts, and vice versa.

Drugs and Dosages Unlike diseases, drugs are consistently classified internationally using the ATC terminology. While some countries employ additional national classification systems for diverse purposes, they are complementary to the international standard. The availability of drugs throughout different countries may differ, however, causing interoperability issues when transferring drug records between countries' systems.

FIGURE 3: **UML class diagram depicting terminologies of drug dosages and their interrelations.**



For the notation of drug dosages international standards do not exist. Countries have often implemented national approaches, which are usually created bottom-up and which differ in their degree of standardization. Dosage notations exist of the frequency with which certain specified units should be taken, optionally augmented with instructions (such as take two pills once a day, at mealtimes).

The Netherlands is the only country with a fully implemented dosage notation system. It uses the format XTYAB, where X and Y are (ranges of) numbers and T, A and B are standardized strings. XT denotes the frequency (e.g. 2D = twice per day), YA the form (e.g. 1DR = one drop), and B the optional instructions (e.g. ALCO = be careful with alcohol in combination with this drug).

While other countries support de facto standards for the T, A and B values, no recognized national standards are implemented. The UK uses a combination of Latin and English abbreviations as dosage notations (e.g. gtt = drop, gr = grain). In Germany, Latin and German terms are commonly in use (e.g. qs, quantum satis = as many as required, fein = fine), although there are no widely accepted de facto standards to denote frequency.

Denmark, finally, does not support any standardized format for dosage notations, except for specifying the number of units to be used daily (e.g. once a day two [tablets]).

Measurements Mappings between the two major laboratory terminologies, LOINC and NPU, do not exist. Only a limited, incomplete mapping of the Dutch Table 45 to LOINC could be accessed. Since all laboratory terminologies exist of, at least, thousands of entries, creating a custom mapping for this study was not feasible. Consequently, no assessment of laboratory terminologies could be made, and in the remainder of this paper the measurement concept is therefore omitted.

IMPLEMENTATION

Logic

Information exchange between the terminologies is constrained by inference rules, which can be implemented through a rule engine. Objects used as input for the conversion process, such as diseases or dosages, are registered with the rule engine. If the rules are being fired, all objects are iterated through and applicable rules' consequences executed. The conditions of the rules generally require one terminology to have filled-out values for a certain object. If these conditions are met, rules' consequences consist of amending the object by adding equivalent values of other terminologies to it. Drools does not dictate the order in which objects should be iterated through, nor the order in which applicable rules should be run. The process simply continues until no more rules' conditions are satisfied.

To illustrate the logic, the rule which assigns an ICPC2 value to a disease object which already has an ICPC1NL value is shown below as pseudocode. The rule ICPC1NL_to_ICPC2 would be executed for all objects having an ICPC1NL value and omitting an ICPC2 one. The arrays defined in the rule's first two rows represent the database tables containing the ICPC2 terminology, and the mapping between it and ICPC1NL. The rule engine logic allows for certain conditions to take prevalence over others, making it possible to search for a direct equivalent before resorting to the mapping. In this case, ICPC1NL and ICPC2 being close relatives, a direct equivalent of the ICPC1NL unique identifier is sought for in the ICPC2 table. If it is not found, the mapping between the two terminologies is consulted; the mapping will correct for changes between the first and

second version of ICPC, and Dutch customizations. In both cases, a new ICPC2 terminology value is assigned to the disease object, after which the rule terminates.

```
function ICPC1NL_to_ICPC2(disease)
        let String[] all_icpc2_codes = ["A01","A02",...]
        let String[][] icpc1nl_to_icpc2 = ["B01.01",...]["Z01",...]
        let ICPC1NL icpc1nl = disease.terminologies.icpc1nl

        if icpc1nl.code in all_icpc2_codes then
                set disease.terminologies += new ICPC2(
                all_icpc2_codes[icpc1nl.code]
        )
        else if icpc1nl.code in icpc1nl_to_icpc2 then
                set disease.terminologies += new ICPC2(
                        icpc1nl_to_icpc2[icpc1nl.code]>
                )
end
```

Code sample illustrating matching between disease terminologies.

Figure 4 and Figure 5 below graphically depict the decision rules' execution flow. Both processes are iterated for every object in working memory, and its rules are run parallel to each other. The decision nodes in both diagrams represent the conditions; if a condition is not met, either another one is tested or the process is terminated. For the sake of readability, routes leading to terminations are displayed as dashed lines. Consequences are modeled as actions and may incorporate formulae or database queries, which are not explicitly shown.

Figure 4 shows the rules used to convert between different disease terminologies. Most mappings exist between national implementations and either one of the two international standards, ICPC2 or ICD10. All rules follow the same basic structure: first, a specific terminology value should exist which can be used as input to convert to another classification scheme; second, no value for the subsequent resulting terminology should already be assigned to the object; third, the output terminology should contain an equivalent value for the input value; finally, if all conditions are satisfied, the object is modified by assigning to it the equivalent value of a new terminology.

After assigning a new value to an object, all rules are checked again and trigger if their conditions are satisfied. If a condition is not met (and no further decision nodes are available), the process ends. Since all rules are run in parallel, this does not necessary mean the rule engine process terminates. Sequential processes may still be running; new assignments to the object may even trigger an earlier-terminated rule. Eventually, though, no conditions will be satisfied anymore and the rule engine will terminate.

In Figure 5 the rules necessary to convert between dosage notation forms are graphically depicted. The implementation uses the elements isolated in the Dutch SIG notation: number and frequency (X and T), amount and form (Y and A), and instructions (B). Even though most dosage systems do not support all elements, this distinction enables optimal conversion between them; while separate notation values are retained, prevalence can be given to combined values if desired. When converting between SIG-NL and SIG-UK, for example,

preferably a single equivalent value for X+T is assigned (e.g. 2D = twice a day = b.d.s.). Only if none is available are separate values considered.

FIGURE 4: **UML Activity Diagram depicting the disease rules (routes to dead-ends are dashed).**

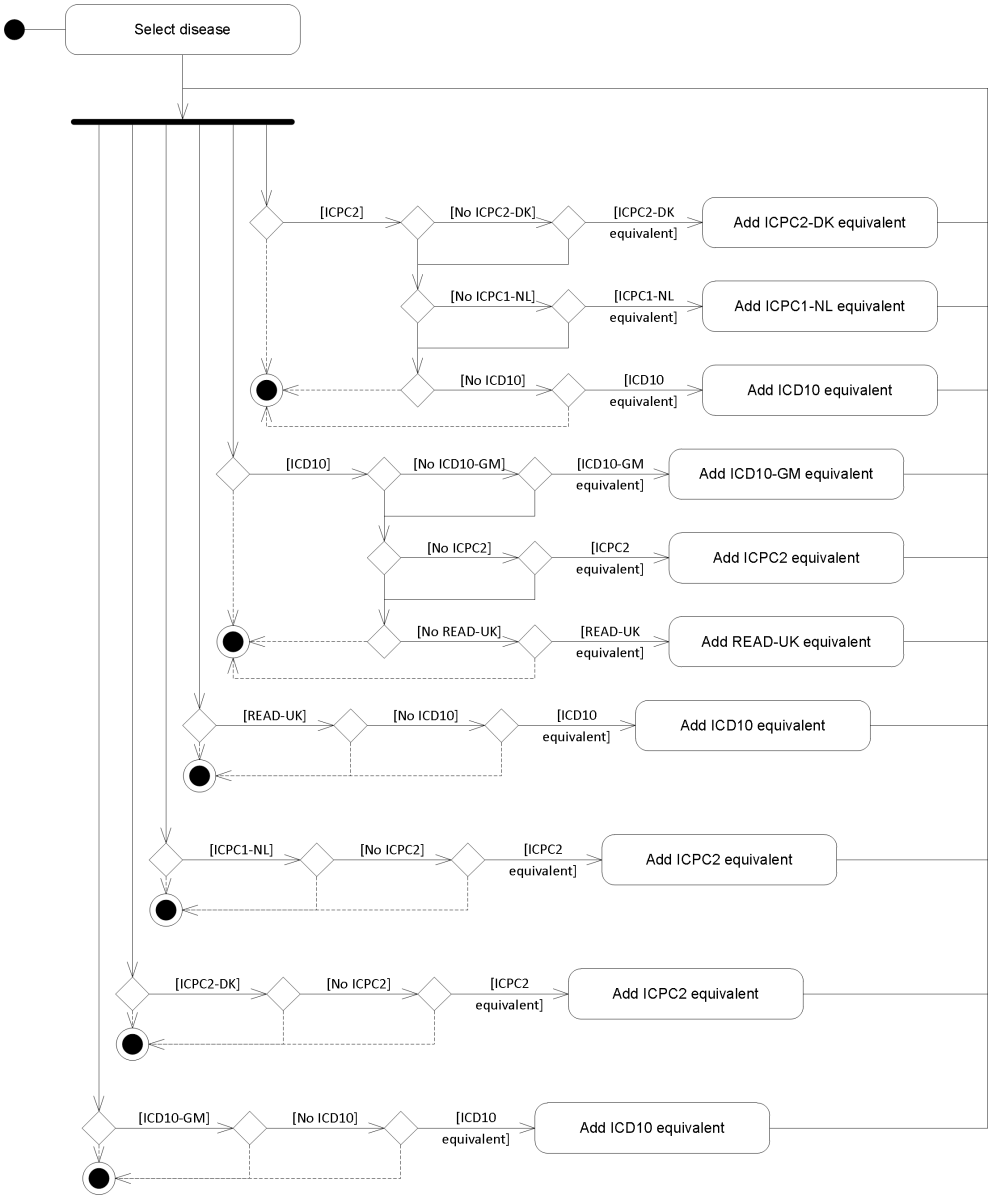
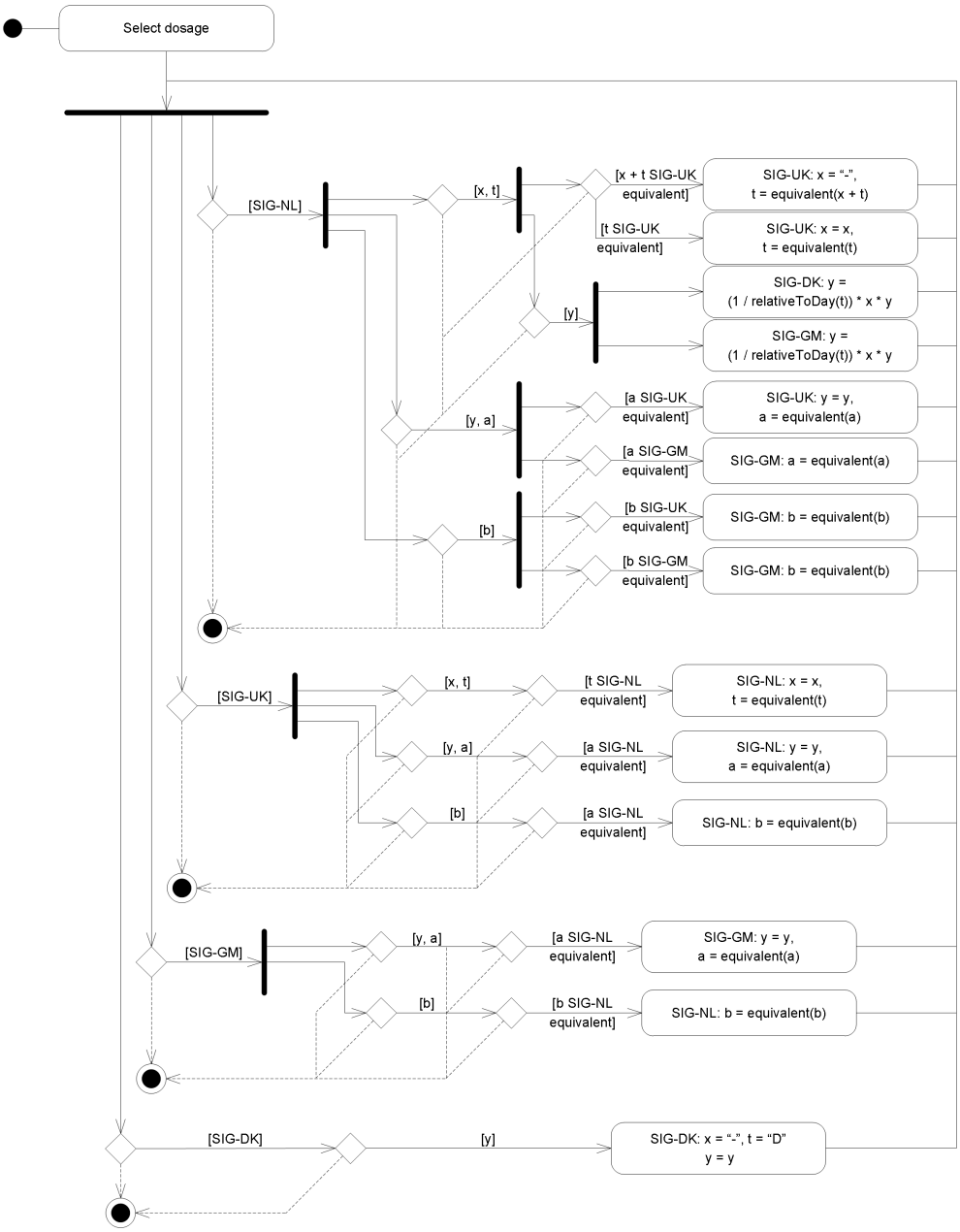


FIGURE 5: **UML Activity Diagram depicting the dosage rules (routes to dead-ends are dashed).**



Consequences for dosage rules can differ. In their simplest forms, only single terminology attributes are assigned values, but often combinations are filled-out (X = '-', T = equivalent(X+T)). In some cases, formulae are used to convert complex frequencies (e.g. twice a week) to daily

amounts. For these calculations, the database mapping tables contain relative values; all frequency tables, for example, contain numbers that indicate their relation to days (e.g. week = 7, hour = 1/24).

Architecture

In order to be able to test the information exchange model described above against semantic interoperability parameters, it should be implemented. A multitude of implementation options exist for rule-based systems, and the approach described here is illustrative. A web service built on a Java backend uses a rule engine to incorporate the logic necessary to exchange data between terminologies. Results of the exchange process are communicated to the frontend through the JSON format.

For every two terminologies for which conversation mappings exist, a MySQL database table has been implemented containing this mapping of unique identifiers. Regarding diseases, tables exist for the conversion of ICPC2 and ICPC1-NL, ICPC2 and ICPC2-DK, ICPC2 and ICD10, ICD10 and ICD10-GM, and ICD10 and READ-UK. To convert between terminologies for which no direct mapping exists, intermediate terminologies are used, processed by decision rules described earlier. All dosage values are mapped to their SIG-NL equivalents, which is in all cases used as an intermediate.

TABLE 3: **Sample output of decision rules.**

ICPC1NL	ICPC2	ICPC2-DK	ICD10	ICD10-GM	READ-UK
Diabetes mellitus (T90)	Diabetes non-insulin dependent (T90)	Diabetes type 2 (T90)	Non-insulin-dependent diabetes mellitus (E11)	Diabetes mellitus, Typ 2 (E11)	Diabetes mellitus (C10..)
			Malnutrition-related diabetes mellitus (E12)	Diabetes mellitus in Verbindung mit Fehl- oder Mangelernährung (E12)	Diabetes insipidus,diabetes mellitus, optic atrophy and deafness (X008t)
			Other specified diabetes mellitus (E13)	Sonstiger näher bezeichneter Diabetes mellitus (E13)	Type 2 diabetes mellitus (X40J5)
			Unspecified diabetes mellitus (E14)	Nicht näher bezeichneter Diabetes mellitus (E14)	Insulin-treated non-insulin-dependent diabetes mellitus (X40J6)
					(20 more)

Drools is a forward-chaining inference based rule engine [173]. The logic necessary to convert objects from one terminology to another was incorporated into Drools as rules, i.e. paired conditions with consequences. These consequences directly modify the object the rule was activated with, thus enabling other rules to fire on the object's new properties. For example, a disease object may have a filled-out ICPC1-NL code and fire a rule converting it to an ICPC2 code. This new ICPC2 property on the object may in turn fire a rule converting it to an ICD10 code.

For example, inputting a disease object with ICPC1-NL code-description combination T90 – Diabetes mellitus will trigger a chain of rules resulting in the object being expanded with the terminology values shown in Table 3. The different cardinalities allow for some terminologies to incorporate multiple values.

EVALUATION

Completeness

As described before, a key determinant of data quality is its completeness, which is the extent to which data's expected attributes are present in a terminology. In a medical scenario, data would be consistently complete if it contained exactly the same information regardless of which terminology was used to describe it. For example, when describing dosages, all additional instructions b should be systematically included in all terminologies.

The model described above was tested for its completeness by measuring the extent to which its terminologies can contain each other's values. Orme, Yao and Etzkorn (2007) introduce many metrics for determining ontologies' complexity and cohesion [161]. These measurement methods were used a starting point to formalize metrics for measuring the degree of completeness between any two vocabularies in an ontological model.

The metric designed for this process is described below. The sections following it include tables containing the results of these equations for the disease and dosage terminologies. The first number in each cell shows how many concepts of one terminology have equivalents in another, the second is the same number expressed as a percentage of the first terminology. For terminologies between which direct mappings exist, this was done by determining how many unique values had equivalents, and relating that to the complete number of terminology entries. For terminology pairs that required intermediate terminologies, results from each intermediate step were used as input for the next as sub queries.

In order to formalize the translations between terminologies, a set theoretical approach is adopted. A terminology then is defined as a pair T= where V describes the set of vocabulary and A denotes the set of axioms that specify the interpretation of the vocabulary set. In this context, translation implies the conversion between the vocabularies in different terminologies.

Let there be some binary relation r between the vocabulary set of two terminologies, $r \subseteq V_1 \times V_2$. We call this relationship a terminology alignment. Our goal is to compose terminologies to

derive new translations from any two arbitrary terminologies. Composition of terminology alignment is defined as follows, given r_{12}= <V_1,V_2> and r_{23}= <V_2,V_3>:

EQUATION 1:

$$r_{13} = r_{12} \cdot r_{23} = \left\{ \left(v_1, v_3\right) \;\middle|\; \left(v_1, v_2 \in r_{12} \wedge \left(v_2, v_3 \in r_{23}\right)\right) \right\}$$

The quality of terminological alignment is then quantified through a function sim, which defines the ratio of similarity of a terminology t_1 to t_2:

EQUATION 2:

$$\mathrm{sim}\left(t_1, t_2\right) = \frac{\left|\left\{ v_1 \;\middle|\; \left(v_1, v_2\right) \in r_{1,2} \right\}\right|}{\left|v_1\right|}$$

In layman's terms, for each term in a terminology the number of equivalent terms in another terminology is determined through a lookup table containing the mapping between the two. A lookup table would take the form of Table 3, showing the concept diabetes mellitus expressed as ICPC1-NL, ICPC2, and ICPC2-DK terms (all T90), their equivalents in ICD10 (E11-E14), etc.

Diseases

TABLE 4: **Degree to which disease terminologies can express other terminologies' concepts. Absolute values denote the number of terms in the vertical terminology that can be expressed in the horizontal terminology, values between brackets are the absolute values expressed as percentages of the vertical terminology.**

		To:				
		ICPC1-NL	ICPC2(-DK)	ICD10	ICD10-GM	READ-UK
From:	ICPC1-NL	-	738 (57,43%)	698 (54,32%)	626 (48,72%)	665 (51,75%)
	ICPC2(-DK)	719 (99,04%)	-	686 (94,49%)	614 (84,57%)	652 (89,81%)
	ICD10	2.832 (23,35%)	2.893 (23,85%)	-	9.975 (82,23%)	10.578 (87,20%)
	ICD10-GM	2.737 (24,16%)	2.779 (24,53%)	9.975 (88,06%)	-	8.665 (76,50%)
	READ-UK	5.658 (9,46%)	5.729 (9,58%)	51.622 (86,31%)	47.731 (79,80%)	-

In Table 4 the results of the completeness test are displayed for the disease concept. Only limited mappings between these terminologies exist, usually between the most adjacent ones in terms of extensiveness. Mappings between ICPC1NL and ICPC2, ICPC2-DK and ICPC2, ICPC2 and ICD10, ICD10 and ICD10-GM, and ICD10 and READ-UK were used.

Noteworthy observations are that the more extensive terminologies (ICD10 and READ-UK) are able to express higher percentages of simpler terminologies (ICPC) than the other way around. Most diseases included in ICPC2 can be expressed in ICD10 and READ-UK as well. Only a limited set of ICPC1-NL sub codes can be expressed in ICPC2 and likewise in more distinguished terminologies. READ-UK concepts can be expressed rather well in ICD10 and vice versa. When attempting to express concepts from these more extensive terminologies in the lower-level ICPC-terminology, only very few can be successfully exchanged.

Dosages

TABLE 5: **Degree to which dosage terminologies can express other terminologies' concepts. Absolute values denote the number of terms in the vertical terminology that can be expressed in the horizontal terminology, values between brackets are the absolute values expressed as percentages of the vertical terminology.**

		To:			
		SIG-NL	SIG-UK	SIG-GM	SIG-DK
From:	SIG-NL	-	89 (11%)	37 (4%)	1 (0%)
	SIG-UK	109 (44%)	-	18 (7%)	1 (0%)
	SIG-GM	45 (55%)	18 (22%)	-	0 (0%)
	SIG-DK	1 (100%)	1 (100%)	0 (0%)	-

Table 5 shows the results of this test for the dosage concept. As is evident from Figure 3, the Dutch SIG-terminology is able to systematically contain all dosage sub concepts (albeit not all possible values). As such, it was chosen as an intermediate to convert values from one terminology to another.

Being the most extensive dosage terminology, SIG-NL can only be partially expressed in others. 11% of its values can be expressed in the British SIG terminology, 4% in the German SIG terminology, and virtually none in the Danish one. Conversely, many values of other terminologies (between 44% and 100%) can be expressed in the Dutch terminology. Amongst the values that are consistent throughout SIG-NL and SIG-UK are especially many (x)+t and a inputs, such as D = d (day) and C = caps (capsule). SIG-GM does not support systematic recording of t, but limited categorization of a, such as T = pil (tablets). The Danish terminology only supports the number of non-specified units per day, enabling only t-value D = d (day).

Eighteen of the a- and b-values in the British dosage terminology can be expressed as their German equivalents. Exactly the same number of concepts in SIG-GM can be expressed in the British terminology. These include tab = pil (tablet) and aq = aqu (add water). SIG-UK does support t-values, so D = d (day), the only Danish SIG code, can be expressed as well. As SIG-DK does not support t-values, it cannot be expressed in the Danish terminology.

