Longitudinal Trajectories of Health Indicators Using Real World Data

Leena Elhussein

Green Templeton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy – Clinical Epidemiology and Medical Statistics

Trinity 2024

Abstract

Longitudinal Trajectories of Health Indicators Using Real World Data

Leena Elhussein, Green Templeton College, DPhil in Clinical Epidemiology and Medical Statistics, Trinity 2024

Older people with complex health needs are often excluded from clinical trials, primarily due to factors such as age, multimorbidity, and polypharmacy. However, they represent a significant portion of healthcare resource consumption and the use of newly authorised medications. Existing guidelines for identifying and treating this population often rely on using cross-sectional values and providing guidance on treating individual conditions rather than addressing the complexities of multimorbidity and treatment combinations. In this thesis, I propose applying novel approaches for identifying and characterising older people with complex health needs, using different health indicators and study designs on real world data. The definitions and cohorts established in this work have the potential to inform decisions for identifying, managing and treating older people with complex health needs.

In the first project, I conducted a cross-sectional analysis to identify three cohorts of older people with high levels of frailty, polypharmacy, or unplanned hospital admissions. Patients in any of these cohorts had high comorbidity burden and preventive therapy use. Although there was considerable overlap between these cohorts, many patients only belonged to one of the three cohorts. This indicates that these health markers are intersectional and complementary to each other.

Frailty and polypharmacy are cumulative conditions that take years to develop, making cross-sectional cohorts unable to describe their progression over time. In projects two and three, I modelled frailty and polypharmacy in older people over 4-5 years of follow-up. I identified subgroups with distinct frailty or polypharmacy trajectories, which demonstrated different

association levels with the risk of mortality. Most of the population belonged to the low-steady/slow (healthy) subgroup. However, important subgroups emerging from these studies started from a seemingly healthy state, deteriorated rapidly over the study follow-up and had the highest mortality risks, indicating their need for more healthcare resources and monitoring. The subgroups were identified in a UK primary care database and then externally validated in two independent national and international databases. They demonstrated generalisability with good external validity, similar trajectories and clinical characteristics.

Previous evidence reported that frailty and polypharmacy could start from middle age, and some of the identified subgroups of older people started from elevated or intermediate levels of frailty and polypharmacy. To understand these health markers' progression from early on, I modelled polypharmacy over time in middle-aged people in the fourth project. I identified three subgroups with distinct polypharmacy trajectories and associated mortality risks. I found that those with the fastest polypharmacy trajectory had the highest mortality risk, followed by those starting at the highest polypharmacy baseline values. Those patients are likely to continue progressing and end up in one of the non-healthy subgroups at older age.

My research demonstrated that monitoring trajectories of frailty and polypharmacy predicts mortality better than cross-sectional values. The identified subgroups were generalisable and had distinctive clinical characteristics. Future research can focus on further generalisability of the identified subgroups, and investigate how polypharmacy and frailty progress over longer periods, together and individually.

Declaration

The work presented in this thesis was undertaken by Leena Elhussein and has not previously been submitted for the award of a degree by any university. All contributions by others have been acknowledged. This work was performed under the supervision of Prof Daniel Prieto-Alhambra, Dr Victoria Y. Strauss, Dr Antonella Delmestri and Dr Edward Burn, and completed within the Centre for Statistics in Medicine (CSM) in the Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS) at the University of Oxford.

Acknowledgements

To the amazing supervisory team, without whom I would not have even dreamed of reaching this point. Dani, whose guidance, care, and support go above and beyond the supervisory duties. A lot has happened over the course of four years: a pandemic, a loss in the family, a coup, and then war in my country, to name a few. Dani's support throughout all of this was instrumental in me continuing my studies and maintaining my well-being. Antonella, whose attention to detail saved me once and again. She continually restored my belief in myself when doubts won the best of me. I could always find humour with Antonella, even in the hardest times. Vicky, who believed in me from day one and opened all the doors for me. Her lessons from early on had taught me to be a better statistician. Ed, who made me a better programmer and researcher. I came out stronger because of Ed's constant push for better coding practices and frequent challenging requests.

To Afsie, Sam and Robert, who always made sure that students come first. The graduate studies team's pivotal support and assistance in various matters ensured I was always on track.

To the Pharmaco- and device- epidemiology team. A team of excellent researchers who support and care for each other, and who always find new ways to have fun. This team has made the intense DPhil years bearable and even, at times, joyful. To the friends I picked along the way from this awesome team and with whom I enjoyed trips, game nights and fun experiences: Annika, Cheryl, Marta, Paloma, Danielle, Marti and Albert.

To my beloved Sudan. We yearn for the day we can hear the songs of your birds in the morning, feel your burning sun in the afternoon, and dance by your Nile in the evening. We miss your desert and palm trees, ripe mangos, sour hibiscus and tasty baobab. We miss the tales and wisdom of our elders, and the joyous laughter of our children. We long for the day we come

back and breathe in your air. Until then, we dedicate everything we have and everything we do to you.

To my fun, loving, amazing family. To the endless loud babble, banter and laughter we create whenever we are together. My dad, who loves us immensely, believes in the power of education, and is always, always in my corner. My mom, the one I always run to in all my moods, be it distressed, happy, confused, bored or just chatty. Waddah, who came to life when I was only 18 months old, became my best friend and has not left my side ever since. Lamees, my beautiful, sweet, caring and fashionable sister, who showed me how much fun life with a sister can be. Ahmad, the youngest and most important member of the family, a beautiful soul, a good cook, and the best laughter buddy. Lastly, Mazar, the beautiful sister-in-law and close friend our family is blessed to have. I dedicate this work to all of you with love.

To the loving, joyous extended family. The ones I grew up with in Riyadh; Aunties Awatif, Fathiya, Nawal, Awadeya, Bidoor and Ola and Uncles Kamal, Ahmed, Hassan, Yassir, Hussein and Amir. Uncles and aunties in Sudan, in no particular order; Souad, Amna, Enas, Eman, Aymen, Ehab, Khalaf Allah, Adel and Yassir. The family I found here; Sidahmed, Howida and their three lovely daughters. To my beautiful granny Om Al-Khair, whose prayers always find me in the most desperate times. Lastly, to my grandfathers Ibrahim and Al-Abbas, and granny Sit Al-Banat, who passed away, but their spirit, kindness and love surround and guide me every day.

To childhood friends and cousins: Amira, Omnia, Khalid, Abdullah, Sameer, Yaqeen, Mazin, Abbas, Ahmed, Abdallah, Mohamed and Nima. To the awesome Uni friends; Mohammed Hassan, Maysa, Noosa, Salma, Amro, Qutoof, Maad, Baraa, Hajir, Mutassim and Hussam. We met young and abroad, life scattered us all over, but the friendship remained tethered.

To Paula, who I met six years ago and has become an important chapter of my Oxford story, telling stories of friendship, happiness and growth. To Mohamed Hassan, a friendship I made when I was 19 that forever brought sunshine to my life. To Maysa and Noosa, we were inseparable in Uni, and despite life forcibly parting us, we remain as close as ever. To Amira, my childhood best friend, with whom I always share my life details, be it shallow or deep. To Sameer, the cool, loving and helpful cousin I am fortunate to have. To Mazin, despite distance and ocean separating us for most of our lives, we always evolve in sync and find time to laugh together. To Marwa and Izzy, living with them was fun and nurturing. To Noon and Abdelrahman, who opened my eyes and awareness to a much deeper understanding of life. To Mohamed Mahmoud, whose regular check-ups and words of encouragement have become a lifeline to me.

To everyone who has ever crossed my path and shown me kindness, I am blessed to have experienced love in all its forms. For that, I am now a better person and researcher.

To note, the systematic literature search conducted in Section 2.3 was performed with the support of Elinor Harriss, a knowledge centre manager and outreach librarian at the Bodleian Health Care Libraries, University of Oxford. Annika M. Jödicke, Ying He, and Danielle E. Robinson provided support to the research completed in Chapter 4. Alan Silman, Danielle E. Robinson and Andrew Clegg provided support to the work completed in Chapter 5. Lastly, the analysis in Chapter 7 was run with the help of Ross D. Williams who was my liaison at the Erasmus University Medical Centre. To everyone who contributed to the work done in this thesis, I have learned a lot from you throughout the process, thank you.

Funding

The work on this dissertation was funded by the European Health Data and Evidence Network (EHDEN). I am grateful to EHDEN for their support throughout the DPhil.

Abbreviations

ATC	Anatomical Therapeutic Chemical
COPD	Chronic Obstructive Pulmonary Disease
CPRD	Clinical Practice Research Datalink
eFI	electronic Frailty Index
EHDEN	European Health Data and Evidence Network
EHR	Electronic Health Records
ELSA	English Longitudinal Study of Ageing
ETL	Extract, transform, and load
GEE	Generalised Estimating Equations
HES	Hospital Episodes Statistics
HR	Hazard Ratio
IR	Incidence Rate
IPCI	Integrated Primary Care Information
IQR	Interquartile range
KmL	K-means for Longitudinal data
LTC facilities	Long-Term Care facilities
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
OHDSI	Observational Health Data Sciences and Informatics
OMOP-CDM	Observational Medical Outcomes Partnership Common Data Model
ONS	Office for National Statistics
PP	Point Prevalence
RWD	Real World Data
RWE	Real World Evidence
SD	Standard deviation
UTS	Up-To-Standard

Table of Contents

Table of Tables

Table of Figures

Table of Appendix Tables

Table of Appendix Figures

1. Introduction

1.1 Chapter summary

In this chapter, I describe the context for the studies conducted in the thesis and lay out the research aims with a brief description of each chapter's content.

Older people with complex health needs are usually avoided when conducting clinical trials due to their age and vulnerable health status, exhibited by multimorbidity, frailty and use of multiple drugs (polypharmacy). However, it is this very population that makes up the highest usage of healthcare resources, including the new treatments authorised after clinical trials. Their exclusion from clinical trials results in guidelines addressing the use of single or closely related treatments, leaving a gap for the multimorbidity patients who accumulate drugs and are at higher risks of further comorbidities, more medications, and adverse health outcomes. To tailor the right treatments for this population, accurate definitions of older people with "complex health needs" need to be formed. These definitions should consider the intersectionality of the health markers commonly used to identify complex health needs. They should also consider how monitoring these health markers over time would offer an advantage over the usual cross-sectional values as they are cumulative measures likely progressing over time.

The first study in this thesis aimed to assess and characterise how three different health markers, hospitalisation, frailty and polypharmacy, overlap with each other. Subsequent studies then aimed to describe the progression of frailty and polypharmacy over time and how this progression is associated with mortality. Lastly, the generalisability of identified clusters was assessed by internally and externally validating these clusters.

1.2 Background

Over the last century, life expectancy has increased in most countries worldwide[1]. According to the 2021 census, the number of people aged 65 and over living in England and Wales rose from 5.3 million (10.8%) in 1950 to more than 11.9 million (18.6%) in 2018[2 3]. While the number of years lived increases, the number of health years remains constant. This means that older people live longer but with a lower quality of life[1].

As people age, the accumulation of chronic and acute conditions results in multimorbidity. Multimorbidity is manifested by phenomena such as polypharmacy and frailty[4]. Polypharmacy is the use of multiple drugs, and frailty is characterised by sarcopenia (age-related decline in muscle strength and function) and gradual deterioration in body systems. They are common health markers of older people with "complex health needs" and are associated with adverse health outcomes.

Older people with complex health needs are often excluded from clinical trials but make up the lion's share of healthcare resources used[4-8]. Healthcare for long-term conditions - typically more common in older ages - accounts for approximately 70% of the UK National Health Service (NHS) expenditures[4]. A study found that those who are systemically excluded from new treatments' clinical trials commonly use these treatments and are at a higher risk of mortality[7]. For example, patients with cardiovascular or psychiatric conditions were the most excluded from anti-dementia drugs' clinical trials but had the highest prevalence of using these drugs. Those patients were also at a higher risk of mortality when compared to other anti-dementia drug users with other conditions/morbidities[7]. This was evident recently with the COVID-19 pandemic; older people accounted for 80% of COVID-19 deaths, were excluded from most vaccine clinical trials, and yet constituted the first and highest priority group for vaccine rollout[9-11].

As a result of the exclusion of people with concomitant medications and conditions from clinical trials, current clinical guidelines are available for treating a single health condition or a group of related health conditions (e.g., diabetes and hypertension)[12 13]. The presence of multimorbidity often leads to an accumulation of treatment prescriptions and, consequently, higher chances of further health complications[14]. There is controversial evidence on whether using multiple treatments, particularly certain preventive therapies, benefits the vulnerable older population with shorter life expectancy. For example, there is no consistent evidence on the harms vs. the benefits of continuing bisphosphonates for osteoporosis beyond three years[12].

The first step in tailoring the right treatment packages for older people with complex health needs is correctly identifying and characterising them. This can then help inform guidelines on the necessary interventions, whether pharmacological or non-pharmacological, to manage, reverse deterioration, or minimise the adverse outcomes associated with complex health needs[12].

The available evidence, while extensively studying the association between multimorbidity, polypharmacy and frailty[4 14-16], has yet to provide a definition of "complex health needs" that encompasses multiple health markers at the same time. Moreover, frailty and polypharmacy are cumulative measures that take time to develop. Their progression can start from early life, which would then increase the likelihood of worse status in later life[17 18]. Despite this, they are often studied as cross-sectional measures. Understanding their intersectionality and development journey, from middle to older age, and their association with adverse health outcomes can aid in the timely and correct identification and characterisation of older patients with complex health needs.

This can be achieved by leveraging real world data (RWD) to produce reliable and generalisable evidence. Real world evidence (RWE) is the evidence produced from using RWD

that goes beyond the available data from Phase III clinical trials, i.e., non-interventional data[19] [20]. RWD includes electronic health records (EHR), medical claims data, medication prescription and dispensation records, mobile health (digital epidemiology), and many other sources[20].

Depending on the purpose of collecting RWD, there are special characteristics, advantages, and challenges to be considered when using this data for research purposes. Claims data, for example, are mainly collected for insurance and payment purposes and may, therefore, contain fraudulent information or lack end-point clinical outcomes[20]. Nonetheless, using RWD offers many opportunities that typical clinical trials or epidemiological studies lack. They can be more representative of the population and produce reliable evidence on treatment effectiveness when used in clinical practice. They offer the means to apply data-driven approaches for finding patterns and classifications of patients in different disease-treatment areas. Furthermore, as these databases follow patients for extended periods, they provide a suitable source for conducting longitudinal analyses[20]. Recently, regulatory authorities have been increasingly using RWD to inform policymakers when making new regulations and recommendations[19].

1.3 Research aims and thesis structure

This thesis aimed to use RWD to characterise older people with complex health needs using different health markers, describing the progression of these health markers in older and middle age and the association between their progression and risk of death. To achieve this, I addressed the following research aims:

- Identify cohorts of older people with "complex health needs" using three cross-sectional measures: frailty, polypharmacy, and hospital admissions.
- Identify and characterise clusters (subgroups) of older people with distinct frailty and polypharmacy progression over time, and study their associated mortality.
- Understand the progression of polypharmacy from middle to older age by identifying and characterising clusters of middle-aged people based on their polypharmacy progression over time and associated mortality risks.
- Assess the generalisability of the findings by internally and externally validating the clusters in national and international databases using federated analytics.

The thesis structure, including chapters and contents, is described in Table 1.1.

Table 1.1 Thesis structure and outline

Chapter	Title	Content
1	Introduction	The background and context for studies carried out in the thesis are summarised, and research aims and objectives are described.
2	Literature review	A literature review of key studies on frailty and polypharmacy is summarised.
3	Data and methods	The data and methods used throughout the thesis are described.
4	Characterising complex health needs and the use of preventive therapies in older people	The results of a cross-sectional population-based cohort study to identify three cohorts of different health markers are described.
5	Longitudinal trajectories of frailty in older people, and their association with mortality	The results of a longitudinal study to identify clusters of older people with distinct frailty trajectories and their associated mortality risks are summarised.
6	Longitudinal trajectories of polypharmacy in older people, and their association with mortality	The results of a longitudinal study to identify clusters of older people with distinct polypharmacy trajectories and their associated mortality risks are described.
7	Generalisability of longitudinal trajectories of polypharmacy in older people: a validation study in the Dutch population	The results of an external validation study to assess the generalisability of the identified polypharmacy trajectories are summarised and compared to the original study.
8	Longitudinal trajectories of polypharmacy in middle-aged people, and their association with mortality	The results of a longitudinal study to identify clusters of middle-aged people with distinct polypharmacy trajectories and their associated mortality are described.
9	Discussion	The key findings of the studies carried out in this chapter are summarised and discussed in context with the currently available evidence. Also, the implications, limitations, scope for future research and conclusions are summarised.

2. Literature Review

2.1 Chapter summary

In this chapter, I summarise the available evidence on frailty and polypharmacy. Literature on both health markers has increased over the past few decades. They are commonly used as a proxy for health status and multimorbidity. The prevalence of both health markers has increased in older people. They are also associated with many adverse health outcomes like falls, fractures, and death. Longitudinal studies, although more established in frailty research than polypharmacy, still lag in addressing a few issues. In frailty, the methods applied either ignore patients who drop out before the end of follow-up, or produce average trajectories failing to address the heterogeneity of older people. In polypharmacy, all studies produced average trajectories, avoiding the possibility that it is a diverse population with different progression patterns over time. Additionally, it is likely that in such an old, multimorbid population, dropping out of studies is due to death. Therefore, it is important to describe the trajectories of frailty and polypharmacy, while accounting for drop-outs and assessing the association between these trajectories and the risk of death.

There was evidence on the overlap between polypharmacy and frailty, and how presence of polypharmacy is associated with an increased risk of frailty, and vice-versa. Studies also described the association between these two health markers and the risk of hospital admission. Characterising and understanding how frailty and polypharmacy overlap with each other, and with other health markers like hospitalisation, and whether one definition can encompass the other can be the first step before describing their variation over time. This would then help in the characterisation of these health markers' trajectories.

2.2 Frailty

Frailty affects around 10% of the older people aged 65 and above[21]. It is a progressive condition typically associated with older age, in which an individual suffers from gradual deterioration in different body systems. This leads to a state of vulnerability to stressors and an inability to recover from them quickly. This would materialise in dramatic changes like going from independent to dependent or losing stability and becoming more prone to falls[22]. In 2001, two landmark studies introduced the main two definitions of frailty that are currently being used in practice: the phenotype model by Fried et al.[23] and the Rockwood and Mitnitski cumulative deficits model[24].

A frailty phenotype, where frailty is a biological syndrome exhibited by a number of physical characteristics. The frail patient meets at least three out of the five following criteria: unintentional weight loss, weakness, exhaustion and tiredness, slow movement, and low activity level[23]. Those with one or two criteria were considered pre-frail, and those without any were non-frail. Frail people had more adverse outcomes compared with the non-frail, with the pre-frail falling in between those two categories[23].

A frailty index based on the cumulative number of deficits in an individual, including signs and symptoms, laboratory measurements, and history of illness and diseases[24]. In this model, it is the cumulative effect (ratio of existing deficits/total index deficits) rather than the nature of the individual deficits that provide the most predictive power[25]. In later studies, a guideline was given on how to consider the deficits making up a frailty index, e.g., the deficit's prevalence is age-related but also is a sign of a health condition and not a general age phenomenon (hypertension but not grey hair)[26]. It was also found that a frailty index comprised of 30-40 deficits can be as good as the original score in predicting the risk of adverse events[27].

A recent study found that both models provide equally good value in predicting the risk of mortality and hospitalisation among older frail patients[28]. Despite the good predictive value of the frailty phenotype, it is difficult to apply in clinical care because it relies on tests that are not commonly used in practice[29]. Furthermore, it does not take into account cognitive impairment, which is highly associated with physical disability and multimorbidity[22]. The cumulative deficits frailty index can be difficult to calculate directly during a patient's visit due to the complex mathematical concepts used to calculate it. However, once an algorithm is developed for a particular setting, such as electronic medical records, it is an effective and pragmatic approach to calculating frailty on broader scales[29]. An example is the electronic frailty index (eFI) developed and validated in UK primary care databases by Clegg et al. in 2016. The eFI was successful in identifying those at higher risks of hospitalisation and mortality[30]. eFI has been rolled out in the UK with recommendations for using it in 65+ patients to identify those with multimorbidity and at higher risk of adverse events[31].

Reviews on frailty

Since the frailty phenotype and frailty index were introduced, a large body of evidence studying frailty was produced. As a result, multiple reviews and systematic reviews focusing on these various aspects were published regularly. The published reviews provided a narrative[32] and a critical appraisal[22 29] of both frailty models, described frailty prevalence[21] and its association with adverse health outcomes[33-37] in older people, and discussed the management and intervention practices to tackle frailty[22 32 38]. They highlighted the lack of longitudinal studies accounting for heterogeneity in older people[28]. They also emphasised the need for tailored, comprehensive interventions for the management of frail multimorbidity patients in place for the current guidelines that only focus on individual diseases[32].

Systematic reviews have shown that frailty is highly prevalent in older people. Collard et al. noted that the prevalence varied widely across the 21 studies included in their review,

especially for those using a frailty index. Moreover, studies utilising a cumulative frailty index had a higher pooled prevalence of 13.6% vs. 9.9% in the frailty phenotype-based studies. This was likely due to the many variations of the Rockwood and Mitnitski index used in individual studies[21]. Systematic reviews consistently reported that frailty was more prevalent in women and more pronounced in older ages[21 25 32].

Many systematic reviews focused on the risk of adverse health outcomes such as hospital/care home admission, falls, fractures, and death in frail patients[33-37]. In their systematic review of 24 population-based studies, Shamliyan et al. found that frailty was consistently associated with a higher mortality risk. The risk varied with the frailty measurement; the average risk increase for frail patients was 50% when using the phenotype model and 15% when using the cumulative deficits model. These associations were strongest at 4 years and remained, albeit becoming weaker, up until 11 years after the measurement of frailty[34]. Another systematic review found that frail and pre-frail patients have a 90% higher risk of hospitalisation[33]. The studies included in these reviews assessed the short- and long-term risk of adverse events in frail patients based on single cross-sectional values, and the risks were expressed in terms of hazard or odds ratios.

According to Hoogendijk and Dent's review of the published studies looking at trajectories, transitions, and trends of frailty in the older population, research on changes in frailty over time has increased in the past 15 years[39]. The 2019 systematic review by Kojima et al. looked at 16 studies concerned with frailty transitions in older people. The systematic review only included studies using a 5-item frailty phenotype. The pooled results showed that 56.5% of the population would remain the same over an average of 3.9 years of follow-up, with 29.1% worsening and 13.7% improving. Of those who were non-frail at baseline, around 45.1% would transition to a worse frailty state during follow-up[40]. Welstead et al. systematic review included 25 studies using a continuous frailty measure and looking at its variation over time. They found

that the most common methods included hierarchical/mixed modelling and generalised estimating equations (GEE). These studies were concerned with how frailty trajectory is affected by different sociodemographic and other factors, and how the rate of frailty changes closer to death[41]. In the next section, I describe some studies on frailty transitions and trajectories, the methods employed, and key findings.

Longitudinal studies on frailty

The transition between different frailty statuses/categories

Studies concerned with the transition of frailty in the older population used the frailty phenotype definition. They found that up to 35% would have a transition in frailty state within 4-5 years[42-44], with the majority transitioning towards a worse frailty state. Previous frailty status was related to the likelihood of transitioning[42]. Factors, such as older age, female sex, smoking, obesity, presence of chronic obstructive disease, diabetes, cardiovascular disease, or osteoarthritis were associated with worsening frailty[43 44].

Trajectories and variation of frailty over time

Frailty trajectory studies used variations of the Rockwood and Mitnitski[24] cumulative frailty index with –mostly- 30 to 40 deficits[41]. The individual indices would have to meet the guidelines published in 2008[26], but were different for each data source, even if within the same study. The study population mostly came from community- or cohort-based data. Similar to other types of studies, longitudinal studies found that women had higher levels of frailty[45 46]. Others found that factors such as partner status, education, occupation class, and region/country of living have effects on frailty trajectories[45 47]. While most studies found that frailty had linear trends over time[41 45 48], some found that other non-linear shapes of frailty[49]. In general, non-linear increases were observed in studies over longer observation periods[50], with inflection points and accelerated trajectories observed after 65[51] and closer to death[49 52].

Longitudinal studies applied standard and novel methods to model the trajectory of frailty indices over time. The standard methods included GEEs[45], variations of random effects models such as linear mixed models[46 47 52], latent growth curve models[48 49 51], and latent class models[50]. More complex methods, such as machine learning-based clustering methods[53 54] or joint models[55 56], were also used.

A study that included participants aged ≥ 65 from the Longitudinal Aging Study Amsterdam (LASA) used GEEs to study frailty trajectory over 17 years of observation. The study found that the average time for frailty score to double (accumulating twice the number of deficits as baseline) was 12.6 years. The study also reported that older age, being female, having a lower education level, and having no partner were associated with higher baseline frailty value, but only partner status affected the change of frailty over time[45]. Although GEEs can give an average estimated trajectory on the population level, they do not consider or give information on the subject-specific effects.

Growth curve models and linear mixed models, while having a few differences and the preference of using one over the other depends on the nature of the data and the research question, provide similar analysis concepts. They allow for both fixed effects (population-level estimates) and random effects (between-subject estimates). The use of either method was popular for modelling frailty trajectories. When using latent growth curve models in a study analysing four cohorts of older people, Jenkins et al. found the linear model to perform better than non-linear in all cohorts with an average frailty increase ranging between 0.002 and 0.009 per year[48]. On the other hand, a cohort-based Swedish longitudinal study including people aged 29-103 followed up for 27 years using a latent growth curve model found an inflection point of frailty at the age of 65, beyond which, the frailty trajectory more than doubled with the average rate increasing from 0.014 to 0.038[51]. Moreover, in their retrospective study on terminal frailty in older adults from the US, Stolz et al. used a piecewise linear mixed model

and found that the rate of frailty was 5 times higher in the 3 years preceding death[52]. The linear vs. non-linear findings can be due to many reasons. The Jenkins et al. study included cohorts of older people (minimum average age 74) bypassing the inflection point of 65[48]. Growth curve models and linear mixed models also operate under the missing at-random assumption. Those who dropped out were more likely to be older with higher frailty levels which, if accounted for, could have led to steeper trajectories[46]. Another drawback of using these methods is that they estimate an average trajectory for the population under study. It is assumed that the older population is heterogeneous and that varying trends of frailty progression over time exist within this group.

The use of traditional clustering–an extension to the above-described models- and machine learning-based clustering methods provided a tool to address the heterogeneity problem. Using latent class trajectory modelling, a study analysing data from two US cohorts of older people over 12 years identified 4 clusters: relatively stable, mild, moderate, and severely frail. The mild cluster contained the highest proportion of the study population. The median time to accumulate an additional deficit in the stable cluster was 31 months vs. 20 months in the severe frailty cluster. Survival analysis was conducted after identifying the clusters and found that higher risks of mortality were associated with the worse frailty-level clusters[50]. Similar findings were found in the Chamberlain et al. study[53], which used K-means for longitudinal data (KmL), a machine-learning-based method for clustering longitudinal data[57]. They modelled frailty separately on decade-long age groups: 60-69, 70-79, 80-89. The model identified 3 clusters in the 60-69 group and 2 clusters for each of the other two groups. The study also found higher mortality and hospitalisation risks in the clusters with worse frailty levels/trajectories[53]. Both traditional and machine-learning clustering methods described here either require the same number of observations for all included patients, or assume that the data is missing at random.

The use of joint models in frailty studies confirmed that it is important to include and appropriately model those who leave the study early. The joint model presents a solution for this. It has two parts; the longitudinal sub-model, which uses a linear mixed model to describe the frailty trajectory over time, and the survival sub-model, assessing the association between rate of change and risk of death[58]. Two studies applied joint models and had consistent results: worse frailty trajectories were associated with higher mortality risks, independent of baseline value[55 56]. The trajectory of frailty over time was more valuable for short-term mortality prediction[55].

The studies summarised here highlight the limitations identified by Welstead et al. of the current evidence on frailty trajectories. The first was failure to address missing data and the use of methods that either impute or remove incomplete data[41]. While joint models successfully addressed the bias arising from not correctly accounting for the censored patients, they produced an average trajectory, ignoring the heterogeneity of the population confirmed by the clustering studies. Most longitudinal studies used cohort data that may not be representative of the overall population[41 50 51]. In studies using multiple cohort data, the lack of a uniform frailty index leads to varying results within and across studies[50]. The use of measures like eFI from UK primary care electronic health records (EHR) would solve both limitations, as eFI has already been validated for use and UK primary care records are representative of the overall population[50].

2.3 Polypharmacy

Polypharmacy is the use of multiple drugs either concurrently or cumulatively. It is common among older people and has been on the rise over the past few decades. Polypharmacy is commonly associated with multimorbidity and has been linked with an increased risk of mortality.

Definitions, epidemiology and prevalence of polypharmacy

There are numerous ways to define/calculate polypharmacy, depending on the study aims, the population being studied, and the nature of the data used. A 2017 systematic review has found 138 definitions of polypharmacy in the published research. The most common definition of polypharmacy is the use of 5 or more drugs within a specific period[59]. The use of 10 or more drugs is sometimes referred to as excessive polypharmacy. As described in the next section, the existing evidence has already established the association between these definitions of polypharmacy and an increased risk of adverse health effects. However, a category/binary-based definition may lead to a loss of information on the cumulative effect of each additional drug prescribed. This can be solved by using a count/continuous measure of polypharmacy[60]. Moreover, as there is an increasing trend of drug prescriptions in clinical practice, an arbitrary cut-off of ≥5 might no longer identify those with polypharmacy in the future[60]. Thus, data-driven approaches would accommodate the changing trends and correctly identify those with the highest levels of polypharmacy in the population while also considering the differences arising from studying polypharmacy in different populations, settings, and countries.

Besides the choice of the cut-off value to define polypharmacy, there are other considerations to calculate polypharmacy. These include the approach to counting drugs (e.g., active ingredient vs. drug class), the decision on what is considered multiple drug use (i.e., consecutive, concurrent or cumulative), and prescription duration (chronic polypharmacy, for example, means the use of medications or medications for 3 or 6 months)[59]. Each definition

offers advantages and disadvantages. Cumulative polypharmacy calculates the total number of drugs taken within a pre-specified time window (e.g., one year). It is more comprehensive and allows to account for the health risk of the excessive number of pharmacological agents entering the patient's body[61 62]. A concurrent definition refers to the simultaneous use of multiple drugs at the same time and offers a better assessment of the associated risks of poor adherence and drug-drug interactions[61]. Chronic polypharmacy can be used as a proxy to identify patients with chronic conditions[59]. Similarly with drug classes as opposed to active ingredients, where treatments targeting a single condition would count as one.

Some studies argued that the naïve approach of defining polypharmacy only by counting drugs/ingredients is not as useful as qualitative approaches such as appropriate/inappropriate polypharmacy according to different guidelines such as Beer's[63] or STOPP/START[64] criteria. While these methods are indeed more informative, the common definition of polypharmacy (≥5) offers a simple and good start to understanding prescription/dispensation patterns in different countries and settings. Nonetheless, the characterisation of the prescribing patterns (e.g., most common drugs/classes prescribed), and of those with polypharmacy (i.e., common comorbidities) is crucial for an overall understanding of how polypharmacy develops, progresses, and whether it can be managed.

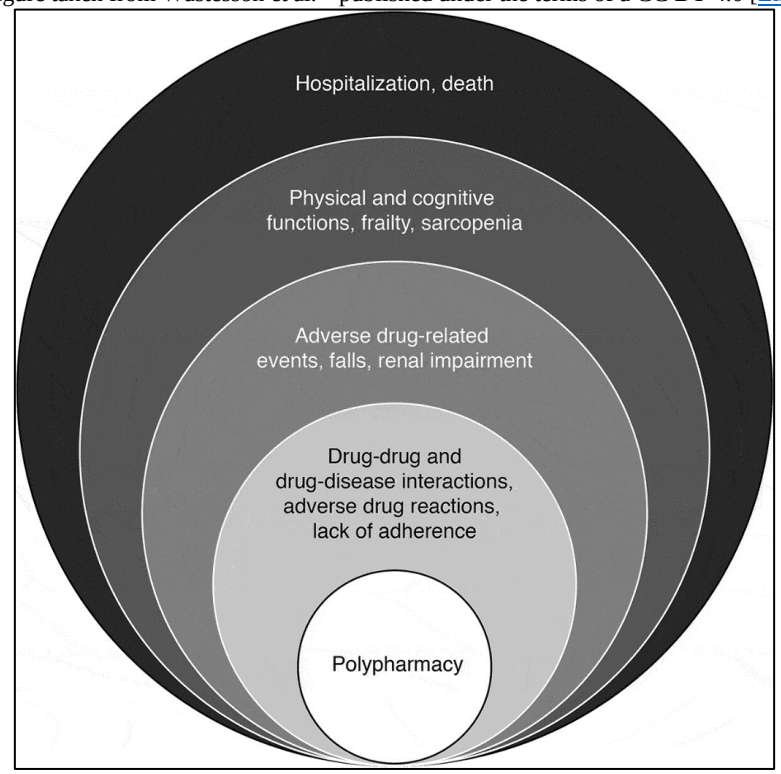
The prevalence of polypharmacy has been continuously increasing over the past 30-40 years. Guthrie et al. found that the proportion of Scottish adults dispensed 5-9 drugs increased from 9.7% in 1995 to 16.3% in 2010. A study of the Flemish-Belgian population found that there was a relative increase of over 80% between 2000 and 2015 in polypharmacy (use of ≥5 drugs) for people aged 75 and above[17]. Overall, it was found, in a 2018 study, that the prevalence of polypharmacy ranged between 26.3 and 39.9% for people aged ≥65 across 17 European nations[65]. In their narrative review, including studies from Europe, the USA, and New Zealand,

Wastesson and colleagues report that studies that include prevalence at two-time points have consistently reported an increasing trend of polypharmacy[66].

Adverse health outcomes associated with polypharmacy

Polypharmacy is associated with many adverse health outcomes, including drug-drug interactions, high-risk prescriptions, falls, and death[66-72]. It is also linked with other ageing phenomena like multimorbidity and frailty[14 16 73]. Figure 2.1 was taken from Wastesson and colleagues' narrative review. It describes how these adverse health outcomes are categorised based on their proximity to polypharmacy. Adverse effects like drug-drug interactions and adverse drug reactions are a direct consequence of polypharmacy[74].

Figure 2.1 Framework for polypharmacy and conceptual classification of outcomes.
Figure taken from Wastesson et al.[66] published under the terms of a CC BY 4.0 [Link]



Many clinical outcomes are indirectly linked to polypharmacy, with additional factors likely contributing to the increased risk[66]. In an analysis of participants from the English Longitudinal Study of Ageing (ELSA), the risk of hospitalisation due to falls increased by 75% for patients with polypharmacy and by three folds for those with excessive polypharmacy[70].

Polypharmacy is an indicator of multimorbidity. Current clinical guidelines provide advice on the treatment of individual diseases, sometimes multiple treatments to treat a single disease. Diseases like diabetes and ischaemic heart diseases lead to multiple drug prescriptions[14]. Polypharmacy is also associated with frailty, although the causal direction of this association has not been established[16 73]. A systematic review found studies reporting increased odds of frailty associated with polypharmacy (ranging between 1.77 and 2.55). The same review argues that, on the other hand, the presence of frailty is associated with unfavourable health outcomes and further morbidity accumulation, which can result in more drug prescriptions[16].

Polypharmacy is associated with mortality. A systematic review of 47 studies assessing the association between polypharmacy and death estimated a 31% increased risk of death for those using ≥5 medications compared to those without any medication use[75]. In the UK, another analysis using ELSA participants has shown an increased risk of all-cause and cardiovascular mortality among patients with polypharmacy and using high-risk central nervous system drugs[69]. However, due to the absence of clinical trials and the inconsistent findings some of the studies produced when attempting to adjust for the confounding by indication posed by polypharmacy, it was used as a proxy for health status in most studies[66 73].

With the increasing trend in polypharmacy among the older population, studies concerned with polypharmacy were frequently published. As a result, multiple reviews concerned with polypharmacy definitions[59], epidemiology[60], association with adverse health outcomes[66 74 75], and other ageing markers[14 16] are available in the literature. Nonetheless, due to the lack of studies modelling variation of polypharmacy over time, there were no reviews/systematic reviews of such studies. In the next section, I summarise the findings from a systematic search I conducted to identify studies on longitudinal polypharmacy.

Longitudinal studies on polypharmacy – a literature search

I conducted a systemic literature search to identify longitudinal observational studies that assessed variation of polypharmacy over time with/without a specific outcome.

The primary purpose of this systematic search was to find previous studies that investigated long-term trends/effects of polypharmacy with the following research questions:

- What were the methods used to assess longitudinal polypharmacy trends and how long did these studies take (follow-up period)?
- What were the findings of the longitudinal polypharmacy trend?
- Which outcomes were associated with longitudinal polypharmacy trends in these studies?

The librarian, Elinor Harriss, searched four databases from inception to 29/07/2021 for this search: PubMed; Ovid Embase; the Cochrane CENTRAL Register of Controlled Trials; and the Web of Science Core Collection (all editions). "A search strategy was constructed to search for the relevant thesaurus headings for polypharmacy and longitudinal studies, and free text terms were applied to the title or abstract fields to search for synonyms for both concepts. All references were exported to (Endnote X9, Thomson Reuters, New York, NY), and duplicates were removed following the method described by Falconer (2018) [link] before being imported into the Rayyan systematic review software for further deduplication and screening."

Polypharmacy-related terms were created with the aid of the narrative review by Pazan & Wehling[73]. A further description of the search strategy used in each database is available in Appendix 2.

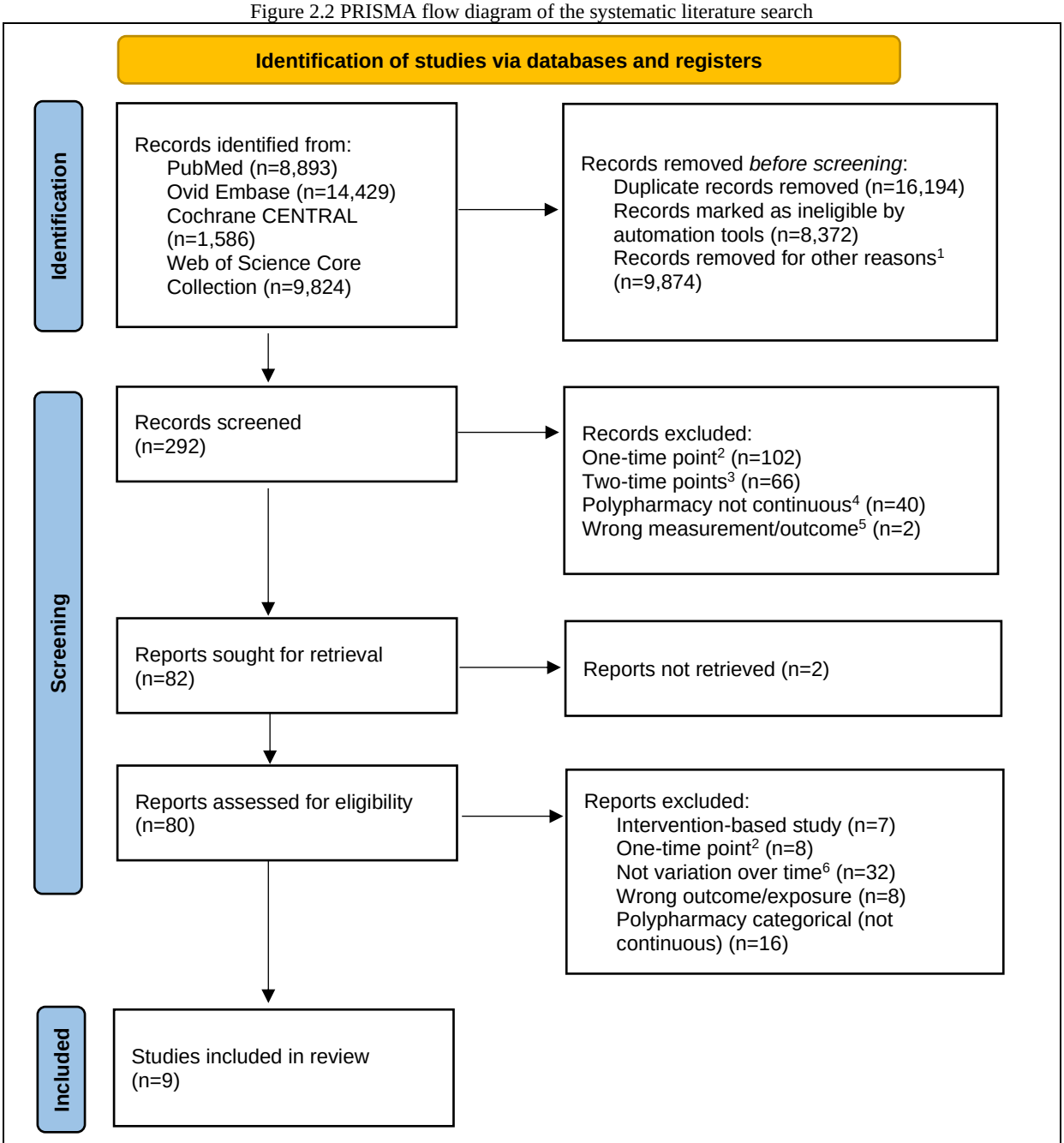
Table 2.1 describes the inclusion/exclusion criteria for including papers in this review.

