Data-driven personalized healthcare

Towards personalized interventions via reinforcement learning for Mobile Health

Alexander Galozy

LICENTIATE THESIS | Halmstad University Dissertations no. 79

Data-driven personalized healthcare – Towards personalized interventions via reinforcement learning for Mobile Health

Halmstad University Dissertations no. 79
ISBN 978-91-88749- 66-6 (printed)
ISBN 978-91-88749-67-3 (pdf)
Publisher: Halmstad University Press, 2021 | www.hh.se/hup

Abstract

Medical and technological advancement in the last century has led to the unprecedented increase of the populace's quality of life and lifespan. As a result, an ever-increasing number of people live with chronic health conditions that require long-term treatment, resulting in increased healthcare costs and managerial burden to the healthcare provider. This increase in complexity can lead to ineffective decision-making and reduce care quality for the individual while increasing costs. One promising direction to tackle these issues is the active involvement of the patient in managing their care. Particularly for chronic diseases, where ongoing support is often required, patients must understand their illness and be empowered to manage their care. With the advent of smart devices such as smartphones, it is easier than ever to provide personalised digital interventions to patients, help them manage their treatment in their daily lives, and raise awareness about their illness. If such new approaches are to succeed, scalability is necessary, and solutions are needed that can act autonomously without costly human intervention. Furthermore, solutions should exhibit adaptability to the changing circumstances of an individual patient's health, needs and goals. Through the ongoing digitisation of healthcare, we are presented with the unique opportunity to develop cost-effective and scalable solutions through Artificial Intelligence (AI).

This thesis presents work that we conducted as part of the project improving Medication Adherence through Person-Centered Care and Adaptive Interventions (iMedA) that aims to provide personalised adaptive interventions to hypertensive patients, supporting them in managing their medication regiment. The focus lies on inadequate medication adherence (MA), a pervasive issue where patients do not take their medication as instructed by their physician. The selection of individuals for intervention through secondary database analysis on Electronic Health Records (EHRs) was a key challenge and is addressed through in-depth analysis of common adherence measures, development of prediction models for MA and discussions on limitations of such approaches for analysing MA. Furthermore, providing personalised adaptive interventions is framed in the contextual bandit setting and addresses the challenge of delivering relevant interventions in environments where contextual information is significantly corrupted.

The contributions of the thesis can be summarised as follows: (1) High-

lighting the issues encountered in measuring MA through secondary database analysis and providing recommendations to address these issues, (2) Investigating machine learning models developed using EHRs for MA prediction and extraction of common refilling patterns through EHRs and (3) formal problem definition for a novel contextual bandit setting with context uncertainty commonly encountered in Mobile Health and development of an algorithm designed for such environments.

To my wife and family

Acknowledgements

I would like to thank my principal supervisor Sławomir Nowaczyk Nowaczyk for his continuing guidance during my PhD studies. He helped me to think not only critically about my work but see value where I sometimes did not, which helped me recognise my continuing growth as a researcher. I'm grateful for the many engaging discussions and ideas, sometimes of personal nature, that keep me interested in science and research.

Furthermore, I would like to thank Anita Sant'Anna for being a motivating presence in the initial phases of my PhD. Her interest and involvement in my work have led to fruitful ideas. In the same token, I would like to thank my current co-supervisor Mattias Ohlsson, for his perspective and health data related expertise, helping me in keeping practical considerations of my research in mind.

I would like to thank both Björn Avgall and Markus Lingman for being always ready to help me with my questions from the clinical side of my research. Through my project work, they provided me with an indispensable source of knowledge and expertise, helping me understand the issues and challenges health care professionals face today.

I also would like to thank all my lab colleagues that provide a friendly and relaxed research environment. Without your commitment, I do not think I would like research as much as I do! Thank you all, Denni, Slawomir, Sepideh, Mahmoud, Peyman, Ece, Pablo, Awais, Stefan, Kevin, Abdallah, Ghaith, Zahra, Shiraz, Josef, Fernando, Kunru, Tiago, and, of course, all others which have temporarily slipped my mind, due to the trying year of 2020!

I would like to thank my sister and mother for being there for me even though a significant distance may separate us. And last but certainly not least, I would like to thank my loving wife for always being there for me and providing encouragement to achieve my goals and taking care of myself. I take great solace in knowing you, and all those years with you have made me, in all, a better human being.

"I'm reaching for the random or whatever will bewilder me."

— Tool - Lateralus

List of Papers

The following papers, referred to in the text by their Roman numerals, are included in this thesis.

PAPER I: **Pitfalls of medication adherence approximation through EHR and pharmacy records: Definitions, data and computation**
Alexander Galozy, Sławomir Nowaczyk, Anita Sant'Anna, Mattias Ohlsson, Markus Lingman. **International Journal of Medical Informatics**, *published* 31 January 2020.

PAPER II: **Prediction and pattern analysis of medication refill adherence through electronic health records and dispensation data**
Alexander Galozy, Sławomir Nowaczyk. **Journal of Biomedical Informatics: X**, *published* 13 June 2020.

PAPER III: **Corrupted Contextual Bandits with Action Order Constraints**
Alexander Galozy, Sławomir Nowaczyk, Mattias Ohlsson. **Artificial Intelligence**, *submitted, under review* 16 November 2020.

Contents

List of Figures

List of Tables

1. INTRODUCTION

The most merciful thing in the world, I think, is the inability of the human mind to correlate all its contents.

— H.P. Lovecraft [3]

Significant improvements in healthcare over the last decades resulted in an increase in lifespan, and with it, an increase in the number of people that live with chronic conditions [4]. This development has lead to the allocation of a significant portion of healthcare resources for the treatment and prevention of chronic illnesses, such as hypertension, stroke, coronary artery disease and heart failure[5]. While knowledge in effective treatment of chronic conditions is ever increasing, a promising, and arguably necessary, approach to improve patient well-being and reduce cost is through self-management, empowering patients to manage their illness by educating and teaching individuals how to identify and solve problems related to their condition [4]. The increasing pervasiveness of digital technologies in modern societies provides fertile ground for the successful growth of new approaches that involve the individual patient in a more holistic healthcare framework towards person-centred-care (PCC). One such approach focuses on using the advances in mobile phone technology to deliver tailored interventions that aim to educate, remind or help patients change their habits and attitudes towards their illness to improve outcomes.

Through the 20th century, the term "Person-Centered Care" has increased in prominence as the de facto mantra of modern healthcare internationally, promising improved outcomes for patients and an increase in care satisfaction [6]. PCC has been described as "understanding the patient as a unique human being" [7] and doing away with the notion that the patient is a passive receiver of care with the sole purpose of medicine being in the diagnosis, treatment and prevention of individual diseases [8]. Through a growing body of evidence, it became clearer for healthcare providers, researchers and policymakers that the premise of PCC constitutes a cultural shift towards the full integration of patients into medical treatment, focusing on their unique needs, goals and experiences [9].

PCC has been linked to success in improving care. It has been found that family physicians and general internists that adopt a practice more along PCC resulted in lower healthcare utilization [10]. The implementation of PCC practices is linked to an increase in quality of primary care [11] and adherence to

and believes about medication has been significantly associated with PCC [12].

This thesis focuses on TeleHealth applications and Mobile Health (mHealth) in particular. The WHO defines TeleHealth as "[The] delivery of healthcare services, where patients and providers are separated by distance. TeleHealth uses information and communication technologies for the exchange of information for the diagnosis and treatment of diseases and injuries, research and evaluation, and for the continuing education of health professionals". MHealth focuses on the delivery of healthcare services and reminders through mobile phone applications and seen a significant increase of interest in the past decade, allowing the easy monitoring and exchange of individual health information at any time and anywhere. Because of the popularity of health applications, researchers have focused their efforts on using this new technology to further aid the vision of PCC, integrating information of the patient's life outside the clinical setting. This holistic view of healthcare opens many new opportunities to provide patients with needed support throughout their daily lives as well as reduce costly human involvement using Artificial Intelligence (AI). Some examples of mHealth applications in healthcare are: Monitoring patients health status [13] allowing early detection of patient deterioration [14]; Wellness applications supporting patients in leading a healthier lifestyle [15]; Applications that deliver behaviour modifications to help patients change habits to improve health outcomes [16] or applications that support patients in staying adherent to their treatment plan[17]. For a comprehensive survey on the topic of mHealth applications, we refer to [18].

While the idea is promising, the actual implementation of effective mobile interventions is challenging. From a patients perspective, intervention fatigue and intervention engagement are important processes for intervention adherence and retention [19]. Intervention fatigue is described as the emotional or cognitive overload or burden associated with treatment engagement. Patients feeling overwhelmed with the constant effort of managing their disease and subsequently become nonadherent to their treatment. Intervention engagement is defined as a "multifaceted state of motivational commitment or investment in the client role over the treatment process" [20]. Both mechanisms play an important role in adherence to interventions and point towards the need to provide individualized support to which the patient is receptive given internal (e.g., current emotional state) and context-specific (e.g., location, time-of-day) factors [19; 21].

In this thesis, we examine the phenomenon of medication nonadherence. Medication nonadherence, commonly defined as the "failure of patients to take their medication as prescribed", is a pervasive issue and significant public health concern, contributing to an increase in levels of morbidity and mortality [22]. It is estimated that the rate of medication nonadherence among people with

chronic diseases lies between 30-50%, increasing the overall cost of care due to avoidable hospitalizations [23]. Through targeted digital interventions and reminders, mHealth applications provide an effective pathway to support patients with their daily medication regiment in a more personalized manner. Many factors are important for successful interventions in mHealth. Many of the factors pertain to organization of the healthcare system, considering the success or failure of interventions depending on costs, system architecture, policies, standardization, reliability, legal aspects and many more. Patient-related factors include: usability, education, patient-provider relationship, user involvement and adherence to treatment [24]. In this work, we focus on the problem of improving a patients adherence to treatment through the use of personalized digital interventions, facilitating improved adherence targeting factors associated with nonadherence, such as: Providing education content, improving self-awareness of the patient about their disease, providing help for more effective self-management, and reducing forgetfulness through reminders [25].

For digital interventions to be most effective, it is important to identify patients in need of support technologies early during treatment. A key challenge is to estimate the level of adherence in an accurate and cost-effective manner. We aim to investigate approximate measures using secondary databases, prescription and pharmacy records, that promise the valid and cost-effective analysis of adherence to medication.

From the technological perspective, systems developed for automatic intervention support need to be robust to issues idiosyncratic to both user and mobile technology. Interest in Reinforcement Learning (RL) as a framework has been growing in recent years, showing great promise for applications in mHealth. Wide scale adoption of RL is stunted due to the sensitivity of contemporary algorithms to assumption-mismatch between environments the algorithms are designed for and the environment of mHealth. For example, when considering survey or self-report data for decision-making, systems have to deal with completeness of the records driven by the engagement level of the users[26]. Using contextual information through mobile phone sensors might be incomplete due to an unstable wireless connection. Sensory information might also be noisy, too costly to acquire or otherwise unavailable due to privacy concerns [27]. These issues can significantly delay learning such that a good intervention strategy might never be learned in a time where it matters. This highlights the need for algorithms and methods that can act under the information uncertainty commonly encountered in the domain of mHealth. We aim to define and investigate a formal problem setting for RL that captures this information uncertainty. We investigate the effect of corrupted contextual information that prohibits the effective selection of relevant interventions and how to deal with this problem for more effective learning.

1.1 Research Questions

As part of the improving Medication Adherence through Person-Centered Care and Adaptive Interventions (iMedA) project, there were several research questions that were of interest to the project. Furthermore, we investigate questions that would contribute to the larger picture of the Doctoral thesis, particularly regarding RL in mHealth. The following research questions are addressed:

- **RI: How are common measures of medication refill adherence affected by pitfalls related to medication adherence approximation through secondary database analysis and how can they be remedied?**

  Before any tailored interventions can be provided to patients, it is necessary that individuals needing treatment support are identified in a timely manner, since for many cardiovascular diseases, timeliness and effectiveness of pharmacological treatment are tightly linked, especially for secondary or tertiary prevention [28]. Current approaches aim to measure the level of Medication Adherence (MA) over the course of treatment and provide interventions as nonadherence is detected. The key challenges lie in the accurate and cost-effective measurement of adherence. Many measures of different types have been developed over the years, here we elaborate some of them.

  The most accurate measurement of medication adherence is through the measurement of metabolite concentration of the drug in body fluids. While accurate, this method is not cost-effective and places a high burden on patients. Recent developments of socalled "digital pills", indigestible sensors that allow monitoring of medication consumption, might provide a less invasive alternative, but may need more wide scale adoption and acceptance from care-providers and patients [29]. Other approaches that are more economic, but consequently less objective, include clinician-patient interviews and self-report through patient-kept diaries [30]. Given that patient might lie about their medication intake or change medication-taking behaviour due to white-coat adherence [31], these measures are potentially misleading and might not accurately reflect real medication-taking patterns. New key technologies are electronic health records and pharmacy dispensation databases, which contain records on the prescriptions and medication pickups. This has promoted the development of indirect measures to approximate medication adherence through so-called medication refill adherence. This data allows valid and cost-effective analysis of refill adherence patterns [32; 33], with the caveat that actual

medication consumption remains unknown, limiting the usefulness for MA estimation. Nevertheless, primary nonadherence to medication, that is, patients that do not fill their prescription, is readily observable given the data source is complete, and appropriate interventions can be provided earlier when detected.

From the data analysis perspective, the measurement of refill adherence using EHRs and pharmacy records is prone to pitfalls that would affect the pool of candidates for intervention significantly [34]. Pitfalls include differences in the definition of medication adherence [35; 36] and refill adherence measures [37; 38]; Handling missing, incorrect or duplicate records [39] and linkage of different data sources [34; 40] as well as data selection for analysis [41]. While efforts have been undertaken to alleviate problems associated with using real-world databases for adherence measurement [37; 38; 42–44], in-depth analysis on the underlying quality of the data and practical realities of real-world databases on how they affect refill adherence estimations is still lacking. Given that future studies may be driven by adherence estimates using EHR databases, it is vital to investigate how data related and measure related issues affect medication adherence approximation and how to alleviate them.

- **RII What is the probability of medication adherence given the available data at prescription time?**

  Interesting from a clinical perspective is the prediction of adherence to medication, i.e., allowing the care provider to estimate the probability of nonadherence to medication in the future. Training prediction models and analysing adherence patterns in EHRs and pharmacy records, would provide physicians with the possibility to intervene early.

  While adherence to treatment is partially determined by individual factors that are not contained in administrative databases, there is growing interest in finding out to what extent clinical predictors contained in EHRs might allow the prediction of adherence. Previous studies utilising sociodemographic factors, clinical factors, or purchasing information had only marginal success in predicting long term adherence to medication [43; 45–47]. While these efforts did not result in accurate prediction models, they point out the importance of analysing patients' refilling behaviours and patterns of healthcare utilization. Given the comprehensive EHR and pharmacies records available to us, we identified a research gap concerning the predictability of adherence using EHRs, focusing specifically on healthcare utilisation factors and analysing temporal patterns of refill medication adherence.