Accuracy

A second key determinant of data quality is accuracy, interpreted as the degree to which real world objects are correctly represented by the data. In the medical scenario explored in this study, data would accurately reflect real world objects if it would retain meaningful information regardless of which terminology was used to express it.

Two ways in which conversions can yield inaccurate results can be distinguished: either erroneous information is added (e.g. when converting diabetes mellitus to diabetes mellitus type A) or correct information is omitted (e.g. when converting testicular cancer to cancer). While qualitative judgment is necessary to determine whether or not a matching between values of two terminologies is partially inaccurate, cardinalities can be used as indicators. Given that terminologies contain no synonyms, any one-to-many relationship would mean that all conversions, except at most one, contain partial inaccuracies, i.e. do add superfluous information. After all, if a concept diabetes mellitus in a certain terminology has relations to three other terminologies in a different classification system, only one of these could be exactly similar; the other two would have to be sub concepts of the disease (such as diabetes mellitus types A and B). Likewise, any many-to-one relationship between terminology values, except at most one, would result in loss of accuracy.

This principle, which states that any relationship other than one-to-one has to result in a certain degree of inaccuracy, is formalized in the following formula. The equiv function represents the qualitative assessment to determine the equivalence of two values. The match function below denotes the equivalent, if any, of a vocabulary term v_1 in an arbitrary terminology t_2.

EQUATION 3:

$$\mathrm{match}\left(v_1, t2\right) = \exists_{v2}^{\leq 1} \cdot \left(v_1, v_2\right) \in r_{12} \wedge \mathrm{equiv}\left(v_1, v_2\right)$$

Diseases The differences between the Dutch version of ICPC1 and the later ICPC2 are relatively few. As ICPC2 was designed to be backwards compatible with its earlier version, all concepts which were included in the original ICPC1 can be expressed in ICPC2, and vice versa. Incompatibility problems arise when trying to express the Dutch sub concepts in ICPC2; a detailed recording of S12.01 – Mite bite can only be expressed in ICPC2 as the more generic S12 – Insect bite or sting, thus losing information. In fewer instances, ICPC2 concepts cannot be mapped to ICPC1-NL (sub) concepts without losing details, or at all; A11 – Chest pain was introduced in ICPC2 and has no ICPC1-NL equivalent.

While a relatively high percentage of ICPC2 concepts can be expressed in ICD10 codes, the difference in perspective between the two classification systems hinders accuracy. While both terminologies strive to categorize diagnoses, ICPC2 also includes subjective complaints as they are experienced by patients. For example, ICPC2 contains a code-description pair H13 – Plugged feeling in ear; the nearest equivalent in ICD10 is H93.8 – Other specified disorders of ear. The latter code is neither a specified disorder nor an accurate description of the complaint. Vice

versa, ICPC2 cannot accurately express many of the more detailed diagnoses included in ICD10. For example, the nearest ICPC2-equivalent for the ICD10 concept E12 – Malnutrition-related diabetes mellitus is T90 – Diabetes mellitus. In this case, the specific causes of the disease are not retained.

The Danish version of ICPC2 is merely a translation and has exactly the same expressive capabilities as the international standard. The same is not true of the German version of ICD10; it is as much an adaptation of the English standard as a translation. Concepts recording causes for morbidity and mortality were deemed unnecessary in primary care practice and therefore excluded; an example is B20.2 – Cytomegalovirus following HIV infection. Additional concepts deemed useful in primary care practice were added, such as D70.0 – Compromised immune system after radiation; no equivalent for this concept in the standard ICD10 issue exists.

The British READ terminology, because of its larger scope, supports a far larger number of concepts than the other terminologies. Consequently, all concepts present in the implemented ICPC and ICD10 versions can be expressed meaningfully in READ code-description combinations, while only a limited number of READ concepts can be reflected in other terminologies. Concepts as detailed as X40JO – Congenital lipoatrophic diabetes cannot be expressed in other terminologies while retaining their specificities.

Dosages When attempting to express dosage values in different terminologies, a distinction between the T, A, and B values sharply distinguishes itself. The first two values, denoting frequency and form, can either be readily expressed or not at all. The instructions described in the B value can often be partially reflected.

No standardized versions to denote frequency exist in the German de facto system, but the Dutch and British terminologies overlap for the more common T-values. Taking a tablet twice daily can be expressed in the Dutch system as 2D or ..., or in the British system as bds, bd, or bid. Highly unusual values in the Dutch standardized system (WT = every two weeks) cannot be mapped to the British system.

The Danish system cannot technically represent A- or B-values, and has a fixed value of one day as T-value. Thus, only the number of units to be taken daily can be expressed. When units are indeed prescribed for daily use this is adequate, but when the frequency differs this results in odd values which do not represent the intended dosage (once per week = 1/7th per day).

The A-values in the terminologies often have clear equivalents in the Dutch system and the British and German SIG codes. The form variant tablet can be expressed in the Dutch system as T, in the British one as tab, and in the German one as pil. Equivalents for common forms such as drops (DR, gtt, gtts) or capsules (C, cap, caps) are also readily available. The more unusual forms available in the Dutch terminology (such as BR = effervescent ('bruistablet')) have no standardized terms in the British or German systems.

Dosage notations can have unlimited numbers of additional instructions, known as B-codes. While most notation forms (i.e. the Dutch, German, and British) support instructions, their

detailed nature makes accurate conversion difficult. While the more common instructions (e.g. MW (SIG-NL) = aq (SIG-UK and SIG-GM) with water) are available in all terminologies, the more specific ones are not at all, or only partially, convertible: the less common Dutch instructions 'do not combine with driving' and 'avoid sunlight' have no standardized equivalents in other terminologies. Others, such as 'add water, not milk', or 'take during or shortly after dinner', can only be partially converted. SIG-UK also contains some specific b-values which cannot be converted into other terminologies, such as CST – Continue same treatment. The British and German systems also contain instructions for pharmacists (e.g. d.t.d., dentur tales doses = give of such doses), which cannot be expressed in different terminologies.

DISCUSSION

In this study the researchers have explored the extent to which primary care terminologies can meaningfully express each other's concepts. In order to determine this, a model containing these terminologies and including axioms to define their relations was developed. Generic metrics were designed to determine the completeness and accuracy of any two arbitrary vocabularies within an ontological model. Tests investigating completeness and accuracy show that simpler disease terminologies, such as ICPC2, can be expressed relatively well in more extensive ones, such as ICD10; mainly concepts which do not equate to diagnoses, such as subjective complaints (e.g. Plugged feeling in ear), cannot be mapped to ICD10 concepts. Vice versa, only a limited number of concepts of extensive terminologies can be mapped to simpler ones and still retain their meaning; the highly specific nature of many ICD10 entries (e.g. Malnutrition-related diabetes mellitus) makes accurate matching with ICPC concepts difficult.

With the exception of the Netherlands, standardizations for dosage notations do not exist. When testing de facto notation standards, it appears that only the most common abbreviations have equivalents in most terminologies. Completeness among these classification systems is thus negligible. Only the most common terms have unambiguous counterparts in most terminologies. The accuracy for more complex or less frequently used terms is compromised; terms can often only be partially matched to others (e.g. add water, not milk, aqua, water).

The inability to meaningfully exchange large numbers of concepts between primary care terminologies creates significant semantic interoperability challenges. These are commonly addressed by advocating the use of an overhauling terminology, such as SNOMED-CT or the UMLS Metathesaurus. Matching concepts of primary terminologies to metaterminologies is not necessarily a flawless process, however. Schulz et al. note structural deficiencies between ICD and SNOMED that impair matching between them [184]. In an attempt to map the Swedish translation of ICD10 to SNOMED-CT, Vikström et al. find that even if extensive mapping rules would exist, "obstacles to high quality mapping remain due to structure and content characteristics in both coding systems" [220]. Thus, even if the majority of primary care concepts have defined meanings in a metaterminology, this does not guarantee complete semantic interoperability. At least two problems persist.

The first issue, and the most practical one, revolves around information exchange between systems using different underlying terminologies. While implementation of a metaterminology would guarantee that a concept has an unambiguous meaning, there is no guarantee that it could be represented in both systems' lower-level terminologies, even if systems would use (different) sub sets of this metaterminology. This makes exchange of patients' health records between different countries difficult. It also greatly hinders communication between primary and secondary care in many nations. Many countries' secondary care systems use ICD10-variants, whereas in primary care ICPC2 is prevalent [53]. This means that, even though the meaning of concepts would not be obfuscated, they could not be directly used as input in the different systems. In practice, this means that family doctors' lists of diseases cannot be entered into secondary care systems, nor can specialists' diagnoses be fed back into general practitioners' systems. Differences in cardinalities of the relations between similar concepts hinders automatic mapping in these cases, making human judgment necessary.

The second issue is of a more conceptual nature and addresses the way in which terminologies limit primary care physicians in their conceptual assessment of diagnoses. The area of cognitive semantics has long held that, considering language to be governed by general cognitive principles, it can "only describe the world as it is organised [sic] within people's conceptual spaces" [45]. From this follows that if language is limited due to dependency on fixed terminologies, people's conceptual spaces may be altered. Consider this example in the primary care domain: to register a patient complaining about headaches, a Dutch physician is limited to the choices offered by the ICPC1-NL terminology. Apart from the generic entry, ICPC1-NL allows for the diagnosis of three further forms of headaches, i.e. migraines, drug-induced headaches, and stress-headaches. According to the Dutch College of General Practitioners (NHG), however, recurring headaches are prime examples of discomforts related to psycho-somatic causes, such as anxiety disorders [61]. As the NHG notes, no ICPC1-NL code for registering psycho-somatic disorders exists, let alone one for psycho-somatic-induced headaches. As a result of ICPC1-NL's limited options, a physician may be prone to overlook this possibility and register the patient's complaints as stress-related headache, without considering further examination.

Both these issues cannot be overcome by mapping existing classification systems to an overhauling metaterminology. The complete replacement of existing terminologies by a new system could potentially solve them, and indeed the exclusive use of SNOMED has been advocated for by opinion leaders in practice [89,207]. Such an approach makes one raise concerns regarding its feasibility and desirability, however. Indeed, the terminologies currently in use offer unique perspectives on diagnosing and prescribing, and have been carefully attuned to adhere to the intricacies of daily practice. Replacing these with an alternative, and still retaining their scopes and perspectives, is not reasonably possible.

A compromising solution to this stalemate would be to employ sub sets of metaterminologies, which are semantically similar to existing lower-level terminologies. In this scenario, physicians would use concepts familiar to them in their specific health care settings,

while simultaneously classifying with standardized metaterminologies. When exchanging data with systems using a different sub set, the receiving party could consult the full metaterminology to unambiguously identify the concept.

In order to enable this next step in achieving semantic interoperability in primary care, existing national terminologies should be standardized where necessary (e.g. in the dosage domain this mostly still needs to be done), then related to a metaterminology such as SNOMED-CT or UMLS, and finally their formal, unaltered use advocated. Standardization of dosage notation forms, while retaining as many nationally used terms as feasible, has been proven successful in the United States [131]. Mapping both disease and dosage terminologies to a metaterminology would be essential in ensuring unambiguous application of terms.

CONCLUSION

This study attempted to explore the extent to which primary care data can be meaningfully exchanged between nationally implemented terminologies in primary care. Literature states that both accuracy and completeness of data suffer from suboptimal matching and mapping between terminologies. Tests on the information exchange model created in this study show that values of terminologies which are closely related can be relatively well-expressed in each other's concepts. Additionally, simpler terminologies holding relatively few values often have equivalents in more extensive classification systems. Concepts of the more comprehensive terminologies, however, can only be expressed in simpler systems to a very limited degree. Interpretation of these results leads the authors to believe that the current state of accuracy and completeness between primary care terminologies does not allow for sufficiently meaningful semantic interoperability.

Limitations & Further Research

Even though the researchers tried to provide a thorough exploration of the subject at hand, some factors limit the results' comprehensiveness and provide angles to be investigated in further research. Firstly, the measurement concept proved too extensive to map and thus was excluded from this study; exploring the differences in perspective and richness between the measurement terminologies (e.g. LOINC and NPU) should be the topic of additional research. A similar case can be made for other disease, drugs, and dosage terminologies used in other countries; their adaptations of international standards and national systems should be tested for their accuracy and completeness. This study only included terminologies used in the prescription process, excluding those applied to other processes in primary care; similar analyses could be conducted on classification systems for, for example, mental disorders (DSM-IV) or surgical interventions (ICD-10-PCS).

Additional emphasis should be placed on exploring the problems associated with implementing an overhauling terminology, such as SNOMED. While this approach may improve

information exchange, it will create new issues regarding the limits of current terminology concepts and physicians' conceptual spaces, which should be addressed.

ACKNOWLEDGEMENTS

The manuscript does not contain clinical studies or patient data. No sources of funding were used to assist in the conduct or preparation of this study. The authors have no conflicts of interest that are directly relevant to the content of this study.

5.2

Risk Mediation in Association Rules

ABSTRACT

Applying discovered association rules in precarious domains – i.e. domains in which association rules' consequences have potentially major impact on their datasets – can be dangerous. Metrics commonly used in association rule mining, such as confidence, lift, and conviction, cannot guarantee their safe application in precarious domains. Applying domain-dependent risk assessment after the mining process can aid in determining the applicability of discovered association rules. We propose a model for the incorporation of risk in association rule application, expressing risk as a function of a rule's probability and the severity of its consequences. This model is implemented in a primary care setting.

Manuscript submitted as: Meulendijk M.C., Spruit M.R., & Brinkkemper S. Risk Mediation in Association Rules: The Case of Decision Support in Medication Review (submitted).

The model was validated using data from a medical recommender system gathered in a randomized controlled trial. The model's outcomes are found to have predictive value when tested against decisions made by physicians on 261 patients' health records. An independent t-test showed a statistical difference in the risk associated with actions proposed by the recommender system which were followed (M = 2.42, SD = 0.57) and the risk of proposed actions which were not followed (M = 2.57, SD = 0.60); t(623) = 3.040, p = .002. Scenarios in which the inclusion of risk assessment in association rule application can be beneficial are briefly discussed.

INTRODUCTION

Association rule mining remains one of the most prominent knowledge discovery methods in existence. Ever since their rise to prominence following Agrawal's algorithm, they have been put to use in a wide variety of contexts [5,76,86,235]. With time, changing requirements, and improving technology, more efficient and extensive algorithms have been developed to perform association rule mining [97]. In essence, however, the core of association rule mining has remained the same: to discover interesting relationships between entities in a database and make them into reusable inference rules.

Association rules have been implemented in countless software applications in a wide variety of domains; the domain of primary care is among those [76,186]. While computerized physician order-entry (CPOE) systems found their legitimacy first and foremost in aiding physicians with administrative tasks, over the last decade they have evolved into full-fledged decision support systems. And, even though decision support systems in primary care often depend at least partially on explicit knowledge bases, association rules have been incorporated into their rule bases as well [144].

The typical example of association rule implementation, the 'market basket', reveals one of its weaknesses [97]. In a market basket analysis, grocery items that are frequently purchased together form an association rule, which in turn can be used to remind shoppers of items they may consider purchasing. In a safe, insensitive domain such as shopping this poses no problems. But in a precarious domain with vulnerable datasets, where association rules can have potentially far-reaching implications, these kinds of associative suggestions are risky; for instance, recommending medicine solely on the basis of what other physicians have prescribed in similar contexts can have dangerous consequences.

Therefore, in this paper we suggest the introduction of the concept of risk in association rule mining. Based on best practices in risk management, we propose a model for the incorporation of risk in association rules. We then validate it in a primary care setting and propose scenarios in which the combination of association rule mining and risk assessment yields useful results in practice.

BACKGROUND

Recommender Systems in Precarious Domains

As illustrated in the introduction, the impact association rules may have depends on the sensitivity of the dataset on which they are applied. This danger involved in applying discovered association rules to sensitive datasets in precarious domains has not received substantial attention. For the remainder of this paper, we define precarious domains as 'domains in which association rules' consequences have potentially major impact on its datasets'. An example of a precarious domain is that of primary care.

Recommender Systems in Primary Care The number and complexity of recommender systems in primary care has increased over the last decade. These clinical decision support systems have been shown to improve the quality of physicians' prescriptions [32].

The impact recommender systems' suggestions have on patients' well-being, however, makes the safety of these systems a major requirement; generating incorrect recommendations or failing to report problems may endanger patients' health. For systems which are not purely based on an explicit knowledge base, the quality and fail-safety of their decision rules are even more difficult to guarantee. For example, decision rules discovered through association rule mining usually have confidence levels far below 100%, indicating that in some of the cases the advice will be inappropriate.

In practice, this problem of quality insurance of clinical decision support systems has manifested itself twofold: alert fatigue and alert overreliance.

In an attempt to ensure their systems are as comprehensive as possible, developers tend to flood their users with warnings and suggestions. These excessive numbers of recommendations have in turn led to users experiencing 'alert fatigue': due to the overabundance of messages and users' time constraints, they have no choice but to ignore them [210].

In other cases, users have shown to fully rely on the recommendations generated by their decision support systems, to the extent that they followed inappropriate advices [121,143]. In these cases, recommender systems introduced problems which would otherwise not have occurred.

Information Sensitivity in Recommender Systems

The issue of information sensitivity in the mining or applying of association rules has been explored to a limited extent. Issues revolving around privacy, sensitivity, and risk have been touched upon by several authors, but do not have unambiguous definitions in the domain of recommender systems. In order to position the concepts proposed in this paper and illustrate our model's niche, we briefly outline related studies that introduce similar concepts.

Several approaches have been taken to address the issue of privacy preservation in association rules. Both the sensitivity of data from which association rules are discovered, and that of the knowledge contained within the association rules themselves, have received

attention. Evfimievski et al. propose a framework for mining association rules where the data has been randomized to preserve privacy of individual transactions [67]. Wu et al. propose an algorithm with which sensitive association rules can be hidden, ensuring that privacy-infringing information is not inadvertently disclosed [237]. Finally, Oliveira and Zaïane address the problem of protecting knowledge contained within discovered association rules, introducing an algorithm that filters strategically sensitive knowledge from databases [160].

The concept of risk has been applied to the domain of recommender systems to a very limited degree. Ali, Manganaris and Srikant propose a formula which tests the predictive value of discovered association rules, and estimates their 'relative risk'. Their goal is "to eliminate patterns that were not predictive even though they were frequent", thereby improving the appropriataness of their system's suggestions [6]. In a more user-centered approach, Bouneffouf proposes the concept of risk in context-aware recommendation systems, using a dynamic context-based risk level to ensure recommendations are as non-intrusive as possible. His prime goal regarding risk management is "to avoid disturbing the user in risky situations" [30], thus improving the safety of recommender systems' use.

This brief summary of earlier work shows that most studies addressing information sensitivity focus on privay preservation and appropriate recommendations. The potentially negative impact association rules may have when executed on sensitive datasets has not yet been explored.

Risk in Recommender Systems

A variety of measures is commonly used when mining association rules. These are all aimed at improving the accuracy with which suggestions in recommender systems can be generated. In this section, we briefly visit the metrics commonly used and their shortcomings. We then introduce the concept of risk as a post-mining metric and relate it to the leading association rule mining measures.

Metrics in Association Rule Mining A number of metrics are commonly used to determine the applicability of association rules. Each of these aims at improving the certainty with which association rules can be inferred.

The support of x in a dataset D is the proportion n of transactions in D which contain the items in set x:

EQUATION 1:

$$\text{support}(x \rightarrow y) = \frac{\sigma(x \cup y)}{n}$$

Together with confidence it is the most used metric in association rule mining [200].

The confidence of an association rule x→y is the proportion of transactions in a dataset D that contain both x and y:

EQUATION 2:

$$\text{confidence}(x \rightarrow y) = \frac{\text{support}(x \cup y)}{\text{support}(x))}$$

A confidence of 1.0 implies that all transactions contain both x and y. In such a scenario infering the association rule is a harmless activity. However, for association rules with confidence levels below 1.0, further metrics are often applied to determine its appropriateness [200].

Lift is the ratio of a rule's confidence to its expected confidence:

EQUATION 3:

$$\text{lift}(x \rightarrow y) = \frac{\text{support}(x \cup y)}{\text{support}(x) * \text{support}(y)))}$$

It has been introduced as a metric to determine the increase in probability of a set of antecedents. If a set of antecedents for a certain rule has a lift larger than 1.0, it can be concluded that the relationship between the antecedent and the consequent is more significant than would be expected if the two sets were independent [200].

Finally, an association rule's conviction is the ratio of the expected frequency that an antecedent occurs without its consequence:

EQUATION 4:

$$\text{conviction}(x \rightarrow y) = \frac{1 - \text{support}(y)}{1 - \text{confidence}(x \rightarrow y)}$$

It is used to determine the probability of the association rule not being applicable, and inferring it being a mistake [200].

The above-mentioned metrics commonly found in association rule literature and recommender systems provide safeguards against the incorrect application of decision rules. In most scenarios, however, they cannot guarantee flawlessness. If, for a specific rule, dataset and set of antecedents, the three metrics were calculated and returned a value below 1.0, its safe inference could not be guaranteed. An additional assessment would still be required to determine its applicability.

Furthermore, the above metrics are domain-independent, meaning that their results are similar for both a shopping basket example and a precarious domain such as primary care. As the real life implications of the inferences in these domains greatly differ, it is sensible to perform a domain-dependent assessment.

Risk as a Post-Mining Metric The domain-independent nature of metrics commonly used in recommender systems leads to the requirement of an additional assessment before association rules can be safely inferred. In this section, the utility of the concept of risk as a mediating factor in the application of association rules is explored.

Risk management is a field of research that has been applied in a wide variety of domains. It involves a set of "coordinated activities to direct and control an organization with regard to

risk" [107]. Risk is typically understood as the "probability of loss" in any given situation, or, in a more generic definition, the "effect of uncertainty on objectives" [107]. In risk management, risk is usually incorporated as a function of these uncertain effects' impact and their likelihood of occurring.

Association rules and risk share the concept of probability; association rules are expressed with a degree of confidence, while risk incorporates the likelihood of dangerous consequences. While association rules' impact is not a standardized concept, the implications their consequences have imply its risk's severity; after all, if a rule with major implications is inferred incorrectly, it will have severe consequences.

These relations between the concepts of association rules and risk are shown in Table 1. When determining their applicability, rules with high confidence and low impact can be safely executed, while rules with low confidence and high impact can be safely discarded. The remaining two types of rules need a risk assessment before they can be safely inferred. Rules with low confidence and low impact may not be dangerous to infer, but cannot be executed without assessing their applicability on domain-specific grounds. Similarly, rules with high confidence and high impact can be safely inferred, but may not be executed automatically because of their domain-specific implications.