Table 2.1 Summary of inclusion and exclusion criteria

Inclusion	**Exclusion**
Reference is in English	Reference is in a foreign language
Published journal articles and PhD theses	Conference abstracts/proceedings and MSc dissertations
In human studies	In animal and in-vitro studies
Multiple polypharmacy measurements	One or two polypharmacy measurements
Polypharmacy is count/continuous	Polypharmacy is a categorical measure
Cohort studies and clinical trials (with multiple data points)	Cross-sectional design (snapshot), case studies, systematic reviews, reports, protocols and case-control studies
Patient level studies	Population-level studies (i.e., incidence and prevalence)

A total of 34,732 references were retrieved. Out of which, 16,194 duplicates were removed. I used Rayyan's automated tools and scanned through the paper titles to exclude 18,246 articles according to the inclusion/exclusion criteria. I excluded 210 more articles after the abstract screening. Lastly, upon retrieval of 80 articles, a full-text screening resulted in the exclusion of 64 papers and 9 papers were included for the final review. Most of the excluded papers in the later stages were studies including polypharmacy association with different events based on a single measure, e.g., hazard and odds ratios, studies comparing between two-time points (before/after an event or an intervention), and studies monitoring the change of polypharmacy as a categorical measure and not as count/continuous measure. The PRISMA[76] diagram in Figure 2.2 describes the references excluded at each stage and the reasons for exclusions.

Figure 2.2 PRISMA flow diagram of the systematic literature search



[1] Reasons included (but were not limited to): wrong publication type, reviews, reports, protocols, wrong study design (e.g., cross-sectional), wrong drug (e.g., polypharmacy as a treatment strategy for lupus), studies not conducted in humans (e.g., in animals or in-vitro), wrong population, etc.

[2] Long-term effects (OR and HR), other types of regression analyses, one measurement across different age groups

[3] Before-after intervention, or before-after certain events (e.g., hospitalisation)

[4] Categorical polypharmacy, prevalence/IR of polypharmacy

[5] E.g., medication regimen complexity

[6] This includes comparing means, two-time points (before, on and after an event, linear regression between baseline and one-point follow-up, etc.)

Search results and discussion

Table 2.2 summarises the information extracted from this review's final set of included studies. The studies used patient information from EHR, cohort data, registers, and individual/multiple centres. Most studies included older people, and the follow-up ranged between 6 months to five years. Characteristics of the included populations varied based on the overarching aims and nature of data used in these studies.

In general, those studies found that being older[77 78], a female[77 78], institutionalised[77], or monitored by specifically trained nurse practitioners[79] was associated with slower trajectories of polypharmacy. Diagnoses with diabetes, hypertension or ischaemic heart disease, on the other hand, were associated with higher levels of polypharmacy[80], with cancer patients having the steepest slopes of polypharmacy in their last year before death[77].

Two studies focused on patients diagnosed with dementia[81 82]. The first followed-up recently diagnosed patients up to 3 years after diagnosis[81], and the second monitored polypharmacy retrospectively during the final year before death in demented patients[82]. They found that levels and progression of polypharmacy depended on the type of diagnosis. They also found that, in the last year before death, the number of prescribed drugs increased by 3% but dropped in the last month[82].

A few studies noted that those who dropped out during follow-up time had different characteristics from those with complete observations. They had higher baseline values, steeper trajectories, and were at a higher risk of death[81 83 84]. The study by Lavikainen and colleagues[85] focused on whether the levels of polypharmacy were associated with the probability of dropping out, and whether the estimates would be different if they accounted for this missing information under different assumptions. While they reported that latent growth analysis under a missing-at-random assumption was robust in generating intercept and slope estimates, they

noted that higher previous polypharmacy values and being older were associated with higher odds of dropping out. They also highlight the fact that low attrition is a recurrent challenge of longitudinal studies, including older people, as this population suffers from higher risks of multimorbidity and mortality. This makes it difficult to always adhere to the missing-at-random assumption in such cases.

The methods for modelling polypharmacy included random effects[78 81 83-85], joint models[81], GEE[77], and variations of regression analyses[79 80 82]. Depending on whether polypharmacy was considered a continuous or count variable, adjustments were made to the assumptions and variation of the model used, like using a Poisson mixed effects model to describe polypharmacy[82].

The longitudinal studies in this review reported important findings regarding the progression of polypharmacy in varying settings and populations and its association with many relevant factors. However, none of these studies conducted a clustering analysis to address the heterogeneous older people. Furthermore, while these studies described polypharmacy progression closer to death or in those who dropped out, they did not provide direct association estimates between polypharmacy progression and the risk of death.

Table 2.2 Summary of included studies

Author, year	Data source, country	Population	Follow-up time and mode of polypharmacy calculation	Methods	Results and main message
Agogo, Ramsey et al. 2018[81]	National Alzheimer's Coordinating Center. The USA.	Patients aged ≥65 with an incident dementia diagnosis (3 types of dementia).	3 years from diagnosis. Polypharmacy was calculated annually.	The analysis considered complete cases and the whole cohort. 1- Two mixed effects models; a Poisson and logistic mixed effects model. 2- A joint model to account for those with censored data points.	The risk of polypharmacy varied based on the type of dementia diagnosis. Point estimates were different when accounting for those who dropped out.
Denholm, Morris et al. 2019[82]	CPRD. The UK.	Patients dying between May 2013 and April 2014 who had incident diagnosis of dementia.	Patients were followed retrospectively up to 12 months prior to death. Polypharmacy was calculated at death and 2 weeks, 1, 2, 4, 6, 9 and 12 months before death.	Counts and means used to describe change over time. Zero-inflated Poisson to describe changes between time points and the effect of clinical and demographic characteristics.	The number of prescriptions increased by 3% between one year to one month before death, then dropped within the last month before death. Palliative care increased. Women, older patients and patients with Alzheimer's diagnosis had lower polypharmacy counts.
Denholm, Morris et al. 2020[78]	CPRD. The UK.	Adult patients admitted to hospital in 2014 (only emergency admissions). Only patients surviving beyond 6 months were included.	Patients were followed until 6 months post-admission. Polypharmacy was calculated at admission, 4 weeks, and 6 months post-admission.	1- Descriptive analysis; mean prescriptions before vs. after hospitalisation, and every data point.	Number of prescriptions increased 4 weeks after admission but then decreased by 6 months. More decreases were observed in patients who were older, diagnosed with

				2- Multi-level linear regression between every two data points.	more than 3 comorbidities or over 7 medications at admission.
Kilpatrick, Tchouaket et al. 2020[79]	Six long-term care facilities (LTC). Canada.	All residents in LTC facilities between September 2015 to August 2016.	Patients were followed up for 12 months, each month constituting a period. The number of prescriptions per day averaged over each period.	Unadjusted bivariate analysis to produce trends of average medicines per day per resident. Analysis was compared between sites.	There was an average decrease of 12% over the follow-up time. Changes varied between sites with more drops observed in sites that had nurse practitioners who regularly monitored the residents, made autonomous decisions and collaborated with pharmacist.
Lavikainen, Leskinen et al. 2015[85]	The Geriatric Multidisciplinary Strategy for the Good Care of the Elderly study	Cohort participants; persons born before 1 November 1928 and living in the city of Kuopio, Finland.	Participants were followed up between 2004 and 2007 with four polypharmacy measurements taken every year	Latent growth curve models, then logistic regression under the ignorable (missing at random (MAR)) and non-ignorable (missing not at random (MNAR)) assumptions. Logistics models calculated the odds of dropping out due to previous (MAR) and current values (MNAR)	Latent growth curve models were robust with the MAR assumptions. When only considering the current measurements, they were significantly associated with higher odds of dropping out. Age increased the risk of dropping out of the study.
Morin, Vetrano et al. 2017[77]	Death certificates linked to prescribed drugs, social services, national patients and education registers. Sweden.	Adults aged ≥65 dying between 2007 and 2013 with available cause of death prescription (3 months before death) information.	Patients were followed retrospectively until 12 months before death. Polypharmacy was calculated monthly.	GEEs to identify factors associated with changes in the number of drugs over time. Factors considered were sex, age, education, place of living, and morbidities.	Patients who died from cancer had the steepest increases. Being older, female, or institutionalised was associated with slower trajectories of drug accumulation.

Veehof, Stewart et al. 2000[80]	Three general practices in North Netherlands. The Netherlands	Adults aged ≥65 with complete observations over the study follow-up.	Patients were followed up for four years. Polypharmacy was the long-term use (≥ 60 days) per quarter per year. The first quarter was a wash-out period.	Means of multiple regression to assess the positive predictive value of age, sex and chronic diseases. No statistical method to describe the trajectory of polypharmacy, means of change over time are instead plotted.	The mean number of drugs increased from 2.6 in the second quarter of the study to 3.6 at the last measurement. Age, number of long-term drugs at baseline, presence of diabetes, hypertension and chronic ischaemic diseases, and use of drugs without clear indication had positive predictive value for the number of long-term medications by the end of the study.
Wastesson, Oksuzyan et al. 2017[83]	The Danish 1905 cohort. Denmark.	All Danish born in 1905 and responded to the first wave (baseline) survey in 1995.	Participants were followed up for five years (1998-2003). Polypharmacy information was collected in 3 waves after baseline: 2000, 2003, and 2005	Multilevel growth models for within-person variation. Ordinary linear regression to estimate the population-level variation. Inverse probability weighting to adjust for drop-outs, but not when drop-out is due to death.	Women had a higher number of prescriptions in every wave. Patients leaving the study early, either due to death or drop-out, had higher baseline prescription levels and steeper trajectories.
Wastesson, Rasmussen et al. 2017[84]	The Danish 1905 cohort. Denmark.	All Danish born in 1905 and responded to the first wave (baseline) survey in 1995.	Participants were followed up for five years (1998-2003). Polypharmacy information was collected in 3 waves after baseline: 2000, 2003, and 2005	Mixed-effects linear regression was used to calculate the within-person variation and compare changes in drug use over time for responders, non-responders, and partial responders.	Similar differences in drug use between the three responding groups. The partial-responders had the highest mortality rate.

2.4 Summary of key findings

In this chapter, I summarised the literature available on frailty and polypharmacy. Both are commonly used health markers to describe multimorbidity and complex health needs.

I identified the two key definitions of frailty. The phenotype is a measure based on the assessment of physical characteristics. The frailty index is a cumulative measure based on the accumulation of different deficits. Numerous studies have focused on both definitions in the last two decades. Individual studies and reviews compared the two definitions, described the prevalence and progression of frailty in older people, and assessed the association between frailty and adverse health outcomes like falls, hospitalisation, and death. The longitudinal studies describing frailty variation over time produced important findings. They found that frailty progression is affected by age, sex, and other lifestyle factors, and that it accelerates at an older age and closer to death. They also identified different trajectories of frailty in older people. Despite the voluminous evidence on frailty trajectory and the various methods applied, most studies used non-representative cohort data. Additionally, the methods applied either produced average trajectories or ignored the patients who drop out early and who are likely to be at higher risks of mortality.

There was more heterogeneity in defining polypharmacy. Although the definition of using five or more drugs was the most commonly used definition, the way of counting drugs and defining multiple drug use varied widely depending on the data source and aims of individual studies. Polypharmacy was also associated with a number of adverse health outcomes. There is still controversy over the extent of these associations and the causal directions between them and polypharmacy. I carried out a systematic search to identify studies describing the trajectory of polypharmacy over time. The final set of studies was comprised of only 9 studies. The included studies described the variation of polypharmacy in certain populations of interest, like people diagnosed with dementia, people in LTC facilities, or during the last year prior to death. These

studies found associations between polypharmacy trajectories and many factors, including age, sex and the presence of certain morbidities. Since the longitudinal research in polypharmacy was far less than in frailty, the same gaps have risen. There was a lack of studies identifying subgroups of polypharmacy trajectories. Also, existing studies recommended further investigating patients who drop out during study follow-up. Lastly, most studies included, and focussed on older people, although polypharmacy can develop from young and middle age[17 18 72].

The literature discussed the association between polypharmacy and frailty and found that the presence of polypharmacy is associated with increased odds of frailty. However, as with other adverse events associated with polypharmacy, the causal direction is still unclear. Moreover, many studies found associations between frailty and polypharmacy, and the risk of hospital admission.

Findings from this chapter highlight the need for longitudinal studies investigating the trajectories of frailty and polypharmacy over time and the association between these trajectories and the risk of death. Before carrying out these studies, it is important to understand how these two health markers overlap with each other and with other health markers and whether the use of one definition encompasses the other. This can then aid in the characterisation of the identified trajectories beyond the single measure used for modelling.

3. Data and methods

3.1 Chapter summary

In this chapter, I describe the data and main methods used throughout the thesis.

I used RWD to carry out all analyses. These included primary care records extracted from the UK Clinical Practice Research Datalink (CPRD) GOLD and Aurum, and the Dutch Integrated Primary Care Information (IPCI). The Observational Medical Outcomes Partnership-Common data model (OMOP-CDM) is a common data model that aims to transform the different data sources into one format with the same structure, content, and vocabularies. Thus, it allows for the replicability of studies across different databases and settings. CPRD GOLD and Aurum, and IPCI have been mapped to OMOP-CDM. Source and mapped instances were used for different analyses in Chapters 4 to 8.

The methods section outlined how the three health markers (unplanned hospitalisation, frailty and polypharmacy) addressed in this thesis were calculated. It also describes considerations made for retrieving death information, as all-cause mortality was the outcome of interest for all longitudinal analyses. Lastly, three longitudinal methods were described and briefly explained. Different combinations of these methods are used in Chapters 5 to 8.

To assess the quality of the data mapped to the OMOP-CDM, I ran a comparison study in the last part of the chapter. In this study, I used source and mapped instances of a CPRD database to identify a polypharmacy cohort. The cohorts identified in the study had good validation metrics and similar demographics. This study proved the reliability of the OMOP-CDM to produce results similar to those coming from the original source format.

3.2 Data

In the UK and in other countries worldwide, primary care is the first line of clinical care for patients. Primary care data are routinely collected from patients, and the records are regularly updated. This makes these data especially rich, allowing for population-based longitudinal studies. All analyses run in this thesis use EHR collected from primary care databases. In this section, I describe the three primary care databases I used. I also explain how these databases were mapped to a common data model that allows for the use of multiple databases to run large-scale network studies.

Clinical Practice Research Datalink (CPRD)

CPRD is a data provider that collects fully coded anonymised EHR from UK primary care practices. These data are representative of the UK population demographics (i.e., sex, age and ethnicity). CPRD records the reasons for all consultations in primary care but also includes information on laboratory tests, other assessments, diagnoses, immunisations and prescriptions[86]. CPRD data can be linked to the Office for National Statistics (ONS) mortality data, small area socioeconomic data, Hospital Episodes Statistics (HES) datasets, and patient registries for disease treatment like the National Cancer Intelligence Network[86]. However, CPRD linkage is only available in England and does not cover patients registered in Scotland, Wales, and Northern Ireland.

There are two databases contributing to CPRD: GOLD and Aurum[86 87]. CPRD GOLD contains data from General practitioners (GP) practices that use Vision® software while Aurum's data come from GP practices using EMIS Web® electronic patient record system software. Aurum's coverage includes practices from England only. When it was launched in 2019, Aurum contained data from >19 million of the current and historical English population (patients acceptable for clinical research), with 7 million active patients representing 13% of the English population. CPRD GOLD -previously known as VAMP, GPRD and CPRD- was

launched in 1987. By September 2019, it contained data from 17.5 million acceptable patients from all four UK nations, with 2.8 million active patients representing 4.3% of the UK population[88]. However, this coverage has changed dramatically in recent years, with many GP practices migrating their data to Aurum, leaving GOLD to cover only seven English practices. As a result of this migration, some GPs have contributed data to both databases, leading to a small proportion of overlapping practices [link].

In CPRD databases, each patient is labelled as 'acceptable' for use in research by a checking process that identifies individuals with reasonable demographics, continuous follow-up and valid clinical records. Moreover, in CPRD GOLD, each practice is associated with an up-to-standard (UTS) date, which represents the starting point when data in the practice are considered to have continuous high quality and, therefore, are suitable for research purposes[89].

CPRD linkages with ONS mortality data

The Office for National Statistics collects records from local registration offices in England and Wales, which carry out a service for registering information on certain life events (including births, civil partnerships and deaths), in partnership with the General Register Office. ONS provides data in line with Codes of Practice for Statistics guidelines [90]. The data held in ONS mortality register come from medical death certificates and include information on death place and date, death registration date, causes of death, age, sex, and area of residence, along with unique identifiers allowing for the linkage to other datasets (e.g., CPRD primary care). The linkage itself is performed by a trusted third party (i.e. NHS England) and the information released to CPRD and to researchers is completely anonymise and comprises only death date and causes of death, when available[91 92].

ONS mortality data is considered the gold standard when linkages are available for CPRD. However, there are a few limitations to consider when using this data. The main limitation is

the delay between death occurrence and registration. While the recommendation is for a death to be registered within 5 days of occurrence, the median reporting period has increased to 7 days in 2022[93]. Longer delays are usually expected when the cause of death is unexpected or suspicious, requiring an investigation (like deaths from accidents, suicides, drug abuse, etc.). These delays can take up to years; in 2021, 15.6% of deaths due to external causes took over a year to be registered[94]. However, these accounted for only 1% of all deaths registered in that year. Other limitations of using ONS mortality data include the data linkages being available to only a subsample of the CPRD population (England only) and over a limited linkage period. Moreover, ONS mortality data are incomplete for a small number of patients[92].

Throughout this thesis, when CPRD data were used with linkages to ONS mortality data, a stringent process was applied following the guidelines for consolidating CPRD with ONS[92].

The Integrated Primary Care Information (IPCI)

IPCI is a Dutch primary care database. It contains anonymised longitudinal data from General practitioner (GP) records. IPCI data contains information on patient demographics, symptoms and diagnoses, laboratory tests, prescriptions, and correspondences with secondary care. IPCI was started in 1992 by the Department of Medical Informatics of the Erasmus University Medical Centre in Rotterdam. The currently available IPCI data come from 2006 onwards, as the number of practices joining IPCI in its first decade was limited. As of July 2021, IPCI contains 2.5 million patient records with 1.4 million active patients, covering 8.1% of the Dutch population. IPCI is representative of the Dutch population, with its practices covering mostly the central parts of the country, but it also has good coverage of non-urban areas[95].

EHDEN, OMOP-CDM and Distributed Network Studies

The European Health Data and Evidence Network (EHDEN) initiative is a European-wide project that aims to harmonise different healthcare databases [link]. To achieve this, EHDEN

uses the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM). OMOP-CDM is a mechanism to standardise different patient-level observational data into one common data format having the same structure, content, and vocabularies. The transformation of data from multiple sites into OMOP-CDM allows for running distributed network studies. In network studies, analytical packages can be developed to run on multiple data sources remotely and in parallel, and the results produced can then be pooled and summarised. This allows health researchers to conduct their studies on different data sources in parallel using the same analytics, sharing only population-level data and protecting patient confidentiality. The OMOP-CDM allows for reproducible data flow, facilitates international collaborations, and produces robust, reliable, and generalisable health research output[96].

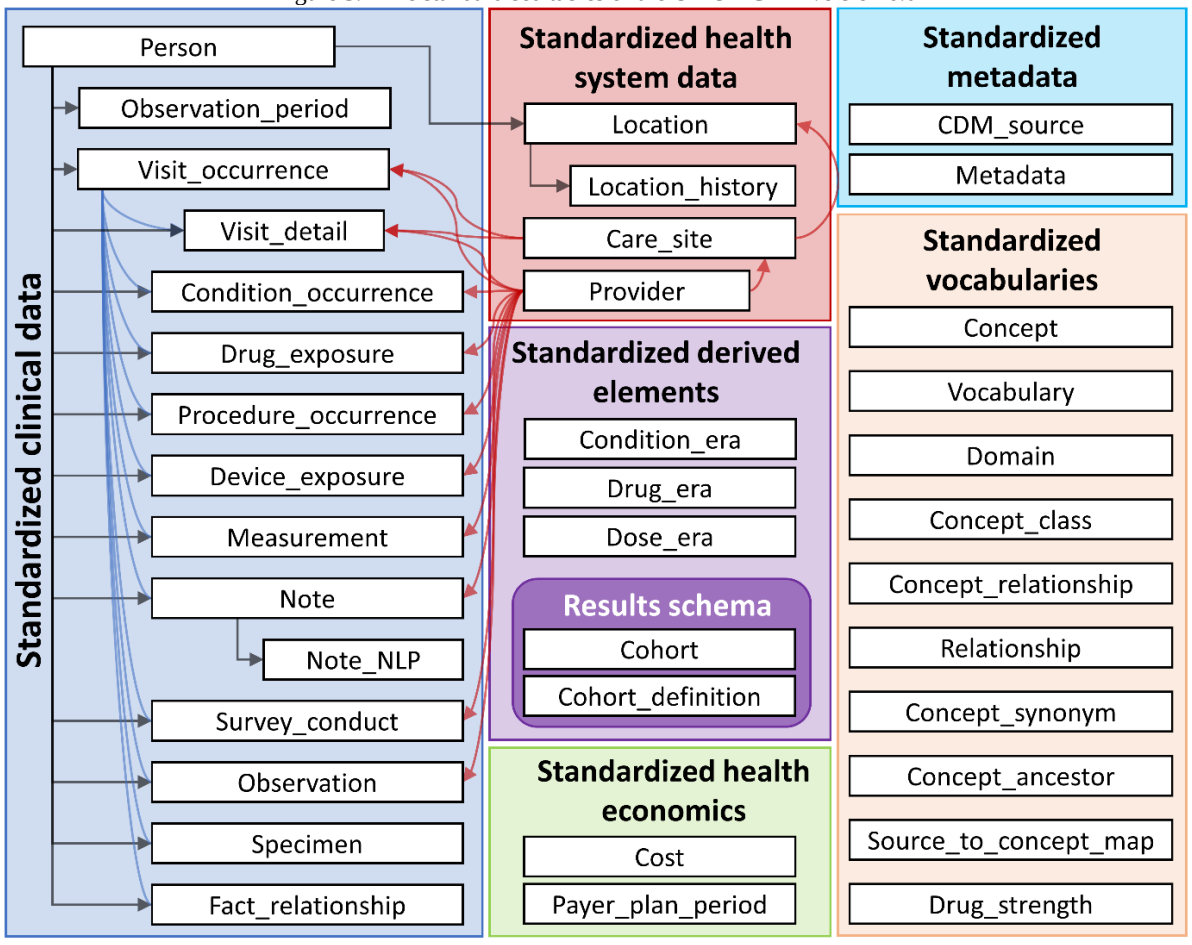
The raw data is extracted, transformed and then loaded (ETL) to the OMOP-CDM following the procedures established by the Observational Health Data Sciences and Informatics (OHDSI). Figure 3.1 describes the standardised tables in the OMOP-CDM. OHDSI provides a wide range of open-source analytical tools, including web-based interactive tools and R packages built for OMOP-CDM users. It also allows users to develop custom codes for specific analysis[96]. Due to the novelty and unique aspects of the studies in this thesis, a mix of OHDSI R packages and customised codes were used for analysis. Some of the main OMOP-CDM tables used for analysis were:

- Person table: includes information on person ID, practice ID, date of birth, sex, and race or ethnicity (depending on information collected in each database).
- Observation period table: includes information on the length of time a person contributed information to the EHR (start and end date).
- Death table: includes mortality information including date and cause of death.
- Drug era table: derived from drug exposure table, combines information about consecutive exposures of the same active ingredient.

- Condition occurrence table: includes information on diagnoses, signs and symptoms, and summary descriptions.

The level of granularity in these tables allows for the mapping of different types of EHR, regardless of the level of its richness. For example, some sources do not have mortality information but can still map their data to other tables in the OMOP-CDM. Additionally, clinical tables need to be used in conjunction with the standardised vocabularies tables to understand and aggregate similar concepts (e.g., ingredients and conditions) together. Lastly, it is important to note that the quality of any OMOP-CDM data relies on the quality of the source data and an error-free ETL process. More information about the OMOP-CDM can be found in the Book of OHDSI[96].

Figure 3.1 The standardised tables of the OMOP-CDM version 6.0[96]



By 2022, 453 databases from 41 countries were mapped to the OMOP-CDM[97]. With such a large community of researchers sharing their expertise and knowledge, collaborations were fostered more rapidly and efficiently. CPRD GOLD, Aurum and IPCI are mapped to the OMOP-CDM, and their data contribute to international impactful studies. Throughout this thesis, data extracted in traditional/raw format is referred to as source data, while the mapped data is the OMOP-CDM instance.

3.3 Methods

While every study had its own specific aims, unique data considerations and methods applied, there were common aspects throughout the thesis. This section describes how the health markers were calculated. It also describes KmL, joint models and joint latent class models, the methods applied for analysis in Chapters 5 to 8.

Health markers

Unplanned hospital admissions

This health marker was defined by the number of unplanned hospital admissions to emergency rooms (extracted from HES Admitted Patient Care and Accident and Emergency) identified in the linked HES data.

Frailty

Frailty was defined using the validated eFI score developed by Clegg et al.[98] and based on the count of 36 frailty markers/deficits as recorded using a pre-specified list of Read codes[98]. The number of zero was given for individuals who did not have any recorded consultations for any of the deficits. The list of eFI deficits are available in Appendix Table S 3.1. The Read codes were compatible with GOLD, and mapped to SNOMED CT codes for extraction from Aurum.

Polypharmacy

For source data, polypharmacy was defined by the number of different drug substances prescribed in a given year. For each patient, all prescriptions issued by GPs in the target year were identified and for these prescriptions, the drug substance was retrieved using the "PRODUCT" dictionary and "THERAPY" table in CPRD GOLD. Fixed combinations of multiple substances in products were counted as one extra substance.

In OMOP-CDM, polypharmacy was defined as the absolute number of ingredients taken during the target year. This information was extracted from the drug era table. In the 'drug era' table,

different prescriptions of the same ingredient—regardless of the dose or the combination—are joined together when the interval between these prescriptions is less than 30 days[96]. For each follow-up year, any drug era that overlaps is counted, regardless of the duration of the overlap. Multiple drug eras of the same ingredient counted as one.

Outcome

All-cause mortality was the outcome of interest in Chapters 5-8.

In the source data, linkages were available to ONS -the gold standard for UK death records- and mortality information was retrieved following the guidelines for reconciling ONS mortality records with CPRD[92].

For OMOP-CDM, mortality information was extracted from the death information recorded in the primary care databases, as linkages were not yet available in the OMOP-CDM instance. Death information is deemed reliable with ±30 days accuracy in CPRD[99], and are available on a monthly level in IPCI[95].

Statistical models

<u>K-means for longitudinal data</u>

KmL is an extension to the machine-learning classification method, K-means. While the original K-means identifies clusters based on single/cross-sectional values, KmL offers an attractive approach to classify longitudinal data into homogeneous distinct clusters. KmL offers advantages over the traditional longitudinal clustering/classifying methods. It does not require the data to be linear or parametric or to have a priori information regarding the shape of the trajectories. The Calinski & Harabatz criterion is usually assessed to determine the optimal number of clusters for a KmL model. Similar to traditional clustering methods, KmL does not account for censored data, and either excludes or imputes missing data[57].

Joint models

Joint models is a method that jointly models the longitudinal marker/exposure trajectory and the risk of having an adverse event. Joint modelling consists of two sub-models, a linear mixed model for the longitudinal marker, and a proportional hazard model for the risk of event. The hazard function at any time t is affected by the history of the longitudinal exposure up to that time[58]. Joint modelling accounts for the censored data (patients who drop out) and is used when it is assumed that the marker's variation over time is associated with the risk of an event. Joint models produce an average trajectory, assuming a homogeneous population.

Joint latent class models

Joint latent class models are an extension to joint modelling. It is a longitudinal method to describe trajectories of markers/exposures over time and their association with the risk of an adverse event. A joint latent class model assumes a heterogeneous population and uses latent class modelling to divide the population into homogeneous classes (clusters) with similar trajectories and risk of adverse events' occurrence[100 101]. The advantage of using this over other traditional clustering methods is that it accounts for censored data. Therefore, patients who died before the end of the observation period are included in the analysis, minimising survival bias. The joint latent class model has 3 sub-models:

- Class-specific marker trajectory: uses a mixed linear model to describe the longitudinal cluster-specific exposure/marker's trajectories over time. The linear mixed model has a fixed-effects part accounting for the cluster-specific average trajectory (population level), and a random part accounting for the individual-specific variation from the population's average intercept and slope.
- Class-specific risk of event: applies a survival model to calculate the class-specific baseline hazard/ hazard ratio. Only parametrised hazard functions are considered for this part, e.g., Weibull and splines.

- The latent class membership probability: uses multinomial logistic regression to assign individuals to the identified clusters by applying a probabilistic approach based on the observed data.

To determine the optimum number of classes, many considerations were made. These include statistical criteria like convergence and the smallest Bayesian Information Criteria (BIC), but take account of clinical plausibility, meaningful clusters, and size of the smallest cluster (≥1%).

For fitting the model to new data, the R-package used to build the joint latent class models- 'lcmm'[102]- has a function that allocates new data to the best fitting cluster based on the patients' characteristics (marker/exposure progression over time, covariates, and mortality information). Posterior probabilities, i.e., the probability of an individual belonging to any cluster based on their observed information, are usually assessed to inspect the internal/external validity of the joint latent class model once it is built. Good validation performance can be defined as the cluster-specific average posterior probabilities of belonging to the pre-defined clusters being ≥0.7 for at least half (50%) of each cluster's members. The threshold of 0.7 indicates a clear classification of people into clusters[103]. Other aspects should be considered to examine the model's validity, including the cluster-specific risk profile, and the clinical and demographic characteristics of the clusters resulting from fitting the new external data.

To characterise resultant clusters:

1. For frailty models: individual deficits at baseline and end of study follow-up were extracted and compared.
2. For polypharmacy models: comorbidities and drug use were compared across the clusters, at both baseline and last observation. These included the Charlson comorbidity index (and individual morbidities), and drug classes -according to the World Health Organisation's Anatomical Therapeutic Chemical (ATC) Classification level 1; pharmacological or therapeutic groups. Also, the most common conditions and drug exposures were identified using functions from 'PatientProfiles' [link], an R package developed to characterise patients in the OMOP-CDM.

3.4 Quantifying polypharmacy in older people: a comparison between source and mapped data instances of CPRD GOLD

Background

As explained in section 3.2, the use of mapped data in observational health studies has many advantages: these include, allowing for reproducible data flow, encouraging international collaborations, and producing robust, reliable and generalisable health research output.

In the case of identifying a cohort with polypharmacy in CPRD GOLD, certain challenges exist when using the mapped instance. One of these challenges is the method used to identify drugs intake/use, i.e., on the prescription/product level vs. the ingredient level, which could lead to different populations identified. Another obstacle is the availability of data linkages (e.g., hospital data) as HES linkages were not available in the CPRD mapped instance.

In this study, I aimed to:

1. Test the feasibility of identifying a polypharmacy cohort in older population using the OMOP-CDM mapped version of CPRD GOLD data comparing it to the polypharmacy cohort identified using the source data.
2. Compare the cohort characterisation between the two cohorts.

Methods

Data source

For source data, I used data extracted from CPRD GOLD (September 2019 release), with linkages to ONS and HES datasets (described in 3.2). For mapped data, I used CPRD GOLD (July 2022 release) mapped to OMOP-CDM.

Population

A matched cohort of eligible patients existing in both instances was created. For the source instance, all subjects aged >65 years on 1st January 2010, and acceptable for clinical research in CPRD GOLD who were registered with an UTS practice for at least 1 year and were eligible for HES linkage were identified (3.2). The patients whose patient IDs were found in the OMOP-CDM instance and could be matched to the identified source population constituted the eligible study population.

Exposure

Polypharmacy identified as the absolute number of substances/ingredients taken in 2009 was calculated following the definitions described in 3.3. Table 3.1 describes how the two modes of drug calculation (substance vs. ingredients) are considered in the two CPRD GOLD instances.

Table 3.1 Methods applied to identify polypharmacy cohort in Source CPRD GOLD and OMOP-CDM mapping

	Source	**OMOP-CDM**
Drug level	Substance	Ingredient
Example (paracetamol, aspirin, paracetamol + aspirin)	Three different substances	Two different ingredients
Example (paracetamol + aspirin)	One substance	Two different ingredients
Doses	Different doses count the same	Different doses count the same

Cohort identification

A data-driven approach was used to identify the cohorts. People belonging to the top quintile in terms of substance/ingredient counts made up the polypharmacy cohort in each instance.

Statistical analysis

Cut-off values, baseline characteristics and use of oral bisphosphonates as an example of preventive therapies in the year prior to start were summarised and compared between the two cohorts. Cohort validity – with source as the true classification – was examined using cohort evaluation metrics; sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). I used R (Version 4.3.2) for the analysis.

Results

Cohort definitions in source and mapped data

A total of 467,843 patients were matched between source and OMOP-CDM mapped data. The cut-off was ≥15 ingredients in the mapped data and ≥10 substances in the source data. Table 3.2 is a confusion matrix that summarises the cohort overlap between the source and mapped data along with the cohort validation metrics (source data is the true classification).

Table 3.2 Confusion matrix of the identified cohorts in Source and OMOP-CDM data

	Source (true classification)			
OMOP-CDM		**Cohort +**	**Cohort -**	**Total**
	Cohort +	81208	16209	97417
	Cohort -	32846	337580	370426
	Total	114054	353789	467843
Sensitivity = 71.2% Specificity = 95.4% PPV = 83.4% NPV = 91.1%				

Cohort +: Belongs to polypharmacy cohort.
Cohort -: Does not belong to polypharmacy cohort.

Cohort characteristics

Both cohorts had similar baseline characteristics (Table 3.3). One-year prevalence of oral bisphosphonates was 17.9% (95% CI 17.6% to 18.2%) in the mapped data and 17.1% (95% CI 17.4% to 17.9%) in the source data.

Table 3.3 Baseline characteristics of polypharmacy cohorts from both instances

	Source	**OMOP-CDM**
N	114,054	97,417
Age (mean (SD))	78.2 (7.4)	78.2 (7.6)
Female gender (n (%)	69277 (60.7%)	60015 (61.6%)
Charlson morbidity index	0.39 (0.91)	0.52 (0.98)

Discussion

In this study, I assessed the validity of using OMOP-CDM mapped data in comparison to the Source data (reference) for identifying and quantifying a polypharmacy cohort using a data-driven segmentation approach. The polypharmacy cohort identified in the mapped data had high similarity in all validity metrics when compared to the source data. People belonging to the polypharmacy cohort in both instances were similar in terms of demographics, morbidity burden and use of preventive therapy.

Sensitivity was slightly lower than other validity metrics. This could be attributed to the different methods applied in calculating polypharmacy in both instances which led to a higher cut-off in the mapped instance.

This study proved the clinical validity of identifying a polypharmacy cohort in the OMOP-CDM data. More research is needed to test the generalisability of using a cut-off of 15 ingredients to define polypharmacy in other OMOP-CDM data sources.

4. Characterising complex health needs and the use of preventive therapies in older people

4.1. Chapter summary

In this chapter, I identified older people with complex health needs based on three definitions: unplanned hospital admissions, frailty, and polypharmacy. The cohorts were identified from CPRD GOLD and were based on the distribution of these health markers recorded within one year. I characterised the three cohorts, and quantified point-prevalence and incidence rates of three common preventive medicines.

The three cohorts identified were older, had higher morbidity and preventive therapy use compared to the rest of the older population. While there was considerable overlap between the three cohorts, frailty and polypharmacy cohorts had the highest pairwise overlap. Most comorbidities such as diabetes and chronic kidney disease were more common in the frailty and polypharmacy cohorts compared to the hospitalisation cohort. As expected, prevalence of preventive medicines use was highest in the polypharmacy cohort compared to the other two cohorts.

This study highlighted the need for more research into the value of different definitions of complex health needs and the risk-benefit of preventive therapies in the older population. Frailty and polypharmacy are likely to accumulate over time, making cross-sectional measures like the ones currently used misleading. It would be beneficial to assess the value of using repeated measures of frailty/polypharmacy over time and how they are associated with adverse outcomes.

Note:

The content, methods, results, and discussions presented in this chapter are based on an associated publication:

Elhussein, L., Jödicke, A.M., He, Y. et al. Characterising complex health needs and the use of preventive therapies in the older population: a population-based cohort analysis of UK primary care and hospital linked data. BMC Geriatr 23, 58 (2023). https://doi.org/10.1186/s12877-023-03770-z

4.2. Introduction

With society growing older, the number of people suffering from multimorbidity, frailty and polypharmacy is rapidly increasing in the UK[72 104]. Around 70% of UK people aged 60-69 have multimorbidity, with that proportion rising to more than 90% for those older than 80[105]. Moreover, prevalence of polypharmacy in adults doubled in recent years, with the number of older patients receiving 10 or more drugs even tripled between 1995 and 2010[72]. People with multimorbidity are more likely to suffer from adverse events, receive more treatments and consume healthcare resources[4].

While the current guidelines address the use of single -or closely related- treatments, the presence of multimorbidity leads to a build-up of prescribed treatments. However, guidance on the simultaneous use of multiple treatments targeting different chronic diseases and considering the impact resulting from the accumulation of these treatments in an already vulnerable population is often scarce[13]. Aside from comorbidities, a patient's treatment requirement is influenced by functional status, preferences and life expectancy. Drug treatments targeting several single conditions often accumulate in older people, increasing the risk for drug-drug interactions[72] and adverse events. The gap in concrete evidence on the benefits and risks of treatment in older people leads to under- and over-prescriptions. For example, there is controversial evidence on whether or not the use of common preventive therapy, such as bisphosphonates, statins and anti-hypertensives is beneficial in the vulnerable older population with limited life expectancy and frailty.

The UK's National Institute for Health and Care Excellence (NICE) recognises the importance of improving health care for multimorbid older people. Its recently published multimorbidity guideline recommended using routinely collected data and electronic health records to identify older patients with complex health needs[12]. Their guideline suggests considering validated tools such as eFI[98] to identify multimorbid patients at risk for adverse events and unplanned

hospital admissions, as well as markers of polypharmacy to identify patients with high treatment burden. Whilst several definitions exist for multimorbidity, polypharmacy, and frailty, it is yet unclear to what extent cross-sectional single healthcare markers capture the complexity of health-related needs in older people in the community.

The objective of this study was to identify and characterise older patients with complex health needs, aiming to enhance further research into this important patient group. I selected patients based on three distinct health markers (unplanned hospitalisation, frailty and polypharmacy), compared their characteristics to the background population, and explored the overlap between these three cohorts. Subsequently, I assessed the utilisation of common preventive drugs in people with complex health needs.

4.3. Materials and methods

Design

In this study, I conducted a cross-sectional cohort analysis to identify three cohorts of older people with complex health needs based on their number of unplanned hospital admissions, frailty, and polypharmacy. I used a data-driven method where people with the top quintile number of records belonged to their respective cohorts. I summarised and compared demographic and clinical characteristics between the identified cohorts, their overlap, and the background population. Lastly, I described the use of three preventive therapies in the study population.

Data source

I used data extracted from CPRD GOLD (September 2019 release), with linkages to ONS and HES datasets (described in 3.2). These HES datasets will be referred to as CPRD-HES.

Population

All CPRD GOLD (September 2019) subjects with linkages to CPRD-HES at study start (01/01/2010), aged >65 years, acceptable for clinical research, who were registered with a UTS practice for at least 1 year comprised the source population (Refer to section 3.2). Patients exited the study at the earliest of practice last collection date, patient transfer-out of practice date, or death date[92].

Cohort definitions

Among the source population, I identified three cohorts of older patients with 'complex health needs' in CPRD-HES, as recommended by NICE guidelines[12]. The definition of the cohorts were based on the health-care markers described in 3.3, as follows:

a) Hospitalisation cohort: defined by the number of unplanned hospital admissions recorded in 2009.

b) Frailty cohort: defined based on the count of eFI deficits as recorded during 2009.
c) Polypharmacy cohort: defined by the number of different drug substances prescribed in 2009.

After examining the data distribution of these markers, patients in the top quintile (20%) constituted the respective cohorts. If there was no clear cut-off to determine the top quintile due to count data, the nearest division based on the data distribution, or a pre-existing clinical cut-off was chosen. Patients could be included in multiple cohorts at the same time. Patients who belonged to all three cohorts constituted the overlap group, whereas those who did not belong to any of the three cohorts comprised the background population.

Subjects included in any of these three cohorts of 'older patients with complex health needs' were eligible for the subsequent analyses of key preventive therapies use in their respective cohorts.

Utilisation of preventive therapies

The NICE multimorbidity guideline recommended research into continuing and stopping of three preventive therapies, namely bisphosphonates, statins, and anti-hypertensives, as those may potentially be avoidable in patients with limited life expectancy and frailty[12]. I evaluated the use of oral bisphosphonates, statins, and anti-hypertensives, separately in each cohort. Prescriptions of preventive treatments were identified based on product codes in CPRD GOLD. Anti-hypertensives were analysed as a group, and separately by drug class according to the WHO-ATC classification. I calculated treatment episodes by combining individual prescriptions, allowing for a maximum 90-day gap between prescriptions and with adding a 90-day washout period at the end of each episode.

Statistical analyses

I calculated one-month, three-month, and one-year point prevalence (PP) prior to start date to identify prevalent users of oral bisphosphonates, statins, and anti-hypertensives in each cohort, the overlap group, and background population, separately. Furthermore, I calculated treatment incidence rates (IR) in the first three years after start date having excluded one-year prevalent patients of each drug class for the same groups. I assumed a Poisson distribution for both analyses. Incidences were counted as the first prescription after start date and IRs were calculated per 1,000 person-years. For IR analysis, a patient would contribute to the sum of person-years until a prescription of the preventive therapy is reported. I used STATA (Version 15.1) for cohort identification, and R (Version 3.6.0) for cohort characterisation and calculation of PPs and IRs.