- **RIII: How can we provide adaptive and personalized intervention in mHealth using reinforcement learning?**

  Reinforcement learning is a mature and powerful framework for sequential decision making, providing an ideal fit for mHealth application due to its ability to learn intervention strategies that are individually tailored to patients. Furthermore, the capability of RL methods of lifelong learning equips autonomous agents with the ability to quickly adapt to changing circumstances of the patient, promising higher patient engagement and better health outcomes through more relevant interventions.

  While promising, there are several challenges in mHealth that make the straightforward application of contemporary RL-algorithms difficult. Some of the challenges include the requirement of a good initial policy avoiding too frequent or irrelevant interventions; assessing the usefulness of features for decision-making; robustness to failure of algorithmic assumptions and dealing with noisy or missing data [27]. In this thesis, we address the problem of missing and noisy data related to mHealth applications in the framework of RL. The research conducted in the thesis addressing this research question has just started and provides the basis for future work.

1.2 Contributions

In the following, we list the individual contributions of this thesis towards the project work and future work with the doctoral thesis:

- We have shown, through extensive comparative experimentation, how data quality influences refill adherence estimates and provided recommendations on how to remedy data quality issues (Paper I).
- We have demonstrated that minor variations in how patients pick up their medication can significantly influence refill adherence estimates and shown what common measures of adherence are particularly susceptible estimation error (Paper I).
- We have shown that including historical information of past levels of adherence improves the performance of MA prediction models for a variety of practical scenarios (Paper II).
- We show that while predictive performance is high in the context of selecting patients for intervention, using EHR and dispensation records for refill adherence prediction introduces a data bias towards patients

with high healthcare utilisation. Choosing patients for intervention based on prediction models that use these data sources is potentially unreliable (Paper II).

- We extracted common longitudinal medication pickup patterns through cluster analysis. Furthermore, we cross-correlated these patterns using simple simulation models of mediation consumption. We show that several different consumption patterns can result in similar pickup patterns, which can be potentially misleading when determining the necessity and type of interventions (Paper II).
- We formulate a new problem setting for mHealth in a RL-framework, exhibiting action order that can be exploited in case of significant context corruption (Paper III).
- We develop a meta-algorithm for this setting that shows superior empirical performance compared to state-of-the-art algorithms (Paper III).

1.3 Contribution to the Doctoral Thesis

The overall theme of the doctoral thesis is framed from the perspective of tackling the issue of providing personalised adaptive interventions in the domain of mHealth. The thesis should contribute to both knowledge and technical solutions in the form of data analysis, novel algorithms and methods. Focus is placed on the problem of providing these interventions under domain-specific constraints that need special consideration when developing such solutions. We argue that the Reinforcement Learning (RL) framework in AI exhibits a great fit for sequential decision making problems and promises to deliver state-of-the-art solutions in optimal decision-making for mHealth. While powerful, contemporary algorithmic solutions in RL face a plethora of issues that make their practical implementation often difficult. Recently, four major types of challenges have been identified [48]: (C1) Long-term influence of actions on patient behaviour, (C2) fast learning in noisy contexts, (C3) accommodation of model misspecification and non-stationary and (C4) learning policies that allow offline evaluation. While work will be carried out addressing all these challenges, efforts in the near feature will be directed towards addressing challenges C2-C4.

The licentiate contributes to the doctoral thesis in several ways. The work was carried out in the context of the project iMedA. As such, a significant portion of time as been allotted to the initial phases of the project, exploring and analysing the problem of MA from the clinical perspective and MA measurement using EHRs. We explored the feasibility of using secondary

database analysis to understand what clinical and healthcare utilisation factors are indicative and predictive of patients that are nonadherent or may become nonadherent in the future. The results of the initial work were used to select patients for the pilot and large-scale intervention study of the project and contributes knowledge and recommendations for the use of comprehensive EHRs for predictive modelling and measurement of MA. The results of this work are mainly reflected in paper I and paper II. Furthermore, the licentiate provides some initial work on addressing challenges C2 and C3 in RL for mHealth, through the development a meta-algorithm in paper III. Due to the COVID-19 pandemic, experimental validation of our proposed algorithm on real-world mHealth data was significantly delayed and could not be included in this thesis.

1.4 Ethical Approval

All studies using patient data from electronic health records had approval from the Ethics Committee in Lund (Dnr. 2018/294). Prediction studies and adherence analysis were carried out on data between 2012-2019 obtained from the Regional Healthcare Information Platform [49]. Consent from individual patients was obtained through opt-out, that is, an opportunity was given for patients request removal of their data from the analysis.

1.5 Disposition

The remainder of this thesis is organized as follows. Given the strong connection with the iMeda project, this thesis is divided into three parts. Part one, outlined in chapter 2, discusses the problems laid out in the research question RI, focusing on measurement using EHRs and pharmacy records of medication adherence. Part two, chapter 3, addresses research question RII and discusses predictive modelling in EHRs in general and for medication adherence in particular, as well as illustrates the discovery and simulation of longitudinal adherence patterns. The work presented in these two chapters has been the result of the initial phases of the iMedA project. Part three, outlined in chapter 4, addresses the research question RIII, discussing the issues of sequential decision making in mHealth setting and presents the developed solutions for addressing RL challenges C2 and C3. The work of the third part coincides with the final stages of the project that aims to deploy a solution to provide adaptive personalized interventions. The summary of the papers is presented in chapter 5. Finally, chapter 6 draws conclusions, followed by the discussion of ongoing and future research directions.

2. MEDICATION ADHERENCE MEASUREMENT IN EHRs

Medication adherence is a complex multi-dimensional phenomenon that is influenced by a multitude of societal, health system and personal factors, often unique to the individual circumstances of the patient. The WHO defines adherence as the interplay between five different factor or dimensions of adherence: Health system / healthcare team factors, social / economic factors, therapy-related factors, patient-related factors and condition-related factors. The complex problem of MA is not amendable to "one-size-fits-all" solutions and requires a targeted, individual approach taking the profile of the patients along these five dimensions into account when creating the optimal intervention or treatment plan [50].

Significant efforts have been undertaken to develop interventions targeted at nonadherent patients to support and improve outcomes [51]. The challenge is to identify patients at risk of nonadherence as early as possible to maximize interventions' effectiveness. For many cardiovascular diseases, pharmacological treatment's timeliness and effectiveness are tightly linked, especially for secondary or tertiary prevention [28]. The cost-effective and less intrusive measurement of MA remains a key challenge. This thesis investigates indirect measures that approximate medication adherence through medication refill adherence using proxy information such as pharmacy dispensation data. This proxy information provides a low cost, low burden solution, with the caveat that actual medication consumption remains unknown [32; 33]. Many medication refill adherence measures, henceforth called refill adherence, have been developed, with the Medication Possession Ratio (MPR) and Proportion of Days Covered (PDC) being the most popular ones.

MPR is one of the most commonly used adherence measures. MPR is simply computed as the ratio between the number of dispensed pills over the total number of pills to be dispensed in measurement window:

$$MPR = \frac{\#dispensed\ pills}{\#total\ pills} \tag{2.1}$$

The MPR belongs to the Continuous Measures of medication Acquisition (CMA), since the timeliness of dispensations are not considered, only the total

dispensed supply.

The PDC measure is computed as the ratio of the number of days theoretically covered by medication over the total number of days in the measurement window. A common definition is "percentage of days covered by medication" [52; 53]:

$$PDC = 1 - \frac{\#gap\ days}{measure\ window} \tag{2.2}$$

Where $\#gaps\ days$ refers to days without medication and *measure window* is the number of days over which MA is measured.

The PDC measure belongs to the Continuous Measures of medication Gaps (CMG). CMG measures take potential gaps in medication coverage into account and therefore consider the timeliness of medication dispensations.

From the data analysis perspective, evaluation of refill adherence using administrative databases is prone to methodological pitfalls, affecting the resulting adherence values significantly enough to warrant special consideration [34]. Through our studies, we have identified several data and measure-related pitfalls that are of interest. Before discussing the pitfalls in more detail, we first define the adherence measures in the family of CMA and CMG.

Authors in [42] defined variations of CMA and CMG, denoted as CMA1-8. The first four CMA measures do not consider supply gaps and operate on the dispensed supply only being most similar to the MPR (or "MPR-like" measures). They do not explicitly include oversupply in their definitions, unlike the definitions in this thesis. The last four CMAs (5-8) belong to the group that compute supply gaps in the measurement window, constituting variations of the PDC measure (or "PDC-like" measures) [43]. Some of those measures consider oversupply in their definition, although it is entirely possible to include oversupply calculations for all measures. Oversupply refers to the amount of medication supply available at the beginning of the prescriptions, which can be an essential factor for accurately estimating refill adherence. Patients may delay their pickups if supply is available and would be falsely considered nonadherent if oversupply is omitted.

2.1 Medication Adherence Measures

Through work conducted in Paper I, we developed alternative definitions of the CMA1-8 measures previously defined that account for oversupply. Before listing the measures, we define different sets of data required for computation. We define the adherence measures on a per-prescription basis, that is, adherence is measured for single prescriptions. The following quantities are required:

- Prescribed daily dose z in terms of the number of pills per day.
- A set of dispensation dates $T = \{t_1, t_2, \ldots, t_N\}$.
- Dispensation quantities $Q = \{q_0, q_1, q_2, \ldots, q_N\}$. Each dispensation date in T has a corresponding dispensation quantity in Q. q_0 is the supply from a previous prescription available at the start of the measurement window.
- s and e are the start date and the end date of a prescription, respectively.

Additionally, we define important quantities that are relevant for CMG measures. the *initiation gap*, is the time interval (period) $g = t_1 - s$, between the prescription start date s and first dispensation t_1. Correspondingly, the *terminal gap* is the time interval $h = e - t_N$, between the last dispensation t_N and prescription end date e. With these two gaps defined, we define the *supply gaps* in the initiation gap

$$dQ_g = \max\left(g - \frac{q_0}{z}, 0\right), \tag{2.3}$$

and terminal gap

$$dQ_h = \max\left(h - \frac{q_N + dq_{N-1}}{z}, 0\right), \tag{2.4}$$

respectively. The quantity dq_{N-1} refers to oversupply accumulated before the last dispensation in set Q. dq measures "stockpiling" where patients pickup medication before exhausting their available supply. It is computed recursively as: $dq_n = dq_{n-1} + \max\left[q_{n-1} - \left(t_n - t_{n-1}\right) \cdot z, 0\right], n > 1$. Note that at t_1: $dq_0 = q_0$, i.e., the oversupply available at the first dispensation. Figure 2.1 shows an example dispensation pattern illustrating the different gaps and periods.

With the important quantities defined, we present the developed operational definitions of the measures investigated. CMA1 is defined as:

$$\text{CMA1} = \frac{\sum_{n=1}^{N-1} \frac{q_n}{z} + \frac{q_0}{z}}{max\left(t_N - t_1, 1\right)}.$$

Zero or negative measurement windows in the denominator are avoided through the *max* operator. At least two dispensations are required for computation and the supply of the last dispensation is ignored, since it lies outside the

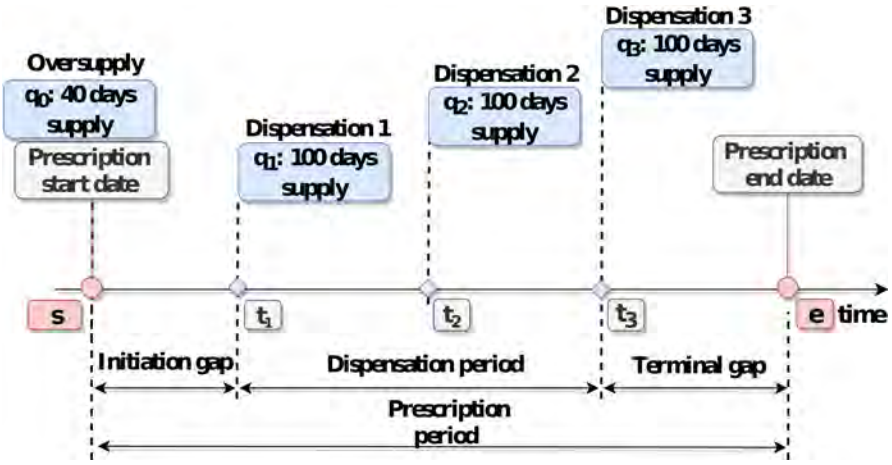

Figure 2.1: Measurement window and gaps for an example dispensation pattern. From [1]

measurement window.
CMA2 is defined as:

$$\text{CMA2} = \frac{\sum_{n=1}^{N} \frac{q_n}{z} + \frac{q_0}{z}}{max\big(e - t_1, 1\big)}.$$

CMA2 measures adherence between the first dispensation and the end of the prescription period e, including all dispensations. CMA3 and CMA4 are defined as: $\text{CMA3} = \min\big(\text{CMA1}, 1\big)$ and $\text{CMA4} = \min\big(\text{CMA2}, 1\big)$. Both CMA3 and CMA4 cap values at 100% and are otherwise computed like CMA1 and CMA2, respectively.
CMA5 is defined as:

$$\text{CMA5} = 1 - \frac{\sum_{n=1}^{N-1} \max\left[\big(t_{n+1} - t_n\big) - \frac{q_n + dq_n}{z}, 0\right]}{max\big(t_N - t_1, 1\big)}, \tag{2.5}$$

The *max* operator avoids "negative-gaps" which would compensate earlier gaps retroactively. Like CMA1, this measure cannot be computed with less than two dispensations. Going forward, we refer to the numerator of equation (2.5) as

$$\delta_q = \sum_{n=1}^{N-1} \max\left[\big(t_{n+1} - t_n\big) - \frac{q_n + dq_n}{z}, 0\right],$$

which is computed the same way for the following measures.

CMA6 is defined as:

$$\text{CMA6} = 1 - \frac{dQ_h + \delta_q}{max\big(e - t_1, 1\big)}$$

CMA6 includes the terminal gap, counting covered days from first dispensation t_1 to the prescription end date e.
CMA7 is defined as:

$$\text{CMA7} = 1 - \frac{dQ_g + dQ_h + \delta_q}{e - s}.$$

The measurement window for CMA7 coincides with the prescription period, between s and e. Note that here dq_0 (see equation (2.5)) is defined as $dq_0 = \max\big[q_0 - g \cdot z, 0\big]$, i.e., the supply carried over into the first dispensation is the oversupply minus the supply consumed in the initiation gap.
CMA8 is defined as:

$$\text{CMA8} = 1 - \frac{dQ_g + dQ_h + \delta_q}{max\big(e - s^*, 1\big)}.$$

Like CMA7, CMA8 counts covered days between s and e. Additionally, CMA8 introduces a "lag-period" equal to q_0/z, postponing the measurement window from s to $s^* = s + q_0/z$.