TABLE 1: **Relations between association rules' confidence and impact, and its risk's probability and severity.**

		Risk Probability		
		High	Low	
Rule Impact	High	-	?	High
	Low	?	+	Low
		Low	High	
		Rule Confidence		

The concept of risk has been applied to the domain of recommender systems to a very limited degree. Ali, Manganaris and Srikant proposed a formula which tested the predictive value of discovered association rules, and estimated their 'relative risk'. Their goal was "to eliminate patterns that were not predictive even though they were frequent" [6].

In a more user-centered approach, Bouneffouf proposed the concept of risk in context-aware recommendation systems, using a dynamic context-based risk level to ensure recommendations are as non-intrusive as possible. His prime goal regarding risk management is "to avoid disturbing the user in risky situations" [30].

Neither one of these papers utilizes risk as a mediating factor to safeguard the application of association rules, as is the focus of this paper.

OBJECTIVE AND APPROACH

The potential impact discovered association rules have on datasets in precarious domains impedes their implementation. In this study, we propose the use of the concept of risk as a factor to determine the safe practicability of association rules. In order to do this, a risk model for association rules is introduced. This model is then implemented in the precarious domain of medication review in primary care and validated to determine to what extent the concept of risk has predictive power. Finally, potential applications of common scenarios in recommender systems in which the risk model could be useful are discussed.

Domain: Medication Review in Primary Care

Growing numbers of older people have more than one chronic disease or medical condition, treated with a multitude of drugs [110]. This chronic use of five or more medicinal drugs, or polypharmacy, has been shown to have detrimental effects on patients' health. Both underprescribing – the non-treatment of diseases – and overtreatment – the prescription of unnecessary medicine – have been identified as causes of adverse health events associated with polypharmacy [122].

Being the principal point of continuing care for patients in most Western-European health care systems, primary care is regarded as the appropriate place to treat polypharmacy. Medication reviews have been shown to be effective means to optimize polypharmacy [143,110]. In this study, medication review in primary care serves as the precarious domain in which the proposed model is implemented and validated.

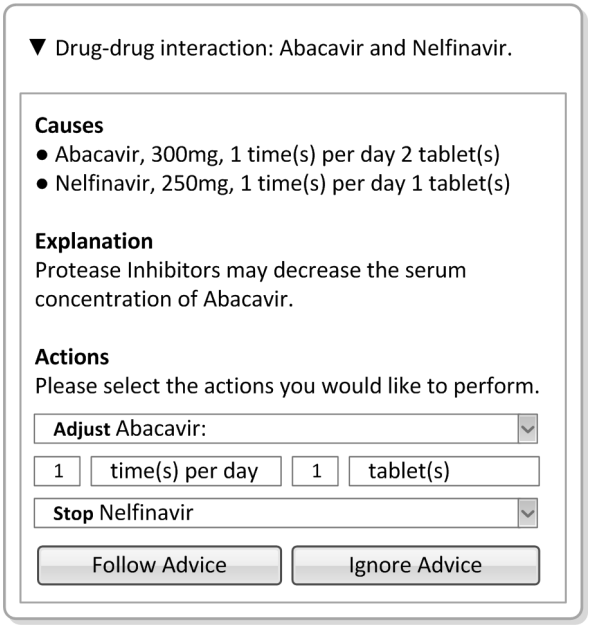
Recommender System: STRIP Assistant

The recommender system which will be used to validate the risk model introduced below is the STRIP Assistant (STRIPA). STRIPA is a web application that aids general practitioners and pharmacists with prescribing medication for vulnerable elderly patients through medication reviews. STRIPA's hybrid rule base consists of (inter)national guidelines and inference rules acquired through association rule mining [144].

When utilizing the system for patients, physicians see their complete health records: the diseases they suffer from, the drugs they are treated with, possible contra-indications and allergies that impede treatment, and relevant measurements and characteristics (such as weight, age, or blood pressure). Based on these factors, the recommender system suggests adding new drugs, discontinuing existing ones, or adjusting their dosages. The system is dynamically context-aware, taking users' actions during the process into account.

figure shows what a recommendation in STRIPA looks like: users are adviced to lower the dosage of one of two drugs causing a clinically relevant interaction, and to discontinue the other one.

FIGURE 1: **Screenshot showing a recommendation caused by a clinically relevant drug-drug interaction.**



RISK MODEL FORMULATION

Recommender systems, whether powered by explicit knowledge bases or implicit content-based or collaborative-based filtering, depend on inference rules. Inference rules are logical functions which analyze premises and, based on their syntax, return conclusions. Recommender systems depend on datasets containing all relevant items to trigger their inference rules. In propositional logic inference rules can be written as $x \rightarrow y$, with a dataset $D=\{ d_1,...,d_n \}$ and $x \in D$. Thus, for a specific rule dataset D contains its premises, along with other items, but never its consequence.

The risk associated with a rule is a function of its unwanted consequences and their likelihood of occurring [107]. The formula to determine the risk of an inference rule $x \rightarrow y$ reads:

EQUATION 5:

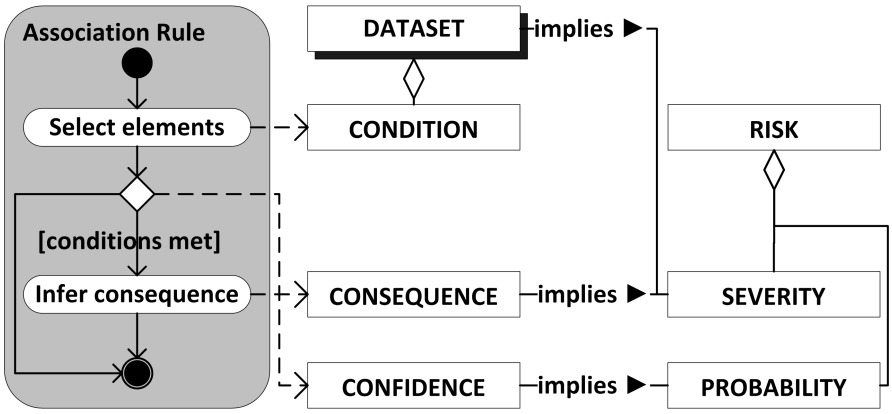
$$\text{risk}(x \rightarrow y) = (1 - \text{probability}(x \rightarrow y)) * \sum_{i = D,\, y} \text{severity}(i)$$

The probability of the consequences being unacceptable is one minus the confidence with which the rule is accepted. In the case of association rules that have been discovered using Agrawal's algorithm or a similar method, the probability in the formula is equal to its confidence.

The severity of a rule's risk is the sum of the impact of the objects associated with it. This comprises not just the danger associated with the inference's consequence y, but also detrimental characteristics of items in dataset D; it may be riskier to perform a certain action on a vulnerable dataset than on a safer one. The danger associated with these items can be estimated through risk formulas as well. As such, when implemented in a domain, the final formula results in the sum of its associated objects' risks, multiplied with its rule's inverse probability.

Figure 2 below graphically depicts an association rule and its subsequent risk, showing how an association rule's confidence, conditions, and consequences determine the probability and severity of its risk.

FIGURE 2: **An association rule's confidence, conditions, and consequences determine risk's probability and severity.**



IMPLEMENTATION

The above-mentioned formulae can be demonstrated and validated by applying them to the STRIP Assistant described earlier. The domain of medication review in primary care, in which the recommender system is used, comes with its characteristics and risk factors. In the following sections, these risk factors are introduced and incorporated into the proposed model.

Health Records' Risk

In the case STRIPA, the dataset consists of a certain patient's health record. The items in this dataset, such as diseases or drugs, serve as premises for the system's inference rules. All inference rules modify one or more drugs by prescribing new medicines, removing existing ones, or adjusting their dosages. A dataset D comprising a certain patient's diseases, drugs, contra-indications, allergies, and measurements can be described as a set:

EQUATION 6:

$$D = \{ \begin{array}{c} \text{disease}_1, \ldots, \text{disease}_k \\ \text{drug}_1, \ldots, \text{drug}_l \\ \text{contraindication}_1, \ldots, \text{contraindication}_m \\ \text{measurement}_1, \ldots, \text{measurement}_n \\ \text{allergy}_1, \ldots, \text{allergy}_p \end{array} \}$$

Following $x \in D$ and the fact that all inference rules adjust drugs, the implemented risk formula reads:

EQUATION 7:

$$\text{risk}(x \rightarrow \text{drug}) = (1 - \text{probability}(x \rightarrow \text{drug})) * (\text{severity}(D) + \text{severity}(\text{drug}))$$

The severity of patients' health records (dataset D) and drugs are explained below.

Health Records' Severity

Determining the risk associated with a patient's health record, or dataset, involves taking into account a multitude of domain-dependent variables, such as his or her age, frailty, physical properties, and cognitive state [62]. Drugs may be too dangerous to be prescribed in high dosages to patients older or younger than a certain age. Patients with cognitive problems may be best served by as little change to their drugs regimen as possible. This results in the severity of a dataset D being the sum of its domain-relevant, patient-specific, risk factors:

EQUATION 8:

$$\text{severity}(D) = \sum_{\text{riskFactor} \in D} \text{riskFactor}$$

To determine risk factors for polypharmacy patients, the Dutch multidisciplinary guideline for polypharmacy in elderly people was consulted [62]. This guideline proposes seven factors that increase patients' risk of harm due to inappropriate drug use: age (over 65 years old), polypharmacy (using four or more drugs simultaneously), impaired renal function, impaired cognition, frequent falling, decreased therapy adherence, and living in a nursing home. In this study's dataset, four of these risk factors were available: age, polypharmacy, impaired renal function, and impaired cognition.

Drug Severity

As will be elaborated upon in the subsequent sections, a drug's severity can be expressed as a function of its dose, or toxicity, and its adverse effects, or harm [64]:

EQUATION 9:

$$\text{severity}(\text{drug}) = \text{toxicity}(\text{drug}) * \text{harm}(\text{drug})$$

Adverse Effects Medicines have always had risks associated with them in the form of adverse effects. An adverse effect, according to Edwards and Aronson, is "a response to a drug that is noxious and unintended and occurs at doses normally used in man" [64]. These effects are usually classified in terms of their likelihood of occurring. Their frequency and potential impact are used to determine drugs' safety for prescribing. As such, the function of the number of adverse effects a drug has and their frequency can be useful to determine a substance's potential harm:

EQUATION 10:

$$\text{harm}(\text{drug}) = \sum_{e \in E} \text{e.frequency}$$

The sets of adverse effects, defined per active substance, were retrieved from a database maintained by the Royal Dutch Pharmacists Association [120]. Adverse effects are classified due to their likelihood of occurring: often (over 30% of patients or more), sometimes (10%-30% of patients), rarely (1%-10% of patients), and very rarely (less than 1% of patients). As such, each drug's adverse effects can be described as a set E={ e_1,...,e_n }, where e_i=(id,frequency). In the case of the relatively safe painkiller paracetamol, the set E contains a single item: e_1=(id="hypersensitivity",frequency=0.01).

Toxicity The definition provided for adverse effects also takes into account the prescribed dosage, something not accounted for in the formula above. Research has shown that the probability of adverse effects generally increases with higher dosages being used [64,151]. For each drug, the World Health Organization has defined an average strength with which it is typically prescribed. This Defined Daily Dose (DDD) is "the assumed average maintenance dose per day for a drug used for its main indication in adults" [233]. Dividing a patient's actual daily dosage of a drug by the DDD provides a factor that can be used to relativize the drug's risk. Its toxicity can thus be calculated as such:

EQUATION 11:

$$\text{toxicity}(\text{drug}) = \frac{\text{prescribedDailyDose}(\text{drug})}{\text{definedDailyDose}(\text{drug})}$$

The defined daily doses of each active substance were retrieved from the WHO [233]. Paracetamol, for example, has a defined daily dose of three grams, which in its most common form constitutes six tablets of 500 milligrams each.

Summary

Combining formulas 4 and 5 leads to a value representing a domain-specific association rule's severity. Incorporating its probability as shown in formula 3 results in a risk assessment formula that can be used for any dataset compatible with formula 2.

VALIDATION

The risk model introduced in formula 1 and implemented in formula 3 can be validated by comparing its predictions with actual actions taken by experts. In this study, the STRIP Assistant was used on real patient cases by dedicated teams of general practitioners and pharmacists for the duration of a year. A randomized controlled trial was performed in 25 general practices located in Amsterdam, the Netherlands, including 500 patients [227]. For the 261 of these patients that were placed in the intervention arm, four teams consisting of one GP and one pharmacist each used the software to optimize their medical records. During their use of the software, the user would respond to patient-specific advice, recommending them to prescribe new drugs for particular diseases. Their responses to advices were gathered; each time a suggestion was heeded or ignored, the instance, along with relevant patient case information, was logged. A total of 776 responses to advices, of which 311 were heeded, has been gathered and will be used to validate the risk model.

During the period in which the data was gathered, the risk model was in no way implemented in the software application. Advices were not adjusted based on their probable harm to patients, nor were the actual risk outcomes of items shown to users. Assuming that users will strive for the minimization of risk, the authors hypothesize that users will have chosen the least risky option whenever possible.

H1. The higher the risk of an action proposed by a recommendation, the least likely an expert is to perform it.

Based on the assumption that riskier patients – i.e. patients who have multiple risk factors in their dataset (or health record) – are best served with as little change to their drug regimen as possible, it was hypothesized that the higher an action's risk was, the least likely it was to be performed by users. To test this, the risk factors of each generated recommendation were calculated; its proposed drug's risk and the relevant patient's risk factors were summed according to the introduced model. An independent t-test affirmed the hypothesis, showing a statistical difference in the risk associated with proposed actions which were followed ($M = 2.42$, $SD = 0.57$) and the risk of proposed actions which were not followed ($M = 2.57$, $SD = 0.60$); $t(623) = 3.040$, $p = .002$.

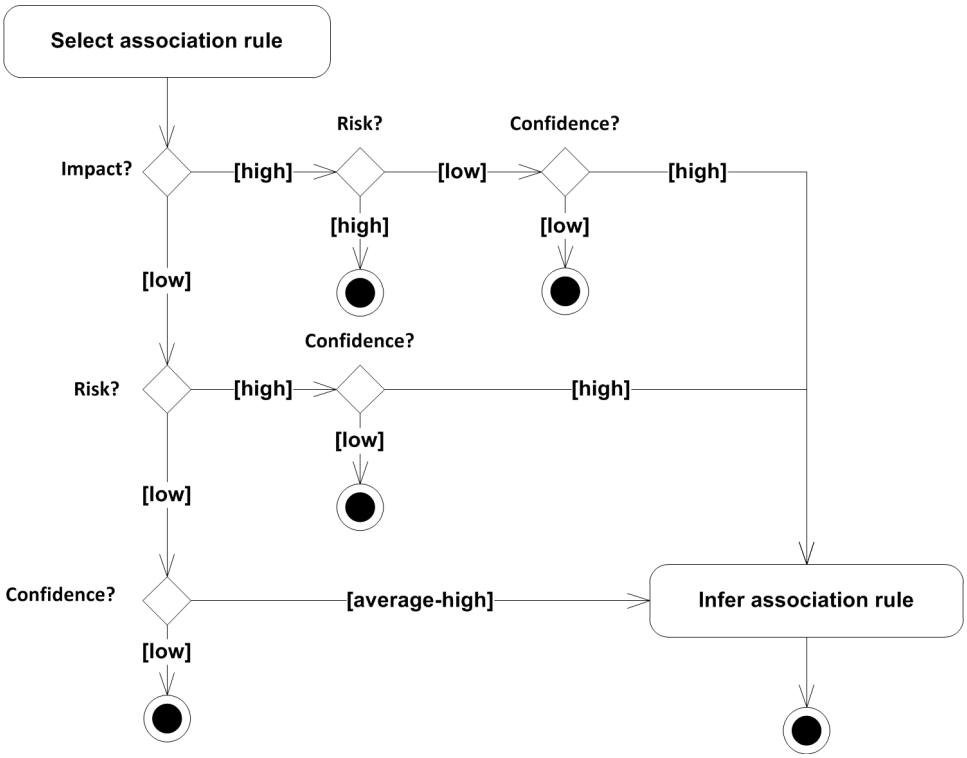
This result confirms that in the domain of medication review in primary care, the risk model has predictive value when prescribing drugs.

APPLICATION

With the risk model validated in the domain of medication review in primary care, its application can be discussed. As shown in Table 1, rules with high impact and high confidence, and those with low impact and low confidence, can be considered for risk assessment. These two scenarios are illustrated with implemented examples from STRIPA. In figure it is clearly

illustrated how generic decisions, taken with domain-dependent values, can be followed to determine whether or not an association rule can be safely inferred.

FIGURE 3: **Activity diagram illustrating the generic conditions with domain-dependent values that determine whether or not an association rule can be safely inferred.**



High Impact and High Confidence Rules

Among the most prominent reasons the authors considered risk a promising concept to use in conjunction with association rules, are their limitations in precarious domains. Implementing association rules by automatically following up on their consequences may lead to dangerous situations. In the case of STRIPA, association rules were discovered with potentially far-reaching implications, where risk assessment was needed in determining their implementation and presentation.

Association rules in this domain can be formulated as follows:

EQUATION 12:

$$\text{cause}_1 \cup \ldots \text{cause}_n \xrightarrow{\text{confidence}} \text{prescribe}(\text{drug})$$

An advice suggesting ace inhibitors for patients suffering from heart failure was followed up on in 61% of the cases:

EQUATION 13:

$$\text{heartFailure} \xrightarrow{0.61} \text{prescribe}(\text{aceInhibitor})$$

Similarly, a recommendation for a vitamin D-supplement for elderly patients with osteoporosis was heeded in 59% of the cases:

EQUATION 14:

$$\text{age} > 65 \cup \text{osteoporosis} \xrightarrow{0.59} \text{prescribe}(\text{vitaminD})$$

These association rules could be presented to the user in a variety of ways for patients matching the criteria. In the recommender system's user interface they could take the form of open suggestions, modal dialogs prompting a response, or even undoable prescriptions. Combining the rules' confidence with risk assessment aids in determining an appropriate implementation option.

Vitamin D-supplements do not have any known side effects associated with them, whereas for ace inhibitors a number of rare and very rare side effects have been reported, leading to a risk outcome for the latter drug of (5*0.1)+(10*0.01)=0.6. When prescribing the default dose, the toxicity factor can be discarded.

When incorporating these values, vitamin D-supplements are safer to prescribe in a more automated fashion than ace inhibitors. The complete risk model can be used to assess specific patients' risk and the safety of either association rule. For a patient having two risk factors in his or her health record, the risk assessment for the vitamin-D rule would yield: (1-0.59)*(2+0)=0.82, while the ace inhibitor rule would result in: (1-0.61)*(2+0.6)=1.01. Using a threshold of 1.0 in STRIPA, for this patient the ace inhibitor-advice was presented as an open suggestion, whereas the vitamin D-rule was shown as a modal dialog, forcing the user to respond.

Low Impact and Low Confidence Rules

While the previous scenario showed how risk can be used to determine the implementation of rules with high confidence and potentially high impact, it can also be used for rules with opposite characteristics. Rules with little impact and a low confidence level can be employed after passing a risk assessment. In the case of STRIPA, this approach was used in the assignment of drugs to diseases. One of the tasks users have to perform is assigning drugs to diseases for which they have been indicated; paracetamol, for example, will often be prescribed for the treatment of pain. This assignment task could be made more efficient by automatically suggesting frequent drug-disease combinations. Association rules were discovered which revealed which drugs and diseases were often combined:

EQUATION 15:

$$\text{disease} \xrightarrow{\text{confidence}} \text{assign}(\text{drug})$$

The confidence with which this assignment could be inferred varied. The sleeping drug temazepam, for example, was prescribed for anxiety in 66% of the cases:

EQUATION 16:

$$\text{disease} \xrightarrow{0.66} \text{assign}(\text{temazepam})$$

For the same disorder, the antidepressant fluoxetin was prescribed in 66% of cases as well:

EQUATION 17:

$$\text{disease} \xrightarrow{0.66} \text{assign}(\text{fluoxetin})$$

Using only this data to automatically assign these drugs to anxiety is error-prone; in a third of the cases, these drugs were prescribed for other diseases instead. Disregarding the data completely, however, hinders a potentially useful functionality.

This problem, too, could be solved using risk as a mediating factor. Only cases in which the possible implications were low were automatically performed. Temazepam and fluoxetin have risk factors of respectively (7*0.1)+(5*0.01)=0.75 and (3*0.3)+(12*0.01)=1.02; incorporating the association rules' probabilities yields risks of 0.25 for the sleeping pill and 0.34 for the antidepressant. Using 0.3 as a tradeoff point in the application, the safer drug of the two (temazepam) was automatically assigned to anxiety, while the riskier one (fluoxetin) was not. Note that in this scenario, patients' risk factors were disregarded, as all drugs being assigned had already been prescribed to them.

CONCLUSION

In this study, the authors explored the potential usefulness of the concept of risk in association rule mining. Based on risk management literature, a risk model was proposed. A sample implementation in the precarious domain of medication review in primary care illustrated its use. Validation of this implementation showed that the risk model indeed has predictive power in this domain. It was shown how in common scenarios, in which the application of discovered association rules is problematic, the introduction of the risk model can be useful.

Further research should focus on applying the risk model in different domains. Implementing and validating it in a variety of contexts can determine its generic applicability.

ACKNOWLEDGEMENT

The authors would like to thank Pieter Meulendijk for his contributions to the conceptual model following his expertise in risk management.

PART VI

General Discussion

Above all, don't lie to yourself. The man who lies to himself and listens to his own lie comes to a point that he cannot distinguish the truth within him.

Fyodor Dostoevesky, *The Brothers Karamazov*

6.1

Conclusion

In the introduction, the motivation for conducting the research in this thesis was drawn from the interrelationships of three separate domains: decision support systems, medication reviews, and primary care. The aim of this dissertation was to create a decision support system to facilitate the conduct of structured medication reviews by physicians and pharmacists in primary care. This aim was formalized in the following main research question:

> MRQ: How can clinical decision support systems contribute to effective and efficient medication reviews in primary care?

In order to answer this question, the research was operationalized through a design science perspective, adopting the design science research framework introduced by Peffers et al [168]. The main research question was divided into four sub questions closely following the steps of this framework; they explored the solution's objectives, its design and development, its demonstration and evaluation, and the new objectives resulting from its evaluation. Each of these sub questions was divided into smaller questions which were answered in chapters. The first section below addressed these sub questions, thereby reiterating the conclusions of the chapters. The sections after, on implications and limitations, synthesize the studies' findings and interpret their relevance for both science and practice.

SUMMARY OF FINDINGS

Objectives

RQ1: What objectives can be identified for clinical decision support systems facilitating medication reviews in primary care from the perspectives of its stakeholders?