4.4. Results

Cohort identification

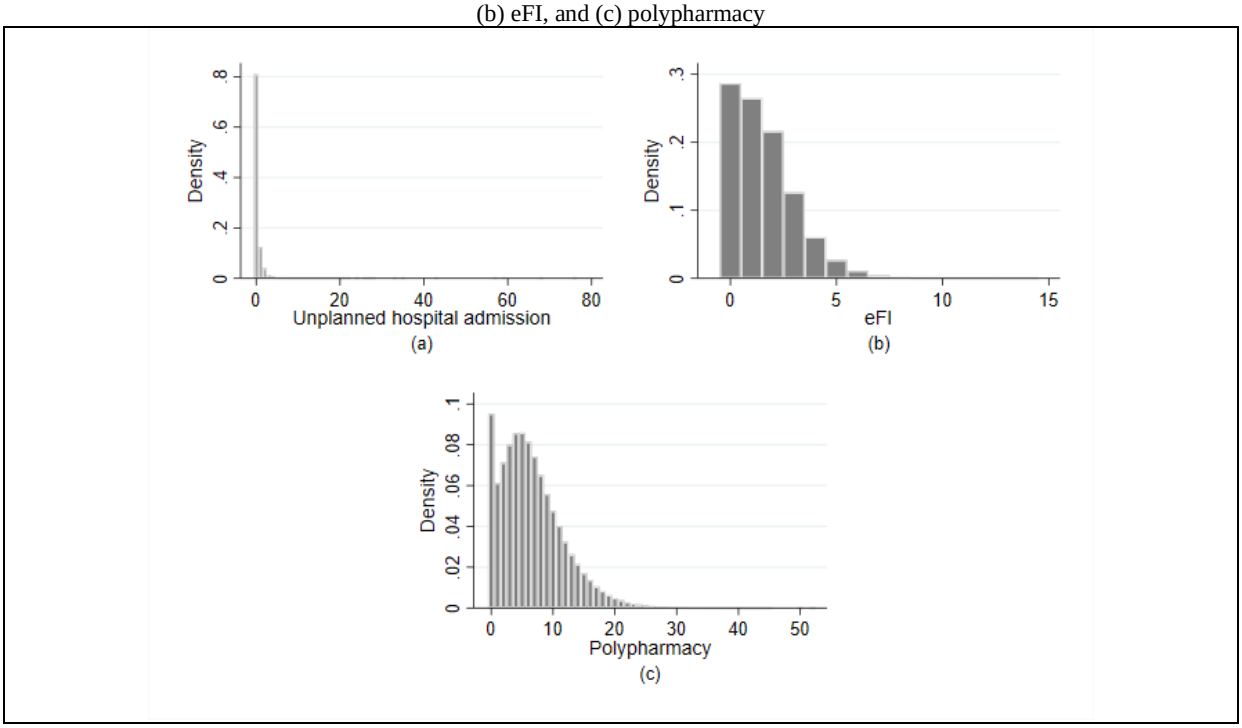
The cut-offs for inclusion in each of the three cohorts were: ≥1 unplanned hospital admission for the hospitalisation cohort, an eFI score ≥3 for the frailty cohort, and ≥10 medicines for the polypharmacy cohort. Table 4.1 and Figure 4.1 describe the distribution of these variables.

Table 4.1 Descriptive statistics of variables used to identify cohorts

Overall population (n=475371)						
Variable	Min	P20	P40	P60	P80	Max
Unplanned hospital admission[a]	0	0	0	0	1	76
eFI[b]	0	1	2	3	4	14
Polypharmacy	0	3	6	8	11	52

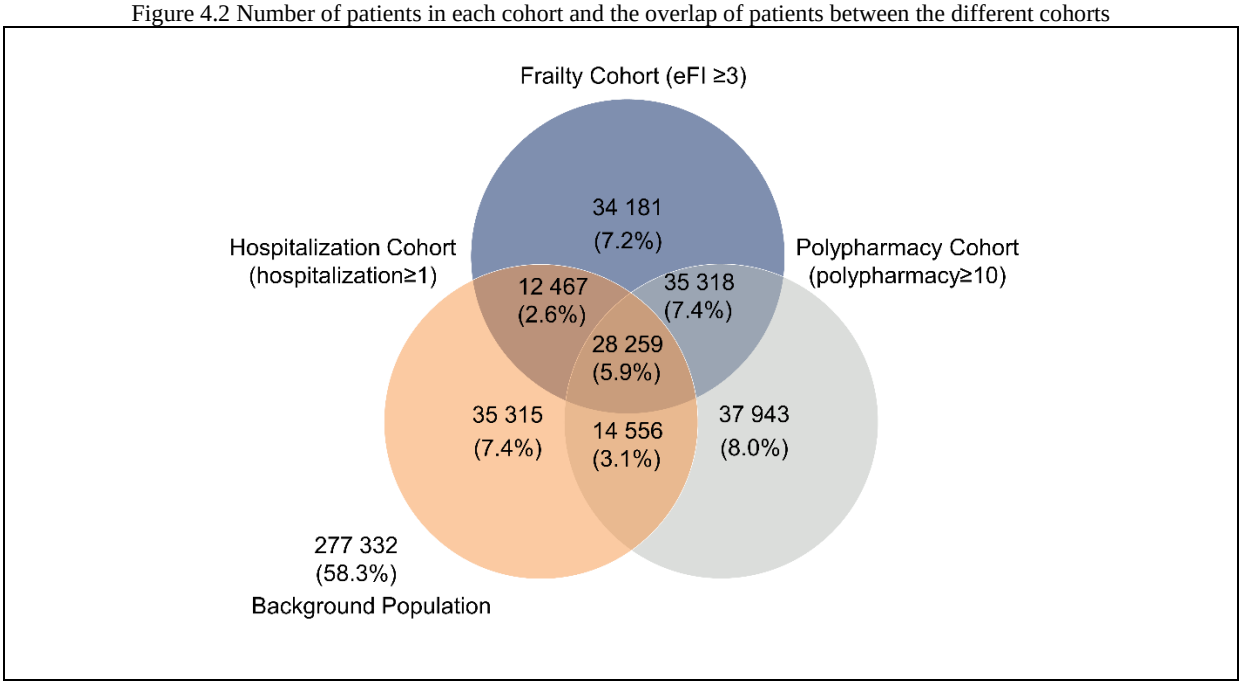
[a] 384 774 (80.9%) of the overall population did not have recorded unplanned hospital admissions in 2009
[b] 50 126 (10.5%) of the overall population had an eFI score ≥4 in 2009. A total of 110 225 (23.2%) had an eFI score ≥3 in 2009, hence, 3 was the cut-off chosen to identify the frailty cohort
P20, P40, P60, and P80: 20th, 40th, 60th, and 80th percentile.

Figure 4.1 Histograms of variables used to identify cohorts of complex health needs: (a) unplanned hospital admission, (b) eFI, and (c) polypharmacy



Of 475,371 participants who met the inclusion criteria, 90,597 (19.1%) were included in the hospitalisation cohort, 110,225 (23.2 %) in the frailty cohort and 116,076 (24.4 %) in the polypharmacy cohort, 28,259 (5.9%) people belonged to all three cohorts and 277,332 (58.3%) did not belong to any and were therefore used as a reference (background) older population with no complex health needs. Frailty and polypharmacy had the highest pairwise overlap (>50%) (Figure 4.2).

Figure 4.2 Number of patients in each cohort and the overlap of patients between the different cohorts



Cohort characteristics

Mean age was similar in the three cohorts and the overlap group (between 78.2 years and 79.7 years), and slightly lower in the background population (74.7 years). The proportion of males was lower in the three cohorts and the overlap group (between 39.3% and 43.3%) compared to the background population, (46.7%). Table 4.2 describes the baseline characteristics of the three complex health needs cohorts, the overlap group and the background population.

Table 4.2 Baseline characteristics of hospitalisation, frailty, polypharmacy cohorts compared to the whole population

	Hospitalisation Cohort	Frailty Cohort	Polypharmacy Cohort	Overlap group	Background population
n	90597	110225	116076	28259	277332
Hospitalisation Cohort (n (%))	90597 (100.0)	40726 (36.9)	42815 (36.9)	28259 (100.0)	0 (0)
Frailty Cohort (n (%))	40726 (45.0)	110225 (100.0)	63577 (54.8)	28259 (100.0)	0 (0)
Polypharmacy Cohort (n (%))	42815 (47.3)	63577 (57.7)	116076 (100.0)	28259 (100.0)	0 (0)
Gender = Male (n (%))	39201 (43.3)	45272 (41.1)	45578 (39.3)	11470 (40.6)	129406 (46.7)
Age (mean (SD))	78.16 (7.76)	78.74 (7.41)	78.19 (7.38)	79.66 (7.41)	74.73 (6.97)
Socio-economic status (n (%))					
1 (least deprived)	20559 (22.7)	24226 (22.0)	25520 (22.0)	5632 (19.9)	78067 (28.1)
2	21641 (23.9)	25823 (23.4)	26496 (22.8)	6340 (22.4)	72208 (26.0)
3	19291 (21.3)	23633 (21.4)	24482 (21.1)	6003 (21.2)	57973 (20.9)
4	16921 (18.7)	21331 (19.4)	22730 (19.6)	5760 (20.4)	44379 (16.0)
5 (most deprived)	12112 (13.4)	15134 (13.7)	16744 (14.4)	4501 (15.9)	24525 (8.8)
Smoking (n (%))					
Ex	34409 (38.0)	46401 (42.1)	48782 (42.0)	12628 (44.7)	85474 (30.8)
No	45804 (50.6)	53183 (48.2)	55991 (48.2)	12973 (45.9)	147173 (53.1)
Yes	8368 (9.2)	9323 (8.5)	9853 (8.5)	2320 (8.2)	27988 (10.1)
Drinking (n (%))					
Ex	3369 (3.7)	4866 (4.4)	5007 (4.3)	1503 (5.3)	5228 (1.9)
No	17779 (19.6)	24343 (22.1)	25945 (22.4)	6832 (24.2)	36386 (13.1)
Yes	50181 (55.4)	62745 (56.9)	64124 (55.2)	15267 (54.0)	159807 (57.6)
BMI[$] * (mean (SD))	27.10 (5.55)	27.95 (5.78)	28.31 (5.86)	27.59 (5.97)	26.73 (4.66)
Charlson index* (mean (SD))	0.37 (0.92)	0.51 (1.02)	0.38 (0.91)	0.64 (1.15)	0.08 (0.42)
eFI[†] * (median [IQR])	2 [1, 3]	3 [3, 4]	3 [2, 4]	4 [3, 5]	1 [0, 2]
Unplanned hospital admissions* (median [IQR])	1 [1, 2]	0 [0, 1]	0 [0, 1]	1 [1, 2]	0 [0, 0]
Number of different drugs* (median [IQR])	9 [6, 13]	10 [8, 14]	13 [11, 15]	14 [12, 18]	5 [3, 9]

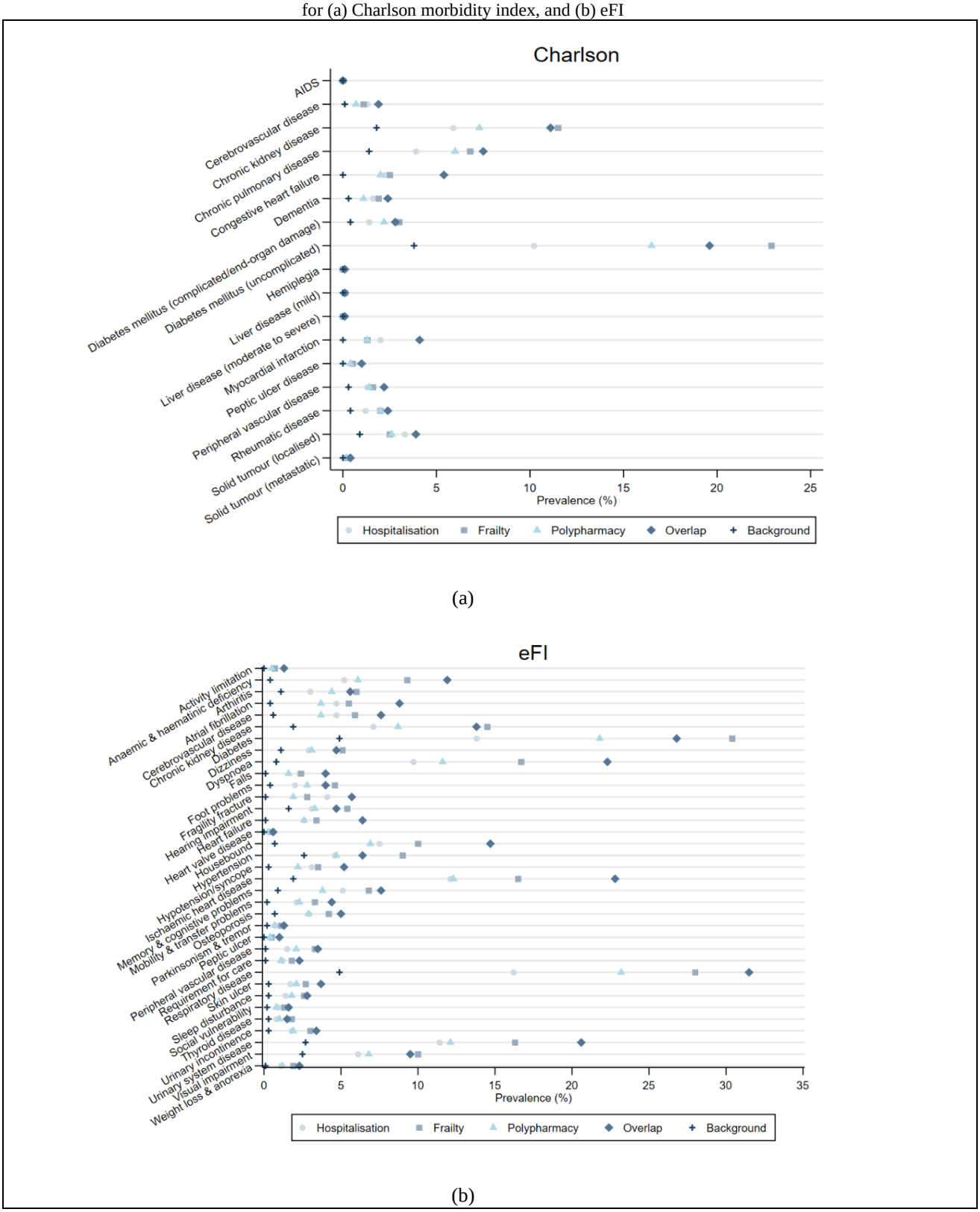
[$] Body mass index
[†] electronic Frailty Index
* In the year prior to start
Missingness was around 0.1% for socio-economic status, ranged between 1.2% to 6.0% for smoking, 16.5% to 27.4% for drinking, and 19.8% to 64.1% for BMI

Comorbidities

Comorbidity burden was higher in the three complex health needs cohorts and the overlap group compared to the background population, especially for diabetes, ischemic heart disease, respiratory disease and urinary system diseases. For most conditions, the frailty cohort had the highest prevalence, followed by the polypharmacy cohort. The hospitalisation cohort had the highest prevalence of myocardial infarction, cancer and fragility fractures. Figure 4.3 shows

the prevalence of each individual condition included in the Charlson Comorbidity Index and eFI score in the year prior to start date.

Figure 4.3 Prevalence of deficits/conditions - in the year prior to start - for all cohorts, overlap and background population for (a) Charlson morbidity index, and (b) eFI



Preventive therapy use

Point prevalence

Figure 4.4 describes one-year, three-month, and one-month PP of bisphosphonates, statins and anti-hypertensives in the three cohorts, the overlap group and the background population. Compared to the other two cohorts, the polypharmacy cohort had higher prevalence, with one-year bisphosphonates PP of 17.6% (95% CI 17.4, 17.9) compared to 13.7% (95% CI 13.5, 13.9) and 14.4% (95% CI 14.2, 14.6) in the hospital and frailty cohorts, respectively. Among the three preventive therapies, the prevalence of anti-hypertensives was substantially higher than oral bisphosphonates and statins in the three cohorts for all timeframes (e.g., ≥70.0% in one-year prevalence). Prevalence was highest in the overlap group and lowest in the background population.

Figure 4.4 Point Prevalence for oral bisphosphonates, statins and anti-hypertensives for all cohorts at (a) One-year, (b) Three-months, and (c) One-month

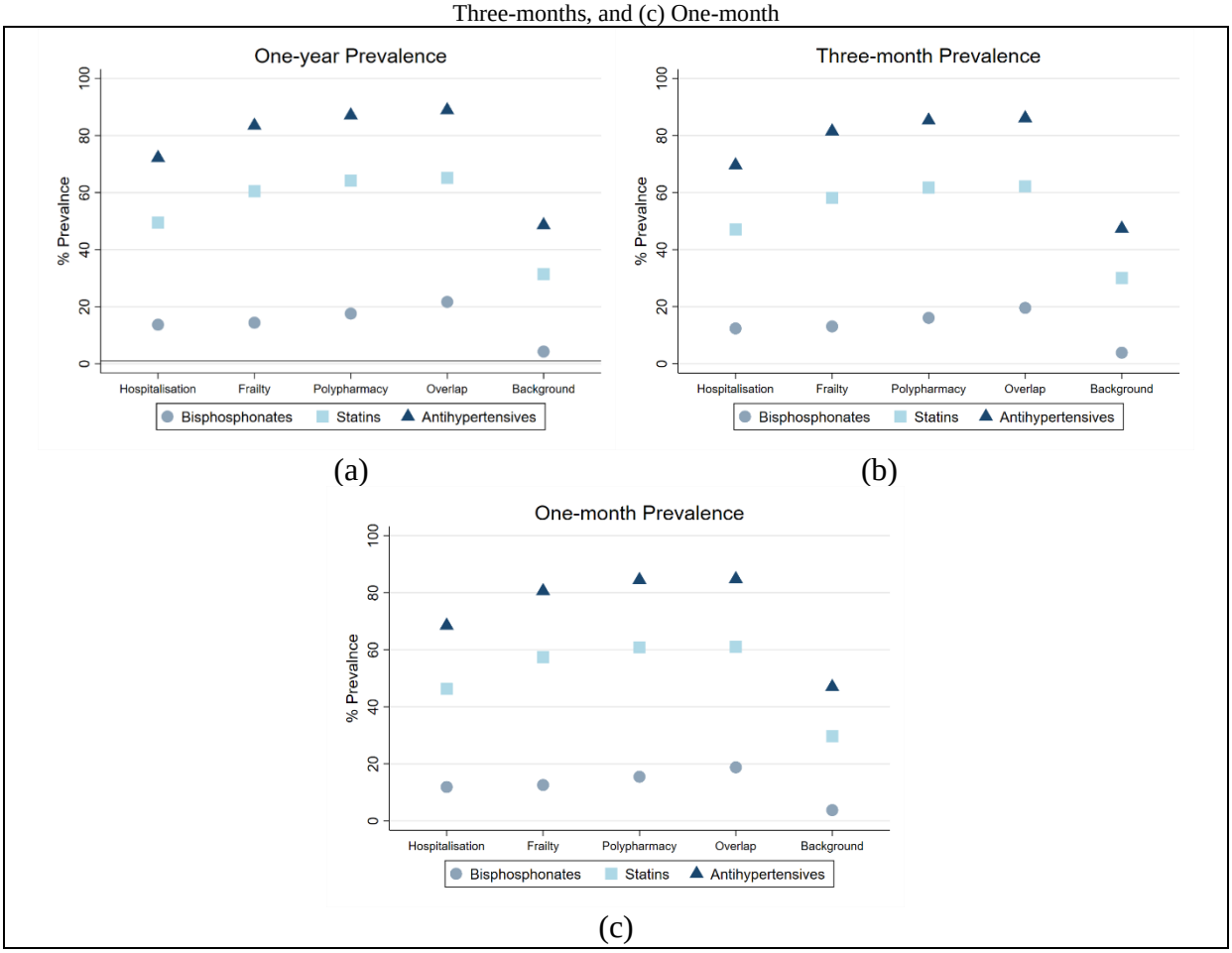


Table S 4.1 to Table S 4.5 in the Appendix report PP of the different WHO-ATC anti-hypertensives classes. ACE inhibitors, calcium-channel blockers, and diuretics had the highest prevalence of the antihypertensive classes. Overall, preventive therapies were commonly prescribed to the population of interest.

Incidence

Table 4.3 describes first, second and third-year IRs of oral bisphosphonates, statins and anti-hypertensives for the three cohorts, the overlap population and the background population.

First-year IR was slightly higher than second and third-year IRs in all analysed cohorts. For example, first-year bisphosphonates incidence rate in the hospitalisation cohort was 31.91 (95% CI 30.58, 33.23), while second and third-year IRs were 23.53 (95% CI 22.28, 24.79) and 21.47 (95% CI 20.16, 22.77), respectively. Both frailty and polypharmacy cohorts had similar IRs for the three preventive therapies; for instance, first-year anti-hypertensives IR was 155.35 (95% CI 149.14, 161.57) in the frailty cohort and 154.55 (95% CI 147.73, 161.37) in the polypharmacy cohort, with hospitalisation cohort having the lowest IRs among the three cohorts in most cases (first-year anti-hypertensives IR was 129.25 (95% CI 124.45, 134.04)). As with point prevalence, the overlap group had generally higher IRs than the three identified cohorts; first-year anti-hypertensives IR was 197.83 (95% CI 180.33, 215.33), while the background population had the lowest IRs; first-year anti-hypertensives IR was 73.06 (95% CI 71.60, 74.52). Appendix Table S 4.6 to Table S 4.10 describe the annual IRs for the different classes of anti-hypertensives. ACE inhibitors, calcium channel blockers and diuretics had the highest IRs compared to the other classes.

Table 4.3 First, second and third-year IRs – excluding 1-year prevalence

Bisphosphonates									
	First-year IR			Second-year IR			Third-year IR		
	N	Person-years	IR (95% CI)	N	Person-years	IR (95% CI)	N	Person-years	IR (95% CI)
Hospitalisation Cohort	2233	69984.6	31.91 (30.58, 33.23)	1355	57580.8	23.53 (22.28, 24.79)	1040	48441.8	21.47 (20.16, 22.77)
Frailty Cohort	2441	85996.6	28.39 (27.26, 29.51)	1752	71311.0	24.57 (23.42, 25.72)	1372	59827.8	22.93 (21.72, 24.15)
Polypharmacy Cohort	2477	87269.8	28.38 (27.27, 29.50)	1908	73105.0	26.10 (24.93, 27.27)	1451	61925.3	23.43 (22.23, 24.64)
Overlap group	780	19081.4	40.88 (38.01, 43.75)	499	14808.9	33.70 (30.74, 36.65)	341	11853.0	28.77 (25.72, 31.82)
Background population	2604	253071.6	10.29 (9.89, 10.69)	2741	226151.1	12.12 (11.67, 12.57)	2581	202646.3	12.74 (12.25, 13.23)
Statins									
	First-year IR			Second-year IR			Third-year IR		
	N incidence	Person-years	IR (95% CI)	N incidence	Person-years	IR (95% CI)	N incidence	Person-years	IR (95% CI)
Hospitalisation Cohort	2472	39902.3	61.95 (59.51, 64.39)	1552	31449.3	49.35 (46.89, 51.81)	1302	25371.0	51.32 (48.53, 54.11)
Frailty Cohort	2520	38216.8	65.94 (63.37, 68.51)	1557	29666.3	52.48 (49.88, 55.09)	1226	23443.6	52.30 (49.37, 55.22)
Polypharmacy Cohort	2327	36515.6	63.73 (61.14, 66.32)	1596	28689.1	55.63 (52.90, 58.36)	1283	22870.1	56.10 (53.03, 59.17)
Overlap group	551	8115.8	67.89 (62.22, 73.56)	334	5834.9	57.24 (51.10, 63.38)	241	4313.9	55.87 (48.81, 62.92)
Background population	8425	177430.1	47.48 (46.47, 48.50)	6551	152759.1	42.885 (41.85, 43.92)	6076	132269.4	45.94 (44.78, 47.09)
Anti-hypertensives									
	First-year IR			Second-year IR			Third-year IR		
	N incidence	Person-years	IR (95% CI)	N incidence	Person-years	IR (95% CI)	N incidence	Person-years	IR (95% CI)
Hospitalisation Cohort	2793	21610.1	129.25 (124.45, 134.04)	1649	16606.1	99.30 (94.51, 104.09)	1288	13221.7	97.42 (92.10, 102.74)
Frailty Cohort	2400	15449.0	155.35 (149.14, 161.57)	1443	11347.4	127.167 (120.60, 133.73)	1040	8648.7	120.25 (112.94, 127.56)
Polypharmacy Cohort	1972	12759.6	154.55 (147.73, 161.37)	1262	9511.4	132.68 (125.36, 140.00)	921	7338.8	125.50 (117.39, 133.60)
Overlap group	491	2481.901	197.83 (180.33, 215.33)	224	1675.2	133.71 (116.20, 151.22)	178	1197.4	148.66 (126.82, 170.50)
Background population	9622	131709.3	73.06 (71.60, 74.52)	7738	110813.2	69.83 (68.27, 71.39)	6734	94064.0	71.59 (69.88, 73.30)

4.5. Discussion

Main findings

I identified three cohorts of older patients with complex healthcare needs based on healthcare resource use (unplanned hospitalizations, polypharmacy) or frailty. Compared to the background population, patients included in either of the three cohorts were older, had a higher comorbidity burden, a higher number of prescribed drugs, and a substantially higher prevalence and incidence of preventive drug use.

Assessment of complex healthcare needs in the older population

Identification of patients with complex health needs

Following the recommendation of the NICE multimorbidity guideline[12], I identified patients with complex health needs based on highest healthcare resource utilisation, highest treatment burden and frailty in primary care data linked to hospital admission records in the UK. In addition to identifying vulnerable patients during routine clinical care, the use of primary care data is a more systematic and comprehensive approach that allows for the identification of potential target groups and for preventive measures at the population level, in turn optimising patient care. I defined patients with "complex health needs" as those individuals belonging to the top quintile of the distribution of the respective health markers. While several different definitions are used side-by-side for polypharmacy and frailty, I opted for a data-driven approach to identify the most vulnerable among older people, allowing for a uniform definition applicable in different healthcare settings, data types and patient collectives. Access to large and representative population-based data allowed to anchor these definitions based on their public health relevance.

Despite the extensive literature available on the association between frailty, polypharmacy and healthcare utilisation[4 16 98 106-111], there is still a lack of direct comparisons between the distinct

patient groups identified by these health markers. Polypharmacy and frailty are important confounders in geriatric pharmaco-epidemiology[112]. Therefore, a comprehensive understanding of these groups of patients and their overlap is crucial.

Polypharmacy

Various methods to determine the number of drugs taken by patients and definitions of polypharmacy are used in the literature[61], with no standard being agreed upon[59 73]. Frequently, the concomitant prescription of 5 or more different substances is considered polypharmacy, with 10 or more prescriptions sometimes referred to as excessive, severe or hyperpolypharmacy[59 72]. However, the method chosen greatly depends on the study setting, target population and assessment method (e.g., survey). I identified a cut-off of ≥10 different substances in the year before the start date for the polypharmacy cohort. While this approach to cumulatively assess drug use might overestimate the number of drugs taken concomitantly, it reflects the patients' treatment burden with respect to both acute and chronic indications over a long period, providing a pragmatic screening tool.

Unplanned hospitalisations

Prevention of unplanned and avoidable hospitalisations are a public health priority, as they present a major burden for the affected patients and substantial costs for the healthcare system. This study found that 19.1% of the older population were admitted to hospital for any unplanned reason in the year before start date, leading to the inclusion of all these patients to the hospitalisation cohort (cut-off of ≥1 unplanned hospitalisation).

Frailty

While drug prescriptions and unplanned hospitalisations are well recorded in real world data, the assessment of a patients' frailty status is more challenging. The concept of frailty has rapidly evolved over the last two decades, with the two major

concepts of "frailty phenotype" and "frailty indexes" being established[16]. Physical and mental fitness varies greatly among patients of the same age and while a patient's frailty will greatly influence a general practitioners treatment decision, such characteristics are typically not recorded in primary care databases. Therefore, frailty is one of the most important confounders in pharmaco-epidemiological studies among the older population. In recent years, different approaches have been developed to identify frailty in routinely collected data including claims databases[113-116] and electronic health records[98]. Clegg et al.[98] developed and validated the electronic frailty index (eFI) using primary care electronic health records from the UK, in which pre-defined deficits are combined to derive a score displaying the level of frailty for an individual patient. Based on population quartiles, categories of fit, mild, moderate and severe frailty were defined. Since 2017, the eFI is used by the NHS to support routine frailty identification[117]. While eFI is a cumulative score based on the deficits occurring at any point in a patient's history, this study calculated the eFI score based on deficits recorded within one year before the start date. This led to a cut-off of an eFI score ≥3 for the frailty cohort.

Association between polypharmacy, frailty and unplanned hospitalisation

Routinely collected health data offer wide possibilities to study the use, safety and effectiveness of preventive treatments in the older population. To identify vulnerable older people from primary health records, different markers are used in pharmaco-epidemiological studies. To my knowledge, this is the first CPRD-based study directly comparing three different markers of complex health needs to characterise patient groups. A previous CPRD-based study, Brilleman and Salisbury[106] compared the predictive validity of multimorbidity measures such as Charlson Index, chronic disease counts and diagnosis clusters towards health service utilisation and mortality, with counts of prescribed drugs being identified as the most powerful measure for predicting future GP consultations.

The overlap of the three individual cohorts identified in this study illustrates the association between polypharmacy, frailty and hospitalisation. A recent systematic review summarised the pairwise association between frailty and polypharmacy[16]: mean drug consumption was reported to be higher among frail older compared to robust patients, and increased likelihood for frailty was described among patients with polypharmacy (odds ratio 1.77–2.55)[107-109] and hyper-polypharmacy (odds ratio 4.47–5.8)[107-109], respectively. In addition, frailty among the UKs community-dwelling older population was shown to be associated with reduced quality of life[110], increased annual GP consultations, emergency hospital admissions and elevated healthcare costs[118]. Chen et al.[111] used Taiwan's National Health Insurance Research Database to assess the combined effects of frailty and polypharmacy on health outcomes in older adults. Compared to non-frail older patients without polypharmacy, increased risks for unplanned hospitalisation (adjusted RR 20.01 [95%CI] 19.30–20.75) were reported for patients with severe frailty plus excess polypharmacy. However, the combined effects appeared to vary in distinct groups. In addition, Clegg et al.[98] described increased risk for hospitalisation in frail patients compared to fit patients with adjusted HR of 1.93 (95% CI 1.86– 2.01), 3.04 (95% CI 2.90–3.19) and 4.73 (95% CI 4.43–5.06) for mild, moderate and severe frailty, respectively.

The results of this study are in line with the associations described in the literature, with 57.7% of patients in the frailty cohort being included in the polypharmacy cohort, and 45.0% of older patients with unplanned hospitalisations being in the frailty cohort. However, I also found a substantial number of patients only included in one of the respective cohorts. This highlights that, while there is a strong association, the markers of complex health needs are distinct and should not be used interchangeably but rather complementing each other. Future pharmaco-epidemiological studies should consider multiple characteristics when designing studies among geriatric patients.

Cohort characteristics and use of preventive treatments in older people

Comorbidity burden

While demographic characteristics only slightly varied between the cohorts of patients with complex health needs, pattern of comorbidities differed more. A recent systematic review[119] highlighted that multimorbidity is often poorly reported and definitions vary greatly between individual studies. Using different measures in this study, comorbidity burden in the complex health needs cohorts was substantially higher than in the background population, highlighting that the cohorts identified patients with high morbidity burden. As counts of deficits were used to define the frailty cohort, it was expected that the highest prevalence of comorbidities was found in the overlap group and frailty cohort. Largest differences in the prevalence of individual conditions were found for diabetes and respiratory diseases with prevalence being substantially lower in the hospitalisation cohort compared to frailty and polypharmacy cohort (diabetes: 13.8% vs. 30.4% and 21.8%, respiratory diseases: 16.2 vs. 28% and 23.2%). Likewise, a relatively large difference was found for chronic kidney disease, with prevalence being 14.5% in the frailty cohort compared to 7.1% and 8.7% in the hospitalisation and polypharmacy cohort.

Preventive treatments

Patients with complex healthcare needs were commonly prescribed preventive treatments, with high prevalence of use in the year before the start date: 17.6%, 64.2% and 87.1% of patients were prescribed oral bisphosphonates, statins and anti-hypertensives in the polypharmacy cohort. Both incidence rates and prevalence were significantly higher in patients with complex health needs compared to the rest of the older population.

Previous observational studies found high use of preventive therapies among older people in UK: O'Keeffe et al.[120] reported annual initiation rates of statins for primary prevention between approx. 30-50/1000 person-years for patients aged 60-84. Statin initiation rates among older

people reported in other European countries were largely comparable, with some variations with respect to patient collective studied, age group definitions and database[121 122]. While direct comparison is difficult with respect to differences in the population studied and statin indication, IRs found in this study were in a similar range (IR for statin initiation between 47.5/1000 patient-years [background population] and 67.9/1000 patient-years [overlap group]) as reported in the literature. A recent CPRD-based study evaluated incidence rates of first anti-osteoporosis treatment prescriptions[123], with IRs between 12.5 and 26.0/1000 person-years for women aged 65-84 years (stratified for age bands). IRs for men were substantially smaller. This study found slightly higher IRs for oral bisphosphonates between 10.3/1000 and 40.9/1000 person-years [background population and overlap group]. UK based studies reporting incidence rates of antihypertensive treatment among the older population are scarce. A CPRD-based study reports annual IRs of approximately 15% in 2001 assessing hypertensive patients aged ≥40 years with ≥3 cardiovascular risk factors[124]. In this study, incidence rates varied across the different cohorts, with IR 73.1/1000 and 197.8/1000 patient years for background population and overlap group.

Strengths, limitations and scope for future work

This study comes with both strengths and limitations. The first strength of this study is the direct comparison of the different definitions of "complex healthcare needs" and their impact on which patients are included into the cohort. I defined the three cohorts with respect to unplanned hospitalisation, polypharmacy and frailty. Polypharmacy, defined as the use of ≥5 medications, was one of the deficits in eFI index. Thus, there is an operational association, which may overestimate the true overlap between cohorts. Lastly, for practical reasons, I assessed all patient information using a baseline period of one year. Therefore, the underreporting of chronic deficits cannot be ruled out.

Frailty and polypharmacy are cumulative conditions that take years to accumulate. For example, a recent study reports that, by age 65, older people would already belong to different groups based on their frailty levels[45]. Those patients would continue to progress/regress in the following years. Relying on cross-sectional values like the ones being used in this study could be misleading. Collecting repeated measures of health markers instead and incorporating variation/trajectory over time would allow to identify those who are stable, improving, or are deteriorating rapidly and in need of additional healthcare. Grouping patients based on their progression over time could also provide an alternative approach to using multiple health markers, especially when it is combined with monitoring the associated risk of adverse events with these markers.

Conclusion

This study provides fundamental work on identifying older people with complex health needs. Cohorts of patients with complex healthcare needs slightly differ with respect to their demographic characteristics as well as regarding drug utilisation. My results highlight that several indicators of complex healthcare needs should be considered in future research on this population, as they may complement each other. I found the use of preventive therapies to be common among older patients with complex healthcare needs. Polypharmacy and frailty are important confounders in geriatric pharmaco-epidemiology. Further studies on how to calculate these confounders (timeframes and medication profiles), how they complement each other, and how they progress over time are needed.

5. Longitudinal trajectories of frailty in older people, and their association with mortality

5.1 Chapter summary

In this chapter, I characterised the progression of frailty over time. It is assumed that frailty is a progressive condition that takes years to develop, and that there are subgroups of the population that may deteriorate faster than the rest. I identified five clusters within the older population that have different frailty trajectories and associated mortality risk in a primary care database. I evaluated these subgroups in a different UK primary care database.

The patients included in this study were aged >65 on January 1st 2010, identified from CPRD GOLD and Aurum, and linked to mortality data. I used eFI to model frailty trajectories. eFI scores were calculated at baseline (2009) and annually in subsequent years (2010-2013). I applied joint latent class models to divide the population into subgroups (clusters) with different trajectories and associated hazard ratios (HR). I built the model in Gold and validated it in Aurum. Identified clusters were described, and validation metrics in Aurum were assessed.

The five clusters identified were: low-slow, low-moderate, low-rapid, high-slow and high-rapid. The high-rapid cluster had the highest intercept, while the low-rapid cluster had the steepest slope. Taking the low-slow cluster as a reference, low-rapid and high-rapid had the highest HRs. Good validation was found in the Aurum population.

This study showed that there are vulnerable subgroups of the older population who are currently frail or have rapid frailty progression and need more healthcare monitoring and resources.

Note: the content, methods, results and discussion presented in this chapter are based on the manuscript:

Elhussein L, Robinson DE, Delmestri A, et al. Longitudinal trajectories of frailty are associated with short-term mortality in older people: A joint latent class models analysis using two UK primary care databases. J Clin Epidemiol. 2024 Jun 26:111442. doi: https://doi.org/10.1016/j.jclinepi.2024.111442

5.2 Introduction

Frailty can be defined as a loss of biological reserve leading to vulnerability to adverse outcomes. These include unplanned hospitalisation, admission to long-term care and death[22 23 36 125]. Frailty is common among older people, with a prevalence of around 10% in that population.

It is beneficial to identify patients with elevated frailty on the individual level. This would help in focusing interventions such as improved physical activity, physiological support, nutritional advice, and even anabolic hormones, which leads to lower rates of negative outcomes[32 38]. Understanding frailty on the population level is also vital for the effective planning of social and health care.

eFI was derived based on the capture of data on the presence of 36 specified deficits in the primary care EHR[30]. Data extracted from those records was used to calculate the eFI score calculated as a proportion of those 36 deficits that had been recorded cumulatively. Based on these scores, people were divided into categories (fit, mild, moderate and severe frailty). These categories were predictive of subsequent morality, emergency hospital and nursing home admission[30]. As a consequence, eFI has now been widely adopted by the NHS in England as a screening test for frailty[31].

While the eFI score was found to be associated with mortality on the population level, using a cross-sectional value had a poor predictive performance on the individual level, even if taken only three months prior to a patient's death[126]. Frailty is not a static phenomenon. Data based on the Rockwood scale suggests the number of deficits on average double between the ages of 50 and 80[127]. In their analysis of 4 survey data for people aged 65+ from Europe and the US, Stolz et al. applied joint modelling to assess the association between frailty trajectory over time

and the risk of death and found that an increasing trend of frailty was a good predictor of mortality[56].

Frailty also does not progress to the same extent in different individuals. In one study, 3 different trajectories of frailty using the Rockwood scale were identified in a study of over 12000 elderly residents[53]. Other studies found similar results; varying distinct frailty trajectories exist in older people with different characteristics and/or near death[49 54]. The methods employed to identify such distinct trajectories included traditional and machine-learning longitudinal clustering. While these methods provide important tools for analysing longitudinal data, they either require a uniform number of data points or fail to account for the non-random dropouts, often occurring due to death in such an old population.

Therefore, this study aims to address whether, by applying this cumulative deficit model of frailty (eFI) to routine medical records, I would be able to identify distinct trajectories of frailty. I tested these trajectories in their ability to predict mortality. Such trajectories could then be used as a more robust predictor over a single baseline score to target interventions.

5.3 Methods

Design

In this study, I conducted a longitudinal cohort analysis to assess frailty change in older people over four years and its association with mortality. I used random samples of UK older people from UK primary care records. Based on eFI, I identified the proportions who were 'fit' or had 'mild', 'moderate' or 'severe' frailty at baseline and determined the transitions over the next 4 years. I then examined how the actual score change over time affected survival. Next, I applied joint latent class models to derive clusters based on the trajectories of the eFI and their associated risk for mortality. Lastly, the clusters were evaluated in an external dataset.

Data source

I used CPRD GOLD (September 2019 release) to build the trajectory clusters and CPRD Aurum (August 2019 release) to evaluate the resultant clusters. CPRD data were also linked to ONS mortality data (described in 3.2).

Population

Patients aged over 65 and alive on 1st January 2010 from both GOLD and Aurum were eligible for inclusion if they had been registered in an 'up-to-standard' (UTS) practice[86] for at least one year before the start date. Duplicate patients belonging to both databases were removed from CPRD Aurum.

The follow-up period was observed from 1st January 2010 to 31st December 2013. Patients exited the study at the earliest of: practice last collection date, patient transfer-out of practice date, death date[92] or study end date.

Exposure

The main exposure was the eFI score. The baseline eFI was calculated based on the counts of deficits recorded at any time prior to 31 December 2009. The eFI was recalculated each year, and all previous deficits were carried forward.

Outcome

The major outcome analysed was all-cause mortality, defined as death recorded in CPRD GOLD and CPRD Aurum during follow-up (mortality calculation in source data, section 3.3)

Statistical analysis

Descriptive analysis

The base demographic and other key morbidity predictors were extracted and compared between the two cohorts: GOLD and Aurum. eFI was calculated at baseline and each year of follow-up, as above. Study participants were allocated to one of the four categories of frailty based on the original eFI scoring system: ‘fit’, ‘mild’, ‘moderate’ and ‘severe’ at baseline and at their last observation time point. I examined the transition between these four states over the follow-up period to understand the underlying pattern.

Joint modelling

I analysed the change in eFI score across the whole cohort over the follow-up period and assessed whether an increase in eFI score impacted survival. Joint modelling (section 3.3) was applied for this. eFI was the longitudinal exposure, and all-cause mortality was used for calculating the risk of event. Age and sex were inserted as covariates for both sub-models. I used joint modelling as a preliminary step to confirm the association between eFI change over time and risk of mortality, and to determine the best way to model eFI. Linear and quadratic time terms were tested in the longitudinal part of the joint modelling, and the best method was taken forward based on Log-Likelihood calculation.

Joint latent class modelling

After choosing the preferred methods to model eFI, I attempted to see if there were clusters of discrete patterns of eFI change and whether allocation to such a cluster was usefully predictive of mortality. For this, I applied joint latent class models, adjusting for age and sex in both sub-models. Two to six-cluster models were fitted, and the optimum number of clusters was based on the criteria described in 3.3.