The original definition by authors in [42] excludes oversupply from the nominator. The definition provided here automatically exclude oversupply since accruing supply-gaps before s^* is not possible as the oversupply is exhausted at this point and cannot affect computations.

2.2 Measure-Related Pitfalls

In this section, we discuss the findings of the studies conducted in Paper I. We consider measure-related pitfalls that can significantly affect how one might interpret the level of adherence and select patients for intervention.

Inconsistent operational definitions As a natural consequence of needing ways to quantify MA, many measures have been developed over the years. These measures are often named similarly, leading to confusion whenever MA values need to be compared between studies. The commonly used MPR measure has been defined in four different ways [37] previously. Similarly, the PDC measure has seen some alternative definitions as well, for example, as "MPR but capped at 100%" [33]. This deviates significantly from what is commonly understood by PDC. Capping MPR at 100% ignores the fact that

significant gaps between medication pickups will not be detectable since MPR only considers accumulated supply. Another definitions that are similar to the definition of PDC is "MPR but surplus medication at the end of the study period excluded" [37]. While unconventional, this definition is a somewhat simpler reformulation of the computational rule for PDC. Avoiding the mixture of nomenclature for common definition such as MPR and PDC is preferable since misinterpretation can significantly impact resulting adherence values.

Inclusion or exclusion of supply available from previous prescriptions For prospective MA studies, it is known who does and does not have an available supply initially. When using secondary database analysis, the factor of oversupply plays a significant role in measuring refill adherence accurately. In general, including oversupply may mask nonadherence patterns in MPR-like measures that do not consider gaps in medication supply. In this case, oversupply may compensate gaps that would have otherwise been detected. On the other hand, excluding oversupply may result in lower refill adherence estimates when patients delay their next refill due to adequate available supply. In any case, for retrospective database analysis and long-term adherence monitoring, exclusion of oversupply can skew refill adherence values, not only on an individual level but the effects may be significant enough to show reduced adherence on a population level. Particularly when considering patients with chronic illnesses, oversupply might likely be available. Undue targeting of such patients for intervention may not only be inefficient from a resource management perspective, but it might also increase patient dissatisfaction and negatively influence the relationship between care-provider and patient.

Measurement windows and gaps Similar to oversupply calculations, for prospective studies that measure MA, the window of observation can be defined beforehand and is usually fixed for all patients. In secondary database analysis, the observation window might be defined depending on the available data source. For instance, the level of adherence can be measured between the first and last dispensation if other information about the prescription is not available, commonly encountered when dispensation records are the only available source. Inadequate refill adherence in the initiation gap is necessarily ignored in this case and may lead to overestimation of refill adherence.

Similarly, ignoring the terminal gap exacerbates overestimation further [42]. If prescription records are available, the measurement window might often be defined as the prescription period. One can choose to include or exclude the initiation and terminal gap from refill adherence computation.

Truncation caused by the operationalisation of adherence measures Refill adherence measures that define their measurement window through dispensations are particularly prone to exclude patients, skewing refill adherence estimates for the whole population to higher values. For instance, using the definition of CMA1, refill adherence for patients who have less than two dispensations cannot be computed for this measure as the measurement window is undefined. Naturally, it is always possible to define a default value of 0 for patients with no dispensations, but the issue remains for patients with one dispensation. Furthermore, special cases for computation might also be necessary for patients with enough oversupply to cover the whole prescription, not needing to dispense medication at all. The number of patients excluded from the analysis due to the sub-optimal choice of measure can be quite significant and warrants careful consideration.

Accumulated supply vs medication supply gaps There is a multitude of ways to compute refill adherence by processing dispensation information contained in pharmacy records. As mentioned in the previous section, measures have been developed that process dispensations in two conceptually different ways. The first approach is operationalised in MPR-like measures, where dispensed supply is accumulated and divided by the total supply. These types of measures are particularly susceptible to overestimation of adherence. If the measurement window is defined by dispensations, patients who refill their prescription early, that is, the supply from the previous dispensation is not yet exhausted, the measurement window might be significantly shorter than the supply would theoretically last, inflating adherence values. A similar scenario might occur for patients with late first dispensations and subsequent early follow up dispensations. For a significant portion of the prescription, the patient has no supply available, but this is not reflected by the measure, resulting in an overestimation of adherence.

PDC-like measures that defined their measurement window through dispensations suffer from the same adherence overestimation issues as MPR-like measures but skew the results much less since adherence values cannot exceed 100%.

2.3 Common Data Related Pitfalls

In this section, we discuss our findings regarding data related pitfalls often encountered in real-world databases where entries are missing, duplicated or implausible that must be corrected before refill adherence can be properly computed. Simple approaches, such as removing missing or incomplete data,

might be reasonable if enough data is available. Specific to adherence analysis, merely removing entries is incorrect. Patients with incomplete records should be removed from the analysis entirely. Otherwise, underestimation of adherence for these individuals may occur to a significant degree [54]. Furthermore, imputation of missing or incorrect values is not as straight forward as using the mean or most common value. The assumption of data being missing at random or missing at completely random is often not satisfied in modern EHR databases, that is, missing values might be a result of care provider policy or practical operations.

Preparing the data for accurate refill adherence computation is the goal. Therefore in paper I, we focused on data that is relevant for adherence computation. We concentrate on **Prescribed daily dose**: The number of pills to be taken each day. **Dispensed pills**: The number of pills dispensed at each dispensation and **Prescription period**: The period the medication should be taken for. While dispensed pills are relevant for computation, this quantity is often recorded in an automatic fashion through modern pharmacy dispensation logging systems. Therefore, we mainly focus on data correction and imputation concerning the prescription period and prescribed daily dose.

Prescribed daily dose All adherence measures require some way of estimating the supply available to the patient, be this through the number of pills or indirectly through the number of days the supply would last. In either case, the prescribed daily dose plays an important role, where missing or incorrect values can significantly affect resulting adherence values since values are usually in the single digits (often 1-3 pills a day). Knowing this can help create simple heuristics to impute missing and find incorrect prescribed daily doses by observing the number of dispensed pills throughout the prescription. For instance, some patients receive more supply per dispensation than is indicated by the prescribed daily dose, consequently inflating adherence estimates. This observation allows for the adjustment of the prescribed daily dose.

Prescription periods A common occurrence is missing prescription end dates in databases. End dates are missing until a new prescription is issued, resulting in missing entries if the prescription has not been renewed or replaced. Using the average or most common prescription length would be the wrong approach in this case, and the imputation strategy will affect measures differently, described next. Measures that define their measurement window through dispensations are not affected by missing end dates, in contrast to measures considering adherence in the prescription period, defined by prescription start and end dates. Depending on when prescription end dates are missing, i.e., for the most recent prescription or an older prescription, the strategy of imputation

might change significantly. For example, the healthcare system policies in Halland, Sweden restrict the maximum validity of a prescription to one year, allowing the imputation of missing end dates for the most recent prescription in many cases.

Exceptions need to be made in cases where treatment is discontinued earlier, but it might be impossible to provide an accurate end date without additional information. The start date of a new prescription might be a natural end date for older prescriptions. While convenient and straightforward, the application of these heuristics is not as simple in some cases, particularly for complicated prescription schemes with multiple medications and changing medication plans during treatment, e.g., when patients experience adverse reactions to specific drugs. Extensive domain knowledge might be required to identify what combination of older prescription and new prescription constitute the current treatment plan allowing the adjustment of prescription periods.

Duplications Given that data entry in EHRs, and particularly for prescription records, is done mostly in a manual fashion, wrong inputs might not be deleted or corrected right away, leaving behind a duplicated entry [42]. Naturally, dispensations need to be assigned uniquely to a prescription, causing the patient to be nonadherent to duplicate prescriptions since these do not "receive" a dispensation. The effect of duplications on adherence estimates may manifest differently depending on the measure employed and computation strategy used for multi-drug therapy. Measures defining the measurement window through dispensations are not affected by duplication since they are not computable without provisions to handle such cases. Duplications can lead to significant downward bias when using other measures, particularly in the case of adherence estimation to multi-drug therapy. In most cases, the identification of duplicate records is relatively straightforward since multiple entries of the same prescribed drug on the same day are easily recognisable and can be excluded from consideration.

3. PREDICTIVE MODELING AND LONGITUDINAL PATTERN EXTRACTION IN EHRs

This chapter shows the process of using administrative databases in healthcare for prediction and pattern mining. As the focus of this thesis is medication adherence through the iMedA project, we describe and exemplify the concepts on the analysis of MA. Nonetheless, the concepts are general enough to be valid in other contexts surrounding data analysis using EHRs.

3.1 Predictive Modelling

EHRs are inherently retrospective data sources, that is, they contain data collected in the contexts of routine clinical operations and do not contain data from controlled clinical trials. As such, they are primarily used in retrospective studies to analyse and predict patient outcomes. In general, a prediction model has three dimensions that need to be addressed by the researcher:

- The outcome data or dependent variable: the outcome that is to be predicted, for example, refill adherence over one year.
- The covariates or predictor data: variables or features that are associate with the outcome in some statistical and sometimes causal manner.
- The statistical model: a mathematical function that maps features to the outcome of interest by representing the relationship between them. The mapping is learned directly from data, as is the case with many machine learning algorithms, for example, Deep Neural Networks, Decision Trees, Logistic Regression, etc.

Outcome Variable The outcome variable is directly extracted or purposefully constructed from EHRs. Direct extraction is done through conditional

statements, for example, if primary medication adherence[1] is the outcome of interest, we might define the target variable in the following way *if prescription date - dispensation date < 30 days; target = 1 (adherent) else target=0*. Some outcomes are not contained natively in EHRs and thus are not directly accessible through conditional statements but may require post-processing after extraction. For instance, refill medication adherence might be defined on individual prescriptions with varying prescription lengths. In this thesis, we define the outcome variable through external measures of adherence, mainly those described in chapter 2.

Features EHRs are usually comprised of information pertaining to all care visits of a patient through time. For each visit, they contain clinical and demographic information. For example, a patient might receive a prescription at a primary care unit. The visit would contain patient and care provider demographic information such as patient's age, gender, primary care unit, prescriber's age and more. The clinical information for a visit is represented in the form of codes representing different clinical concepts such as diagnosis, performed procedures, prescribed medications, lab test and vital signs. Furthermore, each visit entry may contain examination information as free text input.

While the visit data can be used directly, the resulting dimensionality of the data representation is usually very high, and thus result in inefficient learning of statistical models due to the curse of dimensionality [55]. One method for dealing with high dimensional data from EHRs relies on computing features using expert knowledge to summarise or condense the high dimensional representation in human-understandable concepts. These features are then used as input to ML models to predict the desired outcome. Another approach uses the power of modern neural networks to learn so-called machine derived features [56]. While appealing from the perspective of avoiding time-consuming feature-engineering, these representations are often difficult to interpret by a human. In this thesis, we mainly consider human-derived features. We leave the investigation of learned representations to future work.

3.2 Refill Adherence Prediction

Part of the purpose of the study conducted in Paper II was to investigate the predictability of adherence as measured by the PDC (CMA7) via machine learning algorithms. Patients that have a high probability of nonadherence could be considered for early intervention to mitigate the potential impact of long term effects of uncontrolled hypertension [57; 58].

[1] medication is picked up at the pharmacy within a set time window

The cohort consists of patients between 18 and 90 years of age with an essential hypertension diagnosis (ICD10 code i10-). There are a variety of clinically relevant outcomes pertaining to adherence we can investigate for the prediction study. The adherence literature identified two types of adherence:

- *Primary Adherence*

  Patients that fill their first prescription are called *primary adherent*. Treatment initiation is the first important step and one of the instances where refill adherence directly maps to real medication adherence. If the patient did not pick up their medication, we could be somewhat sure that they will not take them, cases aside where the patient might procure the medication through other means. Prediction whether patients will fill their first prescription or not might provide care-providers with the ability to intervene early and address the reasons for primary nonadherence.

- *Secondary Adherence*

  Patients that take their medication as prescribed by the physician are **secondary adherent**. Here we see already a difference compared to being primary adherence. It is not enough to fill the medication, but it has to be taken as prescribed. In the context of refill adherence, patients would fill and refill their prescription regularly and on time. Naturally, the information on whether patients take their medication or not is not available in secondary database analysis. Still, irregular or a refill stop might indicate patterns of real nonadherence. Like primary nonadherence, predicting whether or not patients will inadequately fill their prescription can help to facilitate timely investigations into the reasons for secondary nonadherence.

We predict refill adherence for one-year prescriptions, a common prescription length for patients with chronic hypertension. This naturally excludes patients with shorter prescriptions, where continued pharmacological treatment is not the primary focus. Patterns of nonadherence in this patient subgroup might not be as crucial as for patients that need drug intervention to stay healthy. The prediction problem is framed as a classification problem, where patients above 80% PDC are adherent (1) and nonadherent (0) otherwise.

3.2.1 Data Representation

The features used in this study come from relevant literature on medication adherence. It must be noted that the reasons for nonadherence to medication can be manifold; many of them are not contained in EHRs. This limitation

Table 3.1: Predictors and their description. For the scenario considering the last five prescription (history), two additional predictors are included.

Predictor	Description
PatientAge	Patient Age at Prescription
PatientGender	Patient Gender
NumPrepYear	Number of prescription in particular year
NumDrugClass	Number of different drug classes
DrugYear	Years since first time prescription
PresQuantity	Dose prescribed
NumPolypharm	Number of concurrent prescriptions
OverSupply	Assumed available supply for prescription
PrescriberAge	Age of Prescriber at time of prescription
NumOutvisits	Number of outpatient visits
NumInvisits	Number of inpatient visits
NumEMvisits	Number of emergency visits
DiffOutvisits	Number of outpatient visits since last prescription
DiffInvisits	Number of inpatient visits since last prescription
DiffEMvisits	Number of emergency visits since last prescription
Additional predictors when considering history	
DistToPrev	Distance (days) to previous prescription
PDC	Adherence to previous prescription

makes it significantly harder to predict refill adherence with any specificity. While demographic factors are available, important patient-related factors or behavioural factors such as “stress”, “being busy”, “healthcare satisfaction” or “treatment burden” are not directly available. Some of these individual factors can be approximated to some extent through healthcare utilisation patterns in the form of visits to primary care centres, hospitals and emergency rooms. Treatment burden might be approximated by the number of concurrent prescriptions and drug variety. The predictors and their descriptions for the prediction study are shown in table 3.1.

Model evaluation is commonly carried out under the assumption that samples are independent and identically distributed. Evaluation schemes like random training test splits or k-fold cross-validation are commonly used to evaluate model generalisation on unseen data in these cases. In a realistic deployment scenario using longitudinal data, models need to predict on data collected some time *after* model deployment. The new data might exhibit some form of *dataset shift* defined as “cases where the joint distribution of inputs and outputs differs between training and test stage” [59], leading to significant worse model performance in many cases. While we do not investigate strategies

to mitigate dataset shift in this thesis, we analyse model performance using different realistic data-splitting strategies. We investigate four different splits:

Stratified random split Training data is sampled randomly from the data, stratified by class distribution. This strategy assumes that samples in both training and test sets are independent and identically distributed (iid), i.e., there is no change in distribution between old and new data. This strategy is one of the most commonly applied, especially in combination with random k-fold cross-validation schemes, to estimate the models' performance on unseen data.