Following the design science research framework, the first step pursued was the identification of objectives for the solution. The artefact's primary functionality, i.e. incorporating the STRIP method and subsequent clinical rules into a web-based decision support system, were known. Additional requirements were elicited by exploring (potential) stakeholders' attitudes towards the artefact or its domain: users (physicians), beneficiaries (polypharmacy patients), and academic researchers were consulted to determine their attitudes towards and requirements for the proposed artefact.

First, the attitudes potential users had towards the introduction of a decision support system aiding them with conducting medication reviews were investigated. A survey was distributed amongst 500 Dutch GPs, exploring their experiences with polypharmacy, their attitudes towards current decision support systems, and their opinions towards a possible new one. Virtually all 184 respondents indicated owning a clinical information system, while 21% indicated owning a decision support plug-in. GPs' attitudes towards the newly proposed system were mainly positive (57%); they are likely to adopt the proposed system under the conditions that it improves prescription quality and does not require extensive investments of time or money.

Second, the non-functional requirements polypharmacy patients deemed most important when using medical apps was explored. A grounded theory approach was used to elicit a set of requirements through document study and interviews of GPs, pharmacists, developers, and potential users: accessibility, certifiability, portability, privacy, safety, security, stability, trustability, and usability. Six of these were evaluated with two groups (differing by age) of polypharmacy patients through a vignette study. Trustability, security, and, for the younger group, certifiability, were considered the most important non-functional requirements for medical apps. Furthermore, and contrary to consensus in literature, privacy was considered one of the least important attributes for medical apps by both groups.

Third, the feasibility of introducing a decision support system in primary care facilitating medication reviews was assessed. Through action research, the conceptual, organizational, technological, economic, and societal barriers to widespread implementation of the proposed solution were identified. For each of these hurdles, one or more strategies to overcome them were proposed. Using the running example of a decision support system in primary care, the implementation of each of these strategies was evaluated.

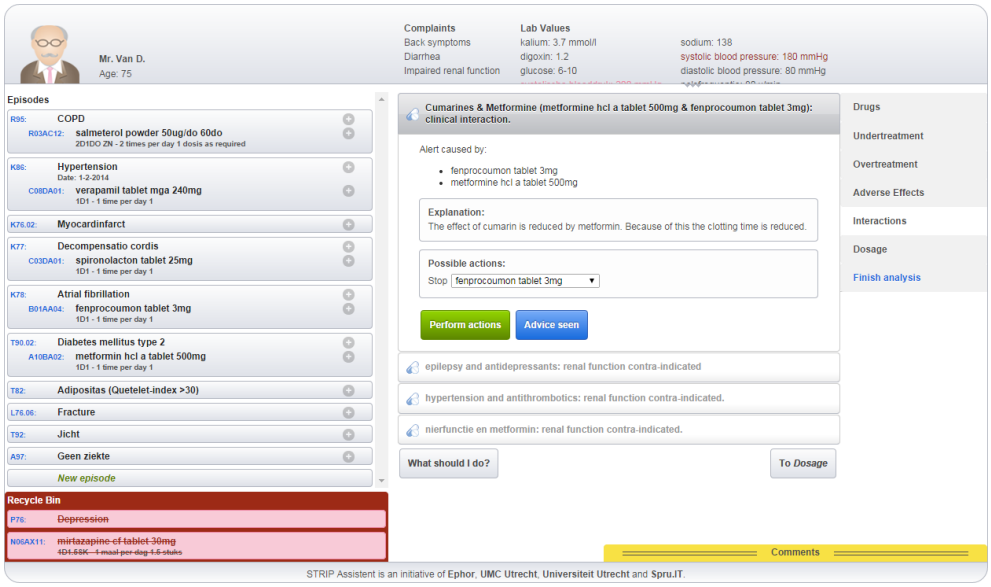
Design & Development

RQ2: What design decisions lead to the development of clinical decision support systems that effectively and efficiently facilitate medication reviews in primary care?

With objectives having been identified, the next phase in the design science research process encompasses the design and development of the artefact. In this project, the artefact is a stand-alone web-based decision support system that advices physicians during the pharmacotherapeutic analysis of patients' health records: the STRIP Assistant (STRIPA). The design decisions that lead to the successful implementation of this artefact have been based on best practices and consensus in literature. The application's development has been conducted through a process of iterative prototyping, involving stakeholders in every step. This leads to thoroughly considered decisions regarding the application's architecture and rule engine, and the design decisions relating to the user interface and semantic interoperability. A screenshot of the prototype is shown in Figure 1.

The STRIP Assistant has been developed as a stand-alone web service, relying on Java and MySQL in the back-end and on JavaScript in the front-end. The communication between the front- and back-ends is facilitated through AJAX, using JSON as data format for its brevity. The expert system in the back-end is powered by the Drools rule engine. A MySQL database holds all patient records and clinical data required to execute the decision rules. A database management system provides the query capabilities necessary to use it.

FIGURE 1: **Screenshot of the STRIP Assistant generating a recommendation to stop fenprocoumon. Heeding this advice would result in fenprocoumon tablet 3mg in the left overview being removed to the Recycle Bin.**



The rules included in STRIPA's decision support system come from a variety of sources, leading to the application incorporating rules with a varying degree of complexity. They could be modelled using three distinguishable types of rules: those based on atomic formulae, conjunctive compound formulae, and disjunctive compound formulae. Atomic formulae depend on a single condition only, containing no deeper propositional structure. Compound formulae, in contrast, do contain logical connectives to incorporate multiple conditions. While conjunctive operators require several conditions to be satisfied before a consequence is implied, disjunctive operators require only one of several conditions to be met. All rules result in advices that can either be heeded or rejected by users. Advices can propose to start new medication, stop specific drugs, or change the dosage or frequency of a drug already used.

Taking into account the literature on user interface mistakes in computerized physician order entry (CPOE) systems, much attention was paid to designing the user interface for STRIPA. The aim was to create an application that had the potential to be accessed through different mediums, i.e. both PCs and mobile devices. After designing an initial wireframe version based on interviews with medical experts and potential users, a prototyping process using prototypes of increasing fidelity was used to refine it. Early prototypes of the application were used in test sessions, where users were invited to comment on its usefulness and user-friendliness. Their remarks were used to further improve the user interface.

Essential to the usefulness and time efficiency of any application relying on external data, is its ability to communicate with third-party applications. In the primary care sector sharing data is notoriously difficult because of its sensitive nature and multitude of (often incompatible) information systems. In STRIPA an old but proven Dutch health exchange format known as 'Medische Overdracht', or Medical Exchange (MEDOVD) has been implemented. MEDOVD is a de facto standard based on the EDIFACT format that can be locally exported and imported by physicians, and which is supported by all Dutch CPOE systems.

Demonstration & Validation

RQ3: To what extent can clinical decision support systems contribute to effective and efficient medication reviews in primary care?

After envisioning and developing the artefact, the design science research process emphasizes demonstrating and evaluating it in practice. The decision support system created in this project has been evaluated in both a controlled experiment and in its real-life environment.

First, a controlled experiment was conducted to validate STRIPA's usability as a tool for physicians optimizing medical records of polypharmacy patients. In an online experiment, forty-two caretakers were asked to optimize two comparable medical records of polypharmacy patients, one in their usual manner and one using STRIPA. Changes in effectiveness were measured by comparing respondents' optimized medicine prescriptions with medication prepared by an expert panel of two geriatrician-pharmacologists. Efficiency was

operationalized by recording the time respondents took to optimize the two cases. Satisfaction was measured with the System Usability Scale. Medication optimization significantly improved with the STRIP Assistant. Appropriate decisions increased from 58% without help to 76% with the STRIP Assistant; p<0.0001. Inappropriate decisions decreased from 42% in the unassisted case to 24% in the assisted one; p<0.0001. Participants spent significantly more time optimizing medication with the STRIP Assistant (24 minutes) than without it (13 minutes); p<0.0001. The respondents filled out a below-average SUS-score of 63.25.

The experiment's results showed improvements in effectiveness when performing medication reviews with the STRIPA decision support system, but decreases in efficiency. Hypothesizing that this lack of efficiency was due to the study not taking into account experience participants gain over time, a longitudinal study was set up. Four expert teams consisting of a physician and a pharmacist conducted structured medication reviews on patients in 13 general practices located in Amsterdam, the Netherlands. During thirteen months, the expert teams performed 261 medication reviews. Analysis of the acquired data showed that the amount of time users needed to perform medication reviews decreased significantly as they gained experience over time; during the first half of the study they needed 15.7 minutes on average to conduct one review, while during the second half the mean time had decreased to 10.7 minutes; p<0.0001.

Refined Objectives

RQ4: What objectives can be identified for the integration and optimization of clinical decision support systems facilitating medication reviews in primary care?

The evaluation of the artefact led to insights into its weaknesses and presented opportunities for refinement. Thus, refinements of the solution's objectives and its development decisions could be explored, starting a new iteration of the design science research process. Two areas of improvement were selected for additional research and refinement of the artefact: optimization of advices through association rule mining, and enhanced semantic interoperability.

The diversity of terminologies used in primary care causes significant challenges regarding semantic interoperability. Attempts to address these challenges usually focus on the creation of metaterminologies, with the peculiarities of national variations of terminologies being overlooked. By focusing on national implementations, the extent to which primary care data can be meaningfully exchanged between nationally implemented terminologies was assessed. A model comprising primary care terminologies and including axioms to define their relations was developed. Assessments of data quality and completeness showed that values of terminologies which are closely related can express each other's concepts relatively well. While less extensive terminologies' concepts often have equivalents in larger classification systems, concepts of more comprehensive terminologies can only be expressed in simpler terminologies to a very limited degree.

Association rule mining is one of the most prominent knowledge discovery methods in existence. The application of discovered association rules in a sensitive domain with potentially far-reaching implications, however, can be risky. Thus, a model for the incorporation of risk in association rules was proposed and implemented in a primary care setting. Validation of this implementation showed that the risk model indeed has predictive power in parts of this domain. It was shown how in common scenarios, in which the application of discovered association rules is problematic, the introduction of the risk model can be useful.

CONTRIBUTIONS & IMPLICATIONS

Scientific Contributions

As has been explicated upon in the introduction of this dissertation, the relationship between theory and practice in design science is complicated. In design science, theory and practice are generally understood to form a symbiont relationship, where research should be evaluated in light of its practical implications [95]. The degree to which theorizing is a part of design science is disputed, but there is consensus that evaluated artefacts do retain knowledge in the form of models or instantiations [134]. This tacit knowledge contained in artefacts can be made available to the scientific community through utility theories, which assert that a particular type of technology has (some kind of) utility in improving a problematic situation [215].

The following scientific contributions presented through the artefact created in this dissertation, a decision support system for medication reviews in primary care, can be distinguished:

1. Demonstrating how a decision support system for medication reviews in primary care can be developed;
2. Evaluating its utility in terms of effectiveness, efficiency, and satisfaction;
3. Demonstrating how mixed methods research incorporating positivist and interpretivist methods can yield complementary results.

Decision Support System Development in Primary Care As shown in the introduction, the artefact that is created and evaluated in this dissertation is situated in the intersection between primary care, medication reviews, and clinical decision support systems. The chapters answering the first research question investigate the objectives of the envisioned system, while the chapter on the second research question documents the design decisions leading to its implementation. Together they demonstrate how a process of requirements elicitation, feasibility analysis, and prototyping involving relevant stakeholders can lead to a system whose utility can be proven.

When conducting the studies after the intended solution's objectives, we started from the assumption that involving (potential) stakeholders in the development process leads to a broadly shared set of objectives. When surveying GPs about their attitudes towards the proposed application, our results were largely in line with consensus in literature, confirming

that they especially regard output quality, time investment, and financial stimuli as influential factors [39,238]. In our study after non-functional requirements of medical apps, the discovered requirements mostly matched those in industry standards. In line with the health domain's precariousness, identified aspects such as privacy, trustability, and, following from those, certifiability, were uncommon in requirements engineering literature [84,108,106].

Utility Evaluation The artefact's utility theory regards polypharmacy as its problem space, medication reviews as its solution space, and the envisioned decision support system as the linking artefact, the utility of which is expressed in terms of its usability, i.e. its effectiveness, efficiency, and satisfaction. The demonstration and validation chapters included in this dissertation prove the benefits in utility the solution has compared to normal approaches. In the experiment described in chapter 6, the benefits in effectiveness the application has over unassisted approaches are unequivocally demonstrated. The longitudinal study described in chapter 7 subsequently shows how users' proficiency with the application increases as they gain experience over time, showing efficiency improvement over regular methods.

The two studies validating the solution's utility in terms of effectiveness, efficiency, and satisfaction, produced results generally in line with consensus in literature. The decision support system tested was able make general practitioners and pharmacists perform better medication reviews than using no system at all, similar to other studies' findings [32]. Even though more appropriate decisions were made with the software application, a non-negligible number of mistakes respondents made can be attributed to software suggestions. This result highlights an area of decision support systems that has so far got little attention [14,66,121]. Regarding efficiency, our study shows that longitudinal research may produce different outcomes than experimental methods. Findings similar to the efficiency improvement observed in our study are scarce, as there is no consensus on decision support systems' effects on efficiency [32]. Limited research has been done into the effects of temporality on usability in general, and efficiency in particular [139,113].

Mixed Methods Application The incorporation of contextual factors in evaluating a system's utility is one of information science's core principles. As a result, interpretivist methods that observe systems in their environments, such as case studies, have long been part of the scientific toolset. These embedded approaches, however, make determining and controlling independent variables, and ultimately reproducibility, difficult. Traditional positivist experiments, on the other hand, enable isolation of variables but cannot simulate real life use. The experiment and the longitudinal case study in this dissertation illustrate this dichotomy, but also demonstrate how the two approaches can yield complementary meaningful results. While the approach of triangulation, or the use of multiple methods to test the same hypothesis, is often used in information science research, our results illustrate the consecutive use of multiple methods building on top of each other to explore a phenomenon [228]. Thus, our study demonstrates how the framework proposed by Edmondson and McManus on methodological fit can be used for information systems research; mature fields benefit from testing formal

hypotheses resulting in evaluative quantitative results, while intermediate fields gain meaningful results from a mix of exploratory and evaluative methods [63].

Societal Implications

Polypharmacy amongst elderly people is a growing problem leading to major challenges in public health [111]. A multitude of initiatives has been developed to assess the appropriateness of drugs prescribed for individual patients. Explicit methods, consisting of lists of clinical interactions or contraindications, include the Beers criteria and the START and STOPP criteria. Implicit methods use patient-specific information, combined with medical knowledge, to determine medication appropriateness, and include the Medication Appropriateness Index and the pharmacotherapy review focused on drugs' use, indication, safety, and effectiveness (GIVE). The effectiveness of these interventions varies; generally they appear beneficial in terms of reducing inappropriate prescribing and medication-related problems, but they have not been proven to lead to clinically significant improvement [167]. All-encompassing approaches to incorporate both implicit and explicit methods have resulted in the POM and the STRIP. In a 2009 study a paper version of the POM was validated, showing that it significantly improved GPs' prescriptions. They also prescribed +/- 10 percent fewer medicines per patient [57].

A variety of barriers is impeding the widespread adoption of structured medication reviews in daily practice. Most factors revolve around physicians, and include inertia (his or her attitudes towards discontinuation, such as fearing negative consequences), self-efficacy (his or her knowledge and available information on the topic), and awareness (his or her having poor insight or discrepant beliefs). Non-physician-related factors include a lack of resources, patients resisting changes to their medication, and practical and cultural factors [10]. A separate study focusing on barriers regarding pharmacist-led medication reviews reported lack of time and lack of self-confidence as the most commonly perceived barriers [157].

The results from chapters 6 and 7, documenting the experiment and the longitudinal case study, show that a clinical decision support system facilitating GPs and pharmacists with conducting medication reviews leads to significant improvements in prescribing quality. Participants made more appropriate decisions and fewer inappropriate ones. After gaining experience with time, users' efficiency increased to conducting a medication review in ten minutes. Thus, the implemented solution has the ability to remove some of the barriers mentioned above, including a lack of resources and a lack of self-confidence. This enables the widespread adoption of medication reviews in primary care, and subsequently improvements in appropriate prescribing.

These improvements in appropriate prescribing can be expected to lead to a decrease in medication-related hospitalizations [129]. The increased risks of (co)morbidities and mortality that have been associated with polypharmacy can be expected to decrease as well [88].

LIMITATIONS & FURTHER RESEARCH

The artefact arrived at in this dissertation is the result of several phases of research and development, following the design science research framework. The methods used to interview, survey, test, and otherwise involve stakeholders in the formulation of objectives, the development of the solution, and the evaluation of the artefact were all conducted rigorously, taking into account the methods' particular validity constraints. The selection of respondents and participants to partake in the studies, however, could not be completely controlled by the researchers. In many methods, participants volunteered to partake or were otherwise selected through opportunity sampling. Because of this, it cannot be excluded that participants were somehow – either positively or negatively – biased towards the artefact or its problem domain. This hinders the studies' generalizability, and further research is necessary to confirm their results.

A randomized controlled trial, in which a control group would be studied next to the intervention group, and to which participants would be randomly allocated, would be an adequate method to confirm the results arrived at in this dissertation. Outcome measures could be expanded to include not only effectiveness, efficiency, and perceived satisfaction, but also, a.o., cost-effectiveness or clinical relevance. Different environments, such as hospitals or geriatric wards, and subsequently different medical specialists could be included to test the solution's utility in varying settings. The 'OPtimising ThERapy to prevent Avoidable hospital admissions in the Multimorbid elderly' (OPERAM) study aims at validating the benefit of structured medication reviews on patients' quality of life through a randomized controlled trial. For this purpose the STRIP Assistant is used and validated in hospital settings in four Western-European countries.

Two studies included in this dissertation were included after the intended artefact had been validated. Following insights gained from the evaluation, two further obstacles to successful implementation of the proposed solution in practice were identified: recording data in such a way that it is semantically interoperable with other systems, and the enrichment of systems with discovered association rules in sensitive domains.

As many other software developers in the health care domain have experienced, we too were challenged with the requirement of exchanging data with third party systems. The need for a more standardized way to record data in health care has been addressed by introducing a large number of classification systems, including taxonomies to systematically classify diseases, drugs, and laboratory tests, among many others. The inability to meaningfully exchange large numbers of concepts between these terminologies creates significant semantic interoperability challenges. These are commonly addressed by advocating the use of an overhauling terminology, such as SNOMED-CT or the UMLS Metathesaurus. The implementation of these metaterminologies, however, raises questions larger than those on data quality and data completeness. The extent to which metaterminologies are conceptually compatible with systems based on different terminologies is one such question. Another addresses the way

in which terminologies limit their users' conceptual spaces, and subsequently make certain concepts inaccessible. Finally, metaterminologies do not do justice to the unique perspectives existing terminologies offer on diagnosing and prescribing. These problems cannot be simply addressed by 'upgrading' terminologies to metaterminologies, but need additional research.

Association rule mining has been one of the most prominent knowledge discovery methods for the last twenty years, and is currently serving as the backbone for many recommender systems. The application of discovered association rules in a sensitive domain with vulnerable datasets, where association rules can have potentially far-reaching implications, however, is risky. While automatically adding beer to a smoker's shopping basket may be technologically similar to prescribing metformin to a diabetic, its implications are not. And while much attention has been paid in the scientific community to the optimal discovery of association rules, literature on the risks and (mis)uses of their application are scarce. Additional research is necessary to safely enrich software systems in the health care domain with the results of knowledge discovery techniques.

REFLECTION

The interplay between the domains of information systems and medicine is a complex one. While it may be tempting for laymen in this field to regard the domain of medicine and its practitioners as conservative in their attitudes towards technology, such a simplification does not do them justice. The domain of medicine is infinitely older than that of information systems, and has reached a maturity far beyond that of most other fields. The long history and wide variety of IT implementations in medicine show how medical practice does not shun technological advancements.

In the Netherlands, information systems in primary care have existed since the 1980s. Having often been developed by pioneering general practitioners themselves, this bottom-up development has resulted in a wide variety of software programs, usually with a small but vocal user-base. While this bottom-up development has had many benefits, such as the widespread adoption of electronic prescribing on a national level, it also hinders standardization and top-down implementation of IT policies. As has been explored in chapter 8, Dutch primary care uses its own standards for the recording of drug dosages and laboratory tests, and an adapted version of the international ICPC-standard for the registration of diseases. The chapter's results show how the incompatibility of these terminologies with those of other countries, and similarly those of secondary care practitioners, forms a major obstacle in the exchange of health information.

On the other hand, the small-scale bottom-up approaches of many primary care systems allow for a flexibility that cannot be matched by large-scale top-down ones. In the last decade, a lively debate and accompanying technological proposals have been had over the introduction of a national patient record. After this proposal had been rejected by the highest political organ, an old bottom-up data structure was re-implemented in information systems to allow for a

basic form of communication of patient records between practitioners. As chapter 5 illustrates, the decision support system created in this study successfully uses this data format to exchange patient data with GPs' IT systems.

These examples illustrate the dichotomy that exists in the interplay of information systems and medicine: one the one hand flexible, but substandard, bottom-up approaches, on the other rigid, but theoretically substantiated, top-down ones. Successfully merging these two domains requires one to respect the strengths of each of these approaches, rather than favoring one over the other. As was reasoned in the discussion of chapter 8, the problem of semantic interoperability in primary care cannot be solved by replacing custom-tailored terminologies with generic ones. Instead, using standardized ontologies as intermediates may prove fruitful: mapping concepts of the Dutch ICPC1 adaptation to those of the international SNOMED-CT terminology ensures unambiguous exchange of concepts, while retaining its customized perspective. Similarly, exchanging locally stored patient data for tasks specified in advance rather than relying on centrally stored data or authorizing practitioners independently of their goals may satisfy both functional requirements and privacy concerns.