Model validation

I assessed the performance of the emergent best-fit cluster model (derived from GOLD) in the validation Aurum sample by calculating the posterior probability and hazard ratios for mortality (3.3). Descriptive analysis and baseline characteristics were performed to compare the resultant clusters as a final step in assessing the model's face and content validity. I also examined the contribution of the individual deficits within the eFI and their change in determining cluster membership.

I used STATA (Version 15.1) for data preparation and R (Version 3.6.0) for the subsequent analyses. R packages ‘JM’ version 1.5-2 [link] and ‘lcmm’ version 2.0.2 [link] were used for joint modelling and joint latent class modelling, respectively.

5.4 Results

Model training in CPRD GOLD

Study population

There were 475,503 patients in CPRD GOLD database eligible for the analysis. Table 5.1 describes the baseline characteristics. Mean age was 75.1 (SD 7.4) and majority of the patients were in the "Fit" eFI category, 305,946 (64.3%).

Table 5.1 Baseline characteristics for GOLD

	Building set – GOLD
n	475503
Gender = Male	211751 (44.5%)
Age (mean (SD))	75.1 (7.4)
Socio-economic status	
1 (least deprived)	81714 (17.2%)
2	93243 (19.6%)
3	95679 (20.1%)
4	100448 (21.1%)
5 (most deprived)	104419 (22.0%)
Smoking status	
Ex-smoker	159573 (33.6%)
Non-smoker	240183 (50.5%)
Current smoker	46859 (9.9%)
Drinking status	
Ex-drinker	11349 (2.4%)
Non-drinker	72254 (15.2%)
Current drinker	265256 (55.8%)
BMI (mean (SD))	27.2 (5.7)
eFI* (mean (SD))	3.8 (2.9)
eFI* (median [IQR])	3 [2, 6]
Frailty category*†	
Fit	305946 (64.3%)
Mild	133543 (28.1%)
Moderate	33342 (7.0%)
Severe	2672 (0.6%)

BMI = Body mass index, eFI = electronic Frailty Index
* Baseline category: calculated on 31st December 2009
† Frailty categories are: fit (mean eFI/36: 0-0.12), mild (>0.12-0.24), moderate (>0.24-0.36), severe (>0.36)
Missingness was 6.1% for smoking, 26.6% for drinking, and 32.3% for BMI in GOLD. Missingness was 4.7% for smoking, 15.9% for drinking and 15.7% for BMI in Aurum

Transition in frailty category

Overall, during the 4 years of observation, 114,218 (24.4 %) patients transitioned to a worse eFI category, with 3,748 (<0.1%) transitioning to a better frailty category. Most transitions, 72,298 (15.2%), were from fit to mild, with 33,158 (7.0%) mild to moderate, and 5,086 (1.1%) moderate to severe transitions, with 3,676 (<0.1%) crossing more than one frailty category (Table 5.2).

Table 5.2 Transitions between frailty categories in GOLD

Baseline	Last observation				
Frailty category*†	Fit	Mild	Moderate	Severe	Total
Fit	230528 (48.5%)	72298 (15.2%)	3096 (0.7%)	24 (<0.1%)	305946 (64.3%)
Mild	3104 (0.7%)	96725 (20.3%)	33158 (7.0%)	556 (0.1%)	133543 (28.1%)
Moderate	0	585 (0.1%)	27671 (5.8%)	5086 (1.1%)	33342 (7.0%)
Severe	0	0	59 (<0.1%)	2613 (0.6%)	2672 (0.6%)
Total	233632 (49.1%)	169608 (35.7%)	63984 (13.5%)	8279 (1.7%)	475503 (100%)

* Baseline category: calculated on 31st December 2009
† Frailty categories are: fit (mean eFI/36: 0-0.12), mild (>0.12-0.24), moderate (>0.24-0.36), severe (>0.36)

Joint modelling

Median follow-up was 4 years [IQR 2.5 to 4 years], with 68,977 (14.5%) dying during the observation period. Based on the log-likelihood calculation, the linear model performed better than the one incorporating a quadratic term (Table S 5.1 in the Appendix). Under this model, the cumulative eFI score increased very slowly across the entire GOLD population, with a slope of 0.35 (0.35, 0.35) per year, equivalent to an increase in just one additional deficit every three years. The survival sub-model showed that, for every additional eFI deficit, the risk of death increased by 12% (95% CI 12%, 12%) in the survival model.

Joint latent class modelling

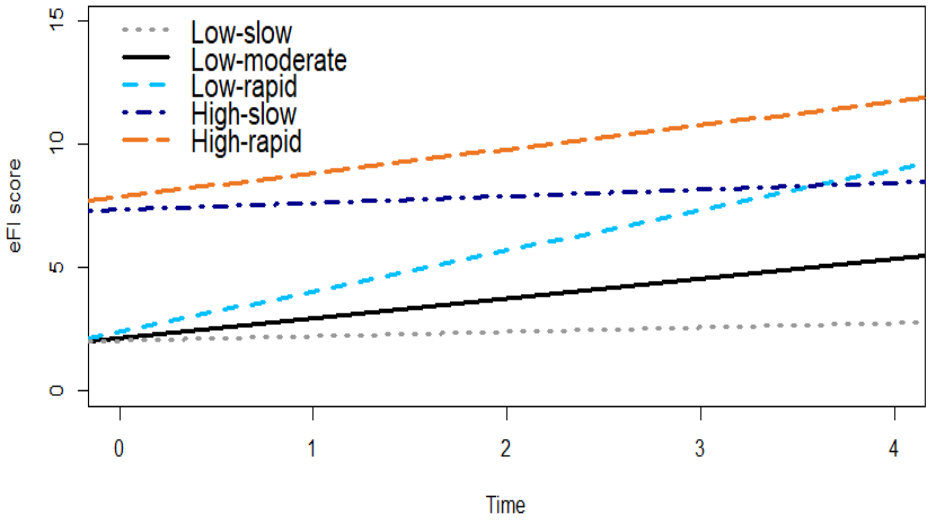
After testing models with up to 6 clusters, the 5-cluster model emerged as the preferred model based on statistical and clinical plausibility. Table S 5.2 in the Appendix describes the models tested. I describe the 5 clusters based on the trends of their trajectories as follows:

- Low-slow (intercept = 2.0, slope = 0.2): individuals who had low eFI value (mostly in the eFI ‘fit-mild categories) at the start of the follow-up and had a very slow progression
- Low-moderate (intercept = 2.1, slope = 0.8): individuals who had low eFI value at the start of the follow-up and had a moderate progression
- Low-rapid (intercept = 2.4, slope = 1.7): individuals who had low eFI value at the start of the follow-up and had a rapid progression
- High-slow (intercept = 7.4, slope = 0.3): individuals who had high eFI value (mostly in the eFI ‘mild-moderate’ categories) at the start of the follow-up and had a very slow progression
- High-rapid (intercept = 7.9, slope = 1.0): individuals who had high eFI value at the start of the follow-up and had a rapid progression in their frailty

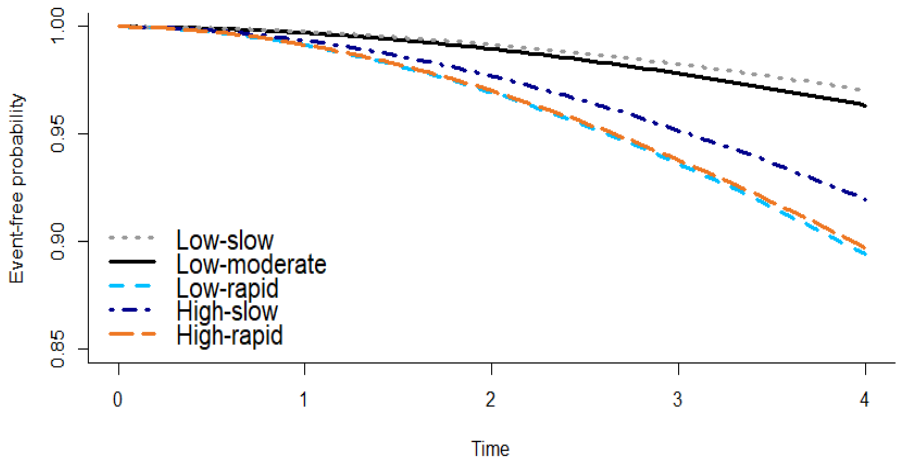
Figure 5.1 shows the trajectories of each of these five clusters and their associated mortality risk profiles. Appendix Table S 5.3 reports the estimates for each cluster.

There was a concordance between the trajectory and the mortality risk. The low-rapid and high-rapid clusters thus had the steepest survival curves, followed by the high-slow cluster. Taking the low-slow cluster as the reference, low-rapid and high-rapid had the highest HR: 3.73 (95%CI 3.71 to 3.76) and 3.63 (95%CI 3.57 to 3.69), respectively. Followed by the high-slow cluster; 2.80 (95%CI 2.74 to 2.85). Low-moderate cluster had the lowest HR after low-slow, 1.24 (95%CI 1.13 to 1.37). Median follow-up, yearly deaths and eFI scores are reported in Appendix Table S 5.4.

Figure 5.1 Results of the 5-cluster model (a) intercept and slope of each cluster (b) Survival curve of each cluster



(a)



(b)

Table 5.3 describes the distribution of the GOLD cohort between these 5 groups. Most of the included patients belonged to the low-slow (71.1%) and low-moderate (14.7%) clusters, with only 4.6% in the 2 rapidly progressive groups combined. The two rapid clusters had the highest mean age.

Table 5.3 Baseline characteristics of each cluster (GOLD)

Cluster	Low-slow	Low-moderate	Low-rapid	High-slow	High-rapid
n	338029 (71.1%)	69794 (14.7%)	12869 (2.7%)	45707 (9.6%)	9104 (1.9%)
Gender = Male (n (%))	153144 (45.3%)	30393 (43.5%)	5790 (45.0%)	18695 (40.9%)	3729 (41.0%)
Age (mean (SD))	74.29 (7.23)	76.48 (7.27)	79.12 (7.18)	76.95 (7.49)	78.21 (7.10)
Socio-economic status (n (%))					
1 (least deprived)	58709 (17.4%)	12041 (17.3%)	2171 (16.9%)	7263 (15.9%)	1530 (16.8%)
2	66758 (19.7%)	14154 (20.3%)	2428 (18.9%)	8269 (18.1%)	1634 (17.9%)
3	69018 (20.4%)	14961 (21.4%)	2698 (21.0%)	7523 (16.5%)	1479 (16.2%)
4	71733 (21.2%)	13149 (18.8%)	2506 (19.5%)	10913 (23.9%)	2147 (23.6%)
5 (most deprived)	71811 (21.2%)	15489 (22.2%)	3066 (23.8%)	11739 (25.7%)	2314 (25.4%)
BMI (mean (SD))	27.02 (5.4)	27.09 (6.2)	26.83 (5.5)	28.28 (6.0)	28.37 (6.0)
eFI score* (mean (SD))	3.1 (2.1)	3.3 (2.2)	3.9 (2.4)	9.4 (2.1)	9.7 (2.2)
eFI score* (median[IQR])	3 [1, 5]	3 [1, 5]	4 [2, 6]	9 [8, 10]	9 [8, 11]
Frailty category*† (n (%))					
Fit	248982 (73.7%)	49124 (70.4%)	7839 (60.9%)	0	1 (0.0%)
Mild	87935 (26.0%)	20066 (28.8%)	4619 (35.9%)	17997 (39.4%)	2926 (32.1%)
Moderate	1112 (0.3%)	604 (0.9%)	410 (3.2%)	25630 (56.1%)	5586 (61.4%)
Severe	0	0	1 (0.0%)	2080 (4.6%)	591 (6.5%)

BMI = Body mass index, eFI = electronic Frailty Index
* Baseline category: calculated on 31st December 2009
† Frailty categories are: fit (mean eFI/36: 0-0.12), mild (>0.12-0.24), moderate (>0.24-0.36), severe (>0.36)
Missingness was 6.1% for smoking, 26.6% for drinking, and 32.3% for BMI in GOLD. Missingness was 4.7% for smoking, 15.9% for drinking and 15.7% for BMI in Aurum

I examined whether there were specific deficits that were more obviously responsible for the change in eFI between the clusters. These are shown Table 5.4 as the prevalence of individual deficits at the start and end of the study for each cluster. Deficits such as polypharmacy, hypertension, ischaemic heart disease, and osteoporosis were already common at baseline for all clusters, and their prevalence was elevated by the end of observation. There were no obvious differences between the clusters in which the new deficits appeared during the follow-up.

Table 5.4 Prevalence of eFI deficits by cluster at baseline and at last observation in GOLD

eFI deficit	At baseline (n (%))					At last observation (n (%))				
	Low-slow N = 338029	Low-moderate N = 69794	Low-rapid N = 12869	High-slow N = 45707	High-rapid N = 9104	Low-slow N = 338029	Low-moderate N = 69794	Low-rapid N = 12869	High-slow N = 45707	High-rapid N = 9104
eFI 1: Activity Limitation	2682 (0.8)	721 (1.0)	218 (1.7)	2266 (5.0)	535 (5.9)	3776 (1.1)	2135 (3.1)	1306 (10.1)	2682 (5.9)	913 (10.0)
eFI 2: Anaemic & haematinic deficiency	20011 (5.9)	4987 (7.1)	1326 (10.3)	13712 (30.0)	2843 (31.2)	27645 (8.2)	13475 (19.3)	4193 (32.6)	16063 (35.1)	4569 (50.2)
eFI 3: Arthritis	37573 (11.1)	8572 (12.3)	1797 (14.0)	15928 (34.8)	3247 (35.7)	45601 (13.5)	14581 (20.9)	2826 (22.0)	17057 (37.3)	3889 (42.7)
eFI 4: Atrial fibrillation	16735 (5.0)	4233 (6.1)	1111 (8.6)	10071 (22.0)	2246 (24.7)	23069 (6.8)	12508 (17.9)	3956 (30.7)	11374 (24.9)	3428 (37.7)
eFI 5: Cerebrovascular disease	15108 (4.5)	3651 (5.2)	904 (7.0)	8862 (19.4)	1952 (21.4)	20147 (6.0)	9493 (13.6)	2886 (22.4)	10104 (22.1)	2922 (32.1)
eFI 6: CKD	50872 (15.0)	11163 (16.0)	2650 (20.6)	20969 (45.9)	4376 (48.1)	61919 (18.3)	20913 (30.0)	5125 (39.8)	22916 (50.1)	5554 (61.0)
eFI 7: Diabetes	36377 (10.8)	8312 (11.9)	1965 (15.3)	16328 (35.7)	3717 (40.8)	42373 (12.5)	13690 (19.6)	3183 (24.7)	17235 (37.7)	4258 (46.8)
eFI 8: Dizziness	37463 (11.1)	8328 (11.9)	1633 (12.7)	16844 (36.9)	3402 (37.4)	44865 (13.3)	14051 (20.1)	2854 (22.2)	17831 (39.0)	4019 (44.1)
eFI 9: Dyspnoea	37492 (11.1)	8839 (12.7)	1980 (15.4)	24365 (53.3)	4879 (53.6)	48081 (14.2)	20876 (29.9)	5805 (45.1)	26567 (58.1)	6344 (69.7)
eFI 10: Falls	6186 (1.8)	1811 (2.6)	471 (3.7)	4531 (9.9)	1033 (11.3)	9527 (2.8)	6066 (8.7)	2244 (17.4)	5730 (12.5)	2167 (23.8)
eFI 11: Foot problems	9477 (2.8)	2394 (3.4)	544 (4.2)	6527 (14.3)	1443 (15.9)	15198 (4.5)	7549 (10.8)	1824 (14.2)	8178 (17.9)	2566 (28.2)
eFI 12: Fragility fracture	9923 (2.9)	2332 (3.3)	565 (4.4)	4655 (10.2)	988 (10.9)	14546 (4.3)	8106 (11.6)	2431 (18.9)	5831 (12.8)	1949 (21.4)
eFI 13: Hearing impairment	36411 (10.8)	8218 (11.8)	1686 (13.1)	12859 (28.1)	2569 (28.2)	49484 (14.6)	17654 (25.3)	3697 (28.7)	14797 (32.4)	3585 (39.4)
eFI 14: Heart failure	5455 (1.6)	1394 (2.0)	431 (3.3)	7045 (15.4)	1528 (16.8)	7483 (2.2)	5323 (7.6)	2424 (18.8)	8064 (17.6)	2630 (28.9)
eFI 15: Heart valve disease	1189 (0.4)	344 (0.5)	99 (0.8)	914 (2.0)	219 (2.4)	1781 (0.5)	1242 (1.8)	468 (3.6)	1089 (2.4)	432 (4.7)
eFI 16: Housebound	17641 (5.2)	4487 (6.4)	1323 (10.3)	10465 (22.9)	2222 (24.4)	23970 (7.1)	12013 (17.2)	4399 (34.2)	12349 (27.0)	3718 (40.8)
eFI 17: Hypertension	111273 (32.9)	21851 (31.3)	4220 (32.8)	27962 (61.2)	5514 (60.6)	122326 (36.2)	31113 (44.6)	5747 (44.7)	28672 (62.7)	5951 (65.4)
eFI 18: Hypotension / syncope	15594 (4.6)	3814 (5.5)	922 (7.2)	10637 (23.3)	2241 (24.6)	20367 (6.0)	8781 (12.6)	2605 (20.2)	11868 (26.0)	3262 (35.8)
eFI 19: Ischaemic heart disease	71035 (21.0)	14760 (21.1)	2793 (21.7)	28704 (62.8)	5722 (62.9)	81015 (24.0)	24587 (35.2)	5378 (41.8)	29910 (65.4)	6561 (72.1)
eFI 20: Memory & cognitive problems	10137 (3.0)	2153 (3.1)	745 (5.8)	4745 (10.4)	959 (10.5)	17278 (5.1)	10123 (14.5)	3483 (27.1)	6608 (14.5)	2516 (27.6)
eFI 21: Mobility and transfer problems	7095 (2.1)	2110 (3.0)	642 (5.0)	6785 (14.8)	1628 (17.9)	9247 (2.7)	5260 (7.5)	2173 (16.9)	7789 (17.0)	2636 (29.0)

eFI deficit	At baseline (n (%))					At last observation (n (%))				
	Low-slow N = 338029	Low-moderate N = 69794	Low-rapid N = 12869	High-slow N = 45707	High-rapid N = 9104	Low-slow N = 338029	Low-moderate N = 69794	Low-rapid N = 12869	High-slow N = 45707	High-rapid N = 9104
eFI 22: Osteoporosis	21452 (6.3)	4712 (6.8)	958 (7.4)	8808 (19.3)	1829 (20.1)	26661 (7.9)	10460 (15.0)	2415 (18.8)	9838 (21.5)	2606 (28.6)
eFI 23: Parkinsonism & Tremor	3568 (1.1)	950 (1.4)	283 (2.2)	2354 (5.2)	537 (5.9)	4702 (1.4)	2146 (3.1)	683 (5.3)	2656 (5.8)	821 (9.0)
eFI 24: Peptic ulcer	3810 (1.1)	865 (1.2)	209 (1.6)	2481 (5.4)	495 (5.4)	4519 (1.3)	1688 (2.4)	525 (4.1)	2677 (5.9)	676 (7.4)
eFI 25: Peripheral vascular disease	5768 (1.7)	1602 (2.3)	477 (3.7)	6013 (13.2)	1504 (16.5)	7249 (2.1)	3417 (4.9)	1121 (8.7)	6732 (14.7)	2100 (23.1)
eFI 26: Polypharmacy	188498 (55.8)	38476 (55.1)	7988 (62.1)	44636 (97.7)	8855 (97.3)	188669 (55.8)	61023 (87.4)	12200 (94.8)	40988 (89.7)	8784 (96.5)
eFI 27: Requirement for care	3388 (1.0)	714 (1.0)	240 (1.9)	2118 (4.6)	420 (4.6)	5180 (1.5)	3443 (4.9)	1654 (12.9)	2928 (6.4)	1203 (13.2)
eFI 28: Respiratory disease	59816 (17.7)	13014 (18.6)	2631 (20.4)	23398 (51.2)	4582 (50.3)	67398 (19.9)	20820 (29.8)	4675 (36.3)	24582 (53.8)	5352 (58.8)
eFI 29: Skin ulcer	10675 (3.2)	2694 (3.9)	720 (5.6)	6570 (14.4)	1514 (16.6)	13696 (4.1)	5949 (8.5)	1784 (13.9)	7510 (16.4)	2302 (25.3)
eFI 30: Sleep disturbance	16583 (4.9)	3977 (5.7)	868 (6.7)	10234 (22.4)	2151 (23.6)	18711 (5.5)	6212 (8.9)	1602 (12.4)	10757 (23.5)	2630 (28.9)
eFI 31: Social vulnerability	6691 (2.0)	1707 (2.4)	455 (3.5)	3788 (8.3)	847 (9.3)	8772 (2.6)	4157 (6.0)	1561 (12.1)	4436 (9.7)	1498 (16.5)
eFI 32: Thyroid disease	9741 (2.9)	2052 (2.9)	389 (3.0)	3704 (8.1)	754 (8.3)	12040 (3.6)	3953 (5.7)	825 (6.4)	4136 (9.0)	1049 (11.5)
eFI 33: Urinary incontinence	10743 (3.2)	2622 (3.8)	646 (5.0)	7230 (15.8)	1562 (17.2)	14279 (4.2)	6756 (9.7)	2162 (16.8)	8250 (18.0)	2436 (26.8)
eFI 34: Urinary system disease	77855 (23.0)	15593 (22.3)	3065 (23.8)	27168 (59.4)	5077 (55.8)	94565 (28.0)	29513 (42.3)	6623 (51.5)	29380 (64.3)	6621 (72.7)
eFI 35: Visual impairment	55929 (16.5)	12523 (17.9)	2770 (21.5)	19872 (43.5)	4038 (44.4)	73756 (21.8)	25171 (36.1)	5234 (40.7)	22583 (49.4)	5533 (60.8)
eFI 36: Weight loss and anorexia	6156 (1.8)	1516 (2.2)	442 (3.4)	4578 (10.0)	1018 (11.2)	8908 (2.6)	4709 (6.7)	1826 (14.2)	5498 (12.0)	1815 (19.9)

Model validation in CPRD Aurum

Study population

I included 390,204 patients from Aurum. Table 5.5 describes the baseline characteristics of this population. The distributions of age, gender and socio-economic status were almost identical between GOLD and Aurum. There was an unexpected difference in eFI score between these two otherwise nationally representative populations. At baseline, the Aurum cohort had, on average, a higher eFI score and their distribution of the frailty category was shifted towards a more severe score. I considered whether specific deficits were more prevalent in the Aurum cohort compared to GOLD in Appendix Table S 5.5. I found that a few deficits had a higher prevalence in Aurum, such as arthritis, hypertension and visual impairment.

Table 5.5 Baseline characteristics for Aurum

	Validation set (Aurum)
n	390204
Gender = Male	173854 (44.6%)
Age (mean (SD))	75.0 (7.4)
Socio-economic status	
1 (least deprived)	71057 (18.2%)
2	70429 (18.1%)
3	76085 (19.5%)
4	85591 (21.9%)
5 (most deprived)	87042 (22.3%)
Smoking status	
Ex-smoker	106807 (27.4%)
Non-smoker	214212 (54.9%)
Current smoker	50967 (13.1%)
Drinking status	
Ex-drinker	867 (0.2%)
Non-drinker	108004 (27.7%)
Current drinker	219267 (56.2%)
BMI (mean (SD))	27.3 (10.3)
eFI* (mean (SD))	4.8 (3.4)
eFI* (median [IQR])	4 [2, 7]
Frailty category*†	
Fit	201865 (51.7%)
Mild	132767 (34.0%)
Moderate	44433 (11.4%)
Severe	11139 (2.9%)

BMI = Body mass index, eFI = electronic Frailty Index
* Baseline category: calculated on 31st December 2009
† Frailty categories are: fit (mean eFI/36: 0-0.12), mild (>0.12-0.24), moderate (>0.24-0.36), severe (>0.36)
Missingness was 6.1% for smoking, 26.6% for drinking, and 32.3% for BMI in GOLD. Missingness was 4.7% for smoking, 15.9% for drinking and 15.7% for BMI in Aurum

Transition in frailty category

A slightly higher proportion of 107,376 (27.5 %) patients transitioned to a worse eFI category in Aurum, with 2,911 (0.7%) transitioning to a better frailty category. Most transitions, 53,674 (13.8%), were from fit to mild, with 36,411 (9.3%) mild to moderate, and 12,496 (3.2%) moderate to severe transitions, with 4,795 (1.2%) crossing more than one frailty category (Table 5.6).

Table 5.6 Transitions between frailty categories in Aurum

Baseline	Last observation				
Frailty category*†	Fit	Mild	Moderate	Severe	Total
Fit	145451 (37.3%)	53674 (13.8%)	2660 (0.7%)	80 (<0.1%)	201865 (51.7%)
Mild	2140 (0.6%)	92161 (23.6%)	36411 (9.3%)	2055 (0.5%)	132767 (34.0%)
Moderate	0	615 (0.2%)	31322 (8.0%)	12496 (3.2%)	44433 (11.4%)
Severe	0	0	156 (<0.1%)	10983 (2.8%)	11139 (2.9%)
Total	147591 (37.8%)	146450 (37.5%)	70549 (18.1%)	25614 (6.6%)	390204 (100%)

* Baseline category: calculated on 31st December 2009
† Frailty categories are: fit (mean eFI/36: 0-0.12), mild (>0.12-0.24), moderate (>0.24-0.36), severe (>0.36)

Resultant clusters

In CPRD Aurum, 84.1% of patients had a posterior probability ≥0.7. Table S 5.6 in the Appendix describes the proportion of patients in each cluster with different posterior probability cut-offs. Similar to CPRD GOLD, low-slow had the highest proportion of patients (57.8%), followed by the high-slow cluster (20.8%), instead of the low-moderate cluster (13.9%). Table 5.7 describes the population belonging to each cluster. Table 5.8 describes the prevalence of individual deficits in each cluster at baseline and at the end of observation.

Similar to GOLD, the low-rapid and high-rapid clusters had the steepest survival curves, followed by the high-slow cluster. With the low-slow cluster being the reference, low-rapid and high-rapid had the highest HR: 3.69 (95%CI 3.57 to 3.81) and 3.31 (95%CI 3.22 to 3.41), respectively. Followed by the high-slow cluster; 2.92 (95%CI 2.87 to 2.98). Low-moderate cluster had the lowest HR after low-slow, 1.15 (95%CI 1.12 to 1.18).

Table 5.7 Baseline of each cluster for the validation dataset (CPRD Aurum)

Cluster	Low-slow	Low-moderate	Low-rapid	High-slow	High-rapid
n	225598 (57.8%)	54171 (13.9%)	9464 (2.4%)	81314 (20.8%)	19657 (5.0%)
Gender = Male (n (%))	103260 (45.8%)	23893 (44.1%)	4262 (45.0%)	34464 (42.4%)	7975 (40.6%)
Age (mean (SD))	73.76 (7.04)	75.99 (7.20)	78.94 (7.19)	76.60 (7.55)	78.51 (7.11)
Socio-economic status (n (%))					
1 (least deprived)	43959 (19.5%)	9707 (17.9%)	1372 (14.5%)	13260 (16.3%)	2759 (14.0%)
2	42844 (19.0%)	9665 (17.8%)	1488 (15.7%)	13533 (16.6%)	2899 (14.7%)
3	44919 (19.9%)	10618 (19.6%)	1676 (17.7%)	15391 (18.9%)	3481 (17.7%)
4	47546 (21.1%)	11980 (22.1%)	2367 (25.0%)	18747 (23.1%)	4951 (25.2%)
5 (most deprived)	46330 (20.5%)	12201 (22.5%)	2561 (27.1%)	20383 (25.1%)	5567 (28.3%)
BMI (mean (SD))	26.94 (10.1)	27.19 (9.9)	27.24 (14.5)	28.19 (10.7)	28.37 (7.9)
eFI* (mean (SD))	3.13 (2.0)	3.80 (2.0)	4.52 (2.3)	9.00 (2.4)	10.31 (2.9)
eFI (median [IQR])	3 (2, 5)	4 (2,5)	5 (3, 6)	9 (7, 10)	10 (8, 12)
Frailty category*† (n (%))					
Fit	163732 (72.6%)	33478 (61.8%)	4634 (49.0%)	14 (0.0%)	7 (0.0%)
Mild	61755 (27.4%)	20528 (37.9%)	4457 (47.1%)	40226 (49.5%)	5801 (29.5%)
Moderate	111 (0.0%)	165 (0.3%)	369 (3.9%)	33954 (41.8%)	9834 (50.0%)
Severe	0	0	4 (0.0%)	7120 (8.8%)	4015 (20.4%)

BMI = Body mass index· eFI = electronic Frailty Index

* Baseline category: calculated on 31st December 2009

† Frailty categories are: fit (mean eFI/36: 0-0.12), mild (>0.12-0.24), moderate (>0.24-0.36), severe (>0.36)

Missingness was 6.1% for smoking, 26.6% for drinking, and 32.3% for BMI in GOLD. Missingness was 4.7% for smoking, 15.9% for drinking and 15.7% for BMI in Aurum

Table 5.8 Prevalence of eFI deficits by cluster at baseline and at last observation in Aurum

N (%)	At baseline (n (%))					At last observation (n (%))				
	Low-slow N = 225598	Low-moderate N = 54171	Low-rapid N = 9464	High-slow N = 81314	High-rapid N = 19657	Low-slow N = 225598	Low-moderate N = 54171	Low-rapid N = 9464	High-slow N = 81314	High-rapid N = 19657
eFI 1: Activity Limitation	1021 (0.5)	354 (0.7)	106 (1.1)	2380 (2.9)	752 (3.8)	1844 (0.8)	1560 (2.9)	793 (8.4)	3073 (3.8)	1447 (7.4)
eFI 2: Anaemic & haematinic deficiency	10898 (4.8)	3574 (6.6)	930 (9.8)	18359 (22.6)	5377 (27.4)	15743 (7.0)	10314 (19.0)	3147 (33.3)	22595 (27.8)	8960 (45.6)
eFI 3: Arthritis	51802 (23.0)	15534 (28.7)	2768 (29.2)	44789 (55.1)	11692 (59.5)	56680 (25.1)	19591 (36.2)	3458 (36.5)	46252 (56.9)	12521 (63.7)
eFI 4: Atrial fibrillation	9589 (4.3)	3445 (6.4)	902 (9.5)	15744 (19.4)	5015 (25.5)	14431 (6.4)	10765 (19.9)	3277 (34.6)	18955 (23.3)	7922 (40.3)
eFI 5: Cerebrovascular disease	9716 (4.3)	3325 (6.1)	830 (8.8)	15190 (18.7)	4733 (24.1)	12516 (5.5)	7543 (13.9)	2402 (25.4)	17083 (21.0)	6556 (33.4)
eFI 6: CKD	25910 (11.5)	8090 (14.9)	1785 (18.9)	31973 (39.3)	8783 (44.7)	33466 (14.8)	15876 (29.3)	3788 (40.0)	35949 (44.2)	11525 (58.6)
eFI 7: Diabetes	19301 (8.6)	6112 (11.3)	1500 (15.8)	24208 (29.8)	7362 (37.5)	23538 (10.4)	10497 (19.4)	2493 (26.3)	26139 (32.1)	8606 (43.8)
eFI 8: Dizziness	21009 (9.3)	6293 (11.6)	1263 (13.3)	25715 (31.6)	7004 (35.6)	24423 (10.8)	9926 (18.3)	2021 (21.4)	27352 (33.6)	8150 (41.5)
eFI 9: Dyspnoea	14356 (6.4)	5176 (9.6)	1154 (12.2)	29783 (36.6)	8659 (44.1)	18237 (8.1)	11267 (20.8)	3386 (35.8)	32978 (40.6)	11465 (58.3)
eFI 10: Falls	9972 (4.4)	3656 (6.7)	920 (9.7)	17800 (21.9)	5572 (28.3)	13088 (5.8)	8567 (15.8)	3078 (32.5)	20801 (25.6)	8686 (44.2)
eFI 11: Foot problems	3570 (1.6)	1179 (2.2)	289 (3.1)	7640 (9.4)	2546 (13.0)	6115 (2.7)	4144 (7.6)	1118 (11.8)	9812 (12.1)	4238 (21.6)
eFI 12: Fragility fracture	9362 (4.1)	2653 (4.9)	561 (5.9)	9794 (12.0)	2706 (13.8)	12329 (5.5)	7002 (12.9)	2028 (21.4)	11738 (14.4)	4792 (24.4)
eFI 13: Hearing impairment	23772 (10.5)	6986 (12.9)	1496 (15.8)	22640 (27.8)	6053 (30.8)	33445 (14.8)	14862 (27.4)	3155 (33.3)	26785 (32.9)	8534 (43.4)
eFI 14: Heart failure	2526 (1.1)	1002 (1.8)	298 (3.1)	8749 (10.8)	3185 (16.2)	3578 (1.6)	3946 (7.3)	1896 (20.0)	10314 (12.7)	5447 (27.7)
eFI 15: Heart valve disease	1025 (0.5)	426 (0.8)	111 (1.2)	1955 (2.4)	741 (3.8)	1560 (0.7)	1513 (2.8)	549 (5.8)	2483 (3.1)	1380 (7.0)
eFI 16: Housebound	16041 (7.1)	4989 (9.2)	1227 (13.0)	22442 (27.6)	6023 (30.6)	19009 (8.4)	9410 (17.4)	3269 (34.5)	24935 (30.7)	8728 (44.4)
eFI 17: Hypertension	92489 (41.0)	23971 (44.3)	4352 (46.0)	57622 (70.9)	14088 (71.7)	101567 (45.0)	31654 (58.4)	5662 (59.8)	59226 (72.8)	14990 (76.3)
eFI 18: Hypotension / syncope	8668 (3.8)	2781 (5.1)	635 (6.7)	14761 (18.2)	4548 (23.1)	11637 (5.2)	6849 (12.6)	2128 (22.5)	17034 (20.9)	6760 (34.4)
eFI 19: Ischaemic heart disease	29614 (13.1)	9091 (16.8)	1828 (19.3)	38198 (47.0)	10437 (53.1)	34620 (15.3)	15650 (28.9)	3596 (38.0)	40891 (50.3)	12495 (63.6)
eFI 20: Memory & cognitive problems	6261 (2.8)	1817 (3.4)	573 (6.1)	7615 (9.4)	2202 (11.2)	10598 (4.7)	7762 (14.3)	2853 (30.1)	10738 (13.2)	5393 (27.4)
eFI 21: Mobility and transfer problems	5231 (2.3)	2007 (3.7)	655 (6.9)	14021 (17.2)	4797 (24.4)	7188 (3.2)	5682 (10.5)	2595 (27.4)	16749 (20.6)	7926 (40.3)
eFI 22: Osteoporosis	12745 (5.6)	3483 (6.4)	679 (7.2)	12712 (15.6)	3573 (18.2)	16195 (7.2)	8208 (15.2)	1863 (19.7)	14581 (17.9)	5195 (26.4)
eFI 23: Parkinsonism & Tremor	2338 (1.0)	894 (1.7)	221 (2.3)	3814 (4.7)	1313 (6.7)	2923 (1.3)	1713 (3.2)	486 (5.1)	4138 (5.1)	1656 (8.4)
eFI 24: Peptic ulcer	3889 (1.7)	1255 (2.3)	245 (2.6)	5180 (6.4)	1532 (7.8)	4362 (1.9)	1975 (3.6)	493 (5.2)	5564 (6.8)	1919 (9.8)
eFI 25: Peripheral vascular disease	3311 (1.5)	1295 (2.4)	406 (4.3)	9192 (11.3)	3434 (17.5)	4458 (2.0)	3092 (5.7)	1100 (11.6)	10656 (13.1)	4908 (25.0)
eFI 26: Polypharmacy	126643 (56.1)	34297 (63.3)	6485 (68.5)	78842 (97.0)	19165 (97.5)	128155 (56.8)	49086 (90.6)	9108 (96.2)	74255 (91.3)	19175 (97.5)
eFI 27: Requirement for care	4801 (2.1)	1566 (2.9)	495 (5.2)	7005 (8.6)	2337 (11.9)	6365 (2.8)	4485 (8.3)	2015 (21.3)	8510 (10.5)	4263 (21.7)

N (%)	At baseline (n (%))					At last observation (n (%))				
	Low-slow N = 225598	Low-moderate N = 54171	Low-rapid N = 9464	High-slow N = 81314	High-rapid N = 19657	Low-slow N = 225598	Low-moderate N = 54171	Low-rapid N = 9464	High-slow N = 81314	High-rapid N = 19657
eFI 28: Respiratory disease	40611 (18.0)	11367 (21.0)	2164 (22.9)	39629 (48.7)	10164 (51.7)	47674 (21.1)	18636 (34.4)	3957 (41.8)	42527 (52.3)	12182 (62.0)
eFI 29: Skin ulcer	5787 (2.6)	1914 (3.5)	477 (5.0)	9191 (11.3)	2935 (14.9)	7908 (3.5)	4475 (8.3)	1420 (15.0)	11095 (13.6)	4642 (23.6)
eFI 30: Sleep disturbance	7681 (3.4)	2336 (4.3)	493 (5.2)	11509 (14.2)	3245 (16.5)	9148 (4.1)	4222 (7.8)	1153 (12.2)	12626 (15.5)	4394 (22.4)
eFI 31: Social vulnerability	9243 (4.1)	2930 (5.4)	656 (6.9)	11666 (14.3)	3519 (17.9)	12006 (5.3)	6028 (11.1)	1817 (19.2)	13471 (16.6)	5208 (26.5)
eFI 32: Thyroid disease	12795 (5.7)	3599 (6.6)	587 (6.2)	12517 (15.4)	3373 (17.2)	14101 (6.3)	5022 (9.3)	969 (10.2)	13118 (16.1)	3916 (19.9)
eFI 33: Urinary incontinence	5191 (2.3)	1788 (3.3)	377 (4.0)	9274 (11.4)	2753 (14.0)	6811 (3.0)	4465 (8.2)	1607 (17.0)	10775 (13.3)	4557 (23.2)
eFI 34: Urinary system disease	47420 (21.0)	12421 (22.9)	2237 (23.6)	42058 (51.7)	10341 (52.6)	57434 (25.5)	22324 (41.2)	4884 (51.6)	46302 (56.9)	13457 (68.5)
eFI 35: Visual impairment	48796 (21.6)	13203 (24.4)	2712 (28.7)	40904 (50.3)	10661 (54.2)	63379 (28.1)	24320 (44.9)	4807 (50.8)	46221 (56.8)	13646 (69.4)
eFI 36: Weight loss and anorexia	3758 (1.7)	1190 (2.2)	389 (4.1)	6921 (8.5)	2136 (10.9)	5449 (2.4)	3715 (6.9)	1571 (16.6)	8496 (10.4)	3876 (19.7)

5.5 Discussion

In this study, I characterised a cumulative deficit model (eFI) implemented in UK electronic health records in terms of transition between frailty categories, trajectory over four years and association with mortality. I applied joint modelling as an intermediary step to confirm that frailty progresses over time and that progression is associated with death. I then used joint latent class models to identify separate subgroups (clusters) within the older population with different eFI attributes (baseline and trajectory over time). I assessed these clusters' association with mortality risk. Lastly, I validated the resultant subgroups in an external dataset and assessed the final model's performance.

Main findings

Frailty transition between categories

Almost a quarter of the study population transitioned between frailty categories during the study period, mostly changing from fit to mild or mild to moderate. An analysis of eFI scores extracted from the large Welsh health database (SAIL) assessed the average transition times between the 4 eFI categories over 9 years of follow-up. Median times of transition from 'fit' to 'mild' frailty was around 6 years but this reduced to 2.5 years for those transitioning from 'moderate' to 'severe' frailty[128]. However, the shorter follow-up time in this study did not allow for a direct comparison. Most frailty transition studies used the phenotype model and reported similar findings in terms of transition magnitude, direction, and time[42-44 129]. One study based on cumulative frailty score found that most transitions occurred between mild to moderate during the 5 years prior to death[54].

Joint models and joint latent class models

Results of joint models suggest that eFI progresses slowly on average. I observed that the linear model was the best approach for modelling frailty with time. This is attributed to the fact that

frailty is a cumulative score, it can only be increasing over time. Worse eFI status and rate of change were associated with a higher risk of mortality[30 55 56]. Clegg et al.'s study calculated the three-year mortality risk based on an individual's single eFI value and found that a higher eFI baseline value is associated with a higher mortality risk[30]. Moreover, the joint modelling results from a study by Stolz et al. found that yearly change in frailty mean score over time was associated with an increased risk of mortality[56].

I identified five distinct clusters of patients with different eFI baseline/trajectories. In addition to the statistical considerations discussed above, the emergent 5-cluster model was clinically plausible, and the resultant clusters had distinct eFI patterns (baseline and trajectories).

Three groups started from a low eFI value and progressed at different slopes over time. Two started from an elevated eFI value and progressed slowly or at a rapid slope over time. The highest proportion of patients belonged to the low-steady cluster (71.1%), with around 4.6% belonging to the two rapid clusters. The study by Stow et al., although run on a shorter interval and focussing on the terminal frailty trajectories, found similar distributions, with 76.6% belonging to the stable frailty cluster. In general, the majority of older people usually belong to steady or moderately progressing clusters[49].

The risk analysis showed that those who progressed at a faster rate and/or had a high eFI baseline value were at higher risk of death compared to the other clusters. These groups were generally older, in agreement with previous studies which showed that older age was linked to worse frailty levels [39 44 45 129]. Findings from Chamberlain et al. and others -including those focusing on terminal frailty- were similar: steeper trajectories and older age groups had a higher risk of mortality [49 52 53], with the rate of accumulating frailty becoming five times higher in the last three years before death[52]. High frailty baseline value was found to be associated with higher 1-, 3- and 5-year mortality risk in Clegg et al. and other previous studies[23 30 128].