Splitting by patient Patients are randomly divided into training and test sets. The patient set in training and test set are disjoint, that is: $P_{train} \cap P_{test} = \varnothing$. Each dataset contains unique patents and their prescriptions, mimicking the situation where new patients need to be classified, assuming that refilling behaviour can be inferred from other patients' behaviour.

We investigate two different approaches for splitting the data based on time:

Forward prediction: Individual split by latest prescription the latest prescription of each patient is contained in the test set. Patients with less than two prescriptions are split randomly into training and test set. This prediction scenario might be employed in cases where the model is continuously retrained as new data becomes available. We expect higher model performance due to lower refill adherence variability for patients with more extended treatment history.

Forward prediction: Most recent prescription We reserve 15% of the most recent prescriptions for the test set. This scenario is likely to occur in practice after model deployment, with older data being used for model training and evaluation. Good performance is achieved if the trained model can generalise to new and old patients alike. This task can be significantly more challenging due to including new patients that lack an adequate history of treatment while also exhibiting a potently more volatile refilling behaviour making accurate predictions more difficult.

3.2.2 Class Imbalance

There is a slight class imbalance issue since most patients (approx. 80%) are refill adherent to medication. The problem of class-imbalance has been studied extensively in the literature. There principally two ways of tackling the issue. The first approach aims to augment the data such that the number of positive and

negative instances balances. Resampling techniques can do this. For instance, the minority class can be oversampled, duplicating entries of the minority class. This approach might work well if the instances are diverse enough but can often lead to significant over-fitting to the minority class instances resulting in low generalisation performance on unseen data. Undersampling the majority class might remove important instances that characterise certain parts of the feature distribution, depending on the level of class imbalance. Thus, it can become significantly more challenging for the model to differentiate between the two classes resulting in higher generalisation error and lower robustness to noise. Other data augmentation techniques aim to reduce class imbalance by synthesising new instances of the minority class. A famous example is the Synthetic Minority Over-sampling Technique (SMOTE) [60]. SMOTE generates new minority class instances by interpolating features between two minority class instances. The combination of SMOTE and undersampling of the majority class shows better performance than pure undersampling alone. Other approaches use the advances in deep adversarial neural networks to model high dimensional probability distribution that allows the controlled generation of new instances of the minority class [61].

The second approach aims to include some form of regularisation in the model, such that prediction is shifted towards the minority class, avoiding significant overfitting of the model to the majority class. In this thesis, we account for the class imbalance by stronger penalisation of model mistakes on the minority class during training. The penalty is inversely proportional to class frequency, that is, classification mistakes on the minority class (nonadherent) are penalised about five times higher during training, shifting prediction performance in favour of the less frequent class. We employ a variety of different machine learning algorithms that include the penalty terms differently during training. For instance, in Random Forest learning, the penalty is used in weighting the splitting criterion, and the final prediction is made using weighted majority voting [62], and in logistic regression, the cost function is penalising mistakes on the minority class higher during optimisation.

3.3 Longitudinal Refill Adherence Pattern Mining and Simulation

The temporal nature of EHRs opens the possibility of tracking and analysing a patient's journey through the healthcare system, allowing the discovery of patterns that indicate the need for corrective intervention. While predicting adherence might provide a way of selecting patients for intervention, we have discovered that models developed on comprehensive EHRs may be unreliable

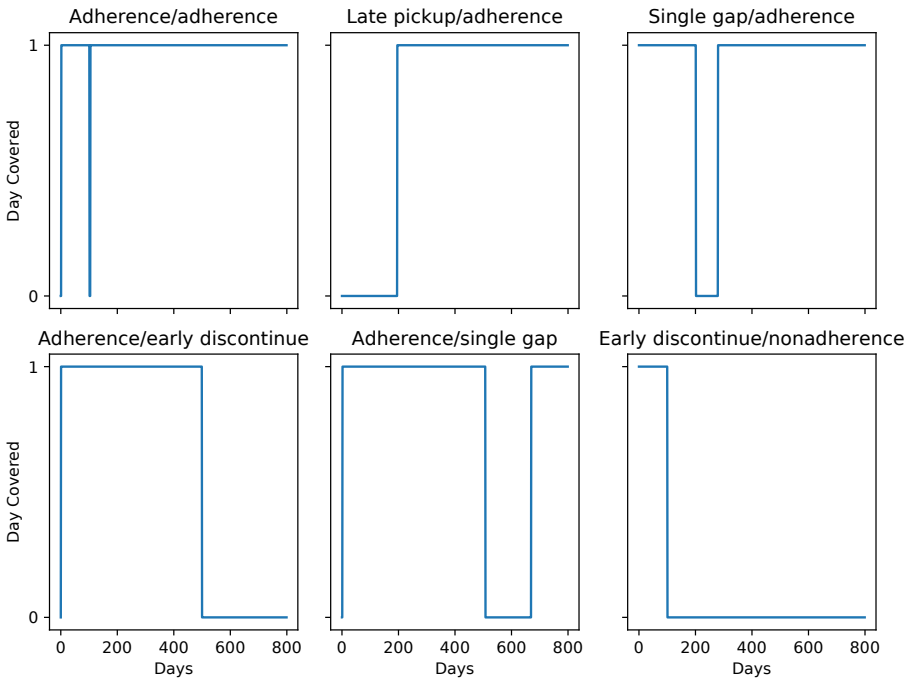

Figure 3.1: Six most common two-year PDC-patterns of patients with index prescription. Four out of six patterns show continuous refilling after the first year, while the other two patterns are showing discontinuation after the first refill or first year. From [2].

for patients without treatment history or sudden changes in adherence. Furthermore, the one-dimensional analysis of adherence based on adherence measures and fixed thresholds leaves out the importance of specific longitudinal patterns of adherence that might be clinically relevant [63].

3.3.1 Discovery of Common Dispensation Patterns

This work specifically focused on discovering typical or common medication refilling patterns, providing baselines for our simulation study later. We primarily focused on patients with monotherapy, that is, single subsequent prescriptions, and extract dispensation patterns from the records represented as binary vectors indicating assumed availability (1) or nonavailability of medication (0). Let $\bar{X} = \{X_1, X_2, \ldots, X_N\}$ be a set of dispensation patterns of n patients with monodrug therapy. We partition $\bar{X}$ int k clusters with $(C_1, C_2, \ldots, C_i), i \in \{1, 2, \ldots, k\}$, representing the *centroid* of cluster i. The *centroid* is the average pattern of all patterns in a cluster.

For clustering, we use the K-means algorithm [64]. Figure 3.1 shows the six

most common PDC-patterns (*central patterns*) discovered in the dispensation records.

3.3.2 Simulation of Longitudinal Medication Consumption Patterns

While the measurement of medication adherence through secondary database analysis provides valuable insights into patients refilling habits, it does not provide any direct insight into if and how patients take their medication. A patient can be perfectly refill adherent, but the medication is not consumed as prescribed or not consumed at all. Nonetheless, it is reasonable to assume and corroborated by previous research [65], that patients that are nonadherent to medication, e.g., discontinue treatment or do not take them as instructed, eventually stop refilling or tend to refill their medication irregularly. Given this link, we might reason about possible consumption patterns that could explain observed dispensation patterns and rank them in terms of "severeness". For example, it has been shown that a sudden stop and restart in medication-taking for anti-hypertensive treatment can be dangerous, leading to spikes in blood pressure [66].

To this end, we investigate several simulation models of patients exhibiting different hypothetical consumption patterns. Reasons for inadequate adherence are manifold. We opted for simple models where the underlying mechanisms that govern medication consumption are "forgetting", that is, patients forget to take their medication at random times, and the concept of "motivation", i.e., the desire or drive of the patient to abide by their regiment. In essence, motivation approximates the patient's general state that drives medication consumption, "summarizing" a multitude of factors, such as stress level, or emotion level. The general model of medication consumption is governed by combining the outcomes of two i.i.d. Bernoulli trials:

$$P(X_c = 1) = P(X_1 = 1) \cup P(X_2 = 1) \tag{3.1}$$

With X_c being a binary random variable indicating medication consumption. The binary random variables X_1 and X_2 are distributed according to:

$$X_1 \sim B(1, M), X_2 \sim B(1, 1 - c_{frob})$$

With $M \in [0,1]$ being the level of motivation expressed as a probability and $c_{frob} \in [0,1]$ being the probability of forgetting to take medication.

While the level of forgetting is fixed, i.e., the level of forgetfulness is somewhat stable, motivation is more dynamic and can fluctuate significantly over some time and may be influenced by environmental factors, such as yearly follow-ups to renew prescriptions leading to temporary white-coat adherence.

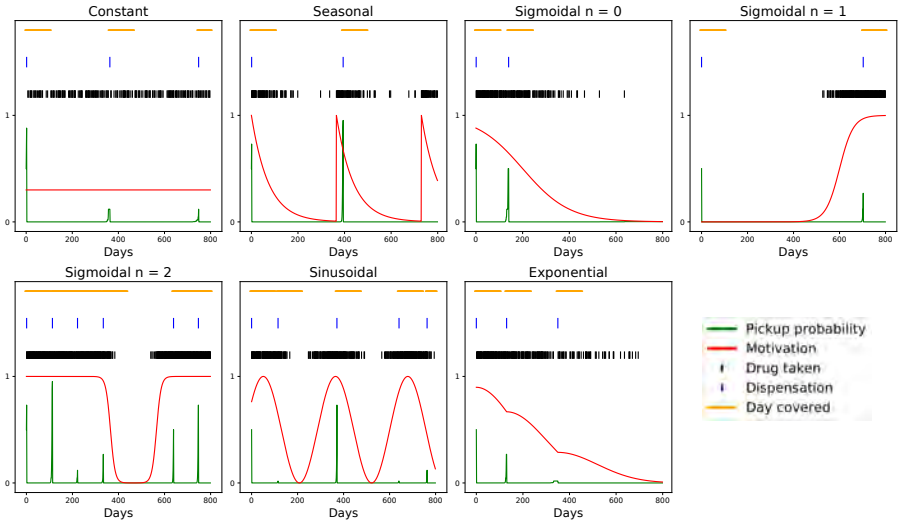

Figure 3.2: Medication-taking (Drug taken), refill (dispensation) and PDC-patterns (day covered) generated by the proposed simulation models. From [2]

Figure 3.2 illustrates simulated medication-taking, refilling and PDC-patterns. All of the common PDC-patterns, shown in figure 3.1 can be reproduced by the simulation models. Any consumption pattern that does not exhaust the initial supply within the two years is compatible with the dispensation pattern of early discontinuation in the first year. Good and near-perfect refill adherence is only possible if consumption patterns are also near perfect, that is, according to the simulation models, perfect refilling in combination with medication nonadherence is unlikely. Furthermore, several different consumption patterns can reproduce similar dispensation patterns. For example, the constant and sinusoidal model can reproduce patterns with significant gaps between dispensations. These two models exhibit significantly different, and clinically relevant, consumption behaviour that might need different types of interventions.

Nevertheless, while the levels of actual medication use are not uniquely retrievable from dispensation records, the potential link between refilling and medication consumption can help us reason about medication consumption behind observed dispensations patterns and devise interventions.

4. SEQUENTIAL DECISION MAKING FOR MOBILE HEALTH

4.1 Sequential Decision Making

Sequential Decision Making is a formalism that allows the modelling of processes that require decisions or "actions" to be made within an evolving environment to achieve a specific goal. These decisions may or may not influence the state of the environment. The agent needs to be aware of what actions lead to desirable states while avoiding actions that lead to undesirable states. Many, if not all, decisions humans do through their daily lives can be viewed as sequential decision-making processes. For example, buying and selling stocks at the stock market with the goal of maximizing short- or long-term profit.

Particularly interesting for the theme of this thesis, we could imagine a hospital setting, where the physician needs to decide on the treatment plan of the patient. Choosing a particular step in the treatment plan will affect the health state of the patient and thus determine the health outcome positively or negatively. The physician needs to be aware of each step's risk and benefits, personalised to the treated individual, and decide appropriately. In a mHealth setting, providing automated and personalised interventions is key to keeping the patient's interest and engagement. The decision-maker or "agent" needs to decide what series of digital intervention is interesting to the individual patient while considering the long-term effects on a performance metric, such as medication adherence level or blood pressure.

Given the ubiquitous nature of decision processes, significant focus has been placed on sequential decision-making to devise methods and algorithms to find optimal or near-optimal decisions automatically. Great success has been achieved in complex and challenging environments, particularly in the domain of competitive games, such as Go [67], Starcraft [68], DotA 2 [69]. These environments are often unpredictable, requiring the player to perform a mixture of short term and long-term strategies under incomplete information to be successful.

In this thesis, we focus on the bandit formulation, specifically *contextual*

bandits, of RL. In this simplified setting, the agent is only concerned with maximising cumulative *immediate* reward. The agent chooses the intervention that would result in the highest immediate reward, contrary to the setting where interventions might provide no (or negative) reward in the immediate but result in future states where more reward can be achieved following the agent's policy. In the following sections, we explain common nomenclature in RL in more detail.

4.1.1 The Agent

The entity or system that interacts with the world and makes decisions is the so-called "agent". The agent takes the information about the environment and decides what actions to take based on its future belief or estimate on the outcome. This information is also known as the "state" that evolves due to the agent's action or outside influences that the agent has no control over. In the context of mobile health, the agent would be the background system that observes the state of each patient. This state could be, for example, demographic information or health-related information. Based on this information, the agent decides which intervention it should deliver at a particular moment in time.

4.1.2 The Environment

The environment the agent finds itself in is, roughly speaking, everything that is external to the agent. The agent may or may not influence the environment through its actions, but in many real-world scenarios, this is often the case, such as in the hospital example described above. If the agent can affect the state of the environment, there are several scenarios in how the environment dynamics may be affected. In a stationary setting, the environment transitions into states according to a fixed set of rules as a response to the agent's actions. These rules can be probabilistic, but the probabilities stay fixed. An agent can discover these rules over time and exploit them to optimise its decision making. In the nonstationary setting, these rules can change over time, making the problem significantly harder since past experience becomes obsolete such that the agent chooses potentially sub-optimal actions.

Nonstationary is not only encountered as a feature of the environment but may manifest itself indirectly through incomplete information available to the agent for decision making. States might look similar but need different actions. In the mHealth setting, the agent has potentially contest with a combination of nonstationarity, incomplete information and varying degrees of influence of actions, requiring complex approaches to action selection to perform optimally.