Bibliography

1. Aanensen, David, Huntley, Derek, Feil, Edward, & Spratt, Brian (2009). EpiCollect: linking smartphones to web applications for epidemiology, ecology and community data collection. *PloS one, 4* (9),
2. Actiz, FNT, GGZ, IGZ, KNMG, KNMP, LEVV, LHV, NFU, NHG, NMT, NICTIZ, NPCF, NVZ, NVZA, Orde, V&VN, VGN, Verenso, VWS, ZN (2008). Richtlijn Overdracht van Medicatiegegevens in de keten (Directive on the Transferal of Medication Information in the Chain, in Dutch).
3. Adlassnig, K. (2009). Automatic mapping of clinical documentation to SNOMED CT. *Medical Informatics in a United and Healthy Europe: Proceedings of MIE 2009, the XXIInd International Congress of the European Federation for Medical Informatics* 228.
4. Adomavicius, G, & Tuzhilin, A (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering, 17* (6), 734-749.
5. Agrawal, R, & Srikant, R (1994). Fast algorithms for mining association rules. *Proc 20th Int Conf Very Large Data Bases VLDB* 487-499.
6. Ali, K, Manganaris, S, & Srikant, R (1997). Partial Classification Using Association Rules. *Proc. 3th Int. Conf. on KDD* 115-118.
7. Ambler, Scott (2004). *The Object Primer: Agile Model-Driven Development with UML 2.0.* (3rd) , Cambridge: Cambridge University Press
8. Ambler, Scott (2005). *The Elements of UML(TM) 2.0 Style.* , Cambridge: Cambridge University Press
9. Ammenwerth, E, Schnell-Inderst, P, Machan, C, & Siebert, U (2008). The effect of electronic prescribing on medication errors and adverse drug events: a systematic review. *Journal of the American Medical Informatics Association, 15* (5), 585-600.
10. Anderson, K, Stowasser, D, Freeman, C, & Scott, I (2014). Prescriber barriers and enablers to minimising potentially inappropriate medications in adults: a systematic review and thematic synthesis. *BMJ open, 4* (12),
11. Angen, M.J (2000). Evaluating interpretive inquiry: Reviewing the validity debate and opening the dialogue. *Qualitative health research, 10* (3), 378-395.

12. Anton, C, Nightingale, P.G, Adu, D, Lipkin, G, & Ferner, R.E (2004). Improving prescribing using a rule based prescribing system. *Quality and Safety in Health Care, 13* (3), 186.
13. Apache Software Foundation (2014). Apache Shiro | Java Security Framework. Retrieved from http://shiro.apache.org/
14. Ash, J.S, Berg, M, & Coiera, E (2004). Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. *Journal of the American Medical Informatics Association, 11* (2), 104-116.
15. Bangor, Aaron, Kortum, Philip, & Miller, James (2008). An empirical evaluation of the system usability scale. *Int J Hum Comput Interact, 24* (6), 574-594.
16. Bangor, Aaron, Kortum, Philip, & Miller, James (2009). Determining what individual SUS scores mean: adding an adjective rating scale. *Journal of Usability Studies, 4* (3), 114-123.
17. Barter, Christine, & Renold, Emma (2000). 'I wanna tell you a story': exploring the application of vignettes in qualitative research with children and young people. *International Journal of Social Research Methodology, 3* (4), 307-323.
18. Baskerville, R.L (1999). Investigating information systems with action research. *Communications of the AIS, 2* (1), 4.
19. Batini, C, & Scannapieco, M (2006). *Data quality: concepts, methodologies and techniques.* Springer
20. Bellahsene, Z, Bonifati, A, & Rahm, E (2011). *Schema Matching and Mapping.* , Berlin: Springer Berlin Heidelberg
21. Benson, T (2010). *Principles of health interoperability HL7 and SNOMED.* , Londen: Springer
22. Bhattacherjee, A, & Hikmet, N (2007). Physicians' resistance toward healthcare information technology: a theoretical model and empirical test. *European Journal of Information Systems, 16* (6), 725–737.
23. Birks, D.F, Fernandez, W, Levina, N, & Nasirin, S (2013). Grounded theory method in information systems research: its nature, diversity and opportunities. *European Journal of Information Systems, 22* (1), 1-8.
24. Björkman, IK, Fastbom, J, Schmidt, IK, & Bernsten, CB (2002). Drug-drug interactions in the elderly. *Ann Pharmacother, 36* (11), 1675-81.
25. Blum, F.H (1955). Action Research--A Scientific Approach?. *Philosophy of Science, 22* (1), 1-7.
26. Bodenreider, O (2008). Issues in mapping LOINC laboratory tests to SNOMED CT. *AMIA Annual Symposium Proceedings*
27. Boonstra, A, Boddy, D, & Fischbacher, M (2004). The limited acceptance of an electronic prescription system by general practitioners: reasons and practical implications. *New technology, work and employment, 19* (2), 128-144.

28. Borneman, W.M, Bastiaanssen, E.H.C, & Westerhof, R.K.D (2014). Publieksversie HIS-Referentiemodel 2014 (Public version CPOE Reference Model, in Dutch). , Utrecht: Dutch College of General Practitioners
29. Boucharas, V, Jansen, S, & Brinkkemper, S (2009). Formalizing software ecosystem modeling. *Proceedings of the 1st international workshop on Open component ecosystems* , New York, 41-50.
30. Bouneffouf, Djallel (2013). DRARS, A Dynamic Risk-Aware Recommender System.
31. Brewer, M.B (2000). Research design and issues of validity. In Reis, Harry, & Judd, Charles, *Handbook of research methods in social and personality psychology* Cambridge University Press
32. Bright, T.J, Wong, A, Dhurjati, R, Bristow, E, Bastian, L, Coeytaux, R.R, & ... Lobach, D (2012). Effect of clinical decision-support systems: a systematic review. *Annals of internal medicine, 157* (1), 29-43.
33. Brooke, John (1996). SUS: A quick and dirty usability scale. In Jordan, Patrick, Thomas, B, McClelland, Ian, & Weerdmeester, Bernard, *Usability Evaluation In Industry* CRC Press
34. Bucklin, R.E, & Sismeiro, C (2009). Click here for Internet insight: Advances in clickstream data analysis in marketing. *Journal of Interactive Marketing, 23* (1), 35-48.
35. Buijink, A.W, Visser, Benjamin, & Marshall, Louise (2013). Medical apps for smartphones: lack of evidence undermines quality and safety. *Evidence-based medicine, 18* (3), 90-2.
36. Campanelli, C.M (2012). American Geriatrics Society updated Beers criteria for potentially inappropriate medication use in older adults: the American Geriatrics Society 2012 Beers criteria update expert panel. *J Am Geriatr Soc, 60* (4), 616-631.
37. Campbell, D.T, & Stanley, J.C (1966). *Experimental and quasi-experimental designs for research.* , Chicago: McNally
38. Checkland, P, & Holwell, S (2007). Action Research: Its Nature and Validity. In Checkland, P, & Holwell, S, *Information Systems Action Research*
39. Chismar, WG, & Wiley-Patton, S (2003). Does the Extended Technology Acceptance Model Apply to Physicians. *HICSS '03 Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03)* , Washington, 160.1.
40. Chung, L, & Do Prado Leite, J. (2009). On Non-Functional Requirements in Software Engineering. In Borgida, Alexander, Chaudhri, Vinay, Giorgini, Paolo, & Yu, Eric, *Conceptual modeling: Foundations and applications* , Berlin: Springer
41. Claxton, AJ, Cramer, J, & Pierce, C (2001). A systematic review of the association between dose regimens and medication adherence. *Clinical Therapeutics, 23* (8), 1296-310.
42. Corbin, J.M, & Strauss, A (1990). Grounded Theory Research: Procedures, Canons and Evaluative Criteria. *Qualitative sociology, 13* (1), 3-21.

43. Council of European Communities (1993). Council Directive 93/42/EEC concerning medical devices.
44. Creswell, John, & Plano Clark, Vicki (2011). Choosing a Mixed Methods Design. In , *Designing and Conducting Mixed Methods Research* (2nd) Sage Publications Inc.
45. Croft, William, & Cruse, D. (2004). *Cognitive Linguistics.* , Cambridge: Cambridge University Press
46. Croker, Joanne, Swancutt, Dawn, Roberts, Martin, Abel, Gary, Roland, Martin, & Campbell, John (2013). Factors affecting patients' trust and confidence in GPs: evidence from the English national GP patient survey. *BMJ Open, 3* (5),
47. Croll, Jasmine (2009). The Impact of Usability on Clinician Acceptance of a Health Information System.
48. Culén, Alma, & Bratteteig, Tone (2013). Touch-Screens and Elderly users: A Perfect Match?. *ACHI 2013, The Sixth International Conference on Advances in Computer-Human Interactions* , Nice, France, 460-465.
49. Dabney, Matthew, Dean, Brian, & Rogers, Tom (2013). No sensor left behind: enriching computing education with mobile devices. *Proceeding of the 44th ACM technical symposium on Computer science education* , Denver, 627-632.
50. Danish Quality Unit of General Practice (2008). General practice - DAK-E. Retrieved from http://www.dak-e.dk/flx/en/general-practice/icpc/
51. Davis, FD (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly, 13* (3), 319-340.
52. De Groot, D.A, De Vries, M, Joling, K.J, Van Campen, J., Hugtenburg, J.G, Van Marum, R.J., Annemieke M.A. Vermeulen Windsant-Van den Tweel, A.M.A, Elders, P.J.M, & Van Hout, H.J (2014). Specifying ICD9, ICPC and ATC codes for the STOPP/START criteria: a multidisciplinary consensus panel. *Age Ageing, 43* (6), 773-778.
53. De Lusignan, S, Minmagh, C, Kennedy, J, Zeimet, M, Bommezijn, H, & Bryant, J (2001). A survey to identify the clinical coding and classification systems currently in use across Europe. In , *Studies in Health Technology and Informatics* IOS Press
54. Dedhiya, S.D, Hancock, E, Craig, B.A, Doebbeling, C.C, & Thomas, J (2010). Incident use and outcomes associated with potentially inappropriate medication use in older adults. *Am J Geriatr Pharmacother, 8* (6), 562-570.
55. Digitalis Rx B.V. (2011). Over Prescriptor ('About Prescriptor' in Dutch). June 30 2011 Retrieved from http://www.digitalis.nl/prescriptor/over-prescriptor/
56. Donyai, P, O'Grady, K, Jacklin, A, Barber, N, & Franklin, B.D (2008). The effects of electronic prescribing on the quality of prescribing. *British journal of clinical pharmacology, 65* (2), 230-237.
57. Drenth-van Maanen, A, Van Marum, Rob, Knol, Wilma, Van der Linden, Carolien, & Jansen, Paul (2009). Prescribing Optimization Method for Improving Prescribing in Elderly Patients Receiving Polypharmacy. *Drugs & Aging, 26* (8), 1-15.

58. Dutch College of General Practitioners (2004). Tabel 25 'Code gebruiksvoorschrift', versie 4, april 2004. (Table 25 'Code use guidelines', version 4, april 2004, in Dutch).
59. Dutch College of General Practitioners (2013). HIS-Tabel Diagnostische Bepalingen ('CPOE Table Diagnostics', in Dutch).
60. Dutch College of General Practitioners (2013). ICPC | NHG. Retrieved from https://www.nhg.org/themas/publicaties/icpc
61. Dutch College of General Practitioners (2013). NHG-Standaard Somatisch Onvoldoende verklaarde Lichamelijke Klachten ('Standard Somatically Insufficienctly explained physical complaints', in Dutch).
62. Dutch College of General Practitioners ('Nederlands Huisartsengenootschap' in Dutch) (2012). Multidisciplinary guideline Polypharmacy in elderly ('Multidisciplinaire richtlijn Polyfarmacie bij ouderen' in Dutch). , Utrecht:
63. Edmondson, A.C, & McManus, S.E (2007). Methodological fit in management field research. *Academy of management review, 32* (4), 1246-1264.
64. Edwards, I.R, & Aronson, J.K (2000). Adverse drug reactions: definitions, diagnosis, and management. *The Lancet, 356* (9237), 1255-1259.
65. Enck, William, Gilbert, Peter, Chun, Byung-Gon, Cox, Landon, Jung, Jaeyeon, McDaniel, Patrick, & Sheth, Anmol (2010). TaintDroid: An Information-Flow Tracking System for Realtime Privacy Monitoring on Smartphones. In , *OSDI*
66. Eslami, S, Abu-Hanna, A, De Keizer, N., & De Jonge, E (2006). Errors associated with applying decision support by suggesting default doses for aminoglycosides. *Drug Safety, 29* (9), 803-809.
67. Evfimievski, A, Srikant, R, Agrawal, R, & Gehrke, J (2004). Privacy preserving mining of association rules. *Information Systems, 29* (4), 343-364.
68. ExpertDoc B.V. (2011). Wat is NHGDoc ('What is NHGDoc?' in Dutch). June 30 2011 Retrieved from http://www.nhgdoc.nl/
69. Ferranti, J.M, Musser, R.C, Kawamoto, K, & Hammond, W (2006). The clinical document architecture and the continuity of care record: a critical analysis. *Journal of the American Medical Informatics Association, 13* (3), 245-252.
70. Ferrero, Natalie, Morrell, Dean, & Burkhart, Craig (2013). Skin scan: A demonstration of the need for FDA regulation of medical apps on iPhone. *Journal of the American Academy of Dermatology, 68* (3), 515-516.
71. Fox, Ronan, Cooley, James, McGrath, Matthew, & Hauswirth, Manfred (2012). Mobile Health Apps--From Singular to Collaborative. *PHealth 2012: Proceedings of the 9th International Conference on Wearable Micro and Nano Technologies for Personalized Health, June 26-28, 2012, Porto, Portugal* , Porto, Portugal, 158-164.
72. Franch, Xavier (1998). Systematic formulation of non-functional characteristics of software. *1998 Third International Conference on Requirements Engineering* 174-181.
73. Frazier, Suzan (2005). Health outcomes and polypharmacy in elderly individuals; an integrated literature review. *J Gerontol Nurs, 31* (9), 4-11.

74. Gallagher, P, O'Connor, M.N, & O'Mahony, D. (2011). Prevention of potentially inappropriate prescribing for elderly patients: a randomized controlled trial using STOPP/START criteria. *Clinical Pharmacology & Therapeutics, 89* (6), 845-854.
75. Gallagher, P, Ryan, C, Byrne, S, Kennedy, J, & O'Mahony, D (2008). STOPP (screening tool of older person's prescriptions) and START (screening tool to alert doctors to right treatment). Consensus validation. *International journal of clinical pharmacology and therapeutics, 46* (2), 72-83.
76. Gamberger, D, Lavrac, N, & Jovanoski, V (1999). High confidence association rules for medical diagnosis. *Proc. IDAMAP* 42-51.
77. Gellad, W.F, Grenard, J.L, & Marcum, Z.A (2011). A systematic review of barriers to medication adherence in the elderly: looking beyond cost and regimen complexity. *The American journal of geriatric pharmacotherapy, 9* (1), 11-23.
78. German Institute of Medical Documentation and Information (2014). DIMDI - ICD-10-GM. Retrieved from http://www.dimdi.de/static/en/klassi/icd-10-gm/index.htm
79. Glaser, B.G (2002). Constructivist Grounded Theory?. *Forum: Qualitative Social Research, 3* (3),
80. Glaser, B.G, & Strauss, A.L (1990). *The discovery of grounded theory: strategies for qualitative research.* , Piscataway: Transaction Books
81. Glinz, Martin (2007). On non-functional requirements. *Requirements Engineering Conference, 2007. RE'07. 15th IEEE International* 21-26.
82. Golafshani, N (2004). Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report, 8* (4), 597-607.
83. Goulding, C (1998). Grounded theory: the missing methodology on the interpretivist agenda. *Qualitative Market Research: An International Journal, 1* (1), 50-57.
84. Grady, Robert (1992). *Practical software metrics for project management and process improvement.* , Upper Saddle River, NJ, USA: Prentice Hall Englewood Cliffs
85. Gruber, Thomas (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5* (2), 199–220.
86. Gupta, N, Mangal, N, Tiwari, K, & Mitra, P (2006). Mining quantitative association rules in protein sequences. *Proceedings of Australasian Conference on Knowledge* 273-281.
87. Haffey, Faye, Brady, Richard, & Maxwell, Simon (2013). A comparison of the reliability of smartphone apps for opioid conversion. *Drug safety, 36* (2), 111-117.
88. Hajjar, Emily, Cafiero, Angela, & Hanlon, Joseph (2007). Polypharmacy in elderly patients. *The American journal of geriatric pharmacotherapy, 5* (4), 345-351.
89. Handler, Jonathan (2013). Abandoning ICD-10. Retrieved from http://www.healthmgttech.com/articles/201304/abandoning-icd-10.php
90. Hanlon, J.T, Schmader, K.E, Samsa, G.P, Weinberger, M, Uttech, K.M, Lewis, I.K, ..., null, & Feussner, J.R (1992). A method for assessing drug therapy appropriateness. *J Clin Epidemiol, 45* (10), 1045-1051.

91. 't Hart, Harm, Van Dijk, Jan, De Goede, Martijn, Jansen, Wim, & Teunissen, Joop (2003). *Onderzoeksmethoden (Research methods, in Dutch).* (6th) , Amsterdam: Boom
92. Häyrinen, K, Saranto, K, & Nykänen, P (2008). Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *International journal of medical informatics, 77* (5), 291-304.
93. Health and Social Care Information Centre (2014). TRUD - UK Read. Retrieved from https://isd.hscic.gov.uk/trud3/user/guest/group/0/pack/9
94. Heavey, Elizabeth (2010). *Statistics for Nursing: A Practical Approach.* Jones & Bartlett Learning
95. Hevner, Alan, March, Salvatore, Park, Jinsoo, & Ram, Sudha (2004). Design science in information systems research. *MIS quarterly, 28* (1), 75-105.
96. Higgins, S.A, De Laat, M, Gieles, P.M.C, & Geurts, E.M (2002). Managing Product Requirements for Medical IT Products. *Proceedings of the 10th IEEE Joint International Conference on Requirements Engineering* , Essen, Germany, 26-34.
97. Hipp, J, Güntzer, U, & Nakhaeizadeh, null (2000). Algorithms for association rule mining—a general survey and comparison. *ACM SIGKDD Explorations Newsletter, 2* (1), 58-64.
98. Hoagland, Heath, & Williamson, Lionel (2000). Feasibility Studies. , Kentucky: University of Kentucky
99. Holden, RJ, & Karsh, BT (2010). The Technology Acceptance Model: Its past and its future in health care. *Journal of Biomedical Informatics, 43* (1), 159-172.
100. Hor, C.P, O'Donnell, J.M, Murphy, A.W, O'Brien, T, & Kropmans, T.J.B (2010). General practitioners' attitudes and preparedness towards Clinical Decision Support in e-Prescribing (CDS-eP) adoption in the West of Ireland: a cross sectional study. *BMC medical informatics and decision making, 10* (1),
101. Hornbæk, K, & Law, E.L.C (2007). Meta-analysis of correlations among usability measures. *Proceedings of the SIGCHI conference on Human factors in computing systems* 617-626.
102. Hornbaek, Kasper (2006). Current practice in measuring usability: challenges to usability studies and research. *Int J Hum Comput Stud, 64* 79-102.
103. Hull, E, Jackson, K, & Dick, J (2011). *Requirements Engineering.* , London: Springer
104. IFCC-IUPAC (2014). IFCC - Search the NPU terminology. Retrieved from http://www.ifcc.org/ifcc-scientific-division/sd-committees/c-npu/npusearch/
105. International Health Terminology Standards Development Organisation (2014). SNOMED CT. Retrieved from http://www.ihtsdo.org/snomed-ct
106. International Organization for Standardization (ISO) (2001). ISO/IEC 9126-1:2001: Software engineering -- Product quality -- Part 1: Quality model.
107. International Organization for Standardization (ISO) (2009). 31000:2009, Risk management – Principles and guidelines.

108. International Organization for Standardization (ISO) (2011). ISO/IEC 25010:2011: Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models.
109. Israel, Glenn (1992). Determining sample size.
110. Jansen, P.A.F, & Brouwers, J.R.B.J (2012). Clinical Pharmacology in Old Persons. *Scientifica,*
111. Jansen, P.A.F, & Brouwers, J.R.B.J (2012). Clinical pharmacology in old persons. *Scientifica (Cairo),*
112. Jansen, R.L, Finkelstein, A, & Brinkkemper, S (2009). A Sense of Community: A Research Agenda for Software Ecosystems. *31st International Conference on Software Engineering, New and Emerging Research Track* 187-190.
113. Karapanos, E, Zimmerman, J, Forlizzi, J, & Martens, J.B (2013). User experience over time. In , *Modeling Users' Experiences with Interactive Systems* Springer Berlin Heidelberg
114. Kate, R.J (2013). Towards converting clinical phrases into snOMeD cT expressions. *Biomedical informatics insights, 6(Suppl 1)* (29),
115. Kelley, T.L (1927). *Interpretation of educational measurements.* , New York: MacMillan
116. Kesselheim, A.S, Cresswell, K, Phansalkar, S, Bates, D.W, & Sheikh, A (2011). Clinical decision support systems could be modified to reduce 'alert fatigue' while still minimizing the risk of litigation. *Health Affairs, 30* (12), 2310-2317.
117. Ketikidis, Panayiotis, Dimitrovski, Tomislav, Lazuras, Lambros, & Bath, Peter (2012). Acceptance of health information technology in health professionals: An application of the revised technology acceptance model. *Health Informatics Journal, 18* (2), 124-134.
118. Kharrazi, Hadi, Chisholm, Robin, VanNasdale, Dean, & Thompson, Benjamin (2012). Mobile personal health records: An evaluation of features and functionality. *International Journal of Medical Informatics, 81* (9), 579–593.
119. Kierkegaard, P (2013). E-prescription across Europe. *Health and Technology, 3* (3), 205-219.
120. KNMP (Royal Dutch Pharmacists Association) (2015). January 1 2015 Retrieved from http://www.apotheek.nl/medicijnen
121. Koppel, R, Metlay, J., Cohen, A, Abaluck, B, Localio, A., Kimmel, S., & Strom, B. (2005). Role of computerized physician order entry systems in facilitating medication errors. *JAMA, 293* (10), 1197-1203.
122. Kuijpers, MA, Van Marum, RJ, Egberts, AC, & Jansen, PAF (2008). Relationship between polypharmacy and underprescribing. *British Journal of Clinical Pharmacology, 65* (1), 130-3.
123. Kuilboer, M.M, Van Wijk, M.A.M, Mosseveld, M, Van der Does, E, De Jongste, J.C, Overbeek, S.E, Ponsioen, B, & Van der Lei, J (2006). Computed Critiquing Integrated

into Daily Clinical Practice Affects Physicians' Behavior. *Methods of information in medicine, 45* (4), 447-54.