External validation of the models in CPRD Aurum

High posterior probability and similar HRs suggest good validation of the model in Aurum. Despite the good model performance in Aurum, the extracted sample of Aurum had a higher baseline eFI value compared to GOLD. This has led to small differences in the proportion of Aurum population belonging to each cluster. These differences are likely to occur when running studies on different EHR systems where different code lists are created. Additionally, the deficit recording was reliant on the general practitioners' accuracy and completeness.

Strengths, limitations and implications

Using joint modelling and joint latent class models is considered to be a particular strength of this study. Since eFI is directly calculated from routinely collected data, the absence of data is likely to be due to death. Traditional clustering methods require complete observations, leading to the use of a relatively healthier group[46]. The models applied in this study allowed me to account for those who were censored before the end of the study; thus, I included all available patients and minimised the risk of bias.

Individual clusters did not have distinctive deficits, but based on the clusters' starting level, the overall pattern went either up or down as a whole. It has been reported early on that the overall score rather than the individual deficits gives the predictive value for an outcome in an individual[25]. This makes monitoring an overall frailty score more effective, giving a general idea of a patient's health and risk of adverse events.

There are some cautions in interpreting these results. Firstly, the cohort was only followed up for a relatively short period. Although this 4-year period was long enough to observe distinct trajectory clusters, longer follow-up might show different relationships with mortality. The challenge of a longer follow-up is the potential that other confounders become more relevant in influencing mortality risk.

The main implication of this study is that monitoring frailty over time is a superior approach to identifying older people at higher risk of death than relying on a cross-sectional measure. The results have shown that both the baseline value and trajectory of frailty progression over time were associated with higher mortality risks. An important group for health providers to monitor is the low-rapid group, where patients started from a seemingly healthy state and deteriorated quickly within 4-5 years. While frailty usually has worsening trends, a small proportion (15-23%) of the older population's frailty state improves over time. This underscores the need for future research to delve deeper into these trends and patterns[42 43]. Exploring in detail eFI trajectory patterns over longer periods, how stable they are, and whether there are inflection points to identify the rapidly deteriorating groups would aid in tailoring preventive treatment packages for the old, frail population.

Conclusion

This study has laid the groundwork and introduced a fresh approach to modelling frailty over time. It has identified subgroups within the population who are currently frail or have rapid frailty progression, indicating the need for more healthcare monitoring and resources. However, to fully understand the implications of these findings, it is crucial to conduct future research that explores longer frailty monitoring periods and different patterns of morbidities, thereby broadening the scope of these subgroups.

6. Longitudinal trajectories of polypharmacy in older people, and their association with mortality

6.1 Chapter summary

In this chapter, I identified four distinct clusters of older people based on their polypharmacy progression over five years and assessed their mortality risk. I characterised those clusters clinically and described their morbidity burden and drug intake throughout the study period. I used a random sample of patients aged ≥65 extracted from CPRD GOLD to build the model and identify the different clusters. Then, I externally validated the clusters in a random sample of older people registered in CPRD Aurum.

The four clusters identified were: low-steady, intermediate-slow/increasing, intermediate-fast/increasing, and high-decreasing. The low-steady was the healthiest cluster, and over 85% of the older population belonged to that cluster. Compared to this healthy population, the other three clusters had higher drug use, comorbidity burden, and risk of mortality. The intermediate-fast/increasing cluster was the group with the highest risk of mortality: they had the highest baseline prevalence and proportion of cancer diagnosis over follow-up, and over 80% died during the study period. The high-decreasing and intermediate-slow/increasing clusters were characterised by having the highest prevalence of chronic diseases at baseline and accumulation over time, respectively. While the validation set had slightly different characteristics, the model had good validation metrics, and all clusters identified had similar progression over time, comorbidity burden, and mortality.

This study highlighted the importance of monitoring polypharmacy progression over time. It confirmed that both baseline and longitudinal progression are relevant in identifying those most vulnerable in the older population. It would be beneficial to observe the polypharmacy trends in middle aged people to understand how patients have accumulated drugs at a younger age.

Also, assessing the identified clusters in other settings and databases is important for the generalisability of the study findings.

6.2 Introduction

Polypharmacy is the use of multiple drugs either concurrently or cumulatively. It has often been defined based on a cross-sectional value, e.g., use of ≥5 or ≥10 in a given period[73]. Along with polypharmacy, problems like poor adherence, complex medication regimens, and inappropriate polypharmacy arise[130-132]. Furthermore, polypharmacy is associated with multiple adverse events, such as drug-drug interactions, adverse drug reactions, hospitalisation, and mortality[73 75 133-135].

Studies investigating the short- and long-term effects of polypharmacy have previously found an association between polypharmacy and the risk of death, falls, and multimorbidity[67 68 136]. Richardson et al. found that the use of ≥5 drugs increased the risk of 2-year mortality by ≥80% for UK patients aged over 65[68].

There is a scarcity of data modelling polypharmacy longitudinally over time. A Dutch study found that besides age and baseline number of drugs, chronic morbidities such as diabetes and hypertension were positively predictive of polypharmacy after a 4-year follow-up[80]. Others confirmed that polypharmacy increases more rapidly for patients leaving the study early[83] or during the year prior to death[77].

Numerous studies have described population-level polypharmacy trends, associated long-term effects, and variation over time. However, there is a gap in identifying polypharmacy changing over time, accounting for censored data and the heterogeneity of the older population.

In this study, I aimed to identify and characterise subgroups (clusters) of UK older people with distinct polypharmacy trajectories over five years, and to assess the association between these clusters and their risk of death.

6.3 Methods

Design

In this study, I assessed the change of polypharmacy over five years by conducting a longitudinal cohort analysis. I identified a random sample of the UK older population from primary care databases. I calculated polypharmacy in terms of the absolute number of ingredients taken annually during the follow-up period. Finally, I applied joint latent class models to identify clusters with distinct polypharmacy patterns over time and assessed their association with all-cause mortality. The identified clusters were characterised, and the model was validated in an external dataset.

Data source

I used CPRD GOLD (January 2023 release) to build the trajectory clusters and CPRD Aurum (May 2022 release) to validate the resultant clusters. Both databases were mapped to the OMOP-CDM (See 3.2).

Population

Random samples of 300,000 patients aged ≥65 and alive on 1st January 2015 from both GOLD and Aurum were eligible for inclusion if they have been registered for at least 1 year prior to start date. Duplicate patients belonging to both databases were removed from the Aurum sample.

Follow-up period was observed from 1st January 2015 to 31st December 2019. The choice of this end date was to avoid different trends that may have been caused by COVID-19. Patients exited the study at the earliest of: practice last collection date, patient transfer-out of practice date, death date[92] or study end date.

Exposure

Exposure was polypharmacy (3.3), calculated annually at baseline and at the end of each follow-up year.

Outcomes

The adverse outcome analysed was all-cause mortality, defined as death recorded in CPRD GOLD or CPRD Aurum during follow-up (3.3).

Statistical analysis

Descriptive analysis

The baseline demographics were extracted and compared between the two cohorts: GOLD and Aurum.

Identification of polypharmacy trajectories using a machine learning method

In this phase, I tried to capture all possible polypharmacy trends in the data using KmL (3.3). As this method -similar to traditional clustering methods- does not account for censored data, only patients surviving until the end date would be included (survivor cohort). Those patients would have six polypharmacy values: one at baseline and then one by the end of each follow-up year. To minimise the survival bias, I varied the end date of follow-up for modelling polypharmacy trajectories and created different models. I started by including patients who had at least three data points (baseline, 2015 and 2016), i.e., three consecutive values of polypharmacy over the first 2 years of follow-up. The number of data points was gradually increased to six, which is the maximum number of data points patients can have. All possible trends identified in this phase were carried forward to the next analysis. If non-linear trajectories of polypharmacy progression (e.g., quadratic or cubic) were observed in the clusters produced by KmL, their associated terms were incorporated into the subsequent joint latent class models.

Joint latent class modelling using the entire study population

To see if there were clusters of distinct patterns of polypharmacy change and whether allocation to such a cluster was predictive of mortality, I applied joint latent class model (3.3).

Since polypharmacy is count data, different link functions -available in the R package (lcmm)[102]- to handle dependent variables that are not necessarily Gaussian, were tested. These functions rely on the simultaneous normalisation of the variable and estimation of the regression parameters using parameterised link functions. The link functions explored were: linear (default), beta, and splines. Only these link functions were available for building a joint latent class model in 'lcmm'. Other link functions, such as the probit function, were not available for these models using this package. Multiple criteria were assessed to determine the best link function: convergence, difference in Bayesian information criterion (BIC), variance explained, and interpretability and plausibility of the results. Similarly, with survival models, Weibull, splines and piecewise survival models were assessed using the same criteria. After determining the best link function and survival function, two- to six-cluster models were fitted. The optimum number of clusters was based on the criteria described in Section 3.3.

Characterisation of the resulting clusters

To characterise the resulting clusters, comorbidities and drug use were compared across the clusters both, at baseline and last observation (3.3).

Model validation

To assess the performance of the emergent best-fit cluster model (derived from GOLD) in the Aurum validation set, I assessed the posterior probability (3.3). Then, I calculated the risk of mortality for the clusters identified in the validation sample and compared them to the original model.

I used R (Version 4.2.3) for analyses carried out in this study. R packages 'kml' version 2.4.6 [link], 'lcmm' version 2.0.2 [link], and 'PatientProfiles' version 0.4.0 [link] were used for K

means for longitudinal data analysis, joint latent class modelling, and characterisation of resultant clusters, respectively.

6.4 Results

Model Building in CPRD GOLD

Study population

For CPRD GOLD, there were 299,859 patients analysed. Mean age was 75.3 (SD 7.9) with 45.4% male patients. At baseline, the median number of ingredients taken the year prior was 8 [IQR 4 to 13] and mean Charlson morbidity index was 1.2 (1.5). Table 6.1 summarises the baseline characteristics and Table 6.2 and Table 6.3 summarise the individual morbidities diagnoses and drug classes use in the study population.

Table 6.1 Baseline characteristics by cluster and for the overall population in GOLD

	Intermediate-fast/increasing	Low-steady	Intermediate-slow/increasing	High-decreasing	Overall
n	3708 (1.2%)	256923 (85.7%)	19207 (6.4%)	20021 (6.7%)	299859
Sex = Male (n (%))	1838 (49.6%)	117450 (45.7%)	8745 (45.5%)	7987 (39.9%)	136020 (45.4%)
Age (mean (SD))	76.73 (7.56)	75.01 (7.90)	76.77 (7.85)	76.86 (7.88)	75.26 (7.91)
Number of ingredients (median [IQR])*	11 [6, 16]	7 [3, 11]	14 [10, 19]	26 [22, 30]	8.0 [4, 13]
Charlson morbidity index (mean (SD))	1.95 (1.92)	1.03 (1.41)	1.83 (1.80)	2.32 (1.99)	1.18 (1.54)

* In the year prior to start

KmL

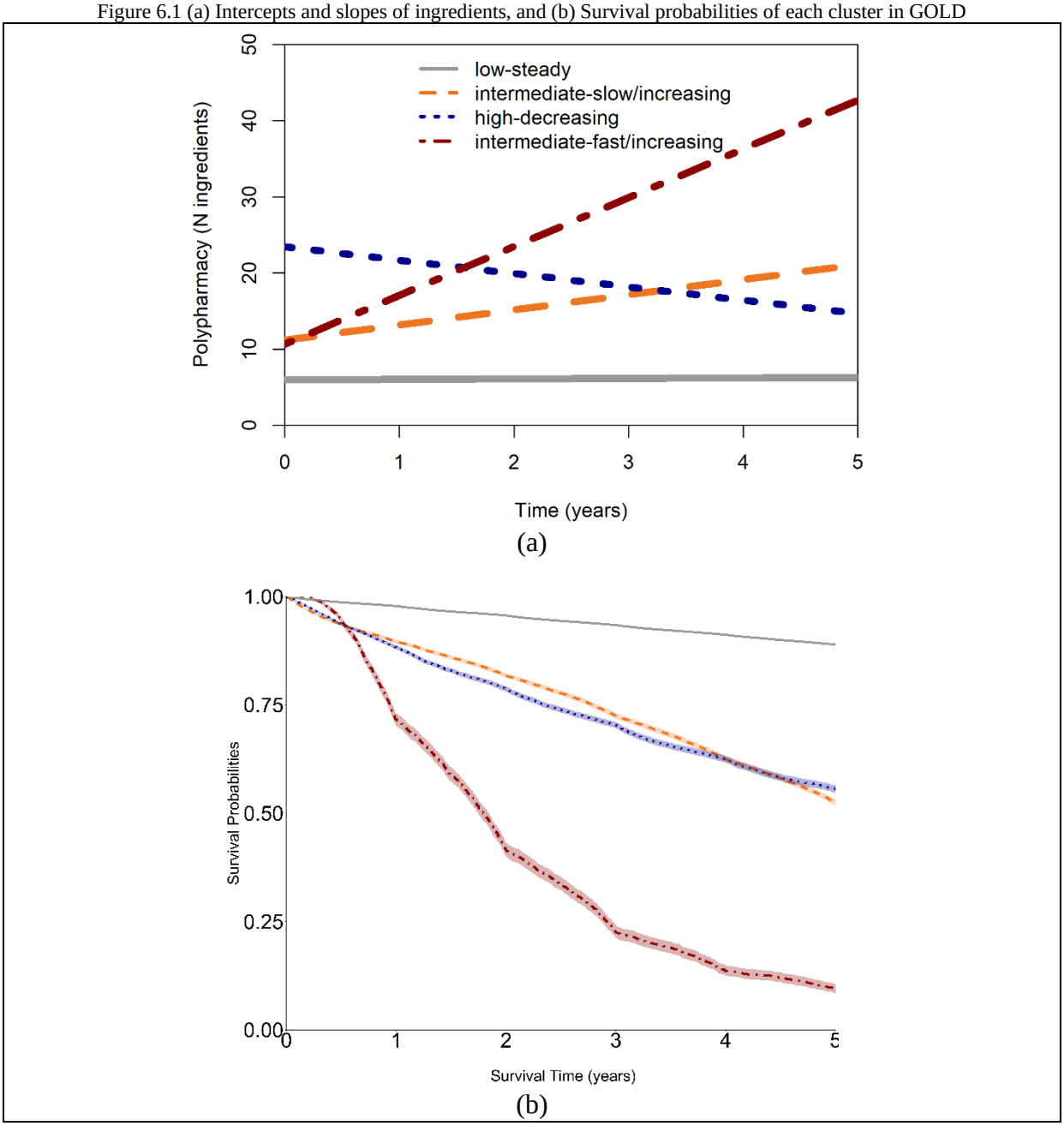
Table S 6.1 and Figure S 6.1 in the Appendix describe the identified clusters with each cut-off point (i.e., 2, 3, 4 and 5 years of observation). No particular shapes were identified, and all trends seemed to be progressing linearly over time.

Joint latent class modelling

The GOLD population was followed up for 4.1 years [IQR 1.4 to 5.0 years] with 38,031 (12.7%) patients dying during the study period. For joint latent class modelling, the linear link function with Weibull proportional hazard were chosen and taken forward. Table S 6.2 and Table S 6.3 in the Appendix compare the performance between the different link and survival functions.

The four-cluster model was the optimum model after assessing BIC, convergence, and clinical plausibility. Table S 6.4 in the Appendix summarises the models tested. The four clusters identified were: low-steady (85.7%, intercept=5.97, slope=0.07), intermediate-slow/increasing (6.4%, intercept=11.22, slope=1.98), intermediate-fast/increasing (1.2%, intercept=10.70, slope=6.40), and high-decreasing (6.7%, intercept=23.35, slope=-1.75). Taking the low-steady class (7.6% deaths) as reference, the intermediate-fast/increasing class had the highest increase in mortality, with a hazard ratio (HR) of 20.5 (95%CI 19.9, 21.2); and 83.8% of patients in this cluster died over the five-year follow-up period. The second and third with highest mortality were the intermediate-slow/increasing and high-decreasing clusters, with HR of 4.95 (95%CI 4.84, 5.07) and 4.64 (95%CI 4.58, 4.73) respectively. Figure 6.1 describes these clusters in terms of progression over time and survival probabilities. Compared to the low-steady, the other clusters had a slightly higher mean age (75.0 vs. 76.7 to 76.9) at baseline. The intermediate-high/decreasing had the highest proportion of males (49.6%). The high-decreasing cluster had the highest baseline mean Charlson morbidity index (2.3 (2.0)) and number of ingredients (27.0 (7.0)). Appendix Table S 6.5 reports the estimates for each cluster. Median follow-up, yearly deaths and polypharmacy are reported in Appendix Table S 6.6.

Figure 6.1 (a) Intercepts and slopes of ingredients, and (b) Survival probabilities of each cluster in GOLD



Characterisation of the identified clusters

Table 6.2 summarises baseline and end of follow-up comorbidities, while Table 6.3 reports on polypharmacy by different ATC classes and use of common drugs use in the different clusters as well as the overall population. Table S 6.7 and Table S 6.8 in the Appendix describe the percentage difference between baseline and at the last observation.

Comorbidities

The low-steady cluster had the lowest Charlson comorbidity index at baseline and by the last observation (mean 1.0 (SD 1.4), Δ=0.4). As a result, the cluster had the least comorbidities and drug use, and was therefore including the healthiest of all study participants. For example, compared to the low-steady cluster, all other clusters had higher -and similar- dementia prevalence by the end of follow-up: ranging from 10.5% to 11.6% in the other clusters vs. 5.6% in the low-steady cluster.

Conversely, the high-decreasing cluster had the highest Charlson comorbidity index at baseline but a slow increase over time (mean 2.3 (SD 2.0), Δ=0.6). Intermediate-high/increasing and intermediate-slow/increasing on the other hand had slightly lower baseline index but bigger increases in Charlson comorbidity index by the last observation; (mean 2.0 (SD 1.9), Δ=1.5) and (mean 1.8 (SD 1.8), Δ=1.1), respectively.

The main characteristic of the intermediate/high-increasing cluster was the high prevalence of cancer: 25.9% of the patients in this cluster had malignancy/ies at baseline, rising to 53.7% at the last observation (Δ=27.8%). Specifically, a diagnosis of metastatic solid tumour was also highest in this cluster: 1.5% at baseline and 9.0% at the last observation (Δ=7.5%). The intermediate-slow/increasing cluster had the second highest cancer onset; 18.5% of patients had prior any malignancy diagnosis at baseline, rising to 33.5% (Δ=15.0%) at the last observation. Additionally, the intermediate-fast/increasing cluster had the highest onset of pneumonia and venous thromboembolism; 8.8% and 6.2% vs. 7.8% and 4.4% in the intermediate-slow/increasing, respectively.

The high-decreasing cluster had the highest burden of non-cancer chronic disease diagnoses at baseline and at the last observation. For example, proportion of patients with chronic obstructive pulmonary disease (COPD) was 32.7% at baseline and 36.6% at the last

observation; this was followed by the intermediate-slow/increasing where 24.2% and 31.6% of patients were diagnosed with COPD at baseline and last observation, respectively. Similar patterns were seen for other chronic diseases, such as diabetes, chronic kidney disease, and osteoporosis. Anxiety and depression were also most common in this cluster; 26.4% and 25.1% at the last observation, respectively.

The intermediate-slow/increasing cluster participants had the highest onset of most -especially chronic- diseases over the study period. While starting at lower prevalence at baseline, the patients belonging to this cluster accumulated diseases rapidly over the study period and had a similar proportion of chronic diagnoses as the high-decreasing population by their last observation. For example, 26.6% had chronic kidney disease at baseline, rising to 33.9% (Δ=7.3%) at the last observation in the intermediate-slow/increasing, compared to 31.9% at baseline to 36.9% (Δ=5.0) at end of follow-up in the high-decreasing cluster.

Medication use

Use of drugs and drug classes followed the overall trend of each cluster, e.g., use of drugs from the musculoskeletal system class decreased from 62.0% at baseline to 46.7% at end of follow-up (Δ=-15.3%) for the high-decreasing cluster.

After assessing the use of common drugs in the study population, an overall increasing trend across all the clusters was observed, with the low-steady and high-decreasing having the smallest changes between baseline and the last observation.

Similar to that observed for comorbidities, the use of drugs for chronic disease treatments, such as drugs for diabetes and beta blockers were highest in the high-decreasing cluster at the last observation, followed by the intermediate-slow/increasing. For example, 29.9% of patients in the high-decreasing cluster used drugs for diabetes in the year prior to last observation vs.

24.6%, 21.0% and 11.6% for the intermediate-slow/increasing, intermediate-fast/increasing, and low-steady clusters, respectively.

The highest rise in patients drug use between baseline and end of follow-up was observed in the intermediate-increasing clusters. There were 20.9% and 12.0% more patients using opioids, for example between baseline and end of follow-up in the intermediate-fast/increasing and the intermediate-slow/increasing clusters, respectively. On the other hand, in low-steady and high-decreasing clusters only 6.3% and 3.3% more patients were using opioids at the end of the study compared to baseline.

Alimentary tract and metabolism (e.g., drugs for acid related disorder), cardiovascular system (e.g., lipid modifying agents), and nervous system (e.g., opioids) were the most common drug classes at baseline and at end of follow-up, with the low-steady cluster having the lowest rate of use.

Table 6.2 Individual morbidities at baseline and the last observation (Including Charlson morbidities) for each cluster and the overall population in GOLD

	At baseline					At the last observation				
	Intermediate-fast/increasing N=3708	Low-steady N=256923	Intermediate-slow/increasing N=19207	High-decreasing N=20021	Overall N=299859	Intermediate-fast/increasing N=3708	Low-steady N=256923	Intermediate-slow/increasing N=19207	High-decreasing N=20021	Overall N=299859
Mean (SD)										
Charlson comorbidity index	1.95 (1.92)	1.03 (1.41)	1.83 (1.80)	2.32 (1.99)	1.18 (1.54)	3.46 (2.54)	1.39 (1.66)	2.95 (2.28)	2.90 (2.20)	1.61 (1.84)
n (%)										
Myocardial infarction	190 (5.1)	7944 (3.1)	1170 (6.1)	1744 (8.7)	11048 (3.7)	331 (8.9)	10518 (4.1)	1949 (10.1)	2119 (10.6)	14917 (5.0)
Congestive heart failure	204 (5.5)	5491 (2.1)	1187 (6.2)	1875 (9.4)	8757 (2.9)	422 (11.4)	9039 (3.5)	2565 (13.4)	2640 (13.2)	14666 (4.9)
Peripheral vascular disease	168 (4.5)	4051 (1.6)	737 (3.8)	1023 (5.1)	5979 (2.0)	217 (5.9)	5489 (2.1)	1143 (6.0)	1249 (6.2)	8098 (2.7)
Cerebrovascular disease	431 (11.6)	14468 (5.6)	2023 (10.5)	2715 (13.6)	19637 (6.5)	603 (16.3)	20598 (8.0)	3121 (16.2)	3386 (16.9)	27708 (9.2)
Dementia	251 (6.8)	7166 (2.8)	1029 (5.4)	1165 (5.8)	9611 (3.2)	421 (11.4)	14297 (5.6)	2227 (11.6)	2102 (10.5)	19047 (6.4)
Chronic obstructive pulmonary disease	808 (21.8)	31383 (12.2)	4645 (24.2)	6544 (32.7)	43380 (14.5)	986 (26.6)	38033 (14.8)	6067 (31.6)	7334 (36.6)	52420 (17.5)
Rheumatologic disease	156 (4.2)	8159 (3.2)	998 (5.2)	1657 (8.3)	10970 (3.7)	195 (5.3)	10475 (4.1)	1534 (8.0)	1891 (9.4)	14095 (4.7)
Peptic ulcer disease	115 (3.1)	5174 (2.0)	625 (3.3)	925 (4.6)	6839 (2.3)	165 (4.4)	6332 (2.5)	896 (4.7)	1090 (5.4)	8483 (2.8)
Mild liver disease	32 (0.9)	685 (0.3)	130 (0.7)	189 (0.9)	1036 (0.3)	63 (1.7)	1024 (0.4)	235 (1.2)	251 (1.3)	1573 (0.5)
Diabetes with chronic complications	381 (10.3)	11909 (4.6)	2140 (11.1)	3234 (16.2)	17664 (5.9)	486 (13.1)	15964 (6.2)	3046 (15.9)	3990 (19.9)	23486 (7.8)
Hemiplegia or paraplegia	9 (0.2)	208 (0.1)	44 (0.2)	54 (0.3)	315 (0.1)	13 (0.4)	242 (0.1)	64 (0.3)	61 (0.3)	380 (0.1)
Renal disease	895 (24.1)	41212 (16.0)	5146 (26.8)	6466 (32.3)	53719 (17.9)	1163 (31.4)	51576 (20.1)	6958 (36.2)	7730 (38.6)	67427 (22.5)
Any malignancy	960 (25.9)	34856 (13.6)	3561 (18.5)	4015 (20.1)	43392 (14.5)	1991 (53.7)	48318 (18.8)	6427 (33.5)	5190 (25.9)	61926 (20.7)
Moderate to severe liver disease	18 (0.5)	184 (0.1)	55 (0.3)	87 (0.4)	344 (0.1)	42 (1.1)	326 (0.1)	122 (0.6)	125 (0.6)	615 (0.2)
Metastatic solid tumour	55 (1.5)	454 (0.2)	105 (0.5)	141 (0.7)	755 (0.3)	333 (9.0)	1262 (0.5)	592 (3.1)	279 (1.4)	2466 (0.8)
AIDS	1 (0)	18 (0)	1 (0)	0 (0)	20 (0)	1 (0)	21 (0)	2 (0)	1 (0)	25 (0)
Hypertension	1254 (33.8)	82262 (32.0)	6965 (36.3)	7683 (38.4)	98164 (32.7)	1334 (36.0)	91691 (35.7)	7835 (40.8)	8126 (40.6)	108986 (36.4)
Heart failure	234 (6.3)	6175 (2.4)	1330 (6.9)	2036 (10.2)	9775 (3.3)	458 (12.4)	10071 (3.9)	2769 (14.4)	2836 (14.2)	16134 (5.4)
Osteoporosis	309 (8.3)	14311 (5.6)	1847 (9.6)	2852 (14.3)	19319 (6.4)	435 (11.7)	18594 (7.2)	2875 (15.0)	3378 (16.9)	25282 (8.4)
Gastroesophageal reflux disease	120 (3.2)	8460 (3.3)	935 (4.9)	1397 (7.0)	10912 (3.6)	161 (4.3)	10466 (4.1)	1250 (6.5)	1618 (8.1)	13495 (4.5)
Chronic kidney disease	882 (23.8)	40772 (15.9)	5100 (26.6)	6390 (31.9)	53144 (17.7)	1062 (28.6)	49666 (19.3)	6501 (33.9)	7385 (36.9)	64614 (21.6)
Venous thromboembolism	233 (6.3)	11996 (4.7)	1329 (6.9)	1972 (9.9)	15530 (5.2)	463 (12.5)	15397 (6.0)	2171 (11.3)	2464 (12.3)	20495 (6.8)

	At baseline					At the last observation				
	Intermediate-fast/increasing N=3708	Low-steady N=256923	Intermediate-slow/increasing N=19207	High-decreasing N=20021	Overall N=299859	Intermediate-fast/increasing N=3708	Low-steady N=256923	Intermediate-slow/increasing N=19207	High-decreasing N=20021	Overall N=299859
Hypothyroidism	288 (7.8)	16937 (6.6)	1792 (9.3)	2267 (11.3)	21284 (7.1)	335 (9.0)	19153 (7.5)	2127 (11.1)	2489 (12.4)	24104 (8.0)
Stroke	218 (5.9)	6784 (2.6)	1058 (5.5)	1354 (6.8)	9414 (3.1)	331 (8.9)	9996 (3.9)	1725 (9.0)	1741 (8.7)	13793 (4.6)
Anxiety	519 (14.0)	29632 (11.5)	2980 (15.5)	4692 (23.4)	37823 (12.6)	648 (17.5)	34024 (13.2)	3788 (19.7)	5275 (26.4)	43735 (14.6)
Asthma	408 (11.0)	19425 (7.6)	2716 (14.1)	4187 (20.9)	26736 (8.9)	456 (12.3)	21893 (8.5)	3214 (16.7)	4546 (22.7)	30109 (10.0)
Pneumonia	182 (4.9)	4416 (1.7)	736 (3.8)	1447 (7.2)	6781 (2.3)	508 (13.7)	8437 (3.3)	2225 (11.6)	2498 (12.5)	13668 (4.6)
Diabetes	630 (17.0)	27625 (10.8)	3755 (19.6)	4922 (24.6)	36932 (12.3)	772 (20.8)	33544 (13.1)	4681 (24.4)	5551 (27.7)	44548 (14.9)
Inflammatory bowel disease	38 (1.0)	1672 (0.7)	190 (1.0)	273 (1.4)	2173 (0.7)	44 (1.2)	1889 (0.7)	243 (1.3)	296 (1.5)	2472 (0.8)
Depressive disorder	499 (13.5)	25187 (9.8)	2848 (14.8)	4571 (22.8)	33105 (11.0)	603 (16.3)	28021 (10.9)	3481 (18.1)	5034 (25.1)	37139 (12.4)

Table 6.3 Drug use by drug classes and common drugs at baseline and the last observation for each cluster and the overall population in GOLD

	At baseline (n (%))					At the last observation (n (%))				
	Intermediate-fast/increasing N=3708	Low-steady N=256923	Intermediate-slow/increasing N=19207	High-decreasing N=20021	Overall N=299859	Intermediate-fast/increasing N=3708	Low-steady N=256923	Intermediate-slow/increasing N=19207	High-decreasing N=20021	Overall N=299859
Drug classes										
Alimentary tract and metabolism	2833 (76.4)	154318 (60.1)	16518 (86.0)	19935 (99.6)	193604 (64.6)	3701 (99.8)	153680 (59.8)	18975 (98.8)	18953 (94.7)	195309 (65.1)
Blood and blood forming organs	1924 (51.9)	86176 (33.5)	11498 (59.9)	15068 (75.3)	114666 (38.2)	2701 (72.8)	93829 (36.5)	14212 (74.0)	14096 (70.4)	124838 (41.6)
Cardiovascular system	2767 (74.6)	167390 (65.2)	16098 (83.8)	18681 (93.3)	204936 (68.3)	3130 (84.4)	171949 (66.9)	17088 (89.0)	17484 (87.3)	209651 (69.9)
Dermatological	1446 (39.0)	78421 (30.5)	9376 (48.8)	15957 (79.7)	105200 (35.1)	2434 (65.6)	68036 (26.5)	12442 (64.8)	12238 (61.1)	95150 (31.7)
Genito-urinary system and sex hormones	710 (19.1)	41727 (16.2)	4724 (24.6)	7338 (36.7)	54499 (18.2)	931 (25.1)	41997 (16.3)	5878 (30.6)	5810 (29.0)	54616 (18.2)
Systemic hormonal preparations, excluding sex hormones and insulins	474 (12.8)	28391 (11.1)	3194 (16.6)	4659 (23.3)	36718 (12.2)	627 (16.9)	29786 (11.6)	3904 (20.3)	4354 (21.7)	38671 (12.9)
Anti-infective for systemic use	1966 (53.0)	99729 (38.8)	11632 (60.6)	17480 (87.3)	130807 (43.6)	3021 (81.5)	81383 (31.7)	14806 (77.1)	13138 (65.6)	112348 (37.5)
Antineoplastic and immunomodulation agents	305 (8.2)	10089 (3.9)	1303 (6.8)	2006 (10.0)	13703 (4.6)	444 (12.0)	11468 (4.5)	1971 (10.3)	1729 (8.6)	15612 (5.2)
Musculoskeletal system	1281 (34.5)	73698 (28.7)	7976 (41.5)	12409 (62.0)	95364 (31.8)	1764 (47.6)	64176 (25.0)	9654 (50.3)	9348 (46.7)	84942 (28.3)
Nervous system	2428 (65.5)	117985 (45.9)	14463 (75.3)	18800 (93.9)	153676 (51.2)	3586 (96.7)	114847 (44.7)	17519 (91.2)	17605 (87.9)	153557 (51.2)
Anti-parasitic products, insecticides and repellents	305 (8.2)	10089 (3.9)	1303 (6.8)	2006 (10.0)	13703 (4.6)	444 (12.0)	11468 (4.5)	1971 (10.3)	1729 (8.6)	15612 (5.2)
Respiratory system	1826 (49.2)	81876 (31.9)	10847 (56.5)	15914 (79.5)	110463 (36.8)	2960 (79.8)	75849 (29.5)	14052 (73.2)	13400 (66.9)	106261 (35.4)
Sensory organs	440 (11.9)	24551 (9.6)	2787 (14.5)	5112 (25.5)	32890 (11.0)	1338 (36.1)	25376 (9.9)	4558 (23.7)	3936 (19.7)	35208 (11.7)
Various	12 (0.3)	156 (0.1)	48 (0.2)	178 (0.9)	394 (0.1)	64 (1.7)	181 (0.1)	151 (0.8)	170 (0.8)	566 (0.2)
Drugs										
Psychostimulants	<5	152 (0.1)	17 (0.9)	40 (0.2)	211 (0.1)	6 (0.2)	168 (0.1)	20 (0.1)	44 (0.2)	238 (0.1)
Opioids	2731 (73.7)	146030 (56.8)	15389 (80.1)	18238 (91.1)	182388 (60.8)	3506 (94.55)	162322 (63.2)	17687 (92.1)	18890 (94.4)	202405 (67.5)
Drugs for obstructive airway disease	1879 (50.7)	98802 (38.5)	11026 (57.4)	14360 (71.7)	126067 (42.0)	2246 (60.6)	114481 (44.6)	13099 (68.2)	15300 (76.4)	145126 (48.4)
Drugs used in diabetes	635 (17.1)	25264 (9.8)	3952 (20.6)	5542 (27.7)	35393 (11.8)	778 (21.0)	29850 (11.6)	4731 (24.6)	5982 (29.9)	41341 (13.8)
Psycholeptics	1914 (51.6)	98049 (38.2)	11065 (57.6)	14782 (73.8)	125810 (42.0)	3193 (86.1)	113820 (44.3)	14672 (76.4)	16149 (80.7)	147834 (49.3)

	At baseline (n (%))					At the last observation (n (%))				
	Intermediate-fast/increasing N=3708	Low-steady N=256923	Intermediate-slow/increasing N=19207	High-decreasing N=20021	Overall N=299859	Intermediate-fast/increasing N=3708	Low-steady N=256923	Intermediate-slow/increasing N=19207	High-decreasing N=20021	Overall N=299859
Agents renin-angiotensin systemic	1927 (52.0)	106426 (41.4)	11310 (58.9)	13568 (67.8)	133231 (44.4)	2136 (57.6)	117443 (45.7)	12776 (66.5)	14190 (70.9)	146545 (48.9)
Anti-thrombotic	1168 (31.5)	48833 (19.0)	6689 (34.8)	8879 (44.4)	65569 (21.9)	1874 (50.5)	67547 (26.3)	10358 (53.9)	10569 (52.8)	90348 (30.1)
Lipid modifying agents	2237 (60.3)	126158 (49.1)	13104 (68.2)	15393 (76.9)	156892 (52.3)	2449 (66.1)	141264 (55.9)	14495 (75.5)	15910 (79.5)	174118 (58.1)
Calcium channel blockers	1598 (43.1)	89837 (35.0)	9428 (49.1)	11424 (57.1)	112287 (37.5)	1794 (48.4)	104550 (40.7)	11039 (57.5)	12185 (60.9)	129568 (43.2)
Diuretics	1957 (52.8)	96627 (37.6)	11327 (59.0)	14192 (70.9)	124103 (41.4)	2523 (68.0)	108436 (42.2)	13786 (71.8)	15201 (75.9)	139946 (46.7)
Antibacterial systemic	3353 (90.4)	213020 (82.9)	18065 (94.1)	19782 (98.8)	254220 (84.8)	3638 (98.1)	225382 (87.7)	19008 (99.0)	19934 (99.6)	267962 (89.4)
Immunosuppressant	101 (2.7)	4390 (1.7)	733 (3.8)	1238 (6.2)	6462 (2.2)	127 (3.4)	5227 (2.0)	1007 (5.2)	1385 (6.9)	7746 (2.6)
Anti-inflammatory and/or anti-rheumatic	2644 (71.3)	168062 (65.4)	14796 (77.0)	17302 (86.4)	202804 (67.6)	3150 (85.0)	180436 (70.2)	16497 (85.9)	17900 (89.4)	217983 (72.7)
Hormonal contraceptives systemic	97 (2.6)	8904 (3.5)	581 (3.0)	806 (4.0)	10388 (3.5)	122 (3.3)	9032 (3.5)	620 (3.2)	835 (4.2)	10609 (3.5)
Anti-epileptics	627 (16.9)	23661 (9.2)	4080 (21.2)	7134 (35.6)	35502 (11.8)	1300 (35.1)	33036 (12.9)	6672 (34.7)	8674 (43.3)	49682 (16.6)
Drugs for acid related disorder	2570 (69.3)	144275 (56.2)	14799 (77.1)	18227 (91.0)	179871 (60.0)	3313 (89.4)	164943 (64.2)	17220 (89.7)	18803 (93.9)	204279 (68.1)
Beta blocking agents	1441 (38.9)	81515 (31.7)	8655 (45.1)	10383 (51.9)	101994 (34.0)	1813 (48.9)	92261 (35.9)	10753 (56.0)	11278 (56.3)	116105 (38.7)
Antidepressants	1691 (45.6)	83376 (32.5)	10015 (52.1)	14142 (70.6)	109224 (36.4)	2338 (63.1)	98832 (38.5)	12860 (67.0)	15452 (77.2)	129482 (43.2)
Antineoplastic agents	281 (7.6)	15320 (6.0)	2210 (11.5)	3242 (16.2)	21053 (7.0)	310 (8.4)	16491 (6.4)	2478 (12.9)	3401 (17.0)	22680 (7.6)

Model validation in CPRD Aurum

Study population

There were 265,101 patients identified in CPRD Aurum. At baseline, Aurum population had similar characteristics to those in GOLD: mean age was 75.3 (SD 8.0) with 45.6% of the study population being male. Median number of ingredients was 8 [IQR 4 to 13]. The Aurum population had a slightly higher comorbidity burden than GOLD; 1.5 (1.7). Table 6.4 summarises the baseline characteristics, Table 6.5 and Table 6.6 summarise the individual morbidities diagnoses and drug classes' use in the study population.

Table 6.4 Baseline characteristics by cluster and for the overall population in Aurum

Aurum	Intermediate-fast/increasing	Low-steady	Intermediate-slow/increasing	High-decreasing	Overall
n	3396 (1.3%)	225383 (85.0%)	20639 (7.8%)	15683 (5.9%)	265101
Sex = Male (n (%))	1760 (51.8%)	103242 (45.8%)	9607 (46.5%)	6275 (40.0%)	120884 (45.6%)
Age (mean (SD))	76.98 (7.52)	74.96 (7.84)	77.20 (7.95)	77.36 (7.96)	75.3 (7.90)
Number of ingredients (median [IQR])*	11 [7, 16]	7 [3, 11]	14 [9, 19]	25 [22, 30]	8 [4, 13]
Charlson morbidity index (mean (SD))	2.48 (2.10)	1.26 (1.54)	2.31 (1.96)	2.83 (2.10)	1.45 (1.69)

* In the year prior to start

Resultant clusters

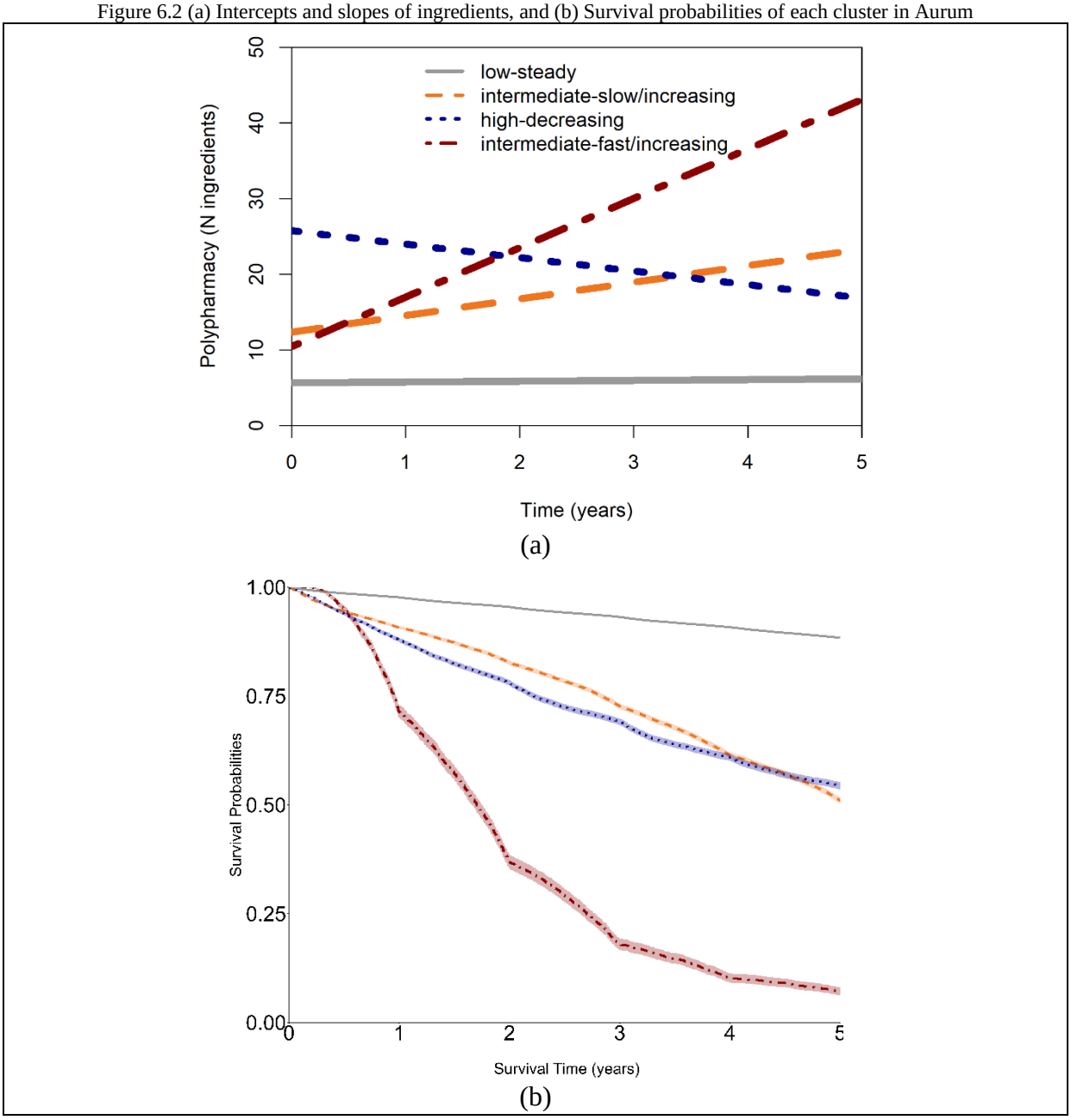
Compared to GOLD's, Aurum patients had a longer median follow-up of 5.0 years [IQR 3.9 to 5.0 years] with 43,346 (16.4%) dying during the follow-up period.