4.1.3 The Reward

For the agent to decide if it needs to change its behaviour, it requires feedback on the "goodness" or "utility" of actions to achieve its goal. This feedback is often encoded in the so-called reward, indicating if actions lead to desirable states to solve a particular problem. Specific to our setting, improvement in blood pressure regulation is the primary goal. Given that blood pressure measurements might be infrequent and prone to errors, we might look at other metrics that we can use to define the reward. For example, the percentage of medication taken. Furthermore, patients might provide feedback for certain types of interventions that are interesting for them, allowing the agent to customise intervention selection further to improve patient engagement.

4.1.4 Behaviour or Policy

The policy is the formal description of the behaviour of the agent. In essence, it operationalises as a set of complicated rules the agent follows at every step of interaction with the environment. There are several ways of generating policies depending on the knowledge of the environment.

One straightforward policy is not necessarily learned at all but provided to the agent through external means. The agent does not need external information and achieves its goal by executing a fixed *plan*. Fixed plans without any form of environmental feedback are not robust in real-world scenarios due to the environment's occasional uncertainties. In a mHealth application, the patient's lifestyle might change over time such that reminders at fixed times may become a nuisance. The agent must occasionally execute a plan to ask the user for an updated reminding schedule before continuing the reminding plan.

If the environmental state can change at every time step, we have the so-called stationary policy, or "universal plan" [70]. The agent uses a stationary policy, either deterministic or stochastic, to evaluate the current state of the environment and performs the action that would maximize the immediate or future reward. Complex real-world environments exhibit randomness, prohibiting exact reward prediction. In such scenarios, it is more beneficial to consider *stochastic policies*, where the agent chooses an action from a probability distribution to maximise the expected reward. As mentioned earlier, in a mHealth setting, we are potentially faced with changing patient behaviours and must adopt a strategy that performs well under nonstationarity and noise.

4.1.5 Exploration and Exploitation

One fundamental issue in most sequential decision making problems is the balance between exploration and exploitation of actions. This need for balancing

both aspects arises due to incomplete reward feedback. The agent only receives the reward of the chosen action; the reward of other actions is not revealed. This type of reward mechanism is often described as *bandit feedback*. This significantly delays learning since the agent needs to explore all actions for all states sufficiently often to be sure that the optimal action has been chosen for the given state. When the agent explores, it deliberately chooses actions that might seem uncompetitive to confirm or revise its belief about the explored actions' utility. When the agent exploits, it selects the action that it beliefs would maximise its reward.

Naturally, the exploration of suboptimal actions for information gathering will result in less reward obtained and carries the risk of affecting the environment in ways that would come with significant penalties. For instance, in mHealth applications, user engagement is paramount for interventions to be successful. Interventions that are irrelevant or timely inconvenient might cause early abandonment of the application. Furthermore, insufficient exploration can lead to habituation that significantly diminishes the effectiveness of the interventions. While the later problem can be dealt with by ensuring diversity among interventions, the former problem requires efficient exploration schemes.

4.2 Multi-Armed Bandits and Contextual Bandits

The multi-armed bandit (MAB) problem in the area of sequential decision making has attracted significant attention due to its applicability in many real-world areas such as clinical trails [71; 72], finance [73; 74], routing networks [75; 76], online-advertising [77; 78] and movie [79] or app recommendation [80]. The agent's goal is to select from a set of available actions (also known as arms) that would maximise the cumulative immediate reward. Furthermore, the bandit formulation assumes no influence of actions on future rewards nor states. While limiting for problems where actions may significantly influence future states and rewards, this simplified setting works well in a variety of practical settings and enjoys provable regret guarantees and good sample complexity. The fact that behaviour change is a long process such that any single intervention has a temporally limited effect [81], gives us some leeway in needing to estimate the long-term rewards for particular interventions. This allows us to consider simpler methods with good convergence and optimally guarantees and better interpretability compared to using more general methods from the full RL setting that lack these properties.

Significant work has been done to design algorithms that provide an optimum or near optimum exploitation/exploration trade-off for various problem settings. Previous works have explored the context-free MAB-setting such as the stochastic variant using upper confidence bounds (UCB) operating under

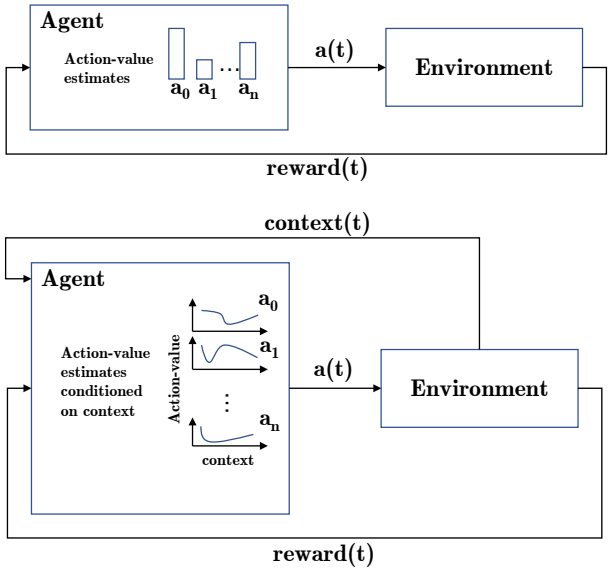

Figure 4.1: Top: Multi-Armed Bandit. Bottom: Contextual Bandit

the principle of "optimism in the face of uncertainty" [82; 83] or Bayesian treatments using Thompson sampling[84; 85] a so-called probability matching technique.

One particular formulation of the MAB has seen significant attention in the past. This extension to the MAB-problem is the Contextual Bandit (CB) problem, also known as the Multi-Armed-Bandit problem with side information or associative reinforcement learning [86]. The agent receives a context, some description of the environmental state in the form of a feature vector, before choosing an action, effectively solving a separate[1] MAB-problem conditioned on each context. Figure 4.1 illustrates the differences between MAB and CB schematically.

4.3 Corrupted Contextual Bandits with Action Order Constraints

In this thesis, we model the task of choosing digital interventions in an mHealth setting as a CB problem. We follow the definition of the CB problem from [87] with slight modifications to include information from several users similar to [88]. At each time step $t \in \{0, 1, \ldots, T\}$ [89]

[1] Strictly speaking, the reward predictor generalises over a space of MAB-problems.

1. The environment reveals a d-dimensional feature vector $\mathbf{x}_{i,t} \in \mathbb{R}^d$ that includes information of a user i.

2. The agent chooses an action $a_{i,t}$ from a set K of alternatives according to its policy. After playing the action $a_{i,t}$, the action's reward $r^i_{a_{i,t}} \in [0,1]$ is revealed.

3. The agent updates its policy using the observations of context $\mathbf{x}_{i,t}$, action $a_{i,t}$ and action-reward $r^i_{a_{i,t}}$ to improve action selection in future rounds.

The action to play at each round is chosen according to the agent's *policy*. A policy is a function that maps its inputs (the current context) to an action or distribution over actions, that is, $\pi : X \to A$. For our problem setting, the rewards the agent receives are binary, also known as the *Bernoulli Bandit* problem. This means the rewards the agent can receive are restricted to the set $r^i_{a_{i,t}} \in \{0,1\}$.

The difference in cumulative reward between the policy that always chooses the optimal action on every time step and the learned policy π of the agent is commonly referred to as the *regret*. We compute the regret as an average over the regret of all users defined as [89]

Definition 1 (Average cumulative regret) The average cumulative regret after T time steps over all users in I.

$$R(T) = \frac{1}{|I|} \sum_{i \in I} \left(\sum_{t=1}^{T} r^i_{t,a^*_{i,t}} - \sum_{t=1}^{T} r^i_{t,\pi(x_{i,t})} \right),$$

where $a^*_{i,t}$ and $a_{i,t} := \pi(x_{i,t})$ denote the optimal action and the action chosen by the agents policy at time step t for user i, respectively.

4.3.1 A Problem Setting for mHealth

The contextual setting we primarily consider is mHealth, where users' needs and wants are partially determined by an underlying evolving state, for example. Depending on the state, different interventions might be required, e.g., some users are significantly affected by stress and may need a particular type of stress coping techniques. In contrast, others may experience stress less severely, where general advice might be enough [90]. The underlying state may induce a natural ordering of actions as the users transition through different "levels" or "stages" that require specific interventions. Additionally, we can see the hidden underlying state as the different stages people may go through when forming

habits, such as initiation, learning or maintenance phase, requiring different interventions or intervention strategies [91].

Furthermore, we have mentioned earlier that the inherent nonstationary in mHealth poses a problem for simple MAB approaches and we presented one solution to this problem using contextual information leading the CB formulation. In our setting of mHealth, we expect users not to know their state perfectly and provide the agent with a noisy estimate of the state or potentially completely irrelevant context that can significantly affect the agent's decision making. This introduced nonstationary through incomplete information, masking changes in the context to reward mapping or misleading the agent to provide an intervention that is not appropriate anymore.

To tackle the issue of decision making in environments that exhibit the aforementioned properties of context uncertainty and action order, we extend a previously described problem setting. In the previous setting, see [92], the agent needs to deal with corrupted contexts where the information content is entirely and irreversibly lost. With probability p, the agent receives a corrupted context. The arbitrary corruption function $\nu : X \to X$ governs how the context is corrupted and is unknown and non-retrievable. The context the agent receives at every time step is defined as

$$\hat{\mathbf{x}}_t = \begin{cases} \nu(\mathbf{x}_t) & \text{with probability} \quad p \\ \mathbf{x}_t & \text{with probability} \quad 1-p. \end{cases}$$

We extend this problem setting to include several users that can provide context at varying degrees of corruption. Additionally, to incorporate the intuition of stages patients might go through, the hidden state evolves in a Markovian manner; that is, the previous state at $t-1$ fully determines the current state at t. Each state is associated with a specific action. Coupled with the Markovian state evolution, this defines a sequence of actions as patients transition through different stages or levels. In protocol 1, we present a high-level description of our problem setting.

Protocol 1 Problem Protocol

procedure PROTOCOL
  for $t = 1, 2, \ldots, T$ **do**
    for user $i \in I$ **do**
      the environment generates context $\mathbf{x}_{i,t}$ from state $s_{i,t}$
      the context is corrupted $\hat{\mathbf{x}}_{i,t} = \nu(\mathbf{x}_{i,t})$ with probability $p_{i,t}$
      the agent chooses an action $a_{i,t} = \pi(\hat{\mathbf{x}}_{i,t})$
      the environment reveals the reward $r_{a_{i,t}}$
      the state s_i is updated: $s_{i,t+1} = \phi(r^i_{a_{i,t}}, \mathbf{s}_{i,t})$ ▷ "Markovian sampling" of next state
      policy π of the agent is updated

At each iteration and for every user, the environment generates the context

$\mathbf{x}_{i,t}$ of user i from the underlying state $s_{i,t}$ and corrupts it with probability $p_{i,t}$. The corrupted context is observed by the agent which chooses an action $a_{i,t}$ to play according to its policy. The environment reveals the action-reward $r^i_{a_{i,t}}$ and updates the state for user i. Finally, the agent updates its policy π [89].

4.3.2 Meta-Algorithm: Competing Bandits with Corrupted Context and Action Correlations

We present the high-level overview of the meta-algorithm COMpeting BandIts with corrupted coNtext and action corrElations (COMBINE) in algorithm 1, and its UCB variant, in the algorithm 2. At each time step, the agent observes the possibly corrupted context $\hat{x}_t$ and decides to use either the CB or MAB policies to select the action to play (line 6). Another bandit policy, the so-called referee, chooses to play the CB or MAB at each round. The complete architecture of the approach is illustrated in figure 4.2.

Algorithm 1 Competing Bandits with Corrupted Context and Action Correlations

```
1:  procedure COMBINE
2:      Input: Algorithm Parameters, Policies: CB, MAB and referee, action set
3:      Initialize: Book-keeping variables for CB, MAB and referee
4:      for t = 1,2,...,T do
5:          for user i ∈ I do
6:              Observe context x̂_{i,t}
7:              Sample Policy π(t) from referee
8:              if π_{i,t} = CB policy then
9:                  choose action using CB policy
10:             else
11:                 choose action using MAB policy from subset 𝒰_i
12:             Observe reward for the chosen action
13:             if previous best action is not equal to the current chosen action then
14:                 Update Adjacency matrix Λ_i
15:             if CB was chosen as a policy then
16:                 Update CB policy
17:             else
18:                 Update MAB policy
19:             Update referee
20:             if reward = 1 then
21:                 Update best current action
22:             else
23:                 Update best previous action
24:             choose action subset 𝒰_i to sample from next
```

Suppose the referee chooses the MAB policy. In that case, the actions are selected from an action-subset $\mathcal{U}$ that is dynamically computed and represents a candidate set of next promising actions. If the CB policy is chosen, the algorithm selects the action according to the context. The action is played, and the agent observes the reward.

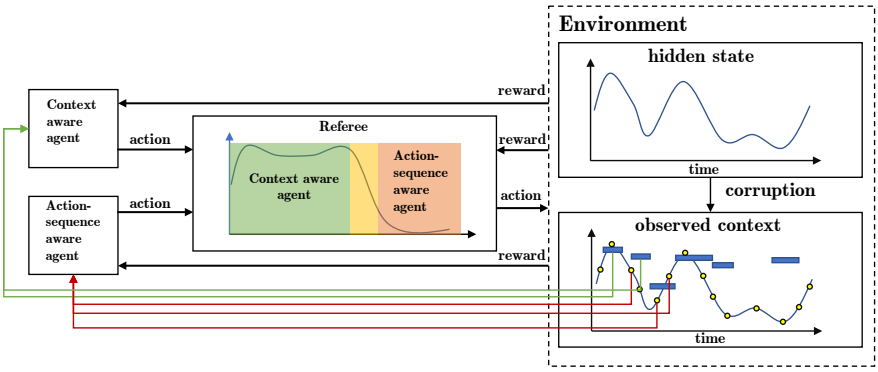

Figure 4.2: COMBINE architecture

If the CB policy was chosen for the action selection, its policy is updated with the new reward (line 15). Otherwise, we update the MAB policy (line 18). The reasoning for this is that if the referee does not trust the context it receives, the CB may not gain anything by learning from that data point. As determined by the referee, discarding useless information is expected to lead to more efficient learning of the context to action-reward mapping. Similarly, if the MAB has difficulty selecting the appropriate action, it may only harm its estimates of true action rewards. The referee is then updated using the reward and the choice of policy.

The algorithm keeps track of action transitions through an action-to-action adjacency matrix that contains the number of observed transitions between actions. The matrix is updated after the reward is observed (line 13). The MAB uses the adjacency matrix to explore the next promising action in cases where contexts are uninformative or highly corrupted. As part of the adjacency matrix update, the best previous action or the current best action is updated (line 20). The best previous action and current best action are used to record what action leads into another such that the adjacency matrix can be properly updated. Finally, the action subset $\mathcal{U}$ is updated to play in the next round (line 24). The UCB instantiation of the meta-algorithm is shown in algorithm 2. Algorithm 3 is used by the MAB to select actions. The procedure dynamically adapts the subset of action to explore on each round.