124. Kuperman, G.J, Bobb, A, Payne, T.H, Avery, A.J, Gandhi, T.K, Burns, G, Classen, D.C, & Bates, D.W (2007). Medication-related clinical decision support in computerized provider order entry systems: a review. *Journal of the American Medical Informatics Association, 14* (1), 29-40.
125. Kuyuma, M, Endo, H, & Umegaki, H (2000). Factors influencing nonadherence with medication regimens in the elderly. *Nihon Ronen Igakkai Zasshi, 37* 363-70.
126. Kvale, S (1989). To validate is to question. In Kvale, S, *Issues of validity in qualitative research* Lund: Studentlitteratur
127. Lankhorst, M (2004). ArchiMate Language Primer: Introduction to the ArchiMate Language for Enterprise Architecture. , Enschede: Telematica Instituut
128. Leendertse, A.J, Koning, G.H.P, Goudswaard, A.N, Belitser, S.V, Verhoef, M, Gier, H.J, Egberts, A.C.G, & Bemt, P.M.L.A (2013). Preventing hospital admissions by reviewing medication (PHARM) in primary care: an open controlled study in an elderly population. *Journal of clinical pharmacy and therapeutics, 38* (5), 379-387.
129. Leendertse, AJ, Egberts, ACG, van den Bemt, PMLA, & Stoker, LJS (2008). Frequency of and Risk Factors for Preventable Medication-Related Hospital Admissions in the Netherlands. *Archives of Internal Medicine, 168* (17), 1890-1896.
130. Leendertse, Anne, De Koning, Fred, Goudswaard, Alex, Jonkhoff, Andries, Van den Bogert, Sander, De Gier, Han, Egberts, Toine, & Van den Bemt, Patricia (2011). Preventing hospital admissions by reviewing medication (PHARM) in primary care: design of the cluster randomised, controlled, multi-centre PHARM-study. *BMC Health Serv Res, 11* (1),
131. Liu, H, Burkhart, Q, & Bell, D.S (2011). Evaluation of the NCPDP Structured and Codified Sig Format for e-prescriptions. *Journal of the American Medical Informatics Association,* 645-651.
132. Mallet, L, Spinewine, A, & Huang, A (2007). The challenge of managing drug interactions in elderly people. *The Lancet, 370* (9582), 185-191.
133. Marceglia, Sara, Bonacina, Stefano, Zaccaria, Vittorio, Pagliari, Claudia, & Pinciroli, Francesco (2012). How might the iPad change healthcare?. *Journal of the Royal Society of Medicine, 105* (6), 233-241.
134. March, S, & Smith, G (1995). Design and Natural Science Research on Information. *Decision Support Systems, 15* (4), 251-266.
135. Martens, J.D, Van der Weijden, T, Severens, J.L, De Clercq, P.A, De Bruijn, D.P, Kester, A.D.M, & Winkens, R.A.G (2007). The effect of computer reminders on GPs' prescribing behaviour: a cluster-randomised trial. *International journal of medical informatics, 76* (3), S403--S416.
136. Matson, James, Brockhouse, John, & Wadsworth, James (2010). Vital Steps: A cooperative feasibility study guide. U.S. Department of Agriculture

137. McInnes, D.K, Saltman, D.C, & Kidd, M.R (2006). General practitioners' use of computers for prescribing and electronic health records: results from a national survey. *Medical journal of Australia, 185* (2), 88-91.
138. MedCom (2009). "Den gode XML recept" ("The good XML prescription", in Danish).
139. Mendoza, V, & Novick, D.G (2005). Usability over time. *Proceedings of the 23rd annual international conference on Design of communication: documenting & designing for pervasive information* 151-158.
140. Mensink, Gerda (2013). Veilig elektronisch verhuizen - een tussenstand (Safely transfering electronically - an intermediate score, in Dutch). *SynthesHis, 1* (12), 4-14.
141. Meulendijk, M.C, Meulendijks, E, Jansen, P.A.F, Numans, M.E, & Spruit, M.R (2014). What concerns users of medical apps? Exploring non-functional requirements of medical mobile applications. *Proceedings of the European Conference on Information Systems (ECIS) 2014* , Tel Aviv, Israel,
142. Meulendijk, M.C, Spruit, M.R, Drenth-van Maanen, A.C, Numans, M, Brinkkemper, S, & Jansen, P.A.F. (2013). General practitioners' attitudes towards decision-supported prescribing: an analysis of the Dutch primary care sector. *Health Informatics J, 19* (4), 247-263.
143. Meulendijk, M.C, Spruit, M.R, Drenth-van Maanen, A.C, Numans, M.E, Brinkkemper, S, Jansen, P.A.F, & Knol, W (2015). Computerized decision support improves medication review effectiveness: an experiment evaluating the STRIP Assistant's usability (submitted). *Drugs & Aging,*
144. Meulendijk, M.C, Spruit, M.R, Jansen, P.A.F, Numans, M.E, & Brinkkemper, S (2015). STRIPA: a rule-based decision support system for medication reviews in primary care (accepted). *Proceedings of the European Conference on Information Systems (ECIS) 2015* , Münster, Germany,
145. Meulendijk, M.C. (2014). STRIP Assistant demonstrating video. Retrieved from http://videodemo.stripa.eu/english/
146. Meulendijk, Michiel, Van de Wijngaert, Lidwien, Brinkkemper, Sjaak, & Leenstra, Herbert (2011). AmI in good care? Developing design principles for ambient intelligent domotics for elderly. *Informatics for Health and Social Care, 36* (2), 75-88.
147. Meyer, C.H, Stanzel, B.V, Moqaddem, S, & Brohlburg, D (2012). Healthcare "Apps" for Smartphones: Relief or toy for patients and the visual impaired?. *Der Ophthalmologe: Zeitschrift der Deutschen Ophthalmologischen Gesellschaft, 109* (1),
148. Morris, C.J, Savelyich, B.S.P, Avery, A.J, Cantrill, J.A, & Sheikh, A (2005). Patient safety features of clinical computer systems: questionnaire survey of GP views. *Quality and Safety in Health Care, 14* (3), 164.
149. Musen, M.A, Middleton, B, & Greenes, R.A (2014). Clinical decision-support systems. In Shortliffe, Edward, & Cimino, James, *Biomedical informatics* Springer London

150. Nandigam, David, Symonds, Judith, Kayes, Nicola, & McPherson, Kathryn (2010). Proceedings of the 11th International Conference of the NZ Chapter of the ACM Special Interest Group on Human-Computer Interaction. 69-72.
151. Naranjo, C.A, Busto, U, Sellers, E.M, Sandor, P, Ruiz, I, Roberts, E.A, Janecek, E, Domecq, C, & D.J, Greenblatt (1981). A method for estimating the probability of adverse drug reactions. *Clinical Pharmacology & Therapeutics, 30* (2), 239-245.
152. Newell, A,, Gregor, P, Morgan, M, Pullin, G, & Macaulay, C (2011). User-Sensitive Inclusive Design. *Universal Access in the Information Society, 10* (3), 235-243.
153. Nictiz (2008). Patiënt Overdracht Bericht (Patient Transfer Message, in Dutch). , Den Haag:
154. Nielsen, J (1994). Heuristic evaluation. In Nielsen, J., & Mack, R.L, *Usability Inspection Methods* , New York: John Wiley & Sons
155. Nielsen, J (1994). *Usability engineering.* Elsevier
156. Nilsson, E.G, & Følstad, A (2012). Effectiveness and Efficiency as Conflicting Requirements in Designing Emergency Mission Reporting. *I-UxSED,* 20-25.
157. Niquille, A, Lattmann, C, & Bugnon, O (2010). Medication reviews led by community pharmacists in Switzerland: a qualitative survey to evaluate barriers and facilitators. *Pharmacy practice, 8* (1), 35–42.
158. O'Mahony, D, O'Sullivan, D, Byrne, S, O'Connor, M.N, Ryan, C, & Gallagher, P (2014). STOPP/START criteria for potentially inappropriate prescribing in older people: version 2. *Age Ageing, 44* (2), 213-218.
159. Okkes, I.M, Oskam, S.K, & Lamberts, H (2005). ICPC in the Amsterdam Transition Project. , Amsterdam: Academic Medical Center/University of Amsterdam, Department of Family Medicine
160. Oliveira, S.R.M, & Zaïane, Osmar (2003). Protecting sensitive knowledge by data sanitization. *Proceedings of the Third IEEE International Conference on Data Mining* , Melbourne, 613-617.
161. Orme, Anthony, Yao, Haining, & Etzkorn, Letha (2007). Indicating Ontology Data Quality, Stability, and Completeness Throughout Ontology Evolution: Research Articles. *Journal of Software Maintenance and Evolution: Research and Practice, 19* (1), 49-75.
162. Osheroff, J.A, Teich, J.M, Middleton, B, Steen, E.B, Wright, A, & Detmer, D.E (2007). A roadmap for national action on clinical decision support. *Journal of the American medical informatics association, 14* (2), 141.
163. Ozdalga, Errol, Ozdalga, Ark, & Ahuja, Neera (2012). The smartphone in medicine: a review of current and potential use among physicians and students. *Journal of medical Internet research, 14* (5),
164. Pachidi, S, Spruit, M, & Weerd, I. (2014). Understanding users' behavior with software operation data mining. *Comput Human Behav, 30* 583-594.

165. Paech, B, & Kerkow, D (2004). Non-Functional Requirements Engineering - Quality is Essential. *Proceedings of the 10th International Workshop of Requirements Engineering: Foundation for Software Quality* , Riga, Latvia, 27-40.
166. Paré, G, Sicotte, C, & Jacques, H (2006). The effects of creating psychological ownership on physicians' acceptance of clinical information systems. *Journal of the American Medical Informatics Association, 13* (2), 197-205.
167. Patterson, S.M, Hughes, C, Kerse, N, Cardwell, C.R, & Bradley, M.C (2012). Interventions to improve the appropriate use of polypharmacy for older people. Cochrane Database of Systematic Reviews
168. Peffers, K, Tuunanen, T, Gengler, C.E, Rossi, M, Hui, W, Virtanen, V, & Bragge, J (2006). The design science research process: a model for producing and presenting information systems research. *Proceedings of the first international conference on design science research in information systems and technology (DESRIST 2006)* 83-106.
169. Perera, Gihan, Holbrook, Anne, Thabane, Lehana, Foster, Gary, & Willison, Donald (2011). Views on health information sharing and privacy from primary care practices using electronic medical records. *International journal of medical informatics, 80* (2), 94-101.
170. Pevnick, J.M, Asch, S.M, Adams, J.L, Mattke, S, Patel, M.H, Ettner, S.L, & Bell, D.S (2010). Adoption and use of stand-alone electronic prescribing in a health plan-sponsored initiative. *The American journal of managed care, 16* (3), 182-191.
171. Rasmussen, H.M.S, Sondergaard, J, Kampmann, J.P, & Andersen, M (2005). General practitioners prefer prescribing indicators based on detailed information on individual patients: a Delphi study. *European journal of clinical pharmacology, 61* (3), 237-241.
172. Reason, P, & Bradbury, H (2006). *Handbook of action research: the concise paperback edition.* SAGE
173. Red Hat Inc. (2014). Drools - Business Rules Management System (Java™, Open Source). Retrieved from http://www.drools.org/
174. Regenstrief Institute (2014). LOINC Background - LOINC. Retrieved from http://loinc.org/background
175. Renaud, Karen, & Van Biljon, Judy (2008). Predicting technology acceptance and adoption by the elderly: a qualitative study. *Proceedings of the 2008 annual research conference of the South African Institute of Computer Scientists and Information Technologists on IT research in developing countries: riding the wave of technology* 210-219.
176. Ricci, F, Rokach, L, & Shapira, B. (2011). Introduction to recommender systems handbook. In Ricci, F, Rokach, L, Shapira, B, & Kantor, P.B, *Recommender systems handbook* Springer US

177. Rolfe, G (2006). Validity, trustworthiness and rigour: quality and the idea of qualitative research. *Journal of advanced nursing, 53* (3), 304-310.
178. Ron, W (2004). The Rhetoric of Positivism Versus Interpretivism. *MIS Quarterly, 28* (1),
179. Roshanov, P.S, Fernandes, N, Wilczynski, J.M, Hemens, B.J, You, J.J, Handler, S.M, & ... Haynes, R.B (2013). Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. *British Medical Journal, 346*
180. Rothman, K.J (1990). No adjustments are needed for multiple comparisons. *Epidemiology, 1* (1), 43-46.
181. Sarasohn-Kahn, J (2010). How smartphones are changing health care for consumers and providers. , Oakland:
182. Saville, D.J (1990). Multiple comparison procedures: the practical solution. *American Statistician, 44* (2), 174-180.
183. Sayers, Y.M, Armstrong, P, & Hanley, K (2009). Prescribing errors in general practice: A prospective study. *European Journal of General Practice, 15* (2), 81--83.
184. Schulz, S, Rector, A, Rodrigues, J.M, & Spackman, K (2012). Competing interpretations of disorder codes in SNOMED CT and ICD. *AMIA Annual Symposium Proceedings* 819-827.
185. Seidling, H.M, Schmitt, S.P.W, Bruckner, T, Kaltschmidt, J, Pruszydlo, M.G, Senger, C, Bertsche, T, Walter-Sack, I, & Haefeli, W.E (2010). Patient-specific electronic decision support reduces prescription of excessive doses. *Quality and Safety in Health Care, 19* (5), 1-7.
186. Serban, G, Czibula, I.G, & Campan, A (). A Programming interface for medical diagnosis prediction. *Studia Universitatis, Babes-Bolyai, Informatica, LI* (1), 21-30.
187. Sergi, G, De Rui, M, Sarti, S, & Manzato, E (2011). Polypharmacy in the elderly: can comprehensive geriatric assessment reduce inappropriate medication use?. *Drugs Aging, 28* (7), 509–518.
188. Shamliyan, T., Duval, S, Du, J, & Kane, R.L (2008). Just What the Doctor Ordered. Review of the Evidence of the Impact of Computerized Physician Order Entry System on Medication Errors. *Health services research, 43* (1 Pt), 32–53.
189. Shaughnessy, J.J, & Zechmeister, E.B (1985). *Research methods in psychology.* , New York : McGraw Hill
190. Shi, S, Mörike, K, & Klotz, U (2008). The clinical implications of ageing for rational drug therapy. *Eur J Clin Pharmacol, 64* (2), 183-99.
191. Shvaiko, P, & Euzenat, J (2013). Ontology matching: state of the art and future challenges. *IEEE Transactions on Knowledge and Data Engineering* 158-176.
192. Sloane, PD, Gruber-Baldini, AL, Zimmerman, S, Roth, M, Watson, L, Boustani, M, Magaziner, J, & Hebel, JR (2004). Medication undertreatment in assisted living settings. *Arch Intern Med, 164* (18), 2031-2037.
193. Smith, Eric (2010). iPhone applications & privacy issues: An analysis of application transmission of iPhone unique device identifiers (UDIDs). , Lewisburg:

194. Spinewine, A, Fialová, D, & Byrne, S (2012). The role of the pharmacist in optimizing pharmacotherapy in older people. *Drugs & aging, 29* (6), 495-510.
195. Spinewine, A, Schmader, K.E, Barber, N, Hughes, C, Lapane, K.L, Swine, C, & Hanlon, J.T (2007). Appropriate prescribing in elderly people: how well can it be measured and optimised?. *The Lancet, 370* (9582), 173-184.
196. Stair, Ralph, & Reynolds, George (2008). *Fundamentals of Information Systems: A Managerial Approach.* Thomson Course Technology
197. Steinman, MA, Landefeld, CS, Rosenthal, GE, Berthenthal, D, Sen, S, & Kaboli, PJ (2006). Polypharmacy and prescribing quality in older people. *Journal of the American Geriatrics Society, 54* (10), 1516-23.
198. Stichting Farmaceutische Kengetallen (2005). Polyfarmacie. *Pharmaceutisch Weekblad, 140* (32),
199. Swanwick, T (2010). *Understanding medical education.* Wiley Online Library
200. Tan, Pang-Ning, Michael, Steinbach, & Kumar, Vipin (2005). Chapter 6. Association Analysis: Basic Concepts and Algorithms. In , *Introduction to Data Mining* Addison-Wesley
201. Tarn, D.M, Paterniti, D.A, Wenger, N.S, Williams, B.R, & Chewning, B.A (2012). Older patient, physician and pharmacist perspectives about community pharmacists' roles. *International Journal of Pharmacy Practice, 20* (5), 285-293.
202. Thomas, G (2011). A typology for the case study in social science following a review of definition, discourse and structure. *Qualitative Inquiry, 17* (6), 511–521.
203. Thompson, Alan (2005). Business Feasibility Study Outline. In Thompson, Alan, *Enterpreneurship and business innovation: The art of successful business start-ups and business planning* , Guildford: Vineyard Publishing
204. Timpka, Toomas (1989). Introducing hypertext in primary health care: on the feasibility of decision support for practitioners. *Computer Methods and Programs in Biomedicine, 29* (1), 1-13.
205. Topinková, E, Baeyens, J.P, Michel, J.P, & Lang, P.O (2012). Evidence-based strategies for the optimization of pharmacotherapy in older people. *Drugs Aging, 29* (6), 477-494.
206. U.S. National Library of Medicine (2014). UMLS Quick Start Guide. Retrieved from http://www.nlm.nih.gov/research/umls/quickstart.html
207. Van Bruwaene, B, Mies, K, Schots, J, & Van de Velde, R (2004). Can SNOMED CT replace ICD-coding?. Retrieved from http://users.telenet.be/benny.vanbruwaene/0510%20SNOMED%20CT.pdf
208. Van den Berg, M, De Bakker, D, Spreeuwenberg, P, Westert, G, Braspenning, J, Van der Zee, J, & Groenewegen, P (2009). Labour intensity of guidelines may have a greater effect on adherence than GPs' workload. *BMC family practice, 10* (1), 74-86.
209. Van der Lugt, J, & Klapwijk, S (2008). Overzicht in beeld ('Overview in focus' in Dutch). , *7* (4),

210. Van der Sijs, H (2009). Drug safety alerting in computerized physician order entry: unraveling and counteracting alert fatigue.
211. van Oostenrijk, A (2004). Screen scraping web services. , Nijmegen:
212. Van Schaik, P, Flynn, D, Van Wersch, A, Douglass, A, & Cann, P (2004). The acceptance of a computerised decision-support system in primary care: a preliminary investigation. *Behaviour & information technology, 23* (5), 321-326.
213. Veenma, K, Batenburg, R, & Breedveld, E (2004). De vignetmethode: Een praktische handreiking bij beleidsonderzoek. , Tilburg, The Netherlands:
214. Velo, Giampaolo, & Minuz, Pietro (2009). Medication errors: prescribing faults and prescription errors. *British Journal of Clinical Pharmacology, 67* (6), 624–628.
215. Venable, J (2006). The role of theory and theorising in design science research. *Proceedings of the 1st International Conference on Design Science in Information Systems and Technology (DESRIST 2006)* 1-18.
216. Venkat, A, Chan-Tompkins, NH, Hegde, GG, Chuirazzi, DM, Hunter, R, & Szczesiul, JM (2010). Feasibility of integrating a clinical decision support tool into an existing computerized physician order entry system to increase seasonal influenza vaccination in the emergency department. *Vaccine, 28* (37), 1873-2518.
217. Venkatesh, V, & Davis, F.D (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science, 46* (2), 186-204.
218. Venkatesh, V, Morris, M.G, Davis, G.B, & Davis, F.D (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly, 27* (3), 425-478.
219. Venkatesh, V, Sykes, T.A, & Zhang, X (2011). 'Just What the Doctor Ordered': A Revised UTAUT for EMR System Adoption and Use by Doctors. *44th Hawaii International Conference on System Sciences* , Kauai, Hawaii, USA, 1-10.
220. Vikström, A, Skånér, Y, Strender, L.E, & Nilsson, G.H. (2007). Mapping the categories of the Swedish primary health care version of ICD-10 to SNOMED CT concepts: Rule development and intercoder reliability in a mapping trial. *BMC medical informatics and decision making, 7* (9), 9.
221. Visvanathan, A, Hamilton, A, & Brady, RRW (2012). Smartphone apps in microbiology—is better regulation required?. *Clinical Microbiology and Infection, 18* (7), E218-E220.
222. Walker, Jan, Ahern, David, Le, Lan, & Delbanco, Tom (2009). Insights for internists:"I want the computer to know who I am". *Journal of general internal medicine, 24* (6), 727-732.
223. Walker, M, Takayama, L, & Landay, J.A (2002). High-fidelity or low-fidelity, paper or computer? Choosing attributes when testing web prototypes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 661-665.
224. Wand, Y, & Wang, R. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM, 39* (11), 86-95.

225. Weiskopf, Nicole, & Weng, Chunhua (2013). Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association, 20* (1), 144-151.
226. Werner, Franz, Werner, Katharina, & Oberzaucher, Johannes (2012). Tablets for Seniors--An Evaluation of a Current Model (iPad). In Wichert, Reiner, & Eberhardt, Birgid, *Ambient Assisted Living*
227. Willeboordse, F, Hugtenburg, J.G, van Dijk, L, Bosmans, J.E, de Vries, O.J, Schellevis, F.G, & Elders, P.J (2014). Opti-Med: the effectiveness of optimised clinical medication reviews in older people with 'geriatric giants' in general practice; study protocol of a cluster randomised controlled trial. *BMC geriatrics, 14* (1),
228. Wohlin, C, Runeson, P, Höst, M, Ohlsson, M.C, Regnell, B, & Wesslén, A (2012). *Experimentation in software engineering*. Springer Science & Business Media
229. Wolf, Joel, Moreau, Jacqueline, Akilov, Oleg, Patton, Timothy, English, Joseph, Ho, Jonhan, & Ferris, Laura (2013). Diagnostic Inaccuracy of Smartphone Applications for Melanoma DetectionMelanoma Detection Using Smartphone Applications. *Jama Dermatology, 149* (4), 422-426.
230. World Health Organization (1990). WHO | The Anatomical Therapeutic Chemical Classification System with Defined Daily Doses (ATC/DDD). Retrieved from http://www.who.int/classifications/atcddd/en/
231. World Health Organization (1994). WHO | International Classification of Diseases (ICD). Retrieved from http://www.who.int/classifications/icd/en/
232. World Health Organization (1998). WHO | International Classification of Primary Care, Second edition (ICPC-2). Retrieved from http://www.who.int/classifications/icd/adaptations/icpc2/en/
233. World Health Organization (2009). Retrieved from http://www.whocc.no/ddd/definition_and_general_considera/
234. Wright, RM, Sloane, R, Pieper, CF, Ruby-Scelsi, C, Twersky, J, Schmader, KE, & Hanlon, JT (2009). Underuse of indicated medications among physically frail older US veterans at the time of hospital discharge: results of a cross-sectional analysis of data from the geriatric evaluation and management drug study. *Am J Geriatr Pharmacother, 7* (5), 271-280.
235. Wu, R.C, Chen, R.S, & Chen, C.C (2005). Data mining application in customer relationship management of credit card business. *29th Annual International on Computer Software and Applications Conference* 39-40.
236. Wu, Wei, Ng, Wee, Krishnaswamy, Shonali, & Sinha, Abhijat (2012). To Taxi or Not to Taxi? - Enabling Personalised and Real-Time Transportation Decisions for Mobile Users. *2012 IEEE 13th International Conference on Mobile Data Management (MDM)* , Bengaluru, Karnataka, 320-323.
237. Wu, Y.H, Chiang, C.M, & Chen, A.L (2007). Hiding sensitive association rules with limited side effects. *IEEE Transactions on Knowledge and Data Engineering, 19* (1), 29-42.