The proportions of patients belonging to each of the resultant clusters in Aurum were similar to GOLD with an increase of 1.4% in the intermediate-slow/increasing and a reduction of 0.8% in the high-decreasing Aurum clusters.

The Aurum population showed good validation metrics; more than 50% of the Aurum population had a posterior probability ≥ 0.7 (Table S 6.9). Table S 6.10 shows the mean posterior probability for patients in each cluster to belong to another cluster: the highest posterior probability mean was ~0.18-0.19 for intermediate-fast/increasing to belong to the intermediate-slow/increasing and the intermediate-slow/increasing to belong to the low-steady cluster. The low means indicate a low probability of misclassification.

The resultant clusters had similar trajectories of ingredients change over the follow-up period. They also had similar trends of mortality risks. Taking the low-steady cluster as reference (10.6% deaths), intermediate-fast/increasing had the highest HR 16.90 (95%CI 16.27, 17.56) with 89.9% deaths, followed by intermediate-slow/increasing and high-decreasing; 4.40 (95%CI 4.32, 4.51) and 4.34 (95%CI 4.24, 4.43), respectively. See Figure 6.2. Median follow-up, yearly deaths and polypharmacy are reported in Appendix Table S 6.11.

Figure 6.2 (a) Intercepts and slopes of ingredients, and (b) Survival probabilities of each cluster in Aurum



Characterisation of the identified cluster

The morbidity burden in the different clusters and the overall population are summarised in Table 6.5, while Table 6.6 describes polypharmacy by ATC classes and use of common drugs. Table S 6.12 and Table S 6.13 show the difference between baseline and last observation of morbidities and medication use in the study population.

In general, the resulting clusters in Aurum showed similar comorbidity and medications profiles to those in GOLD.

Comorbidities

The low-steady cluster in Aurum had the lowest Charlson morbidity index at baseline and the slowest change over time (mean 1.3 (SD 1.5), Δ=0.5). This meant a lower proportion of diagnosis of individual morbidities at baseline and a lower development of further morbidities over the study period.

The intermediate-fast/increasing cluster had the highest baseline cancer diagnosis and onset throughout the study period. Proportion of patients with any malignancy at baseline was 32.1%, jumping to 56.5% by the end of follow-up (Δ=24.4%). Baseline metastatic solid tumour diagnoses were 2.0%, rising to 10.1% by the last observation (Δ=8.1%). This intermediate-slow/increasing cluster had the second highest onset of cancer (Δ=14.6%) and (Δ=3.0%) for any malignancy and metastatic solid tumour, respectively. The intermediate-fast increasing cluster also had the highest onset of venous thromboembolism (Δ =7.3%) followed by the intermediate-slow/increasing cluster (Δ=4.9%)

The high-decreasing cluster was characterised by having the highest proportion of patients with chronic diseases, both at baseline and the last observation. For example, 36.3% of patients had COPD at baseline, rising to 40.6% at the last observation. This was followed by the intermediate-slow/increasing cluster, where 29.3% patients had COPD at baseline and 37.3%

at the last observation. Many chronic morbidities had the same pattern, i.e., high baseline and end of follow-up prevalence. These included diabetes, chronic kidney disease and osteoporosis.

The intermediate-slow/increasing cluster was the fastest at accumulating chronic conditions during the study period. For example, 7.4% of this cluster's patients had rheumatologic disease at baseline and 10.1% at their last observation (Δ=2.7%).

Medication use

For drug classes, they followed the overall trend of the clusters. Use of musculoskeletal system drugs, for example, at baseline was 62.1%, falling to 45.8% (Δ=-16.3%) for the high-decreasing cluster.

Upon looking at individual drugs of interest in this group, they increased over time for all clusters (and the overall population) during the study period. The low-steady and high-decreasing clusters had the lowest increase over time.

The high-decreasing cluster had the highest proportion of patients using drugs for treating chronic diseases, such as drugs for diabetes and beta blockers at both baseline and the last observation. For example, use of beta blockers in the last year of observation was 60.4% in the high-decreasing cluster vs. 39.2%, 58.6%, and 52.8% in the low-steady, intermediate-slow/increasing, and intermediate-fast/increasing clusters, respectively.

Both the intermediate-increasing clusters had a large rise in use of drugs. For example, 22.2% and 12.1% more patients were using opioids by the last observation compared to baseline for intermediate-fast/increasing and intermediate-slow/increasing clusters, respectively. On the other hand, only a rise of 7.1% and 3.8% in opioids use were observed in the low-steady and high-decreasing clusters, respectively.

Like in GOLD, alimentary tract and metabolism, cardiovascular system, anti-infective for systemic use, and nervous system treatments were the most common drug classes at baseline

and by the end of study follow-up for all patients, with the low-steady cluster having the lowest proportions of patients using drugs from these classes.

Table 6.5 Individual comorbidities at baseline and the last observation in Aurum

	At baseline					At the last observation				
	Medium-fast/increasing N=3396	Low-steady N=225383	Medium-slow/increasing N=20639	High-decreasing N=15683	Overall N=265101	Medium-fast/increasing N=3396	Low-steady N=225383	Medium-slow/increasing N=20639	High-decreasing N=15683	Overall N=265101
Mean (SD)										
Charlson morbidity index	2.48 (2.10)	1.26 (1.54)	2.31 (1.96)	2.83 (2.10)	1.45 (1.69)	3.98 (2.67)	1.78 (1.86)	3.56 (2.41)	3.55 (2.33)	2.05 (2.05)
n (%)										
Myocardial infarction	339 (10.0)	11027 (4.9)	2117 (10.3)	2177 (13.9)	15660 (5.9)	448 (13.2)	13866 (6.2)	2990 (14.5)	2544 (16.2)	19848 (7.5)
Congestive heart failure	289 (8.5)	5843 (2.6)	1550 (7.5)	1931 (12.3)	9613 (3.6)	537 (15.8)	11346 (5.0)	3522 (17.1)	2840 (18.1)	18245 (6.9)
Peripheral vascular disease	180 (5.3)	3917 (1.7)	971 (4.7)	939 (6.0)	6007 (2.3)	231 (6.8)	5596 (2.5)	1435 (7.0)	1194 (7.6)	8456 (3.2)
Cerebrovascular disease	476 (14.0)	15774 (7.0)	2938 (14.2)	2785 (17.8)	21973 (8.3)	619 (18.2)	22832 (10.1)	4237 (20.5)	3433 (21.9)	31121 (11.7)
Dementia	238 (7.0)	7392 (3.3)	1303 (6.3)	1217 (7.8)	10150 (3.8)	399 (11.7)	16084 (7.1)	2705 (13.1)	2243 (14.3)	21431 (8.1)
Chronic obstructive pulmonary disease	938 (27.6)	32567 (14.4)	6047 (29.3)	5697 (36.3)	45249 (17.1)	1094 (32.2)	40644 (18.0)	7706 (37.3)	6362 (40.6)	55806 (21.1)
Rheumatologic disease	210 (6.2)	8882 (3.9)	1536 (7.4)	1556 (9.9)	12184 (4.6)	242 (7.1)	11688 (5.2)	2082 (10.1)	1763 (11.2)	15775 (6.0)
Peptic ulcer disease	272 (8.0)	10874 (4.8)	1670 (8.1)	1531 (9.8)	14347 (5.4)	333 (9.8)	12455 (5.5)	2024 (9.8)	1713 (10.9)	16525 (6.2)
Mild liver disease	37 (1.1)	729 (0.3)	179 (0.9)	162 (1.0)	1107 (0.4)	65 (1.9)	1147 (0.5)	310 (1.5)	233 (1.5)	1755 (0.7)
Diabetes with chronic complications	495 (14.6)	14324 (6.4)	3306 (16.0)	3487 (22.2)	21612 (8.2)	601 (17.7)	19206 (8.5)	4418 (21.4)	4180 (26.7)	28405 (10.7)
Hemiplegia or paraplegia	18 (0.5)	439 (0.2)	93 (0.5)	104 (0.7)	654 (0.2)	22 (0.6)	493 (0.2)	112 (0.5)	110 (0.7)	737 (0.3)
Renal disease	897 (26.4)	38476 (17.1)	5909 (28.6)	5441 (34.7)	50723 (19.1)	1156 (34.0)	53581 (23.8)	8329 (40.4)	6843 (43.6)	69909 (26.4)
Any malignancy	1089 (32.1)	38666 (17.2)	4873 (23.6)	3712 (23.7)	48340 (18.2)	1919 (56.5)	54626 (24.2)	7880 (38.2)	4688 (29.9)	69113 (26.1)
Moderate to severe liver disease	16 (0.5)	248 (0.1)	63 (0.3)	63 (0.4)	390 (0.1)	31 (0.9)	466 (0.2)	130 (0.6)	96 (0.6)	723 (0.3)
Metastatic solid tumour	67 (2.0)	397 (0.2)	121 (0.6)	108 (0.7)	693 (0.3)	343 (10.1)	1442 (0.6)	745 (3.6)	229 (1.5)	2759 (1.0)
AIDS	0 (0)	20 (0)	5 (0)	<5	28 (0)	0 (0)	20 (0)	6 (0)	<5	30 (0)
Hypertension	1766 (52)	102870 (45.6)	11677 (56.6)	9586 (61.1)	125899 (47.5)	1866 (54.9)	117760 (52.2)	12946 (62.7)	10201 (65)	142773 (53.9)
Heart failure	286 (8.4)	6011 (2.7)	1563 (7.6)	1942 (12.4)	9802 (3.7)	547 (16.1)	11922 (5.3)	3625 (17.6)	2897 (18.5)	18991 (7.2)
Osteoporosis	280 (8.2)	13115 (5.8)	1967 (9.5)	2193 (14.0)	17555 (6.6)	385 (11.3)	19805 (8.8)	3329 (16.1)	2792 (17.8)	26311 (9.9)
Gastroesophageal reflux disease	165 (4.9)	8678 (3.9)	1195 (5.8)	1381 (8.8)	11419 (4.3)	212 (6.2)	11650 (5.2)	1648 (8.0)	1705 (10.9)	15215 (5.7)
Chronic kidney disease	836 (24.6)	36490 (16.2)	5558 (26.9)	5027 (32.1)	47911 (18.1)	991 (29.2)	49708 (22.1)	7423 (36.0)	6132 (39.1)	64254 (24.2)
Venous thromboembolism	235 (6.9)	8891 (3.9)	1373 (6.7)	1483 (9.5)	11982 (4.5)	482 (14.2)	12818 (5.7)	2401 (11.6)	1939 (12.4)	17640 (6.7)

	At baseline					At the last observation				
	Medium-fast/increasing N=3396	Low-steady N=225383	Medium-slow/increasing N=20639	High-decreasing N=15683	Overall N=265101	Medium-fast/increasing N=3396	Low-steady N=225383	Medium-slow/increasing N=20639	High-decreasing N=15683	Overall N=265101
Hypothyroidism	339 (10)	17814 (7.9)	2300 (11.1)	2166 (13.8)	22619 (8.5)	390 (11.5)	21335 (9.5)	2819 (13.7)	2433 (15.5)	26977 (10.2)
Stroke	221 (6.5)	6631 (2.9)	1278 (6.2)	1254 (8.0)	9384 (3.5)	337 (9.9)	10828 (4.8)	2183 (10.6)	1692 (10.8)	15040 (5.7)
Anxiety	480 (14.1)	26971 (12.0)	3646 (17.7)	3543 (22.6)	34640 (13.1)	604 (17.8)	32831 (14.6)	4674 (22.6)	4140 (26.4)	42249 (15.9)
Asthma	436 (12.8)	20086 (8.9)	3638 (17.6)	3798 (24.2)	27958 (10.5)	466 (13.7)	22306 (9.9)	4070 (19.7)	4038 (25.7)	30880 (11.6)
Pneumonia	225 (6.6)	5282 (2.3)	1202 (5.8)	1401 (8.9)	8110 (3.1)	594 (17.5)	11664 (5.2)	3498 (16.9)	2714 (17.3)	18470 (7)
Diabetes	791 (23.3)	29705 (13.2)	5406 (26.2)	5246 (33.5)	41148 (15.5)	918 (27.0)	37519 (16.6)	6555 (31.8)	5856 (37.3)	50848 (19.2)
Inflammatory bowel disease	38 (1.1)	1953 (0.9)	304 (1.5)	256 (1.6)	2551 (1.0)	41 (1.2)	2274 (1.0)	366 (1.8)	295 (1.9)	2976 (1.1)
Depressive disorder	467 (13.8)	22493 (10.0)	3362 (16.3)	3427 (21.9)	29749 (11.2)	548 (16.1)	25902 (11.5)	4129 (20)	3894 (24.8)	34473 (13.0)

Table 6.6 Drug use by drug classes and common drugs at baseline and the last observation for each cluster and the overall population in Aurum

	At baseline (n (%))					At the last observation (n (%))				
	Medium-fast/increasing N=3396	Low-steady N=225383	Medium-slow/increasing N=20639	High-decreasing N=15683	Overall N=265101	Medium-fast/increasing N=3396	Low-steady N=225383	Medium-slow/increasing N=20639	High-decreasing N=15683	Overall N=265101
Drug classes										
Alimentary tract and metabolism	2691 (79.2)	133883 (59.4)	17729 (85.9)	15639 (99.7)	169942 (64.1)	3387 (99.7)	141200 (62.6)	20446 (99.1)	14839 (94.6)	179872 (67.9)
Blood and blood forming organs	1846 (54.4)	72501 (32.2)	11875 (57.5)	11817 (75.3)	98039 (37.0)	2501 (73.6)	85037 (37.7)	15355 (74.4)	11102 (70.8)	113995 (43.0)
Cardiovascular system	2653 (78.1)	148365 (65.8)	17496 (84.8)	14762 (94.1)	183276 (69.1)	2906 (85.6)	158402 (70.3)	18545 (89.9)	13807 (88.0)	193660 (73.1)
Dermatological	1391 (41.0)	67817 (30.1)	10332 (50.1)	12346 (78.7)	91886 (34.7)	2064 (60.8)	61692 (27.4)	13056 (63.3)	9003 (57.4)	85815 (32.4)
Genito-urinary system and sex hormones	654 (19.3)	36038 (16.0)	4927 (23.9)	5655 (36.1)	47274 (17.8)	817 (24.1)	39008 (17.3)	6189 (30.0)	4401 (28.1)	50415 (19.0)
Systemic hormonal preparations, excluding sex hormones and insulins	478 (14.1)	24534 (10.9)	3400 (16.5)	3552 (22.6)	31964 (12.1)	590 (17.4)	27414 (12.2)	4234 (20.5)	3302 (21.1)	35540 (13.4)
Anti-infective for systemic use	1836 (54.1)	86716 (38.5)	12512 (60.6)	13615 (86.8)	114679 (43.3)	2774 (81.7)	78796 (35.0)	16023 (77.6)	10193 (65.0)	107786 (40.7)
Antineoplastic and immunomodulation agents	310 (9.1)	8762 (3.9)	1451 (7.0)	1456 (9.3)	11979 (4.5)	400 (11.8)	11193 (5.0)	2180 (10.6)	1195 (7.6)	14968 (5.6)
Musculoskeletal system	1162 (34.2)	64731 (28.7)	8840 (42.8)	9733 (62.1)	84466 (31.9)	1557 (45.8)	58211 (25.8)	10218 (49.5)	7184 (45.8)	77170 (29.1)
Nervous system	2202 (64.8)	99736 (44.3)	15178 (73.5)	14615 (93.2)	131731 (49.7)	3298 (97.1)	100508 (44.6)	18680 (90.5)	13433 (85.7)	135919 (51.3)
Anti-parasitic products, insecticides and repellents	310 (9.1)	8762 (3.9)	1451 (7.0)	1456 (9.3)	11979 (4.5)	400 (11.8)	11193 (5.0)	2180 (10.6)	1195 (7.6)	14968 (5.6)
Respiratory system	1644 (48.4)	70774 (31.4)	11792 (57.1)	12438 (79.3)	96648 (36.5)	2741 (80.7)	68829 (30.5)	15175 (73.5)	10261 (65.4)	97006 (36.6)
Sensory organs	451 (13.3)	22257 (9.9)	3347 (16.2)	4294 (27.4)	30349 (11.4)	1244 (36.6)	24932 (11.1)	5248 (25.4)	3351 (21.4)	34775 (13.1)
Various	7 (0.2)	123 (0.1)	38 (0.2)	104 (0.7)	272 (0.1)	18 (0.5)	125 (0.1)	77 (0.4)	86 (0.5)	306 (0.1)
Drugs										
Psychostimulants	6 (0.2)	111 (0)	22 (0.1)	28 (0.2)	167 (0.1)	8 (0.2)	134 (0.1)	26 (0.1)	29 (0.2)	197 (0.1)
Opioids	2451 (72.2)	128988 (57.2)	16289 (78.9)	14051 (89.6)	161779 (61.0)	3205 (94.4)	144919 (64.3)	18785 (91.0)	14652 (93.4)	181561 (68.5)
Drugs for obstructive airway disease	1834 (54.0)	90238 (40.0)	12329 (59.7)	11438 (72.9)	115839 (43.7)	2151 (63.3)	107380 (47.6)	14584 (70.7)	12244 (78.1)	136359 (51.4)
Drugs used in diabetes	701 (20.6)	24224 (10.7)	4779 (23.2)	4889 (31.2)	34593 (13.0)	828 (24.4)	29754 (13.2)	5711 (27.7)	5309 (33.9)	41602 (15.7)
Psycholeptics	1746 (51.4)	87757 (38.9)	11693 (56.7)	11274 (71.9)	112470 (42.4)	2927 (86.2)	103292 (45.8)	15481 (75.0)	12458 (79.4)	134158 (50.6)
Agents renin-angiotensin systemic	1868 (55.0)	96993 (43.0)	12616 (61.1)	11139 (71.0)	122616 (46.3)	2045 (60.2)	110243 (48.9)	14183 (68.7)	11686 (74.5)	138157 (52.1)
Anti-thrombotic	1167 (34.4)	43490 (19.3)	7327 (35.5)	7215 (46.0)	59199 (22.3)	1801 (53.0)	66111 (29.3)	11620 (56.3)	8844 (56.4)	88376 (33.3)

	At baseline (n (%))					At the last observation (n (%))				
	Medium-fast/increasing N=3396	Low-steady N=225383	Medium-slow/increasing N=20639	High-decreasing N=15683	Overall N=265101	Medium-fast/increasing N=3396	Low-steady N=225383	Medium-slow/increasing N=20639	High-decreasing N=15683	Overall N=265101
Lipid modifying agents	2144 (63.1)	112230 (49.8)	13983 (67.8)	12128 (77.3)	140485 (53.0)	2312 (68.1)	133286 (59.1)	15753 (76.3)	12672 (80.8)	164023 (61.9)
Calcium channel blockers	1583 (46.6)	81255 (36.1)	10646 (51.6)	9297 (59.3)	102781 (38.8)	1758 (51.8)	99561 (44.2)	12468 (60.4)	10043 (64.0)	123830 (46.7)
Diuretics	1805 (53.2)	86546 (38.4)	12381 (60.0)	11340 (72.3)	112072 (42.3)	2295 (67.6)	99153 (44.0)	15057 (73.0)	12165 (77.6)	128670 (48.5)
Antibacterial systemic	3143 (92.6)	190797 (84.7)	19526 (94.6)	15454 (98.5)	228920 (86.4)	3328 (98.0)	203148 (90.1)	20465 (99.2)	15610 (99.5)	242551 (91.5)
Immunosuppressant	117 (3.4)	3953 (1.8)	808 (3.9)	926 (5.9)	5804 (2.2)	147 (4.3)	5181 (2.3)	1173 (5.7)	1083 (6.9)	7584 (2.9)
Anti-inflammatory and/or anti-rheumatic	2538 (74.7)	154630 (68.6)	16604 (80.4)	13595 (86.7)	187367 (70.7)	2947 (86.8)	166706 (74)	18110 (87.7)	14084 (89.8)	201847 (76.1)
Hormonal contraceptives systemic	119 (3.5)	11846 (5.3)	1004 (4.9)	853 (5.4)	13822 (5.2)	134 (3.9)	11999 (5.3)	1034 (5.0)	870 (5.5)	14037 (5.3)
Anti-epileptics	550 (16.2)	20166 (8.9)	4252 (20.6)	5308 (33.8)	30276 (11.4)	1107 (32.6)	31894 (14.2)	7474 (36.2)	6958 (44.4)	47433 (17.9)
Drugs for acid related disorder	2390 (70.4)	125830 (55.8)	15750 (76.3)	14226 (90.7)	158196 (59.7)	3048 (89.8)	149789 (66.5)	18564 (89.9)	14718 (93.8)	186119 (70.2)
Beta blocking agents	1502 (44.2)	75609 (33.5)	9718 (47.1)	8573 (54.7)	95402 (36.0)	1793 (52.8)	88270 (39.2)	12096 (58.6)	9479 (60.4)	111638 (42.1)
Antidepressants	1623 (47.8)	76877 (34.1)	11154 (54)	10986 (70.1)	100640 (38)	2158 (63.5)	92516 (41)	13916 (67.4)	11952 (76.2)	120542 (45.5)
Antineoplastic agents	281 (8.3)	11778 (5.2)	2069 (10.0)	2254 (14.4)	16382 (6.2)	326 (9.6)	13395 (5.9)	2492 (12.1)	2402 (15.3)	18615 (7.0)

6.5 Discussion

Main findings

In this study, I used joint latent class models to identify four clusters of older people with distinct polypharmacy trajectories and associated mortality risk. By grouping the older population based on their polypharmacy trends, I was able to identify clinically meaningful clusters. The emerging clusters were low-steady, intermediate-slow/increasing, intermediate-fast/increasing and high-decreasing. Their morbidity burden and drug intake throughout the study period were characterised and compared. Compared to the low-steady cluster, patients with high baseline polypharmacy or increasing trend over time had higher morbidity burden and were at a greater risk of mortality. To my knowledge, this is the first study to group the older population based on their polypharmacy trends over time, while simultaneously accounting for their risk of death.

Choice of the best model

In the KmL analysis, including only survivors introduced a survival bias by limiting the analysis to relatively healthy people as they were the ones expected to survive in this old population. I attempted to minimise this by including more patients and reducing the required data points (years of follow-up). This analysis was performed to capture the potential polypharmacy trends and not to create the final trajectory clusters. Only linear/semi-linear trends were observed in the identified clusters.

Different link functions and transformations for the longitudinal part and distributions of the survival part were tested to improve the performance of the joint latent class model. A joint latent class model with linear link function for the longitudinal part and a Weibull proportional hazard model for the risk analysis part was the model of choice. While the spline-models

performed better for the longitudinal part, the benefits of a simple, interpretable model outweighed the small improvement in the model performance.

Based on BIC, clinical plausibility and convergence, the four-cluster model was the best model. While a three-cluster model showed good performance, plausible polypharmacy trends and associated mortality risks, the new cluster emerging from the four-cluster model (intermediate-fast/increasing) had distinctive clinical characteristics and high mortality risk. Identification of this vulnerable population in a separate group allows for a better understanding of their -and other clusters'- morbidity burden and mortality risk. The hazard ratios were treated as a weighted average for the time-varying hazard of each group[137]. For example, the intermediate-fast-increasing cluster had a seemingly high survival probability in the first year, which then deteriorated rapidly in the following 4 years. The clinical characteristics of this cluster support this observation, i.e., high onset of cancer diagnosis. Similar results are observed with the overlapping hazard ratios in the intermediate-slow/increasing and high-decreasing clusters: they had different characteristics at the start of follow-up but similar morbidity burden by the end of the study period.

Clusters characteristics

The low-steady cluster was the healthiest group. Compared to this cluster, the other groups -by definition- had higher use of different drugs. Use of some of these drugs, such as opioids, antidepressants, and psycholeptics is associated with higher risk of mortality in older people with polypharmacy[69]. The patients belonging to these clusters had higher burden of cognitive impairment (e.g., dementia) and mental problems, which are associated with chronic diseases and polypharmacy[73 138].

The intermediate-fast/increasing cluster had the highest diagnosis of cancer (>50%) and highest risk of mortality, with more than 80% dying during the study period. Patients with

cancer have a higher risk of polypharmacy and potentially inappropriate medications[139 140], especially those with high comorbidity burden[141]. A study monitoring the accumulation of polypharmacy in the year prior to death have found that older people who died from cancer had the highest increase in polypharmacy with an average increase of 3.37 drugs in the last year[77].

The intermediate-slow/increasing cluster had the second-highest onset of cancer diagnosis and the most rapid accumulation of non-cancer chronic diseases. In general, new diagnosis of a chronic disease is accompanied by an increase in polypharmacy[142]. A study looking at the change in number of medications between the year prior to diabetes diagnosis and the year after in adults found that the mean number of drugs increased from 5.0 to 6.6. The increase was mainly due to the use of diabetes treatment, in addition to anti-hypertension and lipid-lowering agents[142].

In the first period after a patient is diagnosed with a chronic disease such as diabetes and hypertension, multiple lines of therapy might be tested before finding the optimal treatment[143 144]. Moreover, physicians tend to decrease the number of drugs for older patients with multimorbidity or at a later stage in life[145]. This could explain the high baseline value and decreasing polypharmacy trend in the high-decreasing cluster, as this cluster has the highest prevalence of chronic diseases at baseline and at the end of study follow-up.

The high-decreasing and intermediate-slow/increasing clusters had overlapping hazard ratios ≥4.0 compared to the healthy cluster. Previous studies produced supporting evidence on the association between older patients with chronic diseases and polypharmacy, and the increased risk of death[146].

External validation of the models in CPRD Aurum

Good validation metrics were found in CPRD Aurum. More than 50% of the Aurum population had a posterior probability ≥0.7 and low probabilities for misclassification. The identified clusters had comparable hazard ratios. The patients in Aurum had slightly higher comorbidity burden. This is likely because GOLD and Aurum cover different areas of the UK and use different coding systems to collect data from patients (3.2). The difference between baseline and last observation in drug intake and disease diagnoses, however, was similar in the Aurum population to that in GOLD. This confirms the consistency of subgroup behaviours in the two databases.

Strengths, limitations and implications

A particular strength of the study is the use of a joint latent class model and its external validity in an independent database, Aurum. Although polypharmacy is a count measure, the linear joint latent class model performed well with good discrimination and generalisable clusters. It identified different groups in terms of polypharmacy variation over time and risk of death across two databases. Traditionally, specific longitudinal methods exist for modelling repeated count data, such as Poisson mixed effects models. However, these models do not account for the censored data and would fail to correctly link between the longitudinal polypharmacy and the associated risk of death. With the polypharmacy counts having medians and means higher than zero, the risk of having predictions below zero when modelling them continuously was avoided. While there is still a possible limitation of inaccurate estimates, the clusters identified were clinically meaningful and had plausible trajectories (intercepts and slopes). They can be used as a starting point for further, more complex methods.

An implication of this study is that monitoring polypharmacy over time can accurately identify the vulnerable and deteriorating subgroups within the older population. Polypharmacy was estimated by calculating the absolute number of ingredients taken annually. This approach

might overestimate the actual number of ingredients a patient is taking at any point in time. Ageing and chronic conditions can affect the pharmacokinetics and pharmacodynamics of treatment[147 148]; therefore, it is important to account for the cumulative effect of polypharmacy on an old, vulnerable population[148]. While a higher mortality risk cannot be solely attributed to the accumulation of drugs intake in a patient, polypharmacy can be used as a proxy to identify vulnerable patients with higher comorbidities burden. This was observed in the high-decreasing cluster, which, despite its decreasing polypharmacy trend, had a high HR compared to the low-steady cluster. Trajectory of polypharmacy is also a sign of incidence multimorbidity as observed in the characteristics of both the intermediate-increasing clusters in this study, and supported by Vos et al.[18]. Multiple guidelines and interventions are available for optimising polypharmacy in older people with cancer and/or chronic diseases[139 149]. Identifying those patients well in advance and understanding their prognosis can give physicians a head start on tailoring the right treatment package. Use of the limited resources would then be optimised by the careful delivery of treatment towards the most vulnerable patient.

Conclusion

This study has shown that grouping the older population based on their polypharmacy progression over time provides a useful tool to identify those most vulnerable and at higher risk of mortality. Polypharmacy acts as an indicator for multimorbidity and a predictor for potential drug-related adverse events. Further analysis to assess the generalisability of the clusters identified is needed.

7. Generalisability of longitudinal trajectories of polypharmacy in older people: a validation study in the Dutch population

7.1 Chapter summary

In this chapter, I externally validated the clusters of polypharmacy trajectories identified in Chapter 6. I tested my previously developed model on older people in a Dutch primary care database, IPCI. Like CPRD GOLD and Aurum in the previous chapter, IPCI is mapped to OMOP-CDM. This has facilitated the replication of my previous analyses using the same study code in a federated manner, without transferring patient-level data.

As the IPCI database is smaller than both CPRD databases, the entire sample of people aged ≥65 was included in the analysis. The IPCI study sample had comparable demographics, and slightly lower baseline polypharmacy levels and slightly different multimorbidity patterns. After allocating the patients to their respective clusters based on their polypharmacy trajectories and risk of death, the results showed good external validity. The clusters had similar mortality hazard ratios. There were differences in disease prevalence, both in the overall population and the individual clusters. Additionally, IPCI had lower prevalence of drug use but greater increases over time in some drugs.

Despite all the differences observed, the resulting clusters had distinct characteristics, following patterns similar to those identified in UK data. The low-steady cluster included most of the population and had the lowest multimorbidity burden and drug consumption. The intermediate-fast/increasing cluster had the highest prevalence and onset of cancer, and the highest risk of death. The high-decreasing cluster had the highest comorbidity burden at baseline and at end of follow-up, while the intermediate-slow/increasing cluster had the fastest rate of chronic disease accumulation, and the second highest risk of cancer throughout the study

period. Both intermediate-slow/increasing and high-decreasing clusters had similar risks of death.

The analyses in this chapter demonstrate the robustness of the polypharmacy clusters identified in UK data, which can be generalised to other populations, here from the Netherlands. The associations identified between the different trajectories and risk of death remained consistent, despite the population and data heterogeneity across the different databases. Further analysis to test the replicability of the analysis and/or generalisability of the clusters in other nations and data types is needed to confirm these findings.

7.2 Introduction

Often, when a statistical model is built in a certain population/setting, it is only valid and accurate for the setting it was built for. This leads to varying evidence answering the same research question. There are a few potential reasons for this. Some are related to different theory-to-practice translations of the research question such as the inclusion criteria, statistical analysis and assumptions, and even interpretations of the produced results. Others include the nature and setting of the data and the populations analysed. Currently, the reproducibility and replicability of evidence are major concerns in the research society[150 151].

Some factors cannot be avoided (e.g., characteristics of different nationalities/races). Others, on the other hand, can be controlled and minimised. The use of a common data model allows for a feasible means to run large-scale network analyses in a database network. To carry out such studies, the researchers are required to share their protocols, methods and codes with their collaborators. This approach is encouraged as it promotes transparency and increases the chances of study reproducibility and replicability[152]. After eliminating the problems of different research practices, running analysis and validation of models can be more efficient, and produce more reliable and generalisable results. Thus, conditional on good quality transformation, any heterogeneity in the results can directly be attributed to the country-specific patient characteristics, treatment guidelines, and data entry practices.

Polypharmacy research has not escaped the replicability problem. For example, conflicting evidence exists on whether polypharmacy has causal links with mortality[73]. It is, therefore, important to be able to assess the generalisability and validity of the model built to study polypharmacy in older people.

In Chapter 6, I used two UK primary care databases mapped to the OMOP-CDM to identify and then validate clusters of older people with distinct longitudinal polypharmacy trajectories,

and their associated risk of mortality. The model performed well in UK data and the resulting clusters of both UK databases had comparable characteristics. In this chapter, I aimed to externally validate these clusters in a Dutch primary care database: IPCI.

7.3 Methods

Design

In this analysis, I validated the model built in Chapter 6. With local support from Ross D Williams (RDW), I prepared a study protocol to be approved by the IPCI committee at the Erasmus University Medical Centre in Rotterdam. After the study approval, I created a GitHub repository that includes the entire study code required for this analysis. In the repository, I included instructions on how to run the analysis in the mapped data. RDW then ran the analysis, reviewed the aggregated results to ensure they met all data governance regulations, and shared them with me. I describe below the methods applied, which were almost identical to those described in Chapter 6.

Data source

The data used in this study were extracted from IPCI and mapped to the OMOP-CDM (3.2).

Population

Patients aged ≥65, alive on 1st January 2015 and with at least one year of prior history in the data were included. They were followed from 1st January 2015 to 31st December 2019. Patients exited the study at the earliest of the following: practice last collection date, patient transfer-out of practice date, death date or study end date.

As IPCI data constitute a far smaller population than those in CPRD databases, we first assessed that the total population meeting the inclusion criteria. If the entire sample of older people was ≤300,000, then everyone meeting the inclusion criteria was included for analysis. Otherwise, a random sample of 300,000 older people was included.

Exposure

I used the same method to calculate polypharmacy as that described in 3.3. Polypharmacy was calculated at baseline and at the end of each follow-up year.

Outcomes

The adverse outcome of interest was all-cause mortality (3.3).

Statistical analysis

Descriptive analysis

Baseline characteristics, including age, sex, comorbidity burden and drug use were described and compared to the UK population.

Model validation

I assessed the performance of the emergent best-fit cluster model -derived from CPRD GOLD in Chapter 6 in the IPCI validation set. I followed the same steps described in 3.3. I calculated the risk of mortality for the clusters identified in the validation sample and compared them to the original model.

Characterisation of the resulting clusters

Characteristics at baseline and by the end of follow-up were summarised and compared between the resulting clusters (3.3).

R (Version 4.3.2) was used for preparation of the codes used in the analysis. R packages 'lcmm' version 2.0.2 [link] and 'PatientProfiles' version 0.4.0 [link] were used for the joint latent class model validation and characterisation of resultant clusters, respectively.

7.4 Results

Model validation in IPCI

Study population

There were 139,307 patients identified in IPCI. They had similar demographics and baseline characteristics to those in CPRD. Mean age was 74.2 (SD 7.3) with 45.4% of the patients being male. Median number of prescribed medicines' ingredients at baseline was slightly lower than in CPRD: 6 [IQR 3 to 10] in IPCI vs. 8 [IQR 4 to 13] in CPRD (Table 7.1).

Table 7.1 Baseline characteristics by cluster and for the overall population in IPCI

	Intermediate-fast/increasing	Low-steady	Intermediate-slow/increasing	High-decreasing	Overall
n	1429 (1.0%)	125835 (90.3%)	8268 (5.9%)	3775 (2.7%)	139307
Sex = Male (%)	814 (57.0%)	56789 (45.1%)	4138 (50.0%)	1565 (41.5%)	63306 (45.4%)
Age (mean (SD))	75.29 (6.96)	74.10 (7.31)	75.09 (7.14)	75.29 (7.13)	74.21 (7.30)
Number of ingredients* (median[IQR])	9 [5, 14]	6 [3, 10]	12 [7, 16]	24 [21, 28]	6 [3, 10]
Charlson morbidity index (mean (SD))	2.17 (1.71)	1.16 (1.39)	1.99 (1.67)	2.64 (1.90)	1.26 (1.46)

* In the year prior to start

Resultant clusters

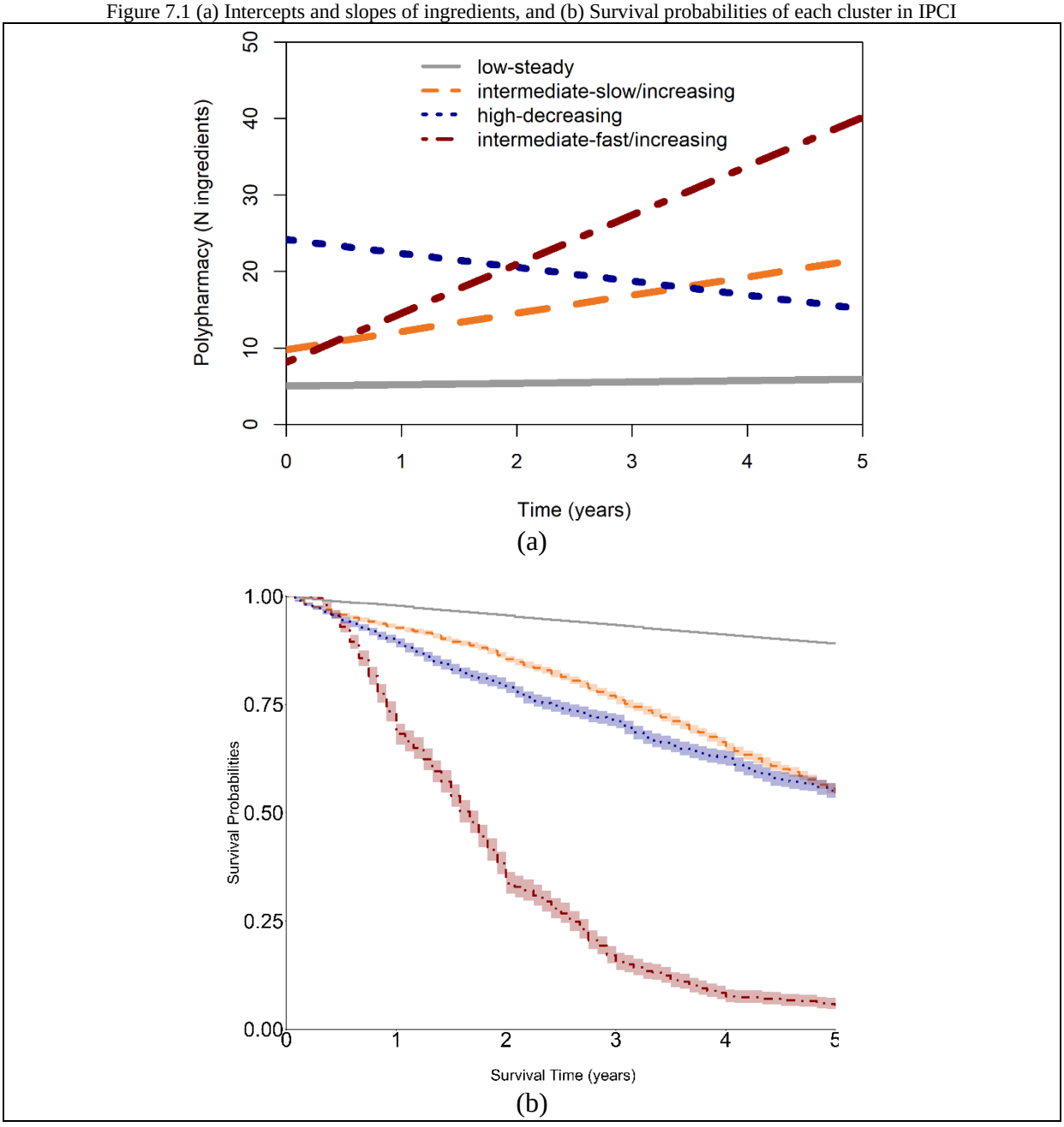
The median follow-up was 5 years [IQR 3.4 to 5] with 18,718 (13.4%) dying during follow-up.

The distribution of patients belonging to each cluster was slightly different in the IPCI population. In general, more patients belonged to the low-steady cluster compared to UK databases; 90.3% in IPCI vs. 85.7 and 85.3% in GOLD and Aurum, respectively. Additionally, a considerably lower proportion of patients belonged to the high-decreasing cluster: 2.7% in IPCI vs. 5-6% in CPRD databases (Table 7.1).

The IPCI participants showed good fit of the data; more than 50% had posterior probability ≥0.7 (Appendix Table S 7.1). The resultant clusters had similar trajectories of ingredients over time and associated risk of mortality to those in UK data. Taking the low-steady cluster as reference (9.8% deaths), the intermediate-fast/increasing cluster had the highest HR of 21.73 (95%CI 21.09, 22.39), followed by the high-decreasing and intermediate-slow/increasing

clusters: 4.93 (95%CI 4.75, 5.13) and 4.50 (95%CI 4.36, 4.65), respectively (See Figure 7.1). Median follow-up, yearly deaths and polypharmacy are reported in Appendix Table S 7.2.