Algorithm 2 Competing UCB with Corrupted Context and Action Correlations

procedure COMBINE-UCB
  Input: $\alpha \in (0,\infty]$, $\alpha_B \in [0,\infty]$, $\gamma \in (0,\infty]$, $\delta_R \in (0,\infty]$, action set K
  Initialize: $H = \{0\}^{|I|\times 2}$, $\bar{R} = \{0\}^{|I|\times|K|}$, $n = \{0\}^{|I|\times|K|}$, $a^+ = \{NaN\}^{|I|\times 1}$, $a^- = \{NaN\}^{|I|\times 1}$, $\beta = \{0\}^{|I|\times 1}$, $\mathcal{U}^{|I|\times|K|} = K$, $\Lambda^i = \{1\}^{|K|\times|K|}$, $\forall a \in K : \mathbf{A}_a = \mathbf{I_d}$, $\forall a \in K : \mathbf{b_a} = \{0\}^{d\times 1}$
  for $t = 1, 2, \ldots, T$ **do**
    for user $i \in I$ **do**
      Observe context $\hat{x}_{i,t}$
      $pb_i = \frac{e^{H_i}}{\sum e^{H_i}}$ ▷ Compute probability of choosing either CB or MAB
      Sample Policy $\pi_{i,t} \sim \mathcal{B}\Big(pb_{i,t}\Big)$
      if $\pi_{i,t} = \pi_0$ **then** ▷ Choose contextual bandit
        for action $a = 1, \ldots, K$ **do**
          $\hat{\mu}_a \leftarrow \mathbf{A}_a^{-1}\mathbf{b}_a$ ▷ Update weight vectors of global CB
        $a_{i,t} = \text{argmax}_{a\in 1,\ldots,K}\, \hat{x}_{i,t}^\top \hat{\mu}_a + \alpha\sqrt{\hat{x}_{i,t}^\top \mathbf{A}_a^{-1}\hat{x}_{i,t}}$
      else ▷ Choose Bandit
        $a_{i,t} = \text{argmax}_{a\in\mathcal{U}_i}\, \theta_{i,a}$
      Observe reward $r^i_{t,a_{i,t}}$ for action $a_{i,t}$
      if $a_i^- \neq a_{i,t}$ **then** ▷ Update Adjacency matrix
        $\Lambda^i_{a_i^-,a_{i,t}} = \Lambda^i_{a_i^-,a_{i,t}} + r^i_{t,a_{i,t}}$
      if $\pi_{i,t} = \pi_0$ **then** ▷ Update global CB if it was chosen as a policy
        $\mathbf{A}_a \leftarrow \mathbf{A}_a + \hat{x}_{i,t}\hat{x}_{i,t}^\top$
        $\mathbf{b}_a \leftarrow \mathbf{b}_a + r^i_{t,a_{i,t}}\hat{x}_{i,t}$
      else
        $\bar{R}_{i,a_{i,t}} = \bar{R}_{i,a_{i,t}} + \gamma\big(r^i_{t,a_{i,t}} - \bar{R}_{i,a_{i,t}}\big)$ ▷ Update average action reward
        $n_{i,a_{i,t}} = n_{i,a_{i,t}} + 1$ ▷ Update action count for chosen action
        $\theta_{i,a_{i,t}} \leftarrow \bar{R}_{i,a_{i,t}} + \alpha_B\sqrt{\frac{2ln(t)}{n_{i,a_{i,t}}}}$ ▷ Update action scores
      $H_{i,\pi_{i,t}} \leftarrow H_{i,\pi_{i,t}} + \delta_R\big(r^i_{t,a_{i,t}} - pb_{i,\pi_{i,t}}\big)$ ▷ Update gradient bandit for $a_{i,t}$
      $H_{i,\neg\pi_{i,t}} \leftarrow H_{i,\neg\pi_{i,t}} - \delta_R 1^{r^i_{t,a_{i,t}}} pb_{i,\neg\pi_{i,t}}$ ▷ Update gradient bandit for all $a \neq a_{i,t}$
      if $r^i_{t,a_{i,t}} = 1$ **then**
        $a_i^+ \leftarrow a_{i,t}$ ▷ Update best current action
      else
        $a_i^- \leftarrow a_i^+$ ▷ Update best previous action
      $\mathcal{U}_i \leftarrow AdjSelect(a_i^-, \Lambda^i, \beta_i, r^i_{t,a_{i,t}})$ ▷ Choose action subset to sample from next

Algorithm 3 Select Action Subset

procedure ADJSELECT
  Input: best previous action a_i^-, Adjacency matrix Λ_i, reach β_i, reward $r_{a_{i,t}}$
  if $\forall k : \theta_{a_{i,t}} = 0$ or $a_i^- = NaN$ **then** ▷ Never have played an action before, or no previous best action
    return A ▷ Return complete action set
  else
    if $r_{a_{i,t}} = 1$ **then**
      $\beta_i \leftarrow 0$ ▷ Reset reach
    else
      $\beta_i \leftarrow \beta_i + 1$ ▷ Increase reach
    Sort row entries of Adjacency matrix $\Lambda^i_{a^-}$ in descending order, creating set $S^i_{a^-}$
    Select top β_i entries $S^i_{a^-,\{1,\ldots,1+\beta\}} := \mathcal{U}_i$
    return $\mathcal{U}_i$

5. Summary of Papers

5.1 Paper I: Pitfalls of medication adherence approximation through EHR and pharmacy records: Definitions, data and computation

Purpose and Conclusion We examined common pitfalls in refill adherence estimations using administrative databases from a theoretical and practical point of view, focusing on definitional and data related aspects. We used data from a comprehensive EHR system that includes near complete prescribing and dispensing information of medication. Through appropriate experimentation, we show that slight changes in definition can lead to significantly under or overestimation of refill adherence compared to the gold standard PDC measure. This has significant implications for, e.g., patient selection for interventions and may lead to false conclusions in real-world pharmacological studies. We analysed different methodological and data-related issues, in particular for data-related issues, we investigated the effects of missing values, duplicate entries and errors in data input, showing a small yet statistically significant effect on population averages, and a large effect in individual cases.

Contribution to the Licentiate Thesis This study was conducted with the goal of selecting the best patients for intervention for the intervention study of the iMedA Project. This study helped to understand how to quantify medication adherence through the use of comprehensive EHRs and highlighted the potential issues that can arise when dealing with real-world databases. Furthermore, this study stressed the consideration of methodological and data related pitfalls that are likely to occur and highlights their significance when selecting patients for intervention. This study's results have shaped the definition of medication adherence for the iMedA project and helped in selecting patients.

5.2 Paper II: Prediction and pattern analysis of medication refill adherence through electronic health records and dispensation data

Purpose and Conclusion We investigated the predictability of refill mediation adherence under various data splitting strategies using a comprehensive EHR system. Various machine learning algorithms were investigated under different predictive scenarios, with tree-based algorithms like Random Forest and Gradient Boosting Trees showing the highest performance. Predictive models have high discriminability on patients with high healthcare utilisation, reaching AUCs of approximately 0.90 and 0.91 on the test sets using baseline predictors and baseline+history predictors, respectively. The models' lowest discriminability is observed in the most realistic scenario of forward-prediction of new prescriptions, with AUCs of 0.77 and 0.80 with baseline predictors and baseline+history predictors, respectively. While model discriminability is relatively high, especially compared to previous studies, these models might still not be suitable for selecting patients for intervention since performance is quite low in "interesting" scenarios such as predicting a sudden change in adherence or predicting adherence of new patients that are at the beginning of treatment, where AUCs range between 0.56-0.65 are achieved on the test set.

We discovered common patterns of refill adherence for patients that start their anti-hypertensive treatment, that is, receive medication for the first time. We observe distinct patterns of refilling and notably the non-unique dependence of the second-year adherence trajectory on adherence in the first year. While patients are more likely to continue treatment after the first year, there is a non-negligible number of patients who discontinue treatment. While some patterns are more likely than others, this highlights the necessity to support patients early, even if they are refill adherent in the first year. To better understand what consumption patterns might cause the patterns observed, we simulated medication taking and correlated the results with refilling patterns. Assuming that medication taking and medication refilling are linked, simulations would allow the reasoning about medication consumption patterns, helping to select potential candidates for intervention. Our simulations show that certain pathological patterns of medication consumption are incompatible with the observed pattern of refill adherent patients, implying that nonadherence can occur suddenly without prior warning signs in refilling patterns. Furthermore, certain medication consumption patterns result in similar pickup patterns, obfuscating potential pathological patterns of medication-taking that can be relevant for intervention.

Contribution to the Licentiate Thesis This study helped in understanding the risk factors associated with refill adherence using clinical and healthcare utilisation information. It was clear from the outset that EHRs do not contain the information required to obtain very accurate prediction models. Given that previous studies had several limitations in terms of available data, the unique opportunity to test the predictability with more comprehensive EHRs has allowed us to underline this point further and draw attention to the fact that factors associated with the patients daily life and perceptions about medication and their illness are more important to proper adherence to treatment. Furthermore, we found common refill patterns showing that refill adherence in the first years of treatment is, to some extent, predictive of long-term refill adherence on population-level, but we often observe deviations from this pattern, resulting in the late detection of real medication nonadherence. These individuals stop treatment after the first year despite a followup prescription and may have required support early on, highlighting the need for monitoring outside the clinical or primary care setting. While the prediction models were not used directly for selecting patients in the iMedA project, the study was instrumental in understanding what the limitations are in terms of predicting MA and usefulness for the project.

5.3 Paper III: Corrupted Contextual Bandits with Action Order Constraints

Purpose and Conclusion We investigate a novel variant of the contextual bandit problem with corrupted context motivated by mHealth applications. MHealth, is a challenging environment for reinforcement learning agent to perform well in given that the information content provided by users is often missing, incomplete and unreliable. Furthermore, keeping user engagement is paramount for the success of interventions and therefore, it is vital to provide relevant recommendations on time. Additionally, users might transition through different treatment stages that require more targeted action selection approaches. The purpose of this study was to formulate a problem setting that would exhibit the aforementioned properties and give an algorithm that can learn and act more effectively than simpler solutions.

We develop a meta-algorithm, called COMBINE, that uses a "referee" that dynamically combines the policies of a contextual bandit, which uses a "context" or feature vector to make decisions, and a multi-armed bandit which aims to find the best action irrespective of context. The multi-armed bandit selects actions through a simple correlation mechanism that captures action to action transition probabilities, effectively learning a dynamics model of

the environment, allowing for more efficient exploration of time-correlated actions than standard bandit algorithms such as LinUCB or LinTS. We evaluate empirically the performance of the developed algorithm on simulated and real-world data.

In most settings where the performance of the combined algorithms differ significantly, the COMBINE approach outperforms single methods, where adjusting to one policy over the other in the short term can result in a reduction in regret compared to using CB and MAB algorithms individually. Given that users might vary in their response to surveys over time, using COMBINE will provide a significant advantage over simpler solutions. On simulated data, we observe that for the extrema of high action fluctuations and low context corruption or low action fluctuations and high context corruption, using a simple agent approach, i.e., either a CB or MAB, might minimise incurred regret. This highlights the necessity to find more efficient exploration and exploitation schemes when combing multiple bandit algorithms such that the overall regret is not significantly larger than the regret of the best single algorithm.

Contribution to the Licentiate Thesis As part of the iMedA project, we aim to provide adaptive, personalised digital interventions to improve medication adherence. This study was conducted as a first step into using reinforcement learning methods to achieve this goal. There are several challenges associated with mHealth application that make the straight forward application of standard RL algorithms difficult. In particular, patients might not provide good contexts for the agent to make good decisions and delay learning good policies. This study was conducted to explore the issue of providing relevant interventions given significant context corruption, affecting the learning and decision making of autonomous reinforcement learning agents.

This work also highlights the difficulty of optimally combining different algorithms, the solution to overcome some of the limitations introduced through context corruption. This study also provided the ground for future work in the area of reinforcement learning, particular for the problem of sequential decision making under uncertainty in the area of mHealth and provided us with valuable insights and ideas.

6. Conclusion

In this thesis, we have presented completed and ongoing work that is part of the iMedA project that aims to improve medication adherence through personalised adaptive interventions in the domain of mHealth. The first part of the project was dedicated to understanding medication adherence from the clinical perspective, prediction and measurement using electronic health records, with the goal of selecting patients for the pilot and large-scale intervention studies. The second part focused on finding ways to provide personalised adaptive interventions to improve MA in an automatic fashion.

This thesis contributed to the first part in the following ways. We described methods to measure MA using EHRs and identified common pitfalls that can significantly affect the accurate estimation of MA. Secondly, we developed prediction models for MA and discussed the implications of using data from EHRs and pharmacy databases to conduct analysis and patient selection for interventions. While an important first step, estimating adherence using these data sources are necessarily incomplete. On the one hand, the true underlying level of adherence to medication is unobserved, and not available in the records and we must be content with approximate methods using dispensations. On the other hand, many factors that influence adherence to medication are insufficiently recorded in EHRs, for example, the attitude towards medication or the perceived medication burden. Additionally, clinical predictors and demographic information do not accurately predict adherence to medication, making early intervention difficult. This highlights the need for the monitoring of adherence outside the clinical or primary care setting to identify patterns and reasons for nonadherence early and provide adaptive tailored interventions.

The thesis contributed to the second part of the project in the following. We present a problem setting in sequential decision making that exhibits proprieties of the mHealth domain, context uncertainty in particular. We developed a meta-algorithm that solves the problem setting, providing the first step towards adaptive interventions in a life-long learning framework. While many of the contemporary RL algorithms can solve very complex domains, their practical application in mHealth is hindered by several challenges, such as non-stationarity in both state and reward, missing or corrupted state information and the requirement for fast learning with small amounts of data and adaptability to individual users, to name a few. These limitations require the development of

novel methods and algorithms that address the challenges providing the basis for future research in the intersection between RL and mHealth.

6.1 Ongoing and Future Work for the Doctoral Thesis

Ongoing and future work will primarily focus on reinforcement learning approaches in the domain of mHealth and related applications that have an emphasis on optimal decision making under uncertainty in both state and reward. We believe that the RL framework fits rather well, providing efficient exploration and exploitation schemes in sequential decision making problems. In particular, we aim to investigate the following topics.

Fast learning of good policies in non-stationary environments due to various uncertainty modalities The provision of optimal interventions in mHealth comes with significant challenges. From the perspective of RL, algorithms need to content with non-stationary due to user habituation or burden or missing and noisy information. From the broader perspective of the doctoral thesis, we aim to develop algorithms and methods that can cope with non-stationary and exhibit better sample complexity in a mHealth setting.