238. Yarbrough, AK, & Smith, TB (2007). Technology Acceptance among Physicians: A New Take on TAM. *Medical Care Research and Review, 64* (6), 650-672.
239. Yin, R.K (2013). *Case study research: Design and methods.* Sage publications
240. Z-Index (2014). December 17 2014 Retrieved from https://www.z-index.nl/g-standaard

SIKS Dissertation Series

2009

2009-01 Rasa Jurgelenaite (RUN)
Symmetric Causal Independence Models

2009-02 Willem Robert van Hage (VU)
Evaluating Ontology-Alignment Techniques

2009-03 Hans Stol (UvT)
A Framework for Evidence-based Policy Making Using IT

2009-04 Josephine Nabukenya (RUN)
Improving the Quality of Organisational Policy Making using Collaboration Engineering

2009-05 Sietse Overbeek (RUN)
Bridging Supply and Demand for Knowledge Intensive Tasks - Based on Knowledge, Cognition, and Quality

2009-06 Muhammad Subianto (UU)
Understanding Classification

2009-07 Ronald Poppe (UT)
Discriminative Vision-Based Recovery and Recognition of Human Motion

2009-08 Volker Nannen (VU)
Evolutionary Agent-Based Policy Analysis in Dynamic Environments

2009-09 Benjamin Kanagwa (RUN)
Design, Discovery and Construction of Service-oriented Systems

2009-10 Jan Wielemaker (UVA)
Logic programming for knowledge-intensive interactive applications

2009-11 Alexander Boer (UVA)
Legal Theory, Sources of Law & the Semantic Web

2009-12 Peter Massuthe (TUE, Humboldt-Universitaet zu Berlin)
Operating Guidelines for Services

2009-13 Steven de Jong (UM)
Fairness in Multi-Agent Systems

2009-14 Maksym Korotkiy (VU)
From ontology-enabled services to service-enabled ontologies (making ontologies work in e-science with ONTO-SOA)

2009-15 Rinke Hoekstra (UVA)
Ontology Representation - Design Patterns and Ontologies that Make Sense

2009-16 Fritz Reul (UvT)
New Architectures in Computer Chess

2009-17 Laurens van der Maaten (UvT)
Feature Extraction from Visual Data

2009-18 Fabian Groffen (CWI)

Armada, An Evolving Database System

2009-19 Valentin Robu (CWI)
Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets

2009-20 Bob van der Vecht (UU)
Adjustable Autonomy: Controling Influences on Decision Making

2009-21 Stijn Vanderlooy (UM)
Ranking and Reliable Classification

2009-22 Pavel Serdyukov (UT)
Search For Expertise: Going beyond direct evidence

2009-23 Peter Hofgesang (VU)
Modelling Web Usage in a Changing Environment

2009-24 Annerieke Heuvelink (VUA)
Cognitive Models for Training Simulations

2009-25 Alex van Ballegooij (CWI)
"RAM: Array Database Management through Relational Mapping"

2009-26 Fernando Koch (UU)
An Agent-Based Model for the Development of Intelligent Mobile Services

2009-27 Christian Glahn (OU)
Contextual Support of social Engagement and Reflection on the Web

2009-28 Sander Evers (UT)
Sensor Data Management with Probabilistic Models

2009-29 Stanislav Pokraev (UT)
Model-Driven Semantic Integration of Service-Oriented Applications

2009-30 Marcin Zukowski (CWI)
Balancing vectorized query execution with bandwidth-optimized storage

2009-31 Sofiya Katrenko (UVA)
A Closer Look at Learning Relations from Text

2009-32 Rik Farenhorst (VU) and Remco de Boer (VU)
Architectural Knowledge Management: Supporting Architects and Auditors

2009-33 Khiet Truong (UT)
How Does Real Affect Affect Affect Recognition In Speech?

2009-34 Inge van de Weerd (UU)
Advancing in Software Product Management: An Incremental Method Engineering Approach

2009-35 Wouter Koelewijn (UL)
Privacy en Politiegegevens; Over geautomatiseerde normatieve informatie-uitwisseling

2009-36 Marco Kalz (OUN)
Placement Support for Learners in Learning Networks

2009-37 Hendrik Drachsler (OUN)
Navigation Support for Learners in Informal Learning Networks

2009-38 Riina Vuorikari (OU)
Tags and self-organisation: a metadata ecology for learning resources in a multilingual context

2009-39 Christian Stahl (TUE, Humboldt-Universitaet zu Berlin)
Service Substitution -- A Behavioral Approach Based on Petri Nets

2009-40 Stephan Raaijmakers (UvT)
Multinomial Language Learning: Investigations into the Geometry of Language

2009-41 Igor Berezhnyy (UvT)
Digital Analysis of Paintings

2009-42 Toine Bogers (UvT)
Recommender Systems for Social Bookmarking

2009-43 Virginia Nunes Leal Franqueira (UT)
Finding Multi-step Attacks in Computer Networks using Heuristic Search and Mobile Ambients

2009-44 Roberto Santana Tapia (UT)
Assessing Business-IT Alignment in Networked Organizations
2009-45 Jilles Vreeken (UU)
Making Pattern Mining Useful
2009-46 Loredana Afanasiev (UvA)
Querying XML: Benchmarks and Recursion

2010

2010-01 Matthijs van Leeuwen (UU)
Patterns that Matter
2010-02 Ingo Wassink (UT)
Work flows in Life Science
2010-03 Joost Geurts (CWI)
A Document Engineering Model and Processing Framework for Multimedia documents
2010-04 Olga Kulyk (UT)
Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments
2010-05 Claudia Hauff (UT)
Predicting the Effectiveness of Queries and Retrieval Systems
2010-06 Sander Bakkes (UvT)
Rapid Adaptation of Video Game AI
2010-07 Wim Fikkert (UT)
Gesture interaction at a Distance
2010-08 Krzysztof Siewicz (UL)
Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments
2010-09 Hugo Kielman (UL)
A Politiele gegevensverwerking en Privacy, Naar een effectieve waarborging
2010-10 Rebecca Ong (UL)
Mobile Communication and Protection of Children
2010-11 Adriaan Ter Mors (TUD)
The world according to MARP: Multi-Agent Route Planning
2010-12 Susan van den Braak (UU)
Sensemaking software for crime analysis
2010-13 Gianluigi Folino (RUN)
High Performance Data Mining using Bio-inspired techniques
2010-14 Sander van Splunter (VU)
Automated Web Service Reconfiguration
2010-15 Lianne Bodenstaff (UT)
Managing Dependency Relations in Inter-Organizational Models
2010-16 Sicco Verwer (TUD)
Efficient Identification of Timed Automata, theory and practice
2010-17 Spyros Kotoulas (VU)
Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications
2010-18 Charlotte Gerritsen (VU)
Caught in the Act: Investigating Crime by Agent-Based Simulation
2010-19 Henriette Cramer (UvA)
People's Responses to Autonomous and Adaptive Systems
2010-20 Ivo Swartjes (UT)
Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative

2010-21 Harold van Heerde (UT)
Privacy-aware data management by means of data degradation
2010-22 Michiel Hildebrand (CWI)
End-user Support for Access to\\ Heterogeneous Linked Data
2010-23 Bas Steunebrink (UU)
The Logical Structure of Emotions
2010-24 Dmytro Tykhonov
Designing Generic and Efficient Negotiation Strategies
2010-25 Zulfiqar Ali Memon (VU)
Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective
2010-26 Ying Zhang (CWI)
XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines
2010-27 Marten Voulon (UL)
Automatisch contracteren
2010-28 Arne Koopman (UU)
Characteristic Relational Patterns
2010-29 Stratos Idreos(CWI)
Database Cracking: Towards Auto-tuning Database Kernels
2010-30 Marieke van Erp (UvT)
Accessing Natural History - Discoveries in data cleaning, structuring, and retrieval
2010-31 Victor de Boer (UVA)
Ontology Enrichment from Heterogeneous Sources on the Web
2010-32 Marcel Hiel (UvT)
An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems
2010-33 Robin Aly (UT)
Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval
2010-34 Teduh Dirgahayu (UT)
Interaction Design in Service Compositions
2010-35 Dolf Trieschnigg (UT)
Proof of Concept: Concept-based Biomedical Information Retrieval
2010-36 Jose Janssen (OU)
Paving the Way for Lifelong Learning; Facilitating competence development through a learning path specification
2010-37 Niels Lohmann (TUE)
Correctness of services and their composition
2010-38 Dirk Fahland (TUE)
From Scenarios to components
2010-39 Ghazanfar Farooq Siddiqui (VU)
Integrative modeling of emotions in virtual agents
2010-40 Mark van Assem (VU)
Converting and Integrating Vocabularies for the Semantic Web
2010-41 Guillaume Chaslot (UM)
Monte-Carlo Tree Search
2010-42 Sybren de Kinderen (VU)
Needs-driven service bundling in a multi-supplier setting - the computational e3-service approach
2010-43 Peter van Kranenburg (UU)
A Computational Approach to Content-Based Retrieval of Folk Song Melodies
2010-44 Pieter Bellekens (TUE)
An Approach towards Context-sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain
2010-45 Vasilios Andrikopoulos (UvT)

A theory and model for the evolution of software services

2010-46 Vincent Pijpers (VU)
e3alignment: Exploring Inter-Organizational Business-ICT Alignment

2010-47 Chen Li (UT)
Mining Process Model Variants: Challenges, Techniques, Examples

2010-48 Withdrawn

2010-49 Jahn-Takeshi Saito (UM)
Solving difficult game positions

2010-50 Bouke Huurnink (UVA)
Search in Audiovisual Broadcast Archives

2010-51 Alia Khairia Amin (CWI)
Understanding and supporting information seeking tasks in multiple sources

2010-52 Peter-Paul van Maanen (VU)
Adaptive Support for Human-Computer Teams: Exploring the Use of Cognitive Models of Trust and Attention

2010-53 Edgar Meij (UVA)
Combining Concepts and Language Models for Information Access

2011

2011-01 Botond Cseke (RUN)
Variational Algorithms for Bayesian Inference in Latent Gaussian Models

2011-02 Nick Tinnemeier(UU)
Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language

2011-03 Jan Martijn van der Werf (TUE)
Compositional Design and Verification of Component-Based Information Systems

2011-04 Hado van Hasselt (UU)
Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference learning algorithms

2011-05 Base van der Raadt (VU)
Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.

2011-06 Yiwen Wang (TUE)
Semantically-Enhanced Recommendations in Cultural Heritage

2011-07 Yujia Cao (UT)
Multimodal Information Presentation for High Load Human Computer Interaction

2011-08 Nieske Vergunst (UU)
BDI-based Generation of Robust Task-Oriented Dialogues

2011-09 Tim de Jong (OU)
Contextualised Mobile Media for Learning

2011-10 Bart Bogaert (UvT)
Cloud Content Contention

2011-11 Dhaval Vyas (UT)
Designing for Awareness: An Experience-focused HCI Perspective

2011-12 Carmen Bratosin (TUE)
Grid Architecture for Distributed Process Mining

2011-13 Xiaoyu Mao (UvT)
Airport under Control. Multiagent Scheduling for Airport Ground Handling

2011-14 Milan Lovric (EUR)
Behavioral Finance and Agent-Based Artificial Markets

2011-15 Marijn Koolen (UvA)

The Meaning of Structure: the Value of Link Evidence for Information Retrieval

2011-16 Maarten Schadd (UM)
Selective Search in Games of Different Complexity

2011-17 Jiyin He (UVA)
Exploring Topic Structure: Coherence, Diversity and Relatedness

2011-18 Mark Ponsen (UM)
Strategic Decision-Making in complex games

2011-19 Ellen Rusman (OU)
The Mind ' s Eye on Personal Profiles

2011-20 Qing Gu (VU)
Guiding service-oriented software engineering - A view-based approach

2011-21 Linda Terlouw (TUD)
Modularization and Specification of Service-Oriented Systems

2011-22 Junte Zhang (UVA)
System Evaluation of Archival Description and Access

2011-23 Wouter Weerkamp (UVA)
Finding People and their Utterances in Social Media

2011-24 Herwin van Welbergen (UT)
Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior

2011-25 Syed Waqar ul Qounain Jaffry (VU))
Analysis and Validation of Models for Trust Dynamics

2011-26 Matthijs Aart Pontier (VU)
Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots

2011-27 Aniel Bhulai (VU)
Dynamic website optimization through autonomous management of design patterns

2011-28 Rianne Kaptein(UVA)
Effective Focused Retrieval by Exploiting Query Context and Document Structure

2011-29 Faisal Kamiran (TUE)
Discrimination-aware Classification

2011-30 Egon van den Broek (UT)
Affective Signal Processing (ASP): Unraveling the mystery of emotions

2011-31 Ludo Waltman (EUR)
Computational and Game-Theoretic Approaches for Modeling Bounded Rationality

2011-32 Nees-Jan van Eck (EUR)
Methodological Advances in Bibliometric Mapping of Science

2011-33 Tom van der Weide (UU)
Arguing to Motivate Decisions

2011-34 Paolo Turrini (UU)
Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations

2011-35 Maaike Harbers (UU)
Explaining Agent Behavior in Virtual Training

2011-36 Erik van der Spek (UU)
Experiments in serious game design: a cognitive approach

2011-37 Adriana Burlutiu (RUN)
Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference

2011-38 Nyree Lemmens (UM)
Bee-inspired Distributed Optimization

2011-39 Joost Westra (UU)
Organizing Adaptation using Agents in Serious Games

2011-40 Viktor Clerc (VU)
Architectural Knowledge Management in Global Software Development
2011-41 Luan Ibraimi (UT)
Cryptographically Enforced Distributed Data Access Control
2011-42 Michal Sindlar (UU)
Explaining Behavior through Mental State Attribution
2011-43 Henk van der Schuur (UU)
Process Improvement through Software Operation Knowledge
2011-44 Boris Reuderink (UT)
Robust Brain-Computer Interfaces
2011-45 Herman Stehouwer (UvT)
Statistical Language Models for Alternative Sequence Selection
2011-46 Beibei Hu (TUD)
Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work
2011-47 Azizi Bin Ab Aziz(VU)
Exploring Computational Models for Intelligent Support of Persons with Depression
2011-48 Mark Ter Maat (UT)
Response Selection and Turn-taking for a Sensitive Artificial Listening Agent
2011-49 Andreea Niculescu (UT)
Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality

2012

2012-01 Terry Kakeeto (UvT)
Relationship Marketing for SMEs in Uganda
2012-02 Muhammad Umair(VU)
Adaptivity, emotion, and Rationality in Human and Ambient Agent Models
2012-03 Adam Vanya (VU)
Supporting Architecture Evolution by Mining Software Repositories
2012-04 Jurriaan Souer (UU)
Development of Content Management System-based Web Applications
2012-05 Marijn Plomp (UU)
Maturing Interorganisational Information Systems
2012-06 Wolfgang Reinhardt (OU)
Awareness Support for Knowledge Workers in Research Networks
2012-07 Rianne van Lambalgen (VU)
When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions
2012-08 Gerben de Vries (UVA)
Kernel Methods for Vessel Trajectories
2012-09 Ricardo Neisse (UT)
Trust and Privacy Management Support for Context-Aware Service Platforms
2012-10 David Smits (TUE)
Towards a Generic Distributed Adaptive Hypermedia Environment
2012-11 J.C.B. Rantham Prabhakara (TUE)
Process Mining in the Large: Preprocessing, Discovery, and Diagnostics
2012-12 Kees van der Sluijs (TUE)
Model Driven Design and Data Integration in Semantic Web Information Systems

2012-13 Suleman Shahid (UvT)
Fun and Face: Exploring non-verbal expressions of emotion during playful interactions
2012-14 Evgeny Knutov(TUE)
Generic Adaptation Framework for Unifying Adaptive Web-based Systems
2012-15 Natalie van der Wal (VU)
Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes.
2012-16 Fiemke Both (VU)
Helping people by understanding them - Ambient Agents supporting task execution and depression treatment
2012-17 Amal Elgammal (UvT)
Towards a Comprehensive Framework for Business Process Compliance
2012-18 Eltjo Poort (VU)
Improving Solution Architecting Practices
2012-19 Helen Schonenberg (TUE)
What's Next? Operational Support for Business Process Execution
2012-20 Ali Bahramisharif (RUN)
Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing
2012-21 Roberto Cornacchia (TUD)
Querying Sparse Matrices for Information Retrieval
2012-22 Thijs Vis (UvT)
Intelligence, politie en veiligheidsdienst: verenigbare grootheden?
2012-23 Christian Muehl (UT)
Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction
2012-24 Laurens van der Werff (UT)
Evaluation of Noisy Transcripts for Spoken Document Retrieval
2012-25 Silja Eckartz (UT)
Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application
2012-26 Emile de Maat (UVA)
Making Sense of Legal Text
2012-27 Hayrettin Gurkok (UT)
Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games
2012-28 Nancy Pascall (UvT)
Engendering Technology Empowering Women
2012-29 Almer Tigelaar (UT)
Peer-to-Peer Information Retrieval
2012-30 Alina Pommeranz (TUD)
Designing Human-Centered Systems for Reflective Decision Making
2012-31 Emily Bagarukayo (RUN)
A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure
2012-32 Wietske Visser (TUD)
Qualitative multi-criteria preference representation and reasoning
2012-33 Rory Sie (OUN)
Coalitions in Cooperation Networks (COCOON)
2012-34 Pavol Jancura (RUN)
Evolutionary analysis in PPI networks and applications
2012-35 Evert Haasdijk (VU)
Never Too Old To Learn -- On-line Evolution of Controllers in Swarm- and Modular Robotics

2012-36 Denis Ssebugwawo (RUN)
Analysis and Evaluation of Collaborative Modeling Processes
2012-37 Agnes Nakakawa (RUN)
A Collaboration Process for Enterprise Architecture Creation
2012-38 Selmar Smit (VU)
Parameter Tuning and Scientific Testing in Evolutionary Algorithms
2012-39 Hassan Fatemi (UT)
Risk-aware design of value and coordination networks
2012-40 Agus Gunawan (UvT)
Information Access for SMEs in Indonesia
2012-41 Sebastian Kelle (OU)
Game Design Patterns for Learning
2012-42 Dominique Verpoorten (OU)
Reflection Amplifiers in self-regulated Learning
2012-43 Withdrawn
2012-44 Anna Tordai (VU)
On Combining Alignment Techniques
2012-45 Benedikt Kratz (UvT)
A Model and Language for Business-aware Transactions
2012-46 Simon Carter (UVA)
Exploration and Exploitation of Multilingual Data for Statistical Machine Translation
2012-47 Manos Tsagkias (UVA)
Mining Social Media: Tracking Content and Predicting Behavior
2012-48 Jorn Bakker (TUE)
Handling Abrupt Changes in Evolving Time-series Data
2012-49 Michael Kaisers (UM)
Learning against Learning - Evolutionary dynamics of reinforcement learning algorithms in strategic interactions
2012-50 Steven van Kervel (TUD)
Ontologogy driven Enterprise Information Systems Engineering
2012-51 Jeroen de Jong (TUD)
Heuristics in Dynamic Sceduling; a practical framework with a case study in elevator dispatching

2013

2013-01 Viorel Milea (EUR)
News Analytics for Financial Decision Support
2013-02 Erietta Liarou (CWI)
MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing
2013-03 Szymon Klarman (VU)
Reasoning with Contexts in Description Logics
2013-04 Chetan Yadati(TUD)
Coordinating autonomous planning and scheduling
2013-05 Dulce Pumareja (UT)
Groupware Requirements Evolutions Patterns
2013-06 Romulo Goncalves(CWI)
The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience
2013-07 Giel van Lankveld (UvT)
Quantifying Individual Player Differences

2013-08 Robbert-Jan Merk(VU)
Making enemies: cognitive modeling for opponent agents in fighter pilot simulators
2013-09 Fabio Gori (RUN)
Metagenomic Data Analysis: Computational Methods and Applications
2013-10 Jeewanie Jayasinghe Arachchige(UvT)
A Unified Modeling Framework for Service Design.
2013-11 Evangelos Pournaras(TUD)
Multi-level Reconfigurable Self-organization in Overlay Services
2013-12 Marian Razavian(VU)
Knowledge-driven Migration to Services
2013-13 Mohammad Safiri(UT)
Service Tailoring: User-centric creation of integrated IT-based homecare services to support independent living of elderly
2013-14 Jafar Tanha (UVA)
Ensemble Approaches to Semi-Supervised Learning Learning
2013-15 Daniel Hennes (UM)
Multiagent Learning - Dynamic Games and Applications
2013-16 Eric Kok (UU)
Exploring the practical benefits of argumentation in multi-agent deliberation
2013-17 Koen Kok (VU)
The PowerMatcher: Smart Coordination for the Smart Electricity Grid
2013-18 Jeroen Janssens (UvT)
Outlier Selection and One-Class Classification
2013-19 Renze Steenhuizen (TUD)
Coordinated Multi-Agent Planning and Scheduling
2013-20 Katja Hofmann (UvA)
Fast and Reliable Online Learning to Rank for Information Retrieval
2013-21 Sander Wubben (UvT)
Text-to-text generation by monolingual machine translation
2013-22 Tom Claassen (RUN)
Causal Discovery and Logic
2013-23 Patricio de Alencar Silva(UvT)
Value Activity Monitoring
2013-24 Haitham Bou Ammar (UM)
Automated Transfer in Reinforcement Learning
2013-25 Agnieszka Anna Latoszek-Berendsen (UM)
Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System
2013-26 Alireza Zarghami (UT)
Architectural Support for Dynamic Homecare Service Provisioning
2013-27 Mohammad Huq (UT)
Inference-based Framework Managing Data Provenance
2013-28 Frans van der Sluis (UT)
When Complexity becomes Interesting: An Inquiry into the Information eXperience
2013-29 Iwan de Kok (UT)
Listening Heads
2013-30 Joyce Nakatumba (TUE)
Resource-Aware Business Process Management: Analysis and Support
2013-31 Dinh Khoa Nguyen (UvT)
Blueprint Model and Language for Engineering Cloud Applications
2013-32 Kamakshi Rajagopal (OUN)

Networking For Learning; The role of Networking in a Lifelong Learner's Professional Development

2013-33 Qi Gao (TUD)
User Modeling and Personalization in the Microblogging Sphere

2013-34 Kien Tjin-Kam-Jet (UT)
Distributed Deep Web Search

2013-35 Abdallah El Ali (UvA)
Minimal Mobile Human Computer Interaction
Promotor: Prof. dr. L. Hardman (CWI/UVA)

2013-36 Than Lam Hoang (TUe)
Pattern Mining in Data Streams

2013-37 Dirk Börner (OUN)
Ambient Learning Displays

2013-38 Eelco den Heijer (VU)
Autonomous Evolutionary Art

2013-39 Joop de Jong (TUD)
A Method for Enterprise Ontology based Design of Enterprise Information Systems