Figure 7.1 (a) Intercepts and slopes of ingredients, and (b) Survival probabilities of each cluster in IPCI



Characterisation of the identified cluster

Table 7.2 summarises morbidity burden, and Table 7.3 summarises prevalence of drug classes and use of common drugs, for all clusters at baseline and by the end of follow-up. Table S 7.3 and Table S 7.4 summarise the difference in percentages between the two time points.

Compared to the UK population, the IPCI population had some differences in disease prevalence. The participants had a lower recorded prevalence of depression, renal and chronic kidney diseases, and asthma, but higher prevalence of pneumonia history. For example, by the end of follow-up, 6.6% and 15.0% in IPCI, vs. 10.0% and 4.6% in CPRD GOLD had asthma and pneumonia, respectively.

Similarly, IPCI population had different prevalence and accumulation rates of drug use. For example, 22.6% of IPCI participants were prescribed opioids in the year prior to start, rising to 39.0% in the last year of observation (Δ=16.4%) vs. 60.8% rising to 67.5% (Δ=6.7%) in CPRD GOLD. This result was similar for psycholeptic, anti-thrombotic, diuretics, anti-bacterial, and anti-inflammatory drugs.

Comorbidities

The low-steady cluster had the lowest Charlson comorbidity index at baseline and the slowest increase over time (mean 1.2 (SD 1.4), Δ=0.5). Subsequently, the baseline and end of follow-up prevalence of individual comorbidities was the smallest. For example, all other clusters had higher prevalence of renal disease by the end of follow-up, ranging from 21.3% to 32.0% vs. 14.6% in the low-steady cluster.

The high-decreasing cluster had the highest baseline Charlson comorbidity index but a slow increase over the study period (mean 2.6 (SD 1.9), Δ=0.8). In turn, the intermediate-fast/increasing and intermediate-slow/increasing clusters had lower comorbidity index but higher increases over the study period: (mean 2.2 (SD 1.7), Δ=1.1) and (mean 2.0 (SD 1.7), Δ=1.2), respectively.

The intermediate-fast/increasing cluster had the highest prevalence and risk of cancer diagnosis over the study period, starting with 44.2% at baseline and rising to 74.8% at end of follow-up (Δ=30.6%). The high-decreasing cluster had the second highest baseline prevalence (36.7%),

while the intermediate-slow/increasing cluster had the second highest risk of cancer (Δ=18.9%). The intermediate-fast/increasing cluster also had the highest risk of venous thromboembolism (Δ=6.4%), followed by the intermediate-slow/increasing cluster (Δ=4.9%).

The intermediate-slow/increasing cluster had the fastest accumulation of most -especially chronic- diseases, such as diabetes, heart failure, COPD and renal diseases. For example, 23.8% had diabetes at baseline rising to 30.8% (Δ=7.0%) in the intermediate-slow/increasing cluster, compared to 31.9% rising to 37.2% (Δ=5.2%) in the high-decreasing cluster.

The high-decreasing cluster had the highest baseline and end-of-study prevalence of non-cancer chronic diseases such as COPD, osteoporosis and diabetes. For example, 39.3% had COPD at baseline, rising to 42.4% by the end of follow-up. This was followed by the intermediate-slow/increasing cluster; 28.4% at baseline and 33.7% by the last observation. The high-decreasing cluster also had the highest prevalence of mental illnesses like anxiety and depression.

Medication use

Overall, the use of drugs and drug classes followed the trajectories of their respective clusters. For example, 73.4% of patients in the high-decreasing cluster were using drugs from the blood and blood-forming organs class in the year prior to start, dropping to 66.4% (Δ=-7.0) in the last year of follow-up.

Use of common drugs increased throughout the study period, with low-steady and high-decreasing clusters having the smallest increment over the study period. For instance, there were 58.4% and 39.1% more patients in the intermediate-fast/increasing and intermediate-slow/increasing clusters, respectively, using opioids by the last year of observation vs. 14.4% in the low-steady and 15.6% in the high-decreasing clusters.

As with comorbidities, the high-decreasing cluster had the highest baseline and end of follow-up use of preventive and chronic disease treatments, such as diuretics, beta-blockers, lipid modifying agents, and drugs used for diabetes and COPD. By the last year of observation, 68.0% of the high-decreasing cluster was using beta-blockers, followed by 64.7% in the intermediate-slow/increasing cluster.

The intermediate-increasing clusters had the highest rise of drug use between baseline and end of follow-up. Use of psycholeptics, for example, increased by 53.0% in the intermediate-fast/increasing and 29.9% in the intermediate-slow/increasing clusters, compared to 12.1% and 10.5% in the high-decreasing and low-steady clusters, respectively.

Drugs belonging to alimentary tract and metabolism (e.g., drugs for acid-related disorders), cardiovascular system (e.g., lipid-modifying agents), and nervous system (e.g., opioids) classes were the most commonly prescribed drugs at baseline and by end of the study.

Table 7.2 Individual comorbidities at baseline and the last observation in IPCI

	At baseline					At the last observation				
	Intermediate-fast/increasing N=1429	Low-steady N=125835	Intermediate-slow/increasing N=8268	High-decreasing N=3775	Overall N=139307	Intermediate-fast/increasing N=1429	Low-steady N=125835	Intermediate-slow/increasing N=8268	High-decreasing N=3775	Overall N=139307
Mean (SD)										
Charlson morbidity index	2.17 (1.71)	1.16 (1.39)	1.99 (1.67)	2.64 (1.90)	1.26 (1.46)	3.29 (1.71)	1.71 (1.67)	3.17 (1.89)	3.46 (2.12)	1.86 (1.76)
n (%)										
Myocardial infarction	162 (11.3)	8536 (6.8)	1039 (12.6)	598 (15.8)	10335 (7.4)	205 (14.3)	10565 (8.4)	1451 (17.5)	680 (18.0)	12901 (9.3)
Congestive heart failure	188 (13.2)	6630 (5.3)	1125 (13.6)	905 (24.0)	8848 (6.4)	350 (24.5)	11001 (8.7)	2219 (26.8)	1231 (32.6)	14801 (10.6)
Peripheral vascular disease	132 (9.2)	4831 (3.8)	727 (8.8)	492 (13.0)	6182 (4.4)	171 (12.0)	7555 (6.0)	1178 (14.2)	686 (18.2)	9590 (6.9)
Cerebrovascular disease	225 (15.7)	12320 (9.8)	1300 (15.7)	776 (20.6)	14621 (10.5)	315 (22.0)	18252 (14.5)	2045 (24.7)	1003 (26.6)	21615 (15.5)
Dementia	35 (2.4)	3871 (3.1)	236 (2.9)	159 (4.2)	4301 (3.1)	81 (5.7)	9022 (7.2)	624 (7.5)	380 (10.1)	10107 (7.3)
Chronic obstructive pulmonary disease	377 (26.4)	15033 (11.9)	2349 (28.4)	1485 (39.3)	19244 (13.8)	432 (30.2)	17448 (13.9)	2786 (33.7)	1602 (42.4)	22268 (16.0)
Rheumatologic disease	104 (7.3)	6353 (5.0)	711 (8.6)	407 (10.8)	7575 (5.4)	120 (8.4)	8106 (6.4)	940 (11.4)	456 (12.1)	9622 (6.9)
Peptic ulcer disease	84 (5.9)	4104 (3.3)	481 (5.8)	267 (7.1)	4936 (3.5)	101 (7.1)	4704 (3.7)	552 (6.7)	299 (7.9)	5656 (4.1)
Mild liver disease	14 (1.0)	314 (0.2)	43 (0.5)	37 (1.0)	408 (0.3)	22 (1.5)	422 (0.3)	73 (0.9)	46 (1.2)	563 (0.4)
Diabetes with chronic complications	49 (3.4)	2378 (1.9)	367 (4.4)	288 (7.6)	3082 (2.2)	67 (4.7)	3478 (2.8)	529 (6.4)	386 (10.2)	4460 (3.2)
Hemiplegia or paraplegia	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Renal disease	200 (14.0)	8327 (6.6)	1097 (13.3)	726 (19.2)	10350 (7.4)	305 (21.3)	18418 (14.6)	2318 (28.0)	1208 (32.0)	22249 (16.0)
Any malignancy	632 (44.2)	30752 (24.4)	2755 (33.3)	1386 (36.7)	35525 (25.5)	1069 (74.8)	42033 (33.4)	4315 (52.2)	1706 (45.2)	49123 (35.3)
Moderate to severe liver disease	4 (0.3)	37 (0)	6 (0.1)	9 (0.2)	56 (0)	6 (0.4)	51 (0)	11 (0.1)	12 (0.3)	80 (0.1)
Metastatic solid tumour	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
AIDS	0 (0)	60 (0)	<5	<5	67 (0)	0 (0)	66 (0.1)	<5	<5	74 (0.1)
Hypertension	492 (34.4)	44158 (35.1)	3008 (36.4)	1494 (39.6)	49152 (35.3)	562 (39.3)	55117 (43.8)	3860 (46.7)	1745 (46.2)	61284 (44)
Heart failure	134 (9.4)	4413 (3.5)	737 (8.9)	688 (18.2)	5972 (4.3)	313 (21.9)	9297 (7.4)	2003 (24.2)	1070 (28.3)	12683 (9.1)
Osteoporosis	83 (5.8)	6222 (4.9)	447 (5.4)	377 (10.0)	7129 (5.1)	122 (8.5)	9273 (7.4)	818 (9.9)	524 (13.9)	10737 (7.7)
Gastroesophageal reflux disease	15 (1.0)	1929 (1.5)	151 (1.8)	82 (2.2)	2177 (1.6)	27 (1.9)	2835 (2.3)	232 (2.8)	119 (3.2)	3213 (2.3)
Chronic kidney disease	5 (0.3)	122 (0.1)	19 (0.2)	11 (0.3)	157 (0.1)	7 (0.5)	204 (0.2)	40 (0.5)	14 (0.4)	265 (0.2)
Venous thromboembolism	49 (3.4)	2854 (2.3)	265 (3.2)	187 (5.0)	3355 (2.4)	140 (9.8)	5503 (4.4)	672 (8.1)	322 (8.5)	6637 (4.8)

	At baseline					At the last observation				
	Intermediate-fast/increasing N=1429	Low-steady N=125835	Intermediate-slow/increasing N=8268	High-decreasing N=3775	Overall N=139307	Intermediate-fast/increasing N=1429	Low-steady N=125835	Intermediate-slow/increasing N=8268	High-decreasing N=3775	Overall N=139307
Hypothyroidism	46 (3.2)	4184 (3.3)	294 (3.6)	243 (6.4)	4767 (3.4)	73 (5.1)	6503 (5.2)	558 (6.7)	343 (9.1)	7477 (5.4)
Stroke	73 (5.1)	4437 (3.5)	462 (5.6)	298 (7.9)	5270 (3.8)	151 (10.6)	9175 (7.3)	1065 (12.9)	488 (12.9)	10879 (7.8)
Anxiety	178 (12.5)	15472 (12.3)	1120 (13.5)	790 (20.9)	17560 (12.6)	284 (19.9)	24162 (19.2)	2056 (24.9)	1134 (30.0)	27636 (19.8)
Asthma	69 (4.8)	5514 (4.4)	591 (7.1)	481 (12.7)	6655 (4.8)	87 (6.1)	7543 (6.0)	907 (11.0)	622 (16.5)	9159 (6.6)
Pneumonia	171 (12.0)	7235 (5.7)	877 (10.6)	758 (20.1)	9041 (6.5)	419 (29.3)	16602 (13.2)	2521 (30.5)	1321 (35)	20863 (15.0)
Diabetes	303 (21.2)	17789 (14.1)	1971 (23.8)	1205 (31.9)	21268 (15.3)	376 (26.3)	22062 (17.5)	2548 (30.8)	1401 (37.1)	26387 (18.9)
Inflammatory bowel disease	9 (0.6)	681 (0.5)	75 (0.9)	59 (1.6)	824 (0.6)	13 (0.9)	991 (0.8)	118 (1.4)	77 (2.0)	1199 (0.9)
Depressive disorder	62 (4.3)	4075 (3.2)	433 (5.2)	363 (9.6)	4933 (3.5)	112 (7.8)	6084 (4.8)	775 (9.4)	505 (13.4)	7476 (5.4)

Table 7.3 Drug use by drug classes and common drugs at baseline and the last observation for each cluster and the overall population in IPCI

	At baseline (n (%))					At the last observation (n (%))				
	Intermediate-fast/increasing N=1429	Low-steady N=125835	Intermediate-slow/increasing N=8268	High-decreasing N=3775	Overall N=139307	Intermediate-fast/increasing N=1429	Low-steady N=125835	Intermediate-slow/increasing N=8268	High-decreasing N=3775	Overall N=139307
Drug classes										
Alimentary tract and metabolism	1083 (75.8)	76827 (61.1)	6943 (84.0)	3768 (99.8)	88621 (63.6)	1426 (99.8)	82120 (65.3)	8211 (99.3)	3516 (93.1)	95273 (68.4)
Blood and blood forming organs	626 (43.8)	34629 (27.5)	4027 (48.7)	2772 (73.4)	42054 (30.2)	1021 (71.4)	44634 (35.5)	6202 (75.0)	2506 (66.4)	54363 (39.0)
Cardiovascular system	1065 (74.5)	80016 (63.6)	6763 (81.8)	3595 (95.2)	91439 (65.6)	1229 (86.0)	83501 (66.4)	7471 (90.4)	3258 (86.3)	95459 (68.5)
Dermatological	494 (34.6)	31145 (24.8)	3489 (42.2)	2692 (71.3)	37820 (27.1)	757 (53.0)	28827 (22.9)	4634 (56.0)	1931 (51.2)	36149 (25.9)
Genito-urinary system and sex hormones	270 (18.9)	16505 (13.1)	1755 (21.2)	1251 (33.1)	19781 (14.2)	407 (28.5)	16922 (13.4)	2362 (28.6)	861 (22.8)	20552 (14.8)
Systemic hormonal preparations, excluding sex hormones and insulins	132 (9.2)	9107 (7.2)	1001 (12.1)	898 (23.8)	11138 (8.0)	276 (19.3)	10134 (8.1)	1514 (18.3)	745 (19.7)	12669 (9.1)
Anti-infective for systemic use	604 (42.3)	40099 (31.9)	3996 (48.3)	3116 (82.5)	47815 (34.3)	1067 (74.7)	39626 (31.5)	5771 (69.8)	2241 (59.4)	48705 (35.0)
Antineoplastic and immunomodulation agents	125 (8.7)	3942 (3.1)	599 (7.2)	491 (13.0)	5157 (3.7)	241 (16.9)	4864 (3.9)	1044 (12.6)	358 (9.5)	6507 (4.7)
Musculoskeletal system	393 (27.5)	28891 (23.0)	2644 (32.0)	2033 (53.9)	33961 (24.4)	674 (47.2)	23802 (18.9)	3397 (41.1)	1259 (33.4)	29132 (20.9)
Nervous system	657 (46.0)	41324 (32.8)	4482 (54.2)	3249 (86.1)	49712 (35.7)	1378 (96.4)	44272 (35.2)	6899 (83.4)	2900 (76.8)	55449 (39.8)
Anti-parasitic products, insecticides and repellents	125 (8.7)	3942 (3.1)	599 (7.2)	491 (13.0)	5157 (3.7)	241 (16.9)	4864 (3.9)	1044 (12.6)	358 (9.5)	6507 (4.7)
Respiratory system	539 (37.7)	29120 (23.1)	3659 (44.3)	2744 (72.7)	36062 (25.9)	822 (57.5)	28370 (22.5)	4805 (58.1)	2160 (57.2)	36157 (26.0)
Sensory organs	131 (9.2)	8451 (6. 7)	921 (11.1)	848 (22.5)	10351 (7.4)	523 (36.6)	10736 (8.5)	1917 (23.2)	594 (15.7)	13770 (9.9)
Various	10 (0.7)	206 (0.2)	57 (0.7)	126 (3.3)	399 (0.3)	74 (5.2)	286 (0.2)	165 (2.0)	88 (2.3)	613 (0.4)
Drugs										
Psychostimulants	6 (0.4)	173 (0.1)	19 (0.2)	28 (0.7)	226 (0.2)	32 (2.2)	272 (0.2)	53 (0.6)	46 (1.2)	403 (0.3)
Opioids	451 (31.6)	25385 (20.2)	3111 (37.6)	2520 (66.8)	31467 (22.6)	1286 (90)	43528 (34.6)	6343 (76.7)	3108 (82.3)	54265 (39.0)
Drugs for obstructive airway disease	563 (39.4)	32776 (26)	3804 (46)	2482 (65.7)	39625 (28.4)	799 (55.9)	46541 (37)	5368 (64.9)	2801 (74.2)	55509 (39.8)
Drugs used in diabetes	344 (24.1)	17305 (13.8)	2317 (28)	1488 (39.4)	21454 (15.4)	423 (29.6)	20416 (16.2)	2770 (33.5)	1588 (42.1)	25197 (18.1)
Psycholeptics	486 (34.0)	32698 (26.0)	3315 (40.1)	2419 (64.1)	38918 (27.9)	1243 (87.0)	45879 (36.5)	5784 (70..0)	2874 (76.1)	55780 (40.0)
Agents renin-angiotensin systemic	733 (51.3)	50313 (40.0)	4785 (57.9)	2618 (69.4)	58449 (42.0)	840 (58.8)	60891 (48.4)	5834 (70.6)	2804 (74.3)	70369 (50.5)
Anti-thrombotic	441 (30.9)	23913 (19.0)	2721 (32.9)	1932 (51.2)	29007 (20.8)	816 (57.1)	40873 (32.5)	5266 (63.7)	2435 (64.5)	49390 (35.5)

	At baseline (n (%))					At the last observation (n (%))				
	Intermediate-fast/increasing N=1429	Low-steady N=125835	Intermediate-slow/increasing N=8268	High-decreasing N=3775	Overall N=139307	Intermediate-fast/increasing N=1429	Low-steady N=125835	Intermediate-slow/increasing N=8268	High-decreasing N=3775	Overall N=139307
Lipid modifying agents	727 (50.9)	49930 (39.7)	4732 (57.2)	2607 (69.1)	57996 (41.6)	836 (58.5)	60585 (48.1)	5749 (69.5)	2769 (73.4)	69939 (50.2)
Calcium channel blockers	370 (25.9)	23439 (18.6)	2566 (31.0)	1595 (42.3)	27970 (20.1)	513 (35.9)	34352 (27.3)	3810 (46.1)	1879 (49.8)	40554 (29.1)
Diuretics	567 (39.7)	34210 (27.2)	3538 (42.8)	2378 (63.0)	40693 (29.2)	876 (61.3)	48029 (38.2)	5575 (67.4)	2804 (74.3)	57284 (41.1)
Antibacterial systemic	844 (59.1)	60194 (47.8)	5260 (63.6)	3273 (86.7)	69571 (49.9)	1264 (88.5)	88149 (70.1)	7631 (92.3)	3632 (96.2)	100676 (72.3)
Immunosuppressant	41 (2.9)	1684 (1.3)	288 (3.5)	234 (6.2)	2247 (1.6)	75 (5.2)	2467 (2)	546 (6.6)	296 (7.8)	3384 (2.4)
Anti-inflammatory and/or anti-rheumatic	737 (51.6)	55675 (44.2)	4781 (57.8)	2956 (78.3)	64149 (46.0)	1208 (84.5)	77965 (62.0)	6942 (84.0)	3330 (88.2)	89445 (64.2)
Hormonal contraceptives systemic	<5	163 (0.1)	22 (0.3)	22 (0.6)	209 (0.2)	5 (0.3)	228 (0.2)	37 (0.4)	23 (0.6)	293 (0.2)
Anti-epileptics	136 (9.5)	5920 (4.7)	872 (10.5)	846 (22.4)	7774 (5.6)	349 (24.4)	10504 (8.3)	1973 (23.9)	1208 (32)	14034 (10.1)
Drugs for acid related disorder	907 (63.5)	63406 (50.4)	5903 (71.4)	3504 (92.8)	73720 (52.9)	1326 (92.8)	85197 (67.7)	7764 (93.9)	3675 (97.4)	97962 (70.3)
Beta blocking agents	606 (42.4)	41594 (33.1)	3971 (48.0)	2336 (61.9)	48507 (34.8)	817 (57.2)	51339 (40.8)	5346 (64.7)	2567 (68)	60069 (43.1)
Antidepressants	226 (15.8)	13396 (10.6)	1645 (19.9)	1356 (35.9)	16623 (11.9)	466 (32.6)	19815 (15.7)	2849 (34.5)	1747 (46.3)	24877 (17.9)
Antineoplastic agents	87 (6.1)	4034 (3.2)	551 (6.7)	386 (10.2)	5058 (3.6)	208 (14.6)	6723 (5.3)	1144 (13.8)	521 (13.8)	8596 (6.2)

7.5 Discussion

Main findings

In this chapter, I externally validated the clusters I identified in Chapter 6. Due to the relatively smaller IPCI population, the analysis was run on the entire sample of older people registered in IPCI during the study period. The clusters identified in IPCI had similar polypharmacy trajectories and mortality risks to those identified in UK data. Their demographic and clinical characteristics were also comparable to those identified in CPRD GOLD and Aurum.

External validation of the models in IPCI

The model showed good validation metrics in IPCI. A slightly higher proportion had a posterior probability ≥0.7 compared to the Aurum population. Consequently, the likelihood of misclassification was lower in the IPCI population. The clusters had similar mortality risks. The proportion of patients belonging to each cluster were slightly different. In general, it seems that the IPCI sample had a higher proportion belonging to the low-steady cluster; 90% vs. ~85% in UK data. This has led to lower proportions of patients belonging to the other more morbid clusters.

Clusters characteristics

Like in GOLD and Aurum, the low-steady cluster in IPCI was the healthiest cluster encompassing most of the study population. Compared to this healthy group, the other clusters had higher comorbidity burden, drug use, and associated risks of mortality. Moreover, the participants accumulated diseases and use of drugs at higher rates. Unlike the UK population, anxiety and dementia burden in the healthy cluster were comparable to the other clusters, but in IPCI, depression was lowest in the low-steady cluster, especially by the end of follow-up.

As in CPRD data, the intermediate-fast/increasing cluster in IPCI had the highest cancer baseline prevalence and cumulative incidence. This resulted in 44% of the intermediate-

fast/increasing cluster having any malignancy at baseline, rising to 75% by end of follow-up, compared to 26% rising to 54% in GOLD and 33% rising to 58% in Aurum. In general, there were higher records of cancer diagnosis in IPCI compared to CPRD data.

Similarly to GOLD and Aurum, the high-decreasing cluster had the highest burden of chronic diseases at baseline and by end of follow-up. While the intermediate-slow/increasing cluster had the fastest accumulation of chronic diseases and second fastest onset of cancer.

Although, the individual clusters in IPCI and CPRD data had different disease and drug use prevalence, they exhibited similar accumulation patterns of comorbidity burden and drug use over the study period. The biggest differences in disease prevalence were exhibited in renal and chronic kidney diseases, which were lower in IPCI, and pneumonia, which was higher in IPCI. Use of most drugs was lower in IPCI. For example, it seems that in IPCI for nervous systems drugs, like opioids, patients in the intermediate-increasing clusters had much lower prevalence at baseline but a great increment over the study period. In general, the difference in disease burden, drug prescriptions, and cluster distribution can be attributed to the different populations, healthcare guidelines, and data entry systems and practices.

Strengths, limitations and implications

The analysis in this chapter comes with both strengths and limitations. The good validation metrics support the generalisability of the model built in the previous chapter. Despite the differences in the study population and health practices, the resulting clusters have shown similar characteristics and polypharmacy trends. Use of a common data model helped to facilitate and expedite the model validation as the OMOP mapped versions of CPRD and IPCI data had the same structure and vocabularies.

An implication of this study is the generalisability of the clusters. The model performed well in two external databases: Aurum and IPCI. This confirms that the identified association

between polypharmacy trajectories and risk of mortality is similar in older people despite the data heterogeneity and varying populations. Both model training and validation were carried out in primary care databases, in neighbouring countries with similar infrastructure and living conditions. Validating the model in more diverse databases with varying structures and populations like US claims could further support the generalisability of this model. Furthermore, due to the complexity and computation time used to build the model, it was only validated in the external database. However, replication of the analysis, i.e., building other models using the same methods/algorithms can provide us with better understanding of the individual datasets and the robustness of the polypharmacy and associations observed.

Conclusion

The analyses carried out in this chapter demonstrate the robustness of my previous analysis of clusters of longitudinal polypharmacy trajectories identified in UK data, and of the observed associations with mortality. These results therefore support the conclusion that longitudinal modelling of polypharmacy should be considered in the assessment of older people.

8. Longitudinal trajectories of polypharmacy in middle-aged people, and their association with mortality

8.1 Chapter summary

In Chapters 6 and 7, I built and validated clusters of polypharmacy trajectories in older people. Some of the clusters included older people who started with intermediate/high polypharmacy levels at baseline. Polypharmacy is a cumulative measure that is likely to develop over the years, and could in some patients develop at younger ages. Moreover, polypharmacy is typically associated with multimorbidity and is often predictive of more morbidities as the patient grows older. In this chapter, I therefore aimed to identify clusters of distinct polypharmacy trajectories and associated mortality risks in the middle-aged population. I hypothesised that understanding the evolution of polypharmacy from an earlier age would provide a better understanding of the patient's later journey.

I used a random sample of patients aged 50-64 on the study start date from CPRD GOLD. I split the sample into training and testing sets, and used joint latent class models to identify the clusters in the training set. Then, I characterised those clusters and compared their demographics, morbidity burden, and annual prevalence of medicine use. Lastly, I used the testing set to validate the model performance.

I found three clusters of distinct polypharmacy trajectories and associated mortality risks: low-steady, high-decreasing, and intermediate-increasing. The low-steady cluster was the healthiest and included most (91%) of the study population. The intermediate-increasing cluster had the highest risk of mortality, the fastest trajectory of polypharmacy progression and accumulation of comorbidities, and was likely driven by a higher prevalence and incidence of cancer. The high-decreasing cluster had the second highest mortality risk and the highest prevalence of other (non-cancer) comorbidities.

The results of this chapter highlight the importance of monitoring polypharmacy over time starting from middle age. Early interventions in the prevention and management of cancer and other co-morbidities in middle-aged adults could lead to better health outcomes in older age. External validation of the identified clusters in independent databases to assess their rigour and generalisability is needed.

8.2 Introduction

Although polypharmacy is a phenomenon typically associated with -and more prevalent in- older compared to younger ages[153], it has become increasingly common across all ages in the last few decades[17 18 72]. Vos et al. found that the prevalence of polypharmacy increased from 16.8% in 2000 to 51.4% in 2014 for females aged 45-64 and 11.4% to 40.3% for their male counterparts[18].

Those who have or accumulate morbidities at an earlier age are more likely to develop further morbidities, polypharmacy and more complex drug regimens later in life[17 18]. For example, a study found that those who have multimorbidity at baseline have 47% higher odds of developing polypharmacy compared to those without multimorbidity[18]. Just like in older age, polypharmacy in middle-aged people is associated with a higher risk of potentially inappropriate medication use[154], falls[155], and mortality[69 156].

Health improvements can occur at any age[157]. However, since the risk of death increases with age and multimorbidity, health improvements from late middle age would guarantee better health in later life, leading to desirable clinical and public health outcomes[157].

A recent systematic review revealed that most research on polypharmacy and multimorbidity focuses on older adults aged 65 and above[15]. While research on polypharmacy trajectories over time is scarce for older people, it is even more limited for middle-aged people.

In Chapters 6 and 7, I identified, characterised, and externally validated clusters of older people with distinct five-year polypharmacy trajectories and associated mortality risk profiles. Around 10-15% of the study populations belonged to clusters that started at intermediate/high baseline values. This means that they have likely developed polypharmacy at a younger age. In this chapter, I aimed to identify clusters of middle-aged people, who have distinct polypharmacy

trajectories and associated mortality risks. Those distinct trajectories could then potentially be indicative of later life polypharmacy progression.

8.3 Methods

Design

In this study, I assessed the change of polypharmacy over five years in middle-aged people by conducting a longitudinal cohort analysis. I extracted a random sample of the UK middle-aged population from a primary care database. I randomly split the study sample 50-50 into training and testing sets. I calculated polypharmacy in terms of the absolute number of ingredients taken annually during the follow-up period. In the training set, I applied joint latent class models to identify clusters with distinct polypharmacy patterns over time and assessed their association with all-cause mortality. The identified clusters were characterised, and the model was internally validated in the testing set.

Data source

Data were extracted from CPRD GOLD (January 2023 release) and mapped to the OMOP-CDM (3.2).

Population

A random sample of 600,000 patients aged 50-64 and alive on 1st January 2015 from CPRD GOLD were eligible for inclusion if they had been registered for at least 1 year prior to the study start date. Patients exited the study at the earliest of the following: practice last collection date, patient transfer-out of practice date, death date[92] or study end date.

The study sample was then randomly split into a training and a testing set (50/50 split), with the partition based on the outcome, to ensure balanced outcome counts in both datasets. The R package 'caret' version 6.0_94 was used to split the data [link].

Exposure

Exposure was polypharmacy (3.3), calculated annually at baseline and at the end of each follow-up year.

Outcomes

The adverse outcome analysed was all-cause mortality (3.3).

Statistical analysis

Descriptive analysis

The baseline demographics were extracted for both training and testing sets.

Joint latent class modelling

To see if there were clusters of distinct patterns of polypharmacy change and whether allocation to such a cluster was predictive of mortality, I applied joint latent class model (3.3). Similar to Chapter 6, I tested linear, spline, and beta link functions for the longitudinal part, as well as Weibull, splines, and piecewise functions for the survival part. The best functions were carried forward for determining the number of clusters. The optimum number of clusters was based on the criteria described in 3.3. Age and sex were adjusted for in both longitudinal and survival sub-models.

Characterisation of the resulting clusters

To characterise the resulting clusters, comorbidities and drug use were compared across the clusters at both baseline and last observation, as described in Section 3.3.

Internal validation

I assessed the performance of the emergent best-fit cluster model (derived from the training set) in the testing set by calculating the posterior probabilities and the risk of mortality for the clusters identified in the testing set and compared them to the original model. I also compared the distribution, demographics and clinical characteristics in the testing set to those in the training set (see 3.3).

I used R (Version 4.3.2) for all analyses carried out in this study. R packages ‘lcmm’ version 2.0.2 [link] and ‘PatientProfiles’ version 0.4.0 [link] were used for the joint latent class modelling and characterisation of resultant clusters, respectively.

8.4 Results

Model training in CPRD GOLD (training set)

Study population

There were 299,936 patients included in the CPRD GOLD training set. Mean age at start was 56.5 (SD 4.3) with 50.2% of the population being male (See Table 8.1).

Table 8.1 Baseline characteristics by cluster and for the overall population in CPRD GOLD training set

	Intermediate-increasing	Low-steady	High-decreasing	Overall
n	9605 (3.2%)	273234 (91.1%)	17097 (5.7%)	299936
Sex = Male (%)	4519 (47.0%)	139270 (51.0%)	6822 (39.9%)	150611 (50.2%)
Age (mean (SD))	57.41 (4.30)	56.41 (4.29)	57.18 (4.31)	56.48 (4.30)
Number of ingredients* (median [IQR])	10 [6, 14]	3 [0, 6]	20 [17, 25]	3 [1, 7]
Charlson morbidity index (mean (SD))	0.88 (1.36)	0.30 (0.73)	1.14 (1.45)	0.37 (0.85)

* In the year prior to start

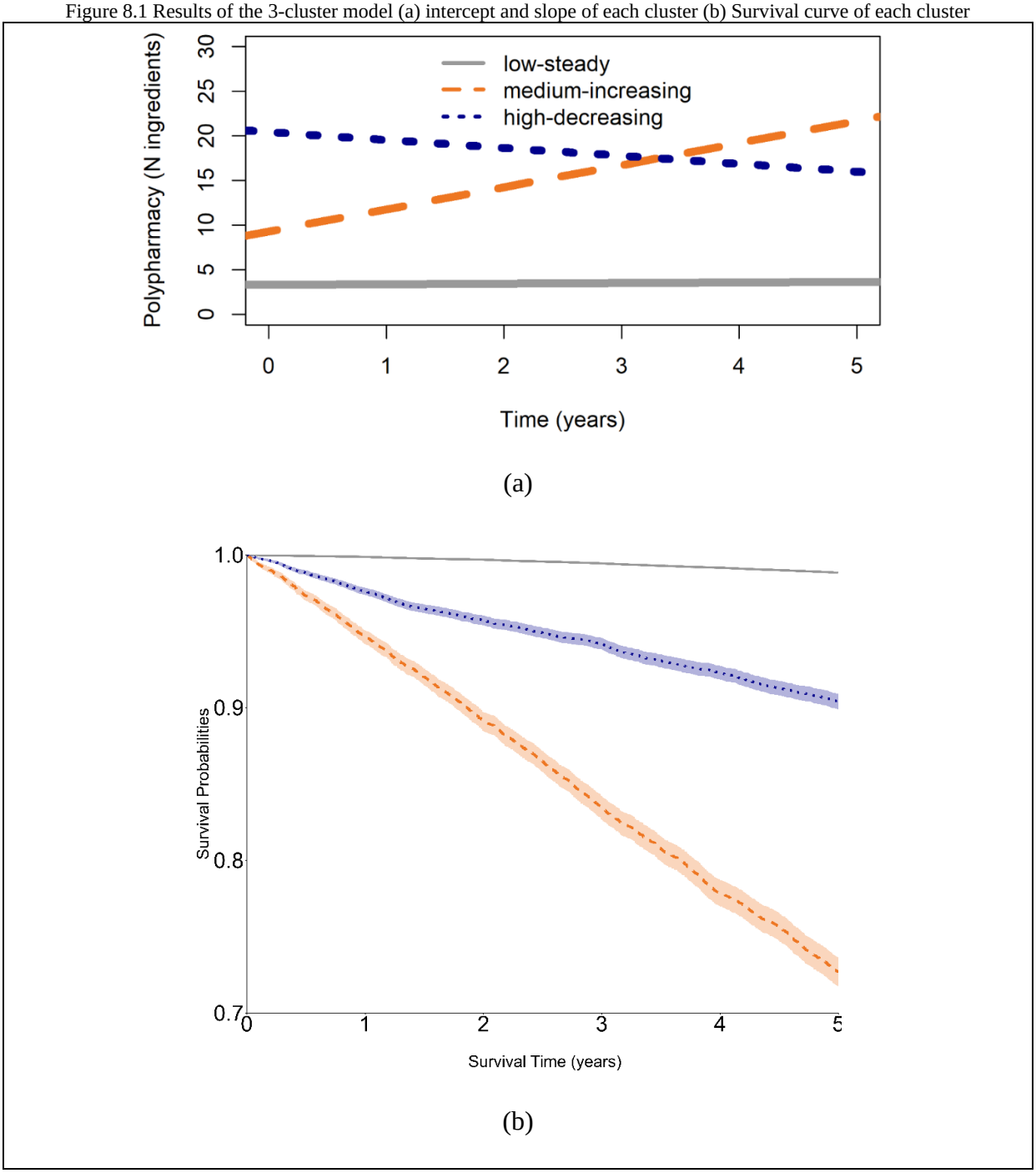
Joint latent class modelling

Median follow-up was 5.0 years [IQR 1.6 to 5.0 years] with 5,833 (1.9%) patients dying during the study period. The linear link function in the longitudinal part and proportional hazard Weibull function in the survival part were taken forward. Table S 8.1 and Table S 8.2 in the Appendix compare the performance between the different link and survival functions.

Based on convergence, BIC and clinical plausibility, the 3-clusters model was the optimal solution. Table S 8.3 in the Appendix summarises the tested models. I identified the three clusters in this model as “low-steady” (91.1%, intercept=3.35, slope=0.06), “intermediate-increasing” (3.2%, intercept=9.33, slope=2.48), and “high-decreasing” cluster (5.7%, intercept=20.47, slope=-0.89). There were 2,168 (0.8%) deaths in the low-steady cluster. Taking that cluster as a reference, the intermediate-increasing cluster had the highest relative mortality (HR 24.6 (95%CI 23.6, 25.7)) with 2,373 (24.7%) deaths, followed by the high-decreasing cluster (HR 8.9 (95%CI 8.5, 9.3)) with 1,292 (7.6%) deaths. Figure 8.1 describes polypharmacy progression and associated mortality probabilities for each cluster. Compared to the low-steady cluster, both intermediate-increasing and high-decreasing clusters had a slightly

older population and a lower proportion of males (Table 8.1). Appendix Table S 8.4 reports the estimates for each cluster. Median follow-up, yearly deaths and polypharmacy are reported in Appendix Table S 8.5.

Figure 8.1 Results of the 3-cluster model (a) intercept and slope of each cluster (b) Survival curve of each cluster



Characterisation of the identified clusters

Table 8.2 and Table 8.3 describe the comorbidity burden and drug use (drug classes and common drugs) at baseline and by the end of follow-up for all the clusters and the overall population. Table S 8.6 and Table S 8.7 in the Appendix describe the differences between the baseline and the last observed proportions.

Comorbidities

The low-steady cluster had the lowest comorbidity burden at baseline and the slowest accumulation of comorbidities over time during the study period. The other two clusters had higher baseline comorbidity burden and more rapid accumulation. Charlson comorbidity index was 0.30 at baseline and increased by (Δ=0.14) end of follow-up in the low-steady cluster, compared to 0.88 (Δ=1.02) and 1.14 (Δ=0.43) in the intermediate-increasing and high-decreasing clusters, respectively.

The high-decreasing cluster had the highest baseline burden of chronic diseases, while the intermediate-increasing cluster had the fastest accumulation of these diseases over the study period. For example, 23.8% had diabetes at baseline in the high-decreasing cluster, rising to 28.7% (Δ=4.6%) at the end of follow-up, while 16.3% had diabetes at baseline in the intermediate-increasing cluster, rising to 23.9% (Δ=7.6%) by the end of the study period. This was similar for hypertension, asthma, COPD and renal diseases. Both clusters had high diagnoses of mental health problems, with over 30% of their populations diagnosed with anxiety or depressive disorder by the end of the study.

The intermediate-increasing cluster also had the highest baseline prevalence and cumulative incidence of cancer throughout the study follow-up, with 9.3% and 24.3% (Δ=15.0%) having any malignancy at baseline and end of follow-up, respectively. This compared to 7.9%

prevalence at baseline rising to 11.6% (Δ =3.7%) at the end of follow-up in the high-decreasing cluster.

Medication use

The clusters followed the overall trend of their respective polypharmacy trajectories when assessing drug use by ATC classes and, to a lesser extent, when looking at specific drugs of interest. Similar to comorbidities, the use of prescribed medicines was highest in the high-decreasing cluster, with the intermediate-increasing cluster having the fastest rate of increase and having similar use of drugs by the end of the study follow-up. For example, 79.0% of patients belonging to the high-decreasing cluster were using drugs from the respiratory system class at baseline, dropping to 70.2% (Δ=-8.8%) by the end of follow-up vs. 52.4% and rising to 74.3% (Δ=21.9%) in the intermediate-increasing cluster. By the end of the study follow-up, over 90.0% of patients belonging to either the high-decreasing cluster or intermediate-increasing cluster were using opioids vs. 51% in the low-steady cluster.