Initial work conducted in this thesis, exploiting action correlations, illustrated that an adaptive combination of algorithms might provide an interesting research direction. While combination approaches show promise, naively combining a set of algorithms can incur more regret than the single best algorithm in the set, particularly for users where algorithms in the set perform similarly [93]. We hypothesise that aside from developing more specialised adaptive exploration/exploitation schemes to combine algorithms, we can improve performance in mHealth by exploiting structure among sets of interventions or actions. We expect groups of users to respond to a potentially evolving set of interventions similarly. This correlation structure among interventions may allow the transfer of feedback information between algorithms by biasing importance weighted returns of arms through arm-to-arm reward regression similar to offline policy evaluation methods such as doubly robust policy evaluation [94].

Another source of delayed learning is the behavioural diversity of patients. While a system that provides perfect adapted intervention to each patient would be the best-case scenario, such an approach would most likely fail to provide good policies in a reasonable amount of interactions. Our experiments in this thesis show the potential to speed-up policy learning by pooling data from multiple users with similar behaviours. We aim to explore the incorporation of mechanisms into our developed algorithms to automatically discover and adapt to different patient groupings through the use of meta-learning techniques [95].

This approach could significantly speed up the training of policies that might perform well "enough" for groups of patients in the short term, but performance may suffer for individuals within groups in the long term. We hypothesise that adjusting group policies to individuals will require many fewer interactions than training an algorithm for specific users from scratch.

Furthermore, incorporating expert knowledge into an RL framework would provide an additional source of information that can be leveraged for faster policy learning by focusing exploration, using the expert decision directly or augmenting it when selecting single or sequences of actions. This would require the efficient translation of concepts such as "Do not provide smoking inventions to non-smokers" or "provide physical activity interventions to sedentary individuals" into policies that an RL agent can learn to imitate, constraints the policy or otherwise incorporate into its action selection scheme. The agent would then rely on the expert's decision in cases where the reward model may be misspecified or the agent is uncertain about the state to reward mapping [96].

References

[1] ALEXANDER GALOZY, SLAWOMIR NOWACZYK, ANITA SANT'ANNA, MATTIAS OHLSSON, AND MARKUS LINGMAN. **Pitfalls of medication adherence approximation through EHR and pharmacy records: Definitions, data and computation**. *International Journal of Medical Informatics*, **136**:104092, 2020. xi, 14

[2] ALEXANDER GALOZY AND SLAWOMIR NOWACZYK. **Prediction and pattern analysis of medication refill adherence through electronic health records and dispensation data**. *Journal of Biomedical Informatics: X*, **6-7**:100075, 2020. xi, 27, 29

[3] HOWARD PHILLIPS LOVECRAFT. **The Call of Cthulhu**. *Weird Tales*, **11**(2), 1928. 1

[4] P. A. GRADY AND L. L. GOUGH. **Self-management: a comprehensive approach to management of chronic conditions**. *American Journal of Public Health*, **104**(8):25–31, Aug 2014. 1

[5] C. DOWRICK, M. DIXON-WOODS, H. HOLMAN, AND J. WEINMAN. **What is chronic illness?** *Chronic Illness*, **1**(1):1–6, Mar 2005. 1

[6] S. S. MCMILLAN, E. KENDALL, A. SAV, M. A. KING, J. A. WHITTY, F. KELLY, AND A. J. WHEELER. **Patient-centered approaches to health care: a systematic review of randomized controlled trials**. *Medical Care Research and Review*, **70**(6):567–596, Dec 2013. 1

[7] E. BALINT. **The possibilities of patient-centered medicine**. *The Journal of the Royal College of General Practitioners*, **17**(82):269–276, May 1969. 1

[8] MARY E TINETTI AND TERRI FRIED. **The end of the disease era**. *The American Journal of Medicine*, **116**(3):179 – 185, 2004. 1

[9] R. M. EPSTEIN AND R. L. STREET. **The values and value of patient-centered care**. *Annals of Family Medicine*, **9**(2):100–103, 2011. 1

[10] K. D. BERTAKIS AND R. AZARI. **Patient-centered care is associated with decreased health care utilization**. *Journal of the American Board of Family Medicine*, **24**(3):229–239, 2011. 1

[11] C. R. JAÈN, R. L. FERRER, W. L. MILLER, R. F. PALMER, R. WOOD, M. DAVILA, E. E. STEWART, B. F. CRABTREE, P. A. NUTTING, AND K. C. STANGE. **Patient outcomes at 26 months in the patient-centered medical home National Demonstration Project**. *Annals of Family Medicine*, **8 Suppl 1**:57–67, 2010. 1

[12] M. C. BEACH, J. KERULY, AND R. D. MOORE. **Is the quality of the patient-provider relationship associated with better adherence and health outcomes for patients with HIV?** *Journal of General Internal Medicine*, **21**(6):661–665, Jun 2006. 2

[13] G. BLUMROSEN, N. AVISDRIS, R. KUPFER, AND B. RUBINSKY. **C-SMART: Efficient seamless cellular phone based patient monitoring system**. In *2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, pages 1–6, 2011. 2

[14] A. MINUTOLO, G. SANNINO, M. ESPOSITO, AND G. DE PIETRO. **A rule-based mHealth system for cardiac monitoring**. In *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pages 144–149, 2010. 2

[15] G. M. TURNER-MCGRIEVY, M. W. BEETS, J. B. MOORE, A. T. KACZYNSKI, D. J. BARR-ANDERSON, AND D. F. TATE. **Comparison of traditional versus mobile app self-monitoring of physical activity and dietary intake among overweight adults participating in an mHealth weight loss program**. *Journal of the American Medical Informatics Association*, **20**(3):513–518, May 2013. 2

[16] J. WEI, I. HOLLIN, AND S. KACHNOWSKI. **A review of the use of mobile phone text messaging in clinical and healthy behaviour interventions**. *Journal of Telemedicine and Telecare*, **17**(1):41–48, 2011. 2

[17] J. W. MCGILLICUDDY, A. K. WEILAND, R. M. FRENZEL, M. MUELLER, B. M. BRUNNER-JACKSON, D. J. TABER, P. K. BALIGA, AND F. A. TREIBER. **Patient attitudes toward mobile phone-based health monitoring: questionnaire study among kidney transplant recipients**. *Journal of Medical Internet Research*, **15**(1):e6, Jan 2013. 2

[18] PHILLIP OLLA AND CALEY SHIMSKEY. **mHealth taxonomy: a literature survey of mobile health applications**. *Health and Technology*, **4**(4):299–308, 2015. 2

[19] I. NAHUM-SHANI, S. N. SMITH, B. J. SPRING, L. M. COLLINS, K. WITKIEWITZ, A. TEWARI, AND S. A. MURPHY. **Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support**. *Annals of Behavioral Medicine*, **52**(6):446–462, 05 2018. 2

[20] GILLIAN KING, MELISSA CURRIE, AND PATRICIA PETERSEN. **Child and parent engagement in the mental health intervention process: a motivational framework**. *Child and Adolescent Mental Health*, **19**(1):2–8, 2014. 2

[21] I. NAHUM-SHANI, E. B. HEKLER, AND D. SPRUIJT-METZ. **Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework**. *Health Psychology*, **34S**:1209–1219, Dec 2015. 2

[22] R. L. CUTLER, F. FERNANDEZ-LLIMOS, M. FROMMER, C. BENRIMOJ, AND V. GARCIA-CARDENAS. **Economic impact of medication non-adherence by disease groups: a systematic review**. *BMJ Open*, **8**(1):e016982, 01 2018. 2

[23] HAYDEN B. BOSWORTH, BRADI B. GRANGER, PHIL MENDYS, RALPH BRINDIS, REBECCA BURKHOLDER, SUSAN M. CZAJKOWSKI, JODI G. DANIEL, INGER EKMAN, MICHAEL HO, MIMI JOHNSON, STEPHEN E. KIMMEL, LARRY Z. LIU, JOHN MUSAUS, WILLIAM H. SHRANK, ELIZABETH WHALLEY BUONO, KAREN WEISS, AND CHRISTOPHER B. GRANGER. **Medication adherence: A call for action**. *American Heart Journal*, **162**(3):412–424, 2011. 3

[24] CONCEIÇÃO GRANJA, WOUTER JANSSEN, AND MONIKA ALISE JOHANSEN. **Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature**. *Journal of Medical Internet Research*, **20**(5):e10235, May 2018. 3

[25] KOBRA ETMINANI, ARIANNA TAO ENGSTRÖM, CARINA GÖRANSSON, ANITA SANT'ANNA, AND SŁAWOMIR NOWACZYK. **How Behavior Change Strategies are Used to Design Digital Interventions to Improve Medication Adherence and Blood Pressure Among Patients With Hypertension: Systematic Review**. *Journal of Medical Internet Research*, **22**(4):e17201, Apr 2020. 3

[26] KAREN CHURCH, DENZIL FERREIRA, NIKOLA BANOVIC, AND KENT LYONS. **Understanding the Challenges of Mobile Phone Usage Data**. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '15, page 504–514, New York, NY, USA, 2015. Association for Computing Machinery. 3

[27] AMBUJ TEWARI AND SUSAN A. MURPHY. *From Ads to Interventions: Contextual Bandits in Mobile Health*. Springer International Publishing, Cham, 2017. 3, 6

[28] F D R HOBBS. **Cardiovascular disease: different strategies for primary and secondary prevention?** *Heart*, **90**(10):1217–1223, 2004. 4, 11

[29] T. PETERS-STRICKLAND, L. PESTREICH, A. HATCH, S. ROHATAGI, R. A. BAKER, J. P. DOCHERTY, L. MARKOVTSOVA, P. RAJA, P. J. WEIDEN, AND D. P. WALLING. **Usability of a novel digital medicine system in adults with schizophrenia treated with sensor-embedded tablets of aripiprazole**. *Neuropsychiatr Dis Treat*, **12**:2587–2594, 2016. 4

[30] FRESCO P. LAM WY. **Medication Adherence Measures: An Overview**. *BioMed research international*, pages 1–12, 2015. 4

[31] THOMAS J. PODSADECKI, BERNARD C. VRIJENS, ERIC P. TOUSSET, RICHARD A. RODE, AND GEORGE J. HANNA. **"White Coat Compliance" Limits the Reliability of Therapeutic Drug Monitoring in HIV-1—Infected Patients**. *HIV Clinical Trials*, **9**(4):238–246, 2008. PMID: 18753118. 4

[32] SUSAN E. ANDRADE, KRISTIJAN H. KAHLER, FERIDE FRECH, AND K. ARNOLD CHAN. **Methods for evaluation of medication adherence and persistence using automated databases.** *Pharmacoepidemiology and drug safety*, **15**:565–574, 2006. 4, 11

[33] WONG L CADARETTE SM. **An Introduction to Health Care Administrative Data**. *The Canadian journal of hospital pharmacy*, **68**:232–237, 2015. 4, 11, 15

[34] ANDREA M. BURDEN, J. MICHAEL PATERSON, ANDREA GRUNEIR, AND SUZANNE M. CADARETTE. **Adherence to osteoporosis pharmacotherapy is underestimated using days supply values in electronic pharmacy claims data**. *Pharmacoepidemiology and Drug Safety*, **24**:67–74, 2014. 5, 12

[35] JOYCE A. CRAMER, ANUJA ROY, ANITA BURRELL, CAROL J. FAIRCHILD, MAHESH J. FULDEORE, DANIEL A. OLLENDORF, AND PETER K. WONG. **Medication Compliance and Persistence: Terminology and Definitions**. *Value in Health*, **11**:44–47, 2008. 5

[36] ALEX J. ADAMS AND SAMUEL F. STOLPE. **Defining and Measuring Primary Medication Nonadherence: Development of a Quality Measure**. *Journal of Managed Care and Specialty Pharmacy*, **22**:516–523, 2016. 5

[37] L. M. HESS, M. A. RAEBEL, D. A. CONNER, AND D. C. MALONE. **Measurement of Adherence in Pharmacy Administrative Databases: A Proposal for Standard Definitions and Preferred Measures**. *Annals of Pharmacotherapy*, **40**:1280–1288. 5, 15, 16

[38] I. ARNET, M. J. KOOIJ, M. MESSERLI, K. E. HERSBERGER, E. R. HEERDINK, AND M. BOUVY. **Proposal of Standardization to Assess Adherence With Medication Records: Methodology Matters**. *Annals of Pharmacotherapy*, **50**:360–368, 2016. 5

[39] BRENDA MOTHERAL, JOHN BROOKS, MARY ANN CLARK, WILLIAM H. CROWN, PETER DAVEY, DAVE HUTCHINS, BRADLEY C. MARTIN, AND PAUL STANG. **A Checklist for Retrospective Database Studies-Report of the ISPOR Task Force on Retrospective Databases**. *Value in Health*, **6**:90–97, 2003. 5

[40] MICHAEL A. FISCHER, MARGARET R. STEDMAN, JOYCE LII, CHRISTINE VOGELI, WILLIAM H. SHRANK, M. ALAN BROOKHART, AND JOEL S. WEISSMAN. **Primary Medication Non-Adherence: Analysis of 195,930 Electronic Prescriptions**. *Journal of General Internal Medicine*, **25**:284–290, 2010. 5

[41] THOMAS WILKE, ANTJE GROTH, SABRINA MUELLER, DALLAS REESE, ROLAND LINDER, SUSANNE AHRENS, AND FRANK VERHEYEN. **How to use pharmacy claims data to measure patient nonadherence? The example of oral diabetics in therapy of type 2 diabetes mellitus**. *The European Journal of Health Economics*, **14**:551–568, 2013. 5

[42] W. M. VOLLMER, M. XU, A. FELDSTEIN, D. SMITH, A. WATERBURY, AND C. RAND. **Comparison of pharmacy-based measure of medication adherence**. *BMC health services research*, **12**:155, 2012. 5, 12, 15, 16, 19

[43] A.V. PROCHAZKA J.F. STEINER. **The Assessment of Refill Compliance Using Pharmacy Records: Methods, Validity, and Applications**. *Journal of Clinical Epidemiology*, **50**:105–116, 1997. 5, 12

[44] SUDEEP KARVE, MARIO A. CLEVES, MARK HELM, TERESA J. HUDSON, DONNA S. WEST, AND BRADLEY C. MARTIN. **Prospective Validation of Eight Different Adherence Measures for Use with Administrative Claims Data among Patients with Schizophrenia**. *Value in Health*, **12**:1098–3015, 2009. 5

[45] ALEXIS A. KRUMME, JESSICA M. FRANKLIN, DANIELLE L. ISAMAN, OLGA S. MATLIN, ANGELA Y. TONG, CLAIRE M. SPETTELL, TROYEN A. BRENNAN, WILLIAM H. SHRANK, AND NITEESH K. CHOUDHRY. **Predicting 1-Year Statin Adherence Among Prevalent Users: A Retrospective Cohort Study**. *Journal of Managed Care & Specialty Pharmacy*, **23**(4):494–502, 2017. 5

[46] J. M. FRANKLIN, W. H. SHRANK, J. LII, A. K. KRUMME, O. S. MATLIN, T. A. BRENNAN, AND N. K. CHOUDHRY. **Observing versus Predicting: Initial Patterns of Filling Predict Long-Term Adherence More Accurately Than High-Dimensional Modeling Techniques**. *Health Services Research*, **51**(1):220–239, Feb 2016.