2013-40 Pim Nijssen (UM)
Monte-Carlo Tree Search for Multi-Player Games

2013-41 Jochem Liem (UVA)
Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning

2013-42 Léon Planken (TUD)
Algorithms for Simple Temporal Reasoning

2013-43 Marc Bron (UVA)
Exploration and Contextualization through Interaction and Concepts

2014

2014-01 Nicola Barile (UU)
Studies in Learning Monotone Models from Data

2014-02 Fiona Tuliyano (RUN)
Combining System Dynamics with a Domain Modeling Method

2014-03 Sergio Raul Duarte Torres (UT)
Information Retrieval for Children: Search Behavior and Solutions

2014-04 Hanna Jochmann-Mannak (UT)
Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation

2014-05 Jurriaan van Reijsen (UU)
Knowledge Perspectives on Advancing Dynamic Capability

2014-06 Damian Tamburri (VU)
Supporting Networked Software Development

2014-07 Arya Adriansyah (TUE)
Aligning Observed and Modeled Behavior

2014-08 Samur Araujo (TUD)
Data Integration over Distributed and Heterogeneous Data Endpoints

2014-09 Philip Jackson (UvT)
Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language

2014-10 Ivan Salvador Razo Zapata (VU)
Service Value Networks

2014-11 Janneke van der Zwaan (TUD)
An Empathic Virtual Buddy for Social Support
2014-12 Willem van Willigen (VU)
Look Ma, No Hands: Aspects of Autonomous Vehicle Control
2014-13 Arlette van Wissen (VU)
Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains
2014-14 Yangyang Shi (TUD)
Language Models With Meta-information
2014-15 Natalya Mogles (VU)
Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare
2014-16 Krystyna Milian (VU)
Supporting trial recruitment and design by automatically interpreting eligibility criteria
2014-17 Kathrin Dentler (VU)
Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability
2014-18 Mattijs Ghijsen (VU)
Methods and Models for the Design and Study of Dynamic Agent Organizations
2014-19 Vinicius Ramos (TUE)
Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support
2014-20 Mena Habib (UT)
Named Entity Extraction and Disambiguation for Informal Text: The Missing Link
2014-21 Kassidy Clark (TUD)
Negotiation and Monitoring in Open Environments
2014-22 Marieke Peeters (UU)
Personalized Educational Games - Developing agent-supported scenario-based training
2014-23 Eleftherios Sidirourgos (UvA/CWI)
Space Efficient Indexes for the Big Data Era
2014-24 Davide Ceolin (VU)
Trusting Semi-structured Web Data
2014-25 Martijn Lappenschaar (RUN)
New network models for the analysis of disease interaction
2014-26 Tim Baarslag (TUD)
What to Bid and When to Stop
2014-27 Rui Jorge Almeida (EUR)
Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty
2014-28 Anna Chmielowiec (VU)
Decentralized k-Clique Matching
2014-29 Jaap Kabbedijk (UU)
Variability in Multi-Tenant Enterprise Software
2014-30 Peter de Cock (UvT)
Anticipating Criminal Behaviour
2014-31 Leo van Moergestel (UU)
Agent Technology in Agile Multiparallel Manufacturing and Product Support
2014-32 Naser Ayat (UvA)
On Entity Resolution in Probabilistic Data
2014-33 Tesfa Tegegne (RUN)
Service Discovery in eHealth
2014-34 Christina Manteli(VU)
The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems.
2014-35 Joost van Ooijen (UU)

Cognitive Agents in Virtual Worlds: A Middleware Design Approach

2014-36 Joos Buijs (TUE)
Flexible Evolutionary Algorithms for Mining Structured Process Models

2014-37 Maral Dadvar (UT)
Experts and Machines United Against Cyberbullying

2014-38 Danny Plass-Oude Bos (UT)
Making brain-computer interfaces better: improving usability through post-processing.

2014-39 Jasmina Maric (UvT)
Web Communities, Immigration, and Social Capital

2014-40 Walter Omona (RUN)
A Framework for Knowledge Management Using ICT in Higher Education

2014-41 Frederic Hogenboom (EUR)
Automated Detection of Financial Events in News Text

2014-42 Carsten Eijckhof (CWI/TUD)
Contextual Multidimensional Relevance Models

2014-43 Kevin Vlaanderen (UU)
Supporting Process Improvement using Method Increments

2014-44 Paulien Meesters (UvT)
Intelligent Blauw. Met als ondertitel: Intelligence-gestuurde politiezorg in gebiedsgebonden eenheden.

2014-45 Birgit Schmitz (OUN)
Mobile Games for Learning: A Pattern-Based Approach

2014-46 Ke Tao (TUD)
Social Web Data Analytics: Relevance, Redundancy, Diversity

2014-47 Shangsong Liang (UVA)
Fusion and Diversification in Information Retrieval

2015

2015-01 Niels Netten (UvA)
Machine Learning for Relevance of Information in Crisis Response

2015-02 Faiza Bukhsh (UvT)
Smart auditing: Innovative Compliance Checking in Customs Controls

2015-03 Twan van Laarhoven (RUN)
Machine learning for network data

2015-04 Howard Spoelstra (OUN)
Collaborations in Open Learning Environments

2015-05 Christoph Bösch(UT)
Cryptographically Enforced Search Pattern Hiding

2015-06 Farideh Heidari (TUD)
Business Process Quality Computation - Computing Non-Functional Requirements to Improve Business Processes

2015-07 Maria-Hendrike Peetz(UvA)
Time-Aware Online Reputation Analysis

2015-08 Jie Jiang (TUD)
Organizational Compliance: An agent-based model for designing and evaluating organizational interactions

2015-09 Randy Klaassen(UT)
HCI Perspectives on Behavior Change Support Systems

2015-10 Henry Hermans (OUN)
OpenU: design of an integrated system to support lifelong learning

2015-11 Yongming Luo(TUE)
Designing algorithms for big graph datasets: A study of computing bisimulation and joins
2015-12 Julie M. Birkholz (VU)
Modi Operandi of Social Network Dynamics: The Effect of Context on Scientific Collaboration Networks
2015-13 Giuseppe Procaccianti(VU)
Energy-Efficient Software
2015-14 Bart van Straalen (UT)
A cognitive approach to modeling bad news conversations
2015-15 Klaas Andries de Graaf (VU)
Ontology-based Software Architecture Documentation
2015-16 Changyun Wei (UT)
Cognitive Coordination for Cooperative Multi-Robot Teamwork
2015-17 André van Cleeff (UT)
Physical and Digital Security Mechanisms: Properties, Combinations and Trade-offs
2015-18 Holger Pirk (CWI)
Waste Not, Want Not! - Managing Relational Data in Asymmetric Memories
2015-19 Bernardo Tabuenca (OUN)
Ubiquitous Technology for Lifelong Learners
2015-20 Loïs Vanhée(UU)
Using Culture and Values to Support Flexible Coordination
2015-21 Sibren Fetter (OUN)
Using Peer-Support to Expand and Stabilize Online Learning
2015-22 Zhemin Zhu(UT)
Co-occurrence Rate Networks
2015-23 Luit Gazendam (VU)
Cataloguer Support in Cultural Heritage
2015-24 Richard Berendsen (UVA)
Finding People, Papers, and Posts: Vertical Search Algorithms and Evaluation
2015-25 Steven Woudenberg (UU)
Bayesian Tools for Early Disease Detection
2015-26 Alexander Hogenboom (EUR)
Sentiment Analysis of Text Guided by Semantics and Structure
2015-27 Sándor Héman (CWI)
Updating compressed column-stores
2015-28 Janet Bagorogoza (TiU)
Knowledge Management and High Performance; The Uganda Financial Institutions Model for HPO
2015-29 Hendrik Baier (UM)
Monte-Carlo Tree Search Enhancements for One-Player and Two-Player Domains
2015-30 Kiavash Bahreini (OUN)
Real-time Multimodal Emotion Recognition in E-Learning
2015-31 Yakup Koç (TUD)
On Robustness of Power Grids
2015-32 Jerome Gard (UL)
Corporate Venture Management in SMEs
2015-33 Frederik Schadd (UM)
Ontology Mapping with Auxiliary Resources
2015-34 Victor de Graaff (UT)
Geosocial Recommender Systems

2016

2016-01 Syed Saiden Abbas (RUN)
Recognition of Shapes by Humans and Machines
2016-02 Michiel Meulendijk (UU)
Optimizing medication reviews through decision support: prescribing a better pill to swallow

List of Publications

1. Meulendijk, M., Van De Wijngaert, L., Brinkkemper, S., & Leenstra, H. (2011). AmI in good care? Developing design principles for ambient intelligent domotics for elderly. Informatics for Health and Social Care, 36(2), 75-88.
2. Meulendijk, M., Spruit, M., Drenth-van Maanen, A., Numans, M., Brinkkemper, S., & Jansen, P. (2013). General practitioners' attitudes towards decision supported prescribing: an analysis of the Dutch primary care sector. Health Informatics Journal, 19(4), 247-263.
3. Meulendijk, M., Meulendijks, E., Jansen, P., Numans, M., & Spruit, M. (2014). What concerns users of medical apps? Exploring non-functional requirements of medical mobile applications. Proceedings of the European Conference on Information Systems (ECIS) 2014. Tel Aviv, Israel.
4. Meulendijk, M., Drenth-van Maanen, A., Jansen, P., Brinkkemper, S., Numans, M., & Spruit, M. (2013). Introducing the COrETeSt Feasibility Analysis in Medical Informatics: A Case Study of a Decision support Knowledge System in the Dutch Primary Care Sector. In M. M. Cruz-Cunha, I. M. Miranda, & P. Gonzalves, Handbook of Research on ICTs and Management Systems for Improving Efficiency in Healthcare and Social Care (pp. 1066-1087). IGI Global.
5. Meulendijk, M., Spruit, M., Jansen, P., Numans, M., & Brinkkemper, S. (2015). STRIPA: A Rule-Based Decision Support System for Medication Reviews in Primary Care. ECIS 2015 Research-in-Progress Papers. Paper 29. Munster, Germany.
6. Meulendijk, M., Spruit, M., Drenth-van Maanen, A., Numans, M., Brinkkemper, S., Jansen, P., et al. (2015). Computerized decision support improves medication review effectiveness: an experiment evaluating the STRIP Assistant's usability. Drugs & Aging, 32(6), 495-503.
7. Meulendijk, M., Spruit, M., Willeboordse, F., Numans, M., Brinkkemper, S., Knol, W., et al. (2015). Efficiency of clinical decision support systems improves over time (submitted).
8. Meulendijk, M., Spruit, M., Lefebvre, A., & Brinkkemper, S. (2015). To what extent can patient data be meaningfully exchanged between primary care terminologies? A case study of four Western European classification systems (submitted).
9. Meulendijk, M., Spruit, M., & Brinkkemper, S. (2015). Risk Mediation in Association Rules: The Case of Decision Support in Medication Review (submitted).

English Summary

The expected growth of the elderly population in the coming decennia presents great challenges for public health. With increasing age comes a greater incidental or continuous dependency on medicine; almost half of all drug expenditures are consumed by elderly people, and of chronically ill people in the Netherlands that use five or more drugs simultaneously, half are over sixty-five years of age [110,198].

There is ample evidence that structured medication reviews could improve polypharmacy patients' medication [57,74], but that there are major barriers impeding the direct implementation of structured medication reviews in general practice [10,157]. Clinical decision support systems have been shown to effectively improve prescribing processes in primary care, leading to the belief that they could be successfully implemented for the conduct of structured medication reviews [32]. Thus, the aim of the research described in this dissertation is to create a decision support system to facilitate the conduct of structured medication reviews by physicians and pharmacists in primary care. The main research question of this dissertation reads:

How can clinical decision support systems contribute to effective and efficient medication reviews in primary care?

The main research question is investigated following the steps of the design science research process, employing both qualitative and quantitative research methods throughout the study.

The objectives that can be identified from the perspectives of its stakeholders are explored. In chapter 2, a questionnaire distributed amongst primary care physicians shows that the majority of them are positively inclined towards using a decision support system for medication reviews. Chapter 3 documents interview sessions with polypharmacy patients, from which it appears that their main concerns with medical applications are trustability, security, and, certifiability. In chapter 4 the feasibility of a decision support system for medication reviews is assessed from conceptual, organizational, technological, economic, and societal perspectives, leading to strategies to minimize the project's risk.

With objectives having been identified, the next phase in the design science research process encompasses the design and development of the artefact. In this project, the artefact has been based on the Systematic Tool to Reduce Inappropriate Prescribing (STRIP), a structured method to perform medication reviews. To facilitate physicians' use of the STRIP method, the STRIP

Assistant (STRIPA) has been developed after the objectives identified in the previous studies. STRIPA is a stand-alone web-based decision support system that advices physicians during the pharmacotherapeutic analysis of patients' health records. In chapter 5 the application's architecture and rule engine, and the design decisions relating to the user interface and semantic interoperability, are described.

After developing the artefact, it was evaluated in both controlled environments and in practice. Chapter 6 describes the results of an experiment validating STRIPA's effectiveness and efficiency as a tool for physicians optimizing medical records of polypharmacy patients. In an online experiment, forty-two physicians were asked to optimize two comparable medical records of polypharmacy patients, one in their usual manner and one using STRIPA. Medication optimization significantly improves with the STRIP Assistant. Appropriate decisions increase from 58% without help to 76% with the STRIP Assistant. Inappropriate decisions decrease from 42% in the unassisted case to 24% in the assisted one. Participants do, however, spend significantly more time optimizing medication with the STRIP Assistant (24 minutes) than without it (13 minutes).

To determine if the system performs well in real life, and if its efficiency increases when it is used over a longer period of time, a case study was conducted. Thus, chapter 7 documents the results of a study to determine if having a group of caretakers perform decision-supported structured medication reviews over a longer period of time will lead to improvements in efficiency. Four expert teams consisting of a physician and a pharmacist conducted structured medication reviews on patients in 13 general practices located in Amsterdam, the Netherlands. During thirteen months, the expert teams performed 261 medication reviews. Analysis of the acquired data shows that the amount of time users need to perform medication reviews decreases significantly as they gain experience over time; during the first half of the study they needed 15.7 minutes on average to conduct one review, while during the second half the mean time had decreased to 10.7 minutes.

The evaluation of the STRIP Assistant has led to insights into its weaknesses and presents opportunities for refinement. Two areas for improvement of the decision support system are explored.

The diversity of terminologies used in primary care causes significant challenges regarding semantic interoperability. Attempts to address these challenges usually focus on the creation of metaterminologies, with the peculiarities of national variations of terminologies being overlooked. In chapter 8 the extent to which primary care data can be meaningfully exchanged between nationally implemented terminologies is assessed. To this end, a model comprising primary care terminologies and including axioms to define their relations is developed and evaluated.

Association rule mining is one of the most prominent knowledge discovery methods in existence. The application of discovered association rules in a sensitive domain with potentially far-reaching implications, however, can be risky. In chapter 9 the concept of risk in association

rule mining is introduced. A model for the incorporation of risk in association rules is proposed and validated in a primary care setting.

Samenvatting (Dutch Summary)

In de komende decennia zal de verwachte groei van de ouderenpopulatie tot grote uitdagingen voor de gezondheidszorg leiden. Hoge leeftijd leidt ertoe dat mensen sterker afhankelijk zijn van medicijnen. Bijna de helft van alle medicijnen wordt gebruikt door ouderen; de helft van de chronisch zieken in Nederland die meer dan vijf medicijnen gebruiken, zijn ouder dan vijfenzestig. [110,198].

Er is voldoende bewijs dat gestructureerde medicicatiebeoordelingen de medicatie van deze polyfarmaciepatiënten kan verbeteren, [57,74], maar er zijn grote opstakels die frequent gebruik van gestructureerde medicatiebeoordelingen in de huisartsenpraktijk belemmeren [10,157]. Er is aangetoond dat klinische beslissingsondersteunende systemen de voorschrijfprocessen in de huisartsenpraktijk kunnen verbeteren, wat leidt tot de aanname dat ze ook geschikt zijn voor het uitvoeren van gestructureerde medicatiebeoordelingen [32]. Het doel van het onderzoek in deze dissertatie is het creëren van een beslissingsondersteunend systeem voor het faciliteren van gestructureerde medicatiebeoordelingen door huisartsen en apothekers. De hoofdvraag van deze dissertatie luidt:

Hoe kunnen klinische beslissingsondersteunende systemen bijdragen aan effectieve en efficiënte medicatiebeoordelingen in de eerstelijnszorg?

Deze hoofdvraag wordt beantwoord via het stappenplan van het *design science research process*, waarbij zowel kwalitatieve als kwantitatieve onderzoeksmethoden toegepast worden.

Eerst worden de doelstellingen die bepaald kunnen worden vanuit het perspectief van de belanghebbenden onderzocht. In hoofdstuk 2 laat een vragenlijst zien dat de meerderheid van huisartsen positief staat tegenover het gebruik van een beslissingsondersteunend systeem voor medicatiebeoordelingen. Hoofdstuk 3 documenteert interviews met polyfarmaciepatiënten, waaruit blijkt dat hun belangrijkste zorgen rond medische applicaties betrekking hebben op vertrouwelijkheid, veiligheid, en certificeerbaarheid. In hoofdstuk 4 wordt de haalbaarheid van een beslissingsondersteunend systeem voor medicatiebeoordelingen getoetst vanuit een conceptueel, organisationeel, technologisch, economisch, en maatschappelijk perspectief, wat leidt tot strategieën om het risico van het project te minimaliseren.

Nadat de doelstellingen bepaald zijn, is de volgende stap in het *design science research process* het ontwerp en de ontwikkeling van het artefact. In dit project is het artefact gebaseerd

op de *Systematic Tool to Reduce Inappropriate Prescribing* (STRIP), een gestructureerde methode om medicatiebeoordelingen uit te voeren. Om huisartsen te faciliteren bij het gebruik van deze methode is de STRIP Assistent (STRIPA) ontwikkeld op basis van de doelstellingen die in de voorgaande studies geïdentificeerd zijn. STRIPA is een op zichzelf staand online beslissingsondersteunend systeem dat huisartsen adviseert tijdens de farmacotherapeutische analyse van de medisch dossiers van patiënten. In hoofdstuk 5 worden de architectuur van de applicatie en de implementatie van de beslisregels, evenals de beslissingen met betrekking tot de gebruikersinterface en semantische interoperabiliteit, beschreven.

Nadat het artefact is ontwikkeld, is het geëvalueerd in zowel een gecontroleerde omgeving als in de praktijk. Hoofdstuk 6 beschrijft de resultaten van een experiment waarin de effectiviteit en efficiëntie van STRIPA als applicatie voor het uitvoeren van medicatiebeoordelingen gevalideerd wordt. In een online experiment werden twee-en-veertig huisartsen gevraagd twee vergelijkbare medische dossiers te optimaliseren: één op hun gebruikelijke wijze, en één via STRIPA. De optimalisatie van de medicatie verbetert significant met STRIPA. Het aantal juiste beslissingen neemt toe van 58% zonder hulp naar 76% met de STRIP Assistent. Onjuiste beslissingen verminderen van 42% op de gebruikelijke wijze naar 24% met STRIPA. Deelnemers hebben echter significant meer tijd nodig om medicatie te optimaliseren met behulp van de STRIP Assistent (24 minuten) dan zonder (13 minuten).

Om te bepalen of het systeem ook in de praktijk goed presteert, en of de efficiëntie verbetert als het gedurende een langere periode gebruikt wordt, is een *case study* uitgevoerd. Hoofdstuk 7 documenteert de resultaten van een studie om te bepalen of het gedurende een langere periode uitvoeren van gestructureerde medicatiebeoordelingen leidt tot verbeteringen in efficiëntie. Vier expertteams, elk bestaande uit een huisarts en een apotheker, voerden gestructureerde medicatiebeoordelingen uit op de dossiers van patiënten van 13 huisartsenpraktijken in Amsterdam. Gedurende dertien maanden voerden de expertteams 261 medicatiebeoordelingen uit. Data-analyse laat zien dat de tijd die gebruikers per medicatiebeoordeling nodig hebben significant vermindert als ze ervaring opbouwen; gedurende de eerste helft van de studie hadden ze gemiddeld 15.7 minuten nodig om een beoordeling uit te voeren, terwijl gedurende de tweede helft van het onderzoek de benodigde tijd afnam naar 10.7 minuten.

De evaluatie van de STRIP Assistent heeft geleid tot inzichten in de zwaktes van het systeem en waar mogelijkheden tot verbetering zijn. Twee deelgebieden waar het beslissingsondersteunende systeem verbeterd kan worden zijn onderzocht.

De diversiteit van terminologieën in eerstelijnszorg veroorzaakt grote uitdagingen met betrekking tot semantische interoperabiliteit. Pogingen om deze uitdagingen aan te pakken spitsen zich meestal toe op het creëren van metaterminologieën, waarbij de specifieke eigenschappen van nationale variaties van terminologieën genegeerd worden. In hoofdstuk 8 wordt de mate waarin data uit eerstelijnszorg betekenisvol uitgewisseld kan worden tussen verschillende nationale terminologieën onderzocht. Een model van verschillende internationale terminologieën uit de eerstelijnszorg en hun relaties wordt ontwikkeld en geëvalueerd.

Het ontdekken van associatieregels is een van de meest vooraanstaande methoden van kennisontsluiting. Het toepassen van ontdekte associatieregels in een gevoelig domein, waarin ze potentieel verrijkende gevolgen hebben, is echter risicovol. In hoofdstuk 9 wordt risico als concept geïntroduceerd in het domein van associatieregels. Een model voor het incorporeren van risico in dit domein wordt voorgesteld en gevalideerd in de eerstelijnszorg.

Curriculum Vitae

Michiel Meulendijk was born on April 5th, 1985, in Capelle aan den IJssel, the Netherlands. He studied information sciences and business informatics at Utrecht University. He graduated in 2009 on the topic of ambient intelligent domotics, which sparked his interest in the field of medical informatics.

In 2011 he started conducting his PhD research at the Center for Organization and Information, Department of Information and Computing Sciences, Utrecht University. During this time, and in addition to his research, he conceived and taught several courses in the business informatics master programme. He currently works as a researcher and software developer in the OPERAM project, a European initiative which came forth out of his PhD work and which aims at conducting a randomized controlled trial to determine the benefit of decision-supported structured medication reviews.

He is an interpretivist in spite of himself. He would love the world to be a perfectly working system waiting to be unraveled by science, but knows better than to believe that. While he is in awe of the perfect predictability of computers, he is nonetheless fascinated by the erratic behavior of their users. The quotes included throughout the dissertation all reflect this dichotomy.