[47] ALEXIS A KRUMME, GABRIEL SANFÉLIX-GIMENO, JESSICA M FRANKLIN, DANIELLE L ISAMAN, MUFADDAL MAHESRI, OLGA S MATLIN, WILLIAM H SHRANK, TROYEN A BRENNAN, GREGORY BRILL, AND NITEESH K CHOUDHRY. **Can purchasing information be used to predict adherence to cardiovascular medications? An analysis of linked retail pharmacy and insurance claims data**. *BMJ Open*, **6**(11):e011015, November 2016. 5

[48] PENG LIAO, KRISTJAN GREENEWALD, PREDRAG KLASNJA, AND SUSAN MURPHY. **Personalized HeartSteps: A Reinforcement Learning Algorithm for Optimizing Physical Activity**. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, **4**(1), March 2020. 7

[49] AWAIS ASHFAQ, STEFAN LÖNN, HÅKAN NILSSON, JONNY A ERIKSSON, JAPNEET KWATRA, ZAYED M YASIN, JONATHAN E SLUTZMAN, THOMAS WALLENFELDT, ZIAD OBERMEYER, PHILIP D ANDERSON, AND MARKUS LINGMAN. **Data resource profile: Regional healthcare information platform in Halland, Sweden, a dedicated environment for healthcare research**. *International Journal of Epidemiology*, 01 2020. 8

[50] EDUARDO SABATE´ AND WOLRD HEALTH ORGANIZATION. **Adherence to long-term therapies: Evidence for action**, 2003. 11

[51] VINAY KINI AND P. MICHAEL HO. **Interventions to Improve Medication Adherence: A Review**. *JAMA*, **320**(23):2461–2473, 12 2018. 11

[52] P. MICHAEL HO, JOHN S. RUMSFELD, FREDERICK A. MASOUDI, DAVID L. MCCLURE, MARY E. PLOMONDON, JOHN F. STEINER, AND DAVID J. MAGID. **Effect of Medication Nonadherence on Hospitalization and Mortality Among Patients With Diabetes Mellitus**. *Archives of internal medicine*, pages 1836–1841, 2006. 12

[53] JASON YEAW, JOSHUA S. BENNER, JOHN G. WALT, SERGEY SIAN, AND DANIEL B. SMITH. **Comparing Adherence and Persistence Across 6 Chronic Medication Classes**. *Journal of Managed Care Pharmacy*, **15**:728–740, 2009. 12

[54] KATHLEEN A. FAIRMAN AND BRENDA MOTHERAL. **Evaluating Medication Adherence: Which Measure Is Right for Your Program?** *Journal of Managed Care Pharmacy*, **6**:499–506, 2000. 18

[55] RICHARD BELLMAN. *Dynamic Programming*. Dover Publications, 1957. 22

[56] AWAIS ASHFAQ. *Predicting clinical outcomes via machine learning on electronic health records*. Number 58 in Halmstad University Dissertations. 2019. 22

[57] D. PRAKASH. **Target organ damage in newly detected hypertensive patients**. *Journal of Family Medicine and Primary Care*, **8**(6):2042–2046, Jun 2019. 22

[58] BEFEKADU ABEBE T. ABEGAZ TM, TEFERA YG. **Target Organ Damage and the Long Term Effect of Nonadherence to Clinical Practice Guidelines in Patients with Hypertension: A Retrospective Cohort Study**. *International Journal of Hypertension*, **2090-0384**(10):749, 2017. 22

[59] JOAQUIN QUIONERO-CANDELA, MASASHI SUGIYAMA, ANTON SCHWAIGHOFER, AND NEIL D. LAWRENCE. *Dataset Shift in Machine Learning*. The MIT Press, 2009. 24

[60] KEVIN W. BOWYER, NITESH V. CHAWLA, LAWRENCE O. HALL, AND W. PHILIP KEGELMEYER. **SMOTE: Synthetic Minority Over-sampling Technique**. *CoRR*, **abs/1106.1813**, 2011. 26

[61] G. D. RANASINGHE, T. LINDGREN, M. GIROLAMI, AND A. K. PARLIKAD. **A Methodology for Prognostics Under the Conditions of Limited Failure Data Availability**. *IEEE Access*, **7**:183996–184007, 2019. 26

[62] CHAO CHEN AND LEO BREIMAN. **Using Random Forest to Learn Imbalanced Data**. *University of California, Berkeley*, 01 2004. 26

[63] MICHEL BURNIER AND BRENT M. EGAN. **Adherence in Hypertension**. *Circulation Research*, **124**(7):1124–1140, 2019. 27

[64] J. MACQUEEN. **Some methods for classification and analysis of multivariate observations**. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, Berkeley, Calif., 1967. University of California Press. 27

[65] RUBY GRYMONPRE, MARY CHEANG, MARJORY FRASER, COLLEEN METGE, AND DANIEL S. SITAR. **Validity of a Prescription Claims Database to Estimate Medication Adherence in Older Persons**. *Medical Care*, **44**(5), 2006. 28

[66] J. A. CRAMER. **Effect of partial compliance on cardiovascular medication effectiveness**. *Heart*, **88**(2):203–206, Aug 2002. 28

[67] DAVID SILVER, AJA HUANG, CHRIS J. MADDISON, ARTHUR GUEZ, LAURENT SIFRE, GEORGE VAN DEN DRIESSCHE, JULIAN SCHRITTWIESER, IOANNIS ANTONOGLOU, VEDA PANNEERSHELVAM, MARC LANCTOT, SANDER DIELEMAN, DOMINIK GREWE, JOHN NHAM, NAL KALCHBRENNER, ILYA SUTSKEVER, TIMOTHY LILLICRAP, MADELEINE LEACH, KORAY KAVUKCUOGLU, THORE GRAEPEL, AND DEMIS HASSABIS. **Mastering the Game of Go with Deep Neural Networks and Tree Search**. *Nature*, **529**(7587):484–489, jan 2016. 31

[68] ORIOL VINYALS, IGOR BABUSCHKIN, JUNYOUNG CHUNG, MICHAEL MATHIEU, MAX JADERBERG, WOJTEK CZARNECKI, ANDREW DUDZIK, AJA HUANG, PETKO GEORGIEV, RICHARD POWELL, TIMO EWALDS, DAN HORGAN, MANUEL KROISS, IVO DANIHELKA, JOHN AGAPIOU, JUNHYUK OH, VALENTIN DALIBARD, DAVID CHOI, LAURENT SIFRE, YURY SULSKY, SASHA VEZHNEVETS, JAMES MOLLOY, TREVOR CAI, DAVID BUDDEN, TOM PAINE, CAGLAR GULCEHRE, ZIYU WANG, TOBIAS PFAFF, TOBY POHLEN, DANI YOGATAMA, JULIA COHEN, KATRINA MCKINNEY, OLIVER SMITH, TOM SCHAUL, TIMOTHY LILLICRAP, CHRIS APPS, KORAY KAVUKCUOGLU, DEMIS HASSABIS, AND DAVID SILVER. **AlphaStar: Mastering the Real-Time Strategy Game StarCraft II**. `https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/`, 2019. 31

[69] OPENAI, :, CHRISTOPHER BERNER, GREG BROCKMAN, BROOKE CHAN, VICKI CHEUNG, PRZEMYSŁAW DĘBIAK, CHRISTY DENNISON, DAVID FARHI, QUIRIN FISCHER, SHARIQ HASHME, CHRIS HESSE, RAFAL JÓZEFOWICZ, SCOTT GRAY, CATHERINE OLSSON, JAKUB PACHOCKI, MICHAEL PETROV, HENRIQUE PONDÉ DE OLIVEIRA PINTO, JONATHAN RAIMAN, TIM SALIMANS, JEREMY SCHLATTER, JONAS SCHNEIDER, SZYMON SIDOR, ILYA SUTSKEVER, JIE TANG,

FILIP WOLSKI, AND SUSAN ZHANG. **Dota 2 with Large Scale Deep Reinforcement Learning**, 2019. 31

[70] THOMAS DEAN, LESLIE PACK KAELBLING, JAK KIRMAN, AND ANN NICHOLSON. **Planning under time constraints in stochastic domains**. *Artificial Intelligence*, **76**(1):35 – 74, 1995. Planning and Scheduling. 33

[71] S. S. VILLAR, J. BOWDEN, AND J. WASON. **Multi-armed Bandit Models for the Optimal Design of Clinical Trials: Benefits and Challenges**. *Statistical Science*, **30**(2):199–215, 2015. 34

[72] HAMSA BASTANI AND MOHSEN BAYATI. **Online Decision Making with High-Dimensional Covariates**. *Operations Research*, **68**(1):276–294, 2020. 34

[73] WEIWEI SHEN, JUN WANG, YU-GANG JIANG, AND HONGYUAN ZHA. **Portfolio Choices with Orthogonal Bandit Learning**. In *International Conference on Artificial Intelligence*, IJCAI'15, page 974–980. AAAI Press, 2015. 34

[74] XIAOGUANG HUO AND FENG FU. **Risk-aware multi-armed bandit problem with application to portfolio selection**. *Royal Society Open Science*, **4**(11):171377, 2017. 34

[75] STEFANO BOLDRINI, LUCA DE NARDIS, GIUSEPPE CASO, MAI LE, JOCELYN FIORINA, AND MARIA-GABRIELLA DI BENEDETTO. **muMAB: A Multi-Armed Bandit Model for Wireless Network Selection**. *Algorithms*, **11**(2):13, Jan 2018. 34

[76] R. KERKOUCHE, R. ALAMI, R. FÉRAUD, N. VARSIER, AND P. MAILLÉ. **Node-based optimization of LoRa transmissions with Multi-Armed Bandit algorithms**. In *2018 25th International Conference on Telecommunications (ICT)*, pages 521–526, 2018. 34

[77] ZHENG WEN, BRANISLAV KVETON, MICHAL VALKO, AND SHARAN VASWANI. **Online Influence Maximization under Independent Cascade Model with Semi-Bandit Feedback**. In I. GUYON, U. V. LUXBURG, S. BENGIO, H. WALLACH, R. FERGUS, S. VISHWANATHAN, AND R. GARNETT, editors, *Advances in Neural Information Processing Systems*, **30**, pages 3022–3032. Curran Associates, Inc., 2017. 34

[78] ERIC SCHWARTZ, ERIC BRADLOW, AND PETER FADER. **Customer Acquisition via Display Advertising Using Multi-Armed Bandit Experiments**. *Marketing Science*, **36**, 04 2017. 34

[79] Q. WANG, C. ZENG, W. ZHOU, T. LI, S. S. IYENGAR, L. SHWARTZ, AND G. Y. GRABARNIK. **Online Interactive Collaborative Filtering Using Multi-Armed Bandit with Dependent Arms**. *IEEE Transactions on Knowledge and Data Engineering*, **31**(8):1569–1580, 2019. 34

[80] LINAS BALTRUNAS, KAREN CHURCH, ALEXANDROS KARATZOGLOU, AND NURIA OLIVER. **Frappe: Understanding the Usage and Perception of Mobile App Recommendations In-The-Wild**. *CoRR*, **abs/1505.03014**, 2015. 34

[81] HUITIAN LEI, AMBUJ TEWARI, AND SUSAN A. MURPHY. **An Actor-Critic Contextual Bandit Algorithm for Personalized Mobile Health Interventions**, 2017. 34

[82] T.L LAI AND HERBERT ROBBINS. **Asymptotically efficient adaptive allocation rules**. *Advances in Applied Mathematics*, **6**(1):4 – 22, 1985. 35

[83] PETER AUER, NICOLO CESA-BIANCHI, AND PAUL FISCHER. **Finite-time Analysis of the Multi-armed Bandit Problem**. *Machine Learning*, **47**:235–256, 05 2002. 35

[84] WILLIAM R. THOMPSON. **On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples**. *Biometrika*, **25**(3/4):285–294, 1933. 35

[85] OLIVIER CHAPELLE AND LIHONG LI. **An Empirical Evaluation of Thompson Sampling**. In J. SHAWE-TAYLOR, R. S. ZEMEL, P. L. BARTLETT, F. PEREIRA, AND K. Q. WEINBERGER, editors, *Neural Information Processing*, pages 2249–2257. Curran Associates, Inc., 2011. 35

[86] RICHARD S. SUTTON AND ANDREW G. BARTO. *Reinforcement Learning: An Introduction*. A Bradford Book, Cambridge, MA, USA, 2018. 35

[87] JOHN LANGFORD AND TONG ZHANG. **The Epoch-Greedy Algorithm for Multi-armed Bandits with Side Information**. In J. PLATT, D. KOLLER, Y. SINGER, AND S. ROWEIS, editors, *Neural Information Processing*, **20**, pages 817–824. Curran Associates, Inc., 2008. 35

[88] LIHONG LI, WEI CHU, JOHN LANGFORD, AND ROBERT E. SCHAPIRE. **A contextual-bandit approach to personalized news article recommendation**. *International conference on World wide web - WWW '10*, 2010. 35

[89] ALEXANDER GALOZY, SLAWOMIR NOWACZYK, AND MATTIAS OHLSSON. **Corrupted Contextual Bandits with Action Order Constraints**, 2020. 35, 36, 38

[90] REDA ABOUSERIE. **Sources and Levels of Stress in Relation to Locus of Control and Self Esteem in University Students**. *Educational Psychology*, **14**(3):323–330, 1994. 36

[91] B. GARDNER, P. LALLY, AND J. WARDLE. **Making health habitual: the psychology of 'habit-formation' and general practice**. *Br J Gen Pract*, **62**(605):664–666, Dec 2012. 37

[92] DJALLEL BOUNEFFOUF. **Online learning with Corrupted context: Corrupted Contextual Bandits**. 2020. 37

[93] ALEKH AGARWAL, HAIPENG LUO, BEHNAM NEYSHABUR, AND ROBERT E. SCHAPIRE. **Corralling a Band of Bandit Algorithms**, 2017. 46

[94] MIROSLAV DUDÍK, JOHN LANGFORD, AND LIHONG LI. **Doubly Robust Policy Evaluation and Learning**. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, page 1097–1104, Madison, WI, USA, 2011. Omnipress. 46

[95] CHELSEA FINN, PIETER ABBEEL, AND SERGEY LEVINE. **Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks**. *CoRR*, **abs/1703.03400**, 2017. 46

[96] MUHAMMAD ALI CHATTHA, SHOAIB AHMED SIDDIQUI, MUHAMMAD IMRAN MALIK, LUDGER VAN ELST, ANDREAS DENGEL, AND SHERAZ AHMED. **KINN: Incorporating Expert Knowledge in Neural Networks**, 2019. 47