Table 8.2 Individual morbidities at baseline and the last observation (Including Charlson morbidities) for each cluster and the overall population in CPRD GOLD training set

	At baseline				At last observation			
	Intermediate-increasing N= 9605	Low-steady N= 273234	High-decreasing N=17097	Overall N=299936	Intermediate-increasing N= 9605	Low-steady N= 273234	High-decreasing N=17097	Overall N=299936
Mean (SD)								
Charlson morbidity index	0.88 (1.36)	0.30 (0.73)	1.14 (1.45)	0.37 (0.85)	1.90 (2.14)	0.44 (0.92)	1.57 (1.74)	0.55 (1.10)
n (%)								
Myocardial infarction	290 (3.0)	2784 (1.0)	884 (5.2)	3958 (1.3)	684 (7.1)	4358 (1.6)	1101 (6.4)	6143 (2.0)
Congestive heart failure	125 (1.3)	663 (0.2)	393 (2.3)	1181 (0.4)	395 (4.1)	1255 (0.5)	601 (3.5)	2251 (0.8)
Peripheral vascular disease	171 (1.8)	875 (0.3)	380 (2.2)	1426 (0.5)	324 (3.4)	1401 (0.5)	536 (3.1)	2261 (0.8)
Cerebrovascular disease	356 (3.7)	2859 (1.0)	849 (5.0)	4064 (1.4)	716 (7.5)	4699 (1.7)	1187 (6.9)	6602 (2.2)
Dementia	47 (0.5)	353 (0.1)	79 (0.5)	479 (0.2)	105 (1.1)	711 (0.3)	165 (1.0)	981 (0.3)
Chronic obstructive pulmonary disease	1734 (18.1)	23103 (8.5)	4469 (26.1)	29306 (9.8)	2645 (27.5)	28570 (10.5)	5415 (31.7)	36630 (12.2)
Rheumatologic disease	226 (2.4)	2324 (0.9)	768 (4.5)	3318 (1.1)	389 (4.0)	3316 (1.2)	975 (5.7)	4680 (1.6)
Peptic ulcer disease	229 (2.4)	2859 (1.0)	550 (3.2)	3638 (1.2)	353 (3.7)	3687 (1.3)	728 (4.3)	4768 (1.6)
Mild liver disease	105 (1.1)	620 (0.2)	214 (1.3)	939 (0.3)	259 (2.7)	1016 (0.4)	324 (1.9)	1599 (0.5)
Diabetes with chronic complications	744 (7.7)	4532 (1.7)	1973 (11.5)	7249 (2.4)	1177 (12.3)	6876 (2.5)	2746 (16.1)	10799 (3.6)
Hemiplegia or paraplegia	22 (0.2)	107 (0)	43 (0.3)	172 (0.1)	27 (0.3)	139 (0.1)	49 (0.3)	215 (0.1)
Renal disease	610 (6.4)	6342 (2.3)	1655 (9.7)	8607 (2.9)	1139 (11.9)	9132 (3.3)	2373 (13.9)	12644 (4.2)
Any malignancy	895 (9.3)	11346 (4.2)	1346 (7.9)	13587 (4.5)	2332 (24.3)	17878 (6.5)	1979 (11.6)	22189 (7.4)
Moderate to severe liver disease	49 (0.5)	178 (0.1)	98 (0.6)	325 (0.1)	139 (1.4)	366 (0.1)	159 (0.9)	664 (0.2)
Metastatic solid tumour	83 (0.9)	161 (0.1)	80 (0.5)	324 (0.1)	429 (4.5)	418 (0.2)	151 (0.9)	998 (0.3)
AIDS	<5	64 (0)	12 (0.1)	78 (0)	<5	76 (0)	13 (0.1)	92 (0)
Hypertension	2443 (25.4)	39577 (14.5)	5141 (30.1)	47161 (15.7)	3166 (33.0)	51171 (18.7)	5834 (34.1)	60171 (20.1)
Heart failure	151 (1.6)	844 (0.3)	449 (2.6)	1444 (0.5)	456 (4.8)	1576 (0.6)	673 (3.9)	2705 (0.9)
Osteoporosis	255 (2.7)	2540 (0.9)	725 (4.2)	3520 (1.2)	255 (2.7)	2540 (0.9)	725 (4.2)	3520 (1.2)
Gastroesophageal reflux disease	470 (4.9)	6468 (2.4)	1207 (7.1)	8145 (2.7)	696 (7.3)	8769 (3.2)	1549 (9.1)	11014 (3.7)
Chronic kidney disease	614 (6.4)	6158 (2.3)	1689 (9.9)	8461 (2.8)	975 (10.2)	8432 (3.1)	2214 (13.0)	11621 (3.9)
Venous thromboembolism	367 (3.8)	4871 (1.8)	972 (5.7)	6210 (2.1)	770 (8.0)	6704 (2.5)	1296 (7.6)	8770 (2.9)

	At baseline				At last observation			
	Intermediate-increasing N= 9605	Low-steady N= 273234	High-decreasing N=17097	Overall N=299936	Intermediate-increasing N= 9605	Low-steady N= 273234	High-decreasing N=17097	Overall N=299936
Hypothyroidism	675 (7.0)	10556 (3.9)	1707 (10.0)	12938 (4.3)	905 (9.4)	12729 (4.7)	1980 (11.6)	15614 (5.2)
Stroke	206 (2.1)	1632 (0.6)	483 (2.8)	2321 (0.8)	427 (4.5)	2643 (1.0)	672 (3.9)	3742 (1.3)
Anxiety	2285 (23.8)	37803 (13.8)	5506 (32.2)	45594 (15.2)	2899 (30.2)	44973 (16.5)	6294 (36.8)	54166 (18.1)
Asthma	1578 (16.4)	19060 (7.0)	4291 (25.1)	24929 (8.3)	2009 (20.9)	22116 (8.1)	4847 (28.4)	28972 (9.7)
Pneumonia	226 (2.4)	2007 (0.7)	592 (3.5)	2825 (0.9)	642 (6.7)	3112 (1.1)	1007 (5.9)	4761 (1.6)
Diabetes	1562 (16.3)	13425 (4.9)	4072 (23.8)	19059 (6.4)	2294 (23.9)	19126 (7.0)	4904 (28.7)	26324 (8.8)
Inflammatory bowel disease	122 (1.3)	1672 (0.6)	311 (1.8)	2105 (0.7)	181 (1.9)	2003 (0.7)	360 (2.1)	2544 (0.9)
Depressive disorder	2559 (26.6)	39257 (14.4)	6236 (36.5)	48052 (16.0)	3046 (31.7)	44141 (16.2)	6899 (40.4)	54086 (18.0)

Table 8.3 Drug use by drug classes and common drugs at baseline and the last observation for each cluster and the overall population in CPRD GOLD training set

	At baseline (n (%))				At last observation (n (%))			
	Intermediate-increasing N= 9605	Low-steady N= 273234	High-decreasing N=17097	Overall N=299936	Intermediate-increasing N= 9605	Low-steady N= 273234	High-decreasing N=17097	Overall N=299936
Drug classes								
Alimentary tract and metabolism	7278 (75.8)	102659 (37.6)	16863 (98.6)	126800 (42.3)	9400 (97.9)	102732 (37.6)	16228 (94.9)	128360 (42.8)
Blood and blood forming organs	2801 (29.2)	25150 (9.2)	8583 (50.2)	36534 (12.2)	5204 (54.2)	29841 (10.9)	8514 (49.8)	43559 (14.5)
Cardiovascular system	5481 (57.1)	80035 (29.3)	13331 (78.0)	98847 (33.0)	7261 (75.6)	92218 (33.8)	13247 (77.5)	112726 (37.6)
Dermatological	4038 (42.0)	58387 (21.4)	12630 (73.9)	75055 (25.0)	6334 (65.9)	52390 (19.2)	10650 (62.3)	69374 (23.1)
Genito-urinary system and sex hormones	2037 (21.2)	34363 (12.6)	6396 (37.4)	42796 (14.3)	3004 (31.3)	34055 (12.5)	5524 (32.3)	42583 (14.2)
Systemic hormonal preparations, excluding sex hormones and insulins	1061 (11.0)	16438 (6.0)	3225 (18.9)	20724 (6.9)	1559 (16.2)	18176 (6.7)	3227 (18.9)	22962 (7.7)
Anti-infective for systemic use	4981 (51.9)	78694 (28.8)	13600 (79.5)	97275 (32.4)	6938 (72.2)	60662 (22.2)	10386 (60.7)	77986 (26.0)
Antineoplastic and immunomodulation agents	483 (5.0)	6131 (2.2)	1648 (9.6)	8262 (2.8)	870 (9.1)	6493 (2.4)	1425 (8.3)	8788 (2.9)
Musculoskeletal system	3505 (36.5)	54424 (19.9)	9945 (58.2)	67874 (22.6)	4740 (49.3)	45763 (16.7)	8062 (47.2)	58565 (19.5)
Nervous system	6644 (69.2)	86620 (31.7)	15832 (92.6)	109096 (36.4)	8610 (89.6)	81078 (29.7)	15018 (87.8)	104706 (34.9)
Anti-parasitic products, insecticides and repellents	483 (5.0)	6131 (2.2)	1648 (9.6)	8262 (2.8)	870 (9.1)	6493 (2.4)	1425 (8.3)	8788 (2.9)
Respiratory system	5032 (52.4)	62079 (22.7)	13513 (79.0)	80624 (26.9)	7133 (74.3)	57538 (21.1)	12010 (70.2)	76681 (25.6)
Sensory organs	668 (7.0)	9375 (3.4)	2740 (16.0)	12783 (4.3)	2112 (22.0)	10615 (3.9)	2323 (13.6)	15050 (5.0)
Various	25 (0.3)	97 (0)	136 (0.8)	258 (0.1)	63 (0.7)	109 (0.0)	120 (0.7)	292 (0.1)
Drugs								
Psychostimulants	24 (0.3)	191 (0.1)	72 (0.4)	287 (0.1)	30 (0.3)	227 (0.1)	79 (0.5)	336 (0.1)
Opioids	7307 (76.1)	121125 (44.3)	15561 (91.0)	143993 (48.0)	8712 (90.7)	139367 (51.0)	16135 (94.4)	164214 (54.8)
Drugs for obstructive airway disease	5302 (55.2)	88653 (32.5)	12046 (70.5)	106001 (35.3)	6710 (69.9)	107030 (39.2)	13268 (77.6)	127008 (42.4)
Drugs used in diabetes	1657 (17.3)	12958 (4.7)	4721 (27.6)	19336 (6.5)	2391 (24.9)	17761 (6.5)	5433 (31.8)	25585 (8.5)
Psycholeptics	5644 (58.8)	88683 (32.5)	12815 (75.0)	107142 (35.7)	7367 (76.7)	103842 (38.0)	13907 (81.3)	125116 (41.7)
Agents renin-angiotensin systemic	3418 (35.6)	47281 (17.3)	8209 (48.0)	58908 (19.6)	4608 (48.0)	58765 (21.5)	9048 (52.9)	72421 (24.2)
Anti-thrombotic	1376 (14.3)	14184 (5.2)	3855 (22.6)	19415 (6.5)	3104 (32.3)	22208 (8.1)	5084 (29.7)	30396 (10.1)
Lipid modifying agents	3977 (41.4)	48408 (17.7)	9730 (56.9)	62115 (20.7)	5407 (56.3)	63686 (23.3)	10830 (63.3)	79923 (26.7)

	At baseline (n (%))				At last observation (n (%))			
	Intermediate-increasing N= 9605	Low-steady N= 273234	High-decreasing N=17097	Overall N=299936	Intermediate-increasing N= 9605	Low-steady N= 273234	High-decreasing N=17097	Overall N=299936
Calcium channel blockers	2360 (24.6)	30230 (11.1)	5917 (34.6)	38507 (12.8)	3552 (37.0)	43770 (16.0)	6977 (40.8)	54299 (18.1)
Diuretics	2699 (28.1)	28064 (10.3)	6991 (40.9)	37754 (12.6)	4087 (42.6)	33961 (12.4)	8036 (47.0)	46084 (15.4)
Antibacterial systemic	8717 (90.8)	207225 (75.8)	16605 (97.1)	232547 (77.5)	9363 (97.5)	222216 (81.3)	16917 (99.0)	248496 (82.9)
Immunosuppressant	349 (3.6)	3410 (1.3)	1412 (8.3)	5171 (1.7)	608 (6.3)	4402 (1.6)	1699 (9.9)	6709 (2.2)
Anti-inflammatory and/or anti-rheumatic	7684 (80.0)	167260 (61.2)	15257 (89.2)	190201 (63.4)	8726 (90.9)	184091 (67.4)	15891 (93.0)	208708 (69.6)
Hormonal contraceptives systemic	1087 (11.3)	26793 (9.8)	2442 (14.3)	30322 (10.1)	1251 (13.0)	29244 (10.7)	2663 (15.6)	33158 (11.1)
Anti-epileptics	2308 (24.0)	19016 (7.0)	7041 (41.2)	28365 (9.5)	4214 (43.9)	27564 (10.1)	8888 (52.0)	40666 (13.6)
Drugs for acid related disorder	6856 (71.4)	110087 (40.3)	15082 (88.2)	132025 (44.0)	8650 (90.1)	137068 (50.2)	15928 (93.2)	161646 (53.9)
Beta blocking agents	2871 (29.9)	41206 (15.1)	6802 (39.8)	50879 (17.0)	4117 (42.9)	50436 (18.5)	7713 (45.1)	62266 (20.8)
Antidepressants	6001 (62.5)	91530 (33.5)	13813 (80.8)	111344 (37.1)	7383 (76.9)	107734 (39.4)	14780 (86.5)	129897 (43.3)
Antineoplastic agents	726 (7.6)	7393 (2.7)	2361 (13.8)	10480 (3.5)	969 (10.1)	8750 (3.2)	2684 (15.7)	12403 (4.1)

Model testing in CPRD GOLD (testing set)

Study population

There were 299,936 patients included in the CPRD GOLD testing set. They had similar characteristics to those in the training set; 50.4% of male participants with a mean age of 56.5 (4.3) (Table 8.4).

Table 8.4 Baseline characteristics by cluster and for the overall population in CPRD GOLD testing set

	Intermediate-slow/increasing	Low-steady	High-decreasing	Overall
n	9415 (3.1%)	273517 (91.2%)	17004 (5.7%)	299936
Sex = Male (%)	4504 (47.8%)	139801 (51.1%)	6727 (39.6%)	151032 (50.4%)
Age (mean (SD))	57.36 (4.34)	56.42 (4.30)	57.17 (4.33)	56.49 (4.31)
Number of ingredients* (median [IQR])	10 [6, 14]	3 [0, 6]	20 [17, 25]	3 [1, 7]
Charlson morbidity index (mean (SD))	0.86 (1.34)	0.30 (0.74)	1.13 (1.41)	0.37 (0.84)

* In the year prior to start

Resultant clusters

Median follow-up was 5.0 years [1.6 to 5.0] with 5,838 (1.9%) patients dying during the follow-up period. The resulting clusters had good validation metrics; 96.6% of the testing population had a posterior probability ≥0.7 (Appendix Table S 8.8 and Table S 8.9). The clusters also had similar mortality profiles. Taking the low-steady cluster as reference (2,134 (0.8%) deaths), the intermediate-increasing cluster had the highest HR of 27.9 (95%CI 27.1, 28.7) with 2,369 (25.2%) deaths, followed by the high-decreasing cluster which had an HR of 9.8 (95%CI 9.4, 10.1) and 1,335 (7.9%) deaths. The resultant clusters also had similar baseline characteristics, i.e., older population and lower proportions of males (See Table 8.4). Median follow-up, yearly deaths and polypharmacy are reported in Appendix Table S 8.10.

Characterisation of the resultant clusters

Table 8.5 and Table 8.6 summarise the comorbidity burden and drug prevalence for all clusters at both, baseline and end of follow-up. Table S 8.11 and Table S 8.12 in the Appendix describe the differences between baseline and end of follow-up proportions.

Comorbidities

The testing set had similar characteristics to the training set. The low-steady cluster had the lowest comorbidity burden at baseline and the slowest accumulation over the study period. Charlson comorbidity index was 0.30 at baseline and increased by 0.15 at the end of follow-up, vs. 0.86 (Δ=1.01) and 1.13 (Δ=0.44) for the intermediate-increasing and high-decreasing clusters, respectively.

The high-decreasing cluster had the highest baseline burden of non-cancer chronic comorbidities, and the intermediate-increasing cluster had the highest accumulation rate of comorbidities. For example, 26.4% had COPD at baseline in the high-decreasing cluster, rising to 31.9% (Δ=5.5%), while 19.3% at baseline and rising to 28.3% (Δ=9.0%) by the end of follow-up for the intermediate-increasing cluster. Similar trends were observed for diabetes, hypertension, and asthma.

The intermediate-increasing cluster had the highest prevalence and incidence of cancer throughout the study period. The prevalence of any malignancy was 9.0% at baseline, rising to 24.1% (Δ=15.1%) by the end of follow-up in the intermediate-increasing cluster. This was followed by the high-decreasing cluster, where 8.3% had any malignancy at baseline, rising to 12.3% (Δ=4.0%) by the end of follow-up.

Medication use

Medication use by ATC classes followed similar trends of the individual clusters' polypharmacy trajectories. A lesser extent of that pattern was observed for specific drugs of interest. The high-decreasing cluster had the highest baseline drug use for most drugs and drug classes, while the intermediate-increasing cluster had the fastest accumulation of these drugs over the study period. For example, 58.3% of patients in the high-decreasing cluster used drugs from the musculoskeletal system class, dropping to 47.0% (Δ=-11.3%) by the end of follow-

up. On the other hand, 37.6% of patients in the intermediate-increasing cluster were using drugs from the same class, rising to 49.3% (Δ=11.7%) by the end of follow-up. Similar to the training set, over 90% of patients in the high-decreasing and intermediate-increasing clusters were using opioids by their last year of follow-up.

Table 8.5 Individual morbidities at baseline and the last observation (Including Charlson morbidities) for each cluster and the overall population in CPRD GOLD testing set

	At baseline				At last observation			
	Intermediate-increasing N=9415	Low-steady N=273517	High-decreasing N=17004	Overall N= 299936	Intermediate-increasing N=9415	Low-steady N=273517	High-decreasing N=17004	Overall N= 299936
Mean (SD)								
Charlson morbidity index	0.86 (1.34)	0.30 (0.74)	1.13 (1.41)	0.37 (0.84)	1.87 (2.12)	0.45 (0.93)	1.57 (1.72)	0.56 (1.10)
n (%)								
Myocardial infarction	299 (3.2)	2830 (1.0)	868 (5.1)	3997 (1.3)	689 (7.3)	4332 (1.6)	1121 (6.6)	6142 (2.0)
Congestive heart failure	140 (1.5)	720 (0.3)	376 (2.2)	1236 (0.4)	397 (4.2)	1359 (0.5)	611 (3.6)	2367 (0.8)
Peripheral vascular disease	151 (1.6)	837 (0.3)	349 (2.1)	1337 (0.4)	317 (3.4)	1384 (0.5)	501 (2.9)	2202 (0.7)
Cerebrovascular disease	316 (3.4)	2895 (1.1)	897 (5.3)	4108 (1.4)	668 (7.1)	4809 (1.8)	1278 (7.5)	6755 (2.3)
Dementia	51 (0.5)	326 (0.1)	90 (0.5)	467 (0.2)	120 (1.3)	668 (0.2)	161 (0.9)	949 (0.3)
Chronic obstructive pulmonary disease	1815 (19.3)	23103 (8.4)	4497 (26.4)	29415 (9.8)	2669 (28.3)	28736 (10.5)	5428 (31.9)	36833 (12.3)
Rheumatologic disease	236 (2.5)	2319 (0.8)	768 (4.5)	3323 (1.1)	440 (4.7)	3317 (1.2)	923 (5.4)	4680 (1.6)
Peptic ulcer disease	197 (2.1)	2728 (1.0)	562 (3.3)	3487 (1.2)	334 (3.5)	3505 (1.3)	715 (4.2)	4554 (1.5)
Mild liver disease	138 (1.5)	582 (0.2)	200 (1.2)	920 (0.3)	285 (3.0)	951 (0.3)	301 (1.8)	1537 (0.5)
Diabetes with chronic complications	658 (7.0)	4544 (1.7)	1928 (11.3)	7130 (2.4)	1077 (11.4)	6909 (2.5)	2687 (15.8)	10673 (3.6)
Hemiplegia or paraplegia	17 (0.2)	119 (0)	51 (0.3)	187 (0.1)	28 (0.3)	148 (0.1)	62 (0.4)	238 (0.1)
Renal disease	543 (5.8)	6473 (2.4)	1620 (9.5)	8636 (2.9)	1023 (10.9)	9429 (3.4)	2291 (13.5)	12743 (4.2)
Any malignancy	846 (9.0)	11454 (4.2)	1403 (8.3)	13703 (4.6)	2272 (24.1)	18134 (6.6)	2091 (12.3)	22497 (7.5)
Moderate to severe liver disease	46 (0.5)	175 (0.1)	88 (0.5)	309 (0.1)	136 (1.4)	331 (0.1)	161 (0.9)	628 (0.2)
Metastatic solid tumour	81 (0.9)	152 (0.1)	59 (0.3)	292 (0.1)	404 (4.3)	463 (0.2)	129 (0.8)	996 (0.3)
AIDS	<5	71 (0)	11 (0.1)	84 (0)	7 (0.1)	90 (0)	12 (0.1)	109 (0)
Hypertension	2342 (24.9)	39478 (14.4)	5013 (29.5)	46833 (15.6)	3064 (32.5)	50912 (18.6)	5743 (33.8)	59719 (19.9)
Heart failure	153 (1.6)	883 (0.3)	436 (2.6)	1472 (0.5)	448 (4.8)	1619 (0.6)	687 (4)	2754 (0.9)
Osteoporosis	218 (2.3)	2615 (1)	690 (4.1)	3523 (1.2)	503 (5.3)	4310 (1.6)	1016 (6)	5829 (1.9)
Gastroesophageal reflux disease	411 (4.4)	6682 (2.4)	1162 (6.8)	8255 (2.8)	615 (6.5)	8978 (3.3)	1443 (8.5)	11036 (3.7)
Chronic kidney disease	542 (5.8)	6199 (2.3)	1618 (9.5)	8359 (2.8)	877 (9.3)	8569 (3.1)	2130 (12.5)	11576 (3.9)

	At baseline				At last observation			
	Intermediate-increasing N=9415	Low-steady N=273517	High-decreasing N=17004	Overall N= 299936	Intermediate-increasing N=9415	Low-steady N=273517	High-decreasing N=17004	Overall N= 299936
Venous thromboembolism	377 (4.0)	4907 (1.8)	962 (5.7)	6246 (2.1)	783 (8.3)	6723 (2.5)	1253 (7.4)	8759 (2.9)
Hypothyroidism	639 (6.8)	10767 (3.9)	1728 (10.2)	13134 (4.4)	833 (8.8)	12918 (4.7)	1987 (11.7)	15738 (5.2)
Stroke	185 (2.0)	1643 (0.6)	520 (3.1)	2348 (0.8)	414 (4.4)	2755 (1.0)	744 (4.4)	3913 (1.3)
Anxiety	2264 (24.0)	37790 (13.8)	5460 (32.1)	45514 (15.2)	2847 (30.2)	44946 (16.4)	6280 (36.9)	54073 (18.0)
Asthma	1530 (16.3)	18947 (6.9)	4345 (25.6)	24822 (8.3)	1973 (21.0)	22195 (8.1)	4903 (28.8)	29071 (9.7)
Pneumonia	226 (2.4)	2090 (0.8)	627 (3.7)	2943 (1)	689 (7.3)	3228 (1.2)	1079 (6.3)	4996 (1.7)
Diabetes	1487 (15.8)	13342 (4.9)	3920 (23.1)	18749 (6.3)	2155 (22.9)	19139 (7.0)	4766 (28)	26060 (8.7)
Inflammatory bowel disease	117 (1.2)	1623 (0.6)	340 (2.0)	2080 (0.7)	150 (1.6)	1957 (0.7)	397 (2.3)	2504 (0.8)
Depressive disorder	2552 (27.1)	39379 (14.4)	6108 (35.9)	48039 (16.0)	3053 (32.4)	44283 (16.2)	6750 (39.7)	54086 (18.0)

Table 8.6 Drug use by drug classes and common drugs at baseline and the last observation for each cluster and the overall population in CPRD GOLD testing set

	At baseline (n (%))				At last observation (n (%))			
	Intermediate-increasing N=9415	Low-steady N=273517	High-decreasing N=17004	Overall N= 299936	Intermediate-increasing N=9415	Low-steady N=273517	High-decreasing N=17004	Overall N= 299936
Drug classes								
Alimentary tract and metabolism	7055 (74.9)	102846 (37.6)	16747 (98.5)	126648 (42.2)	9219 (97.9)	103485 (37.8)	16104 (94.7)	128808 (42.9)
Blood and blood forming organs	2749 (29.2)	24885 (9.1)	8589 (50.5)	36223 (12.1)	5091 (54.1)	29931 (10.9)	8467 (49.8)	43489 (14.5)
Cardiovascular system	5465 (58.0)	80400 (29.4)	13222 (77.8)	99087 (33.0)	7204 (76.5)	92672 (33.9)	13073 (76.9)	112949 (37.7)
Dermatological	4013 (42.6)	58265 (21.3)	12660 (74.5)	74938 (25.0)	6264 (66.5)	52260 (19.1)	10622 (62.5)	69146 (23.1)
Genito-urinary system and sex hormones	1986 (21.1)	34788 (12.7)	6398 (37.6)	43172 (14.4)	2866 (30.4)	34336 (12.6)	5461 (32.1)	42663 (14.2)
Systemic hormonal preparations, excluding sex hormones and insulins	1064 (11.3)	16635 (6.1)	3155 (18.6)	20854 (7.0)	1480 (15.7)	18275 (6.7)	3157 (18.6)	22912 (7.6)
Anti-infective for systemic use	4858 (51.6)	78860 (28.8)	13452 (79.1)	97170 (32.4)	6783 (72.0)	60967 (22.3)	10359 (60.9)	78109 (26.0)
Antineoplastic and immunomodulation agents	492 (5.2)	5904 (2.2)	1598 (9.4)	7994 (2.7)	847 (9.0)	6612 (2.4)	1370 (8.1)	8829 (2.9)
Musculoskeletal system	3538 (37.6)	55022 (20.1)	9919 (58.3)	68479 (22.8)	4640 (49.3)	46285 (16.9)	7992 (47.0)	58917 (19.6)
Nervous system	6548 (69.5)	86740 (31.7)	15707 (92.4)	108995 (36.3)	8395 (89.2)	81232 (29.7)	14897 (87.6)	104524 (34.8)
Anti-parasitic products, insecticides and repellents	492 (5.2)	5904 (2.2)	1598 (9.4)	7994 (2.7)	847 (9.0)	6612 (2.4)	1370 (8.1)	8829 (2.9)
Respiratory system	5060 (53.7)	61963 (22.7)	13358 (78.6)	80381 (26.8)	7109 (75.5)	58015 (21.2)	11879 (69.9)	77003 (25.7)
Sensory organs	642 (6.8)	9500 (3.5)	2847 (16.7)	12989 (4.3)	2026 (21.5)	10430 (3.8)	2356 (13.9)	14812 (4.9)
Various	24 (0.3)	100 (0)	137 (0.8)	261 (0.1)	72 (0.8)	87 (0)	110 (0.6)	269 (0.1)
Drugs								
Psychostimulants	25 (0.3)	217 (0.1)	81 (0.5)	323 (0.1)	33 (0.4)	247 (0.1)	91 (0.5)	371 (0.1)
Opioids	7246 (77.0)	121067 (44.3)	15483 (91.1)	143796 (47.9)	8568 (91)	139294 (50.9)	16037 (94.3)	163899 (54.6)
Drugs for obstructive airway disease	5188 (55.1)	88262 (32.3)	12013 (70.6)	105463 (35.2)	6661 (70.7)	106838 (39.1)	13232 (77.8)	126731 (42.3)
Drugs used in diabetes	1617 (17.2)	12863 (4.7)	4481 (26.4)	18961 (6.3)	2284 (24.3)	17778 (6.5)	5218 (30.7)	25280 (8.4)
Psycholeptics	5546 (58.9)	88722 (32.4)	12868 (75.7)	107136 (35.7)	7235 (76.8)	103795 (37.9)	13963 (82.1)	124993 (41.7)
Agents renin-angiotensin systemic	3455 (36.7)	47475 (17.4)	8053 (47.4)	58983 (19.7)	4593 (48.8)	58775 (21.5)	8917 (52.4)	72285 (24.1)
Anti-thrombotic	1428 (15.2)	14125 (5.2)	3801 (22.4)	19354 (6.5)	3106 (33.0)	22380 (8.2)	5099 (30.0)	30585 (10.2)
Lipid modifying agents	3888 (41.3)	48600 (17.8)	9617 (56.6)	62105 (20.7)	5388 (57.2)	63860 (23.3)	10718 (63.0)	79966 (26.7)

	At baseline (n (%))				At last observation (n (%))			
	Intermediate-increasing N=9415	Low-steady N=273517	High-decreasing N=17004	Overall N= 299936	Intermediate-increasing N=9415	Low-steady N=273517	High-decreasing N=17004	Overall N= 299936
Calcium channel blockers	2341 (24.9)	30404 (11.1)	5831 (34.3)	38576 (12.9)	3526 (37.5)	43610 (15.9)	6982 (41.1)	54118 (18)
Diuretics	2658 (28.2)	28487 (10.4)	6930 (40.8)	38075 (12.7)	4003 (42.5)	34373 (12.6)	8030 (47.2)	46406 (15.5)
Antibacterial systemic	8530 (90.6)	207716 (75.9)	16509 (97.1)	232755 (77.6)	9199 (97.7)	222651 (81.4)	16810 (98.9)	248660 (82.9)
Immunosuppressant	381 (4.0)	3382 (1.2)	1369 (8.1)	5132 (1.7)	654 (6.9)	4322 (1.6)	1608 (9.5)	6584 (2.2)
Anti-inflammatory and/or anti-rheumatic	7528 (80)	167186 (61.1)	15274 (89.8)	189988 (63.3)	8525 (90.5)	184319 (67.4)	15840 (93.2)	208684 (69.6)
Hormonal contraceptives systemic	1013 (10.8)	26914 (9.8)	2461 (14.5)	30388 (10.1)	1150 (12.2)	29317 (10.7)	2691 (15.8)	33158 (11.1)
Anti-epileptics	2249 (23.9)	18980 (6.9)	7044 (41.4)	28273 (9.4)	4066 (43.2)	27510 (10.1)	8901 (52.3)	40477 (13.5)
Drugs for acid related disorder	6663 (70.8)	110633 (40.4)	14958 (88)	132254 (44.1)	8510 (90.4)	137559 (50.3)	15867 (93.3)	161936 (54.0)
Beta blocking agents	2908 (30.9)	41193 (15.1)	6846 (40.3)	50947 (17.0)	4128 (43.8)	50512 (18.5)	7781 (45.8)	62421 (20.8)
Antidepressants	5990 (63.6)	91241 (33.4)	13700 (80.6)	110931 (37.0)	7334 (77.9)	107319 (39.2)	14599 (85.9)	129252 (43.1)
Antineoplastic agents	741 (7.9)	7265 (2.7)	2301 (13.5)	10307 (3.4)	1002 (10.6)	8617 (3.2)	2596 (15.3)	12215 (4.1)

8.5 Discussion

Main findings

In this study, I used joint latent class models to identify three clusters of middle-aged people with distinct polypharmacy trajectories and associated mortality risks. I fitted joint latent class models in a training subset and successfully validated them in the remaining test subset. The three resulting clusters were: low-steady, intermediate-increasing, and high-decreasing. I observed differential mortality across clusters, with the 'intermediate-increasing' cluster having the highest all-cause mortality risk, followed by the 'high-decreasing' and the 'low-steady' groups.

Additionally, I characterised the comorbidity burden and drug use of the emergent clusters. Compared to the low-steady cluster, the other two clusters had higher risks of mortality, comorbidity burden, and drug use. To the best of my knowledge, this is the first study to focus on polypharmacy trajectories in middle-aged people, while accounting for the risk of mortality and the heterogeneity of the study population.

Joint latent class model and characteristics of the emergent clusters

Based on convergence, BIC, variance explained, and the trade-off for a simpler model, the linear link function for the longitudinal part and the Weibull proportional hazard for the survival part were chosen and taken forward.

Based on BIC, convergence, and clinical plausibility, the three-cluster model was the optimal fit. Despite the low proportion of deaths in the study population, the three identified clusters had distinct polypharmacy trajectories, clinical characteristics, and associated mortality risks.

The low-steady cluster was the healthiest cluster and included the majority of the study population (91%). It had the highest proportion of males, lowest comorbidity burden, and lowest prevalence of medicine/s use. The other clusters started from higher baseline

polypharmacy levels, had more comorbidity burden, and more female proportions. This agrees with previous literature; females usually have higher comorbidity and drug use burden at all ages[17 18 72]. Due to this study's unique design and the various ways applied for calculating polypharmacy in research, it is difficult to directly compare the current polypharmacy clusters with other studies. However, in a cross-sectional analysis of data from ELSA, 11.7% of people aged 50-59 had polypharmacy (5-9 drugs), with 2.7% having hyper-polypharmacy (≥10 drugs). These proportions increased with older groups but were similar to my findings, where ~9% of the sample belonged to the polypharmacy clusters[153].

The intermediate-increasing and high-decreasing clusters had higher prevalence of chronic conditions like diabetes and hypertension when compared to the overall population prevalence. Generally, there is an increased risk of polypharmacy among those with multimorbidity and chronic diseases[14 158]. A study found that 50% of patients aged 55-64 and diagnosed with diabetes in 2016 had polypharmacy[159]. The same study found that prescriptions of at least one potentially inappropriate medication was common in the middle-aged diabetic population[159]. Mental illnesses like anxiety and depression were common in the two clusters, which -as previous research confirms- are associated with chronic diseases and polypharmacy[158 160].

Although the total mortality rate was <2% in the overall study sample, more than 24% of the intermediate-increasing cluster population died during the study period. While it was beyond the scope of this study to ascertain the cause of death in the study participants, the high prevalence of cancer, which characterised this cluster, is one potential explanation for this cluster high mortality rate, as it is generally a predictor for increased mortality risks[161].

Internal testing of the models

The test set showed good validation in terms of discrimination, polypharmacy trajectories, and mortality risks of the clusters. Moreover, the identified clusters had similar clinical characteristics, including comorbidity burden and prevalence of drug use.

Strengths, limitations and implications

This study comes with many strengths. The first is using joint latent class models to identify the clusters. Despite the low rate of events in this particular population, there was a considerable proportion of participants dying in the intermediate-increasing cluster, indicating that the use of the traditional longitudinal clustering methods would still have led to a survival bias and a loss of information for those at highest risk of polypharmacy, multimorbidity, and eventually death. Although the 3-knots spline model performed better based on BIC, the variance explained by both models was similar, indicating a minor advantage compared to the simpler, more parsimonious model. Thus, the model with a linear link function, once again, had the advantage of a more interpretable and computationally attractive model. Similar to Chapter 6, the risk of negative predictions was avoided, but inaccurate estimates remain a potential limitation. However, the identified clusters had clinically plausible trajectories and associated mortality risks. These clusters can serve as a starting point for building more complex models.

The large number of the extracted study sample allowed for a 50/50 division without the risk of having low numbers of events or cluster proportions/counts. The results of the internal validation should be treated carefully; good internal validation does not necessarily mean good external validation and generalisability of the model. However, as seen in Chapters 5-7, the joint latent class models performed well also externally, for both frailty and polypharmacy modelling.

The choice of 5 years follow-up was feasible, ensured relative stability in medical and public health practices for the study sample, and was sufficient to identify meaningful clusters and predict short-term mortality. The main limitation of this design was the inability to follow the journey of the same patient as they transition from middle age to older age, but rather making inferences from different studies looking into the two age categories separately. While research for an extended period would provide a better understanding and continuity of the patient's journey, it should be designed while considering the change in time trends, clinical guidelines, and the time-varying confounding effects that could arise in an ageing population.

The main implication of this study was that monitoring polypharmacy trajectories over time and at an earlier age is predictive of multimorbidity and mortality risks. In general, chronic polypharmacy is associated with adverse health outcomes[162]. Research has shown that starting interventions, e.g., physical activity, to lower multimorbidity in middle-aged people can result in desirable outcomes at older age[163]. Hence, understanding comorbidity, polypharmacy burden, and their potential adverse health outcomes can be the first step for optimising these interventions.

Conclusion

In this chapter, I demonstrate the use of joint latent class modelling for the study of polypharmacy in middle-aged people. Filling a gap in the literature, this study has shown that monitoring polypharmacy over time is important in middle-aged people as it is associated with increased short- and long-term multimorbidity and mortality risks. External validation of the identified clusters is needed to confirm their stability and generalisability. Moreover, research for extended follow-up periods can provide a better understanding of the patient's journey.

9. Discussion

9.1 Chapter summary

In this chapter, I summarise the main findings emerging from the studies conducted in this thesis.

The studies addressed four research aims. The first study confirmed that there is an overlap between the health markers used to identify people with multimorbidity and complex health needs. The study identified and characterised three cohorts of complex health needs based on frailty, polypharmacy, and unplanned hospital admissions.

The second study used joint latent class models to identify five clusters of older people with different frailty trajectories and associated mortality risks. External validation of these trajectories confirmed their generalisability.

The third study applied the same method to identify trajectories of polypharmacy. It identified four clusters with distinct polypharmacy trajectories and associated mortality profiles. These trajectories were validated in two external databases, proving generalisability and yielding clusters with similar multimorbidity and drug burden.

The last study confirmed that polypharmacy progression can start from middle age. Three clusters with different polypharmacy trajectories and associated mortality risks were identified. The patients from the non-steady clusters are likely to progress and belong to similar trajectory patterns in older age.

I described the clinical implications of my findings; screening and intervention protocols should consider trajectories of health markers instead of a single value. I summarised the strengths and limitations of the research carried out in this work, especially the advantages and disadvantages of using the UK and the Netherlands primary care data, the methods employed

and the generalisability of the results. Lastly, I gave recommendations for future research, including considering different definitions and longer period studies, assessing the effects of the pandemic on the older population, and examining further generalisability of the research findings.

9.2 Summary of main results

The older population is often excluded from clinical trials[5 6]. This population has the highest usage of health resources and medication intake[4 7]. It is therefore necessary to understand the true burden and associated outcomes with the different health interventions. To achieve this, we need to be able to identify and characterise older people with complex health needs correctly. Health markers like frailty and polypharmacy are frequently used as markers for multimorbidity and complex health needs[4 12].

This thesis was an effort to use RWD to identify older people with complex health needs by using these health markers, then monitoring their variation over time and association with death. Moreover, the thesis aimed to understand the history of these health markers' progression during middle age years and whether they can be indicative of later life poor health status. The key finding of this project was that trajectories of health markers over time are better predictors for mortality than single/cross-sectional measures. Furthermore, characterising these trajectories provides a wealth of information that is crucial for better clinical care and preventive actions in people with multimorbidity and complex health needs.

Key results of individual chapters

The literature review in Chapter 2 identified a gap in the methods employed for describing frailty and polypharmacy progression over time. Although there was larger evidence on frailty progression over time compared to polypharmacy, these studies lacked in making all the considerations for longitudinal analysis in older people, like the heterogeneity and high attrition rates. In Chapter 3, I described the RWD I used to conduct my analyses throughout the thesis. I used three primary care databases from two European countries, the UK and the Netherlands. I also described how using the OMOP-CDM can be of value for carrying out network studies including multiple databases in a federated manner, without transferring patient-level data across institutions or countries. This was later demonstrated in Chapter 7.

Chapter 4 summarised the results from a cross-sectional study aimed to identify cohorts of older people with complex health needs based on three health markers: unplanned hospitalisation, frailty and polypharmacy. Although the study showed that a single measure could lead to an oversight of other patients who might be identifiable with other markers, there was still considerable pairwise and overall overlap between the three identified cohorts. The three cohorts were also characterised by high disease prevalence and the use of preventive therapies. The cumulative nature of these health markers necessitates studying their progression over time.

Chapter 5 introduced the first longitudinal study, in which I identified and externally validated five clusters of older people with distinct frailty trajectories and associated mortality risks. The two rapid clusters had the highest mortality risks, followed by the high-slow (high baseline frailty, slow progression over time). This has shown that the trajectory of frailty is a better predictor for mortality than the cross-sectional/baseline measure.

In Chapters 6 and 7, I identified and externally validated four clusters of older people with different polypharmacy trajectories and associated risk of death. The cluster with the fastest accumulation of drug intake had the highest prevalence and incidence of cancer diagnosis and was the most at risk of mortality, with over 80% dying during follow-up. The high-decreasing and intermediate-slow/increasing had the highest baseline and fastest accumulation of chronic comorbidity burden, respectively. They had similar mortality risks, indicating that both polypharmacy baseline and trajectory are needed to identify those at most risk of mortality, for tailoring treatment packages and targeting appropriate preventive intervention measures.

Chapter 8 allowed for understanding the polypharmacy accumulation journey, those who end up with a high polypharmacy burden later in life have likely accumulated it from middle age or even earlier[17 18]. The analysis has shown that around 9% of middle-aged people have high

polypharmacy or start accumulating polypharmacy during that stage. Similar to older polypharmacy groups, middle-aged people with polypharmacy had a higher comorbidity burden, including chronic diseases and cancer diagnoses, and higher associated mortality risks.

Frailty or polypharmacy?

The overlap and intersectionality between frailty and polypharmacy were evident in all studies conducted in this thesis, which agrees with the already established research[4 16 108 109]. The cross-sectional study in Chapter 4 showed that there was considerable overlap between the frailty and polypharmacy cohorts (>50%). Although polypharmacy was defined with a lower cut-off value when included in eFI score (≥5), it was the most prevalent deficit in the overall population included in Chapter 5, with ≥90% of the patients in the high baseline and rapid trajectory clusters having polypharmacy by the end of the study follow-up. Lastly, many of the deficits comprising the eFI score were common in the non-healthy polypharmacy clusters identified in Chapters 6.

As frailty is a cumulative score based on the count of predefined deficits, the emergent clusters included patients with different baseline frailty and trajectories, but with no specific deficits distinctive of the different clusters. On the contrary, with polypharmacy, the different clusters had distinct clinical characteristics and medication burden in terms of prevalence and cumulative incidence. For example, the intermediate-fast/increasing had the highest proportion of cancer diagnoses, while the high-decreasing had the highest baseline prevalence of non-cancer chronic diseases. The eFI score is limited to a set number of individual deficits, whereas a far greater pool of ingredients is available for polypharmacy, allowing for different combinations. Also, a deficit is only counted once in frailty, but several ingredients may be used to treat the same condition, giving more weight to that condition in polypharmacy clusters. Considering these findings, polypharmacy offers more information when describing older

people with complex health needs, as it can be an indicator of both multimorbidity and drug burden.

However, frailty cannot be overlooked, as it is a general state of physical functionality, not necessarily caused by specific deficits but rather the accumulation of such deficits. Rockwood and Mitnitski argued that mortality prediction was not the driving aim for developing a frailty index but instead acted as a validation tool for the concept. If it were the main purpose, then more weight would have been given to specific deficits known for high mortality rates, like cancer[25]. The design of the studies included in this thesis had a similar underlying rationale: to assess whether the accumulation of deficits/drugs is an indicator of health status, regardless of their nature or type. Both frailty and polypharmacy performed well in answering that question. The choice of one over the other for clinical applications or future research depends solely on the aims they want to address.

Polypharmacy: a journey from middle to older age

When assessing the characteristics of the middle-aged clusters at the end of follow-up, it was difficult to expect where patients would be allocated later in the older clusters. Several reasons could have contributed to this. The unequal number of clusters emerging from the two studies means that patients of one middle-aged cluster could be divided into two or more older-age clusters. The study's design did not lead to a sample that would be exactly 65 by the end (the starting age for the older population in Chapter 6). This heterogeneity in age makes it incorrect to exactly match the younger and older clusters based on their characteristics and polypharmacy trends.

Research studying the prevalence of multimorbidity and polypharmacy over long periods (≥10 years) has shown increasing trends in the same age groups[17 72 164]. This was attributed to several reasons, including different practice guidelines, living conditions, and even coding practices.

Nonetheless, it means that characterising older people at one point in time provides a good indication for future cohorts but does not necessarily mean that the younger population will follow the same trends later in the future.

Despite all these reasons, the middle-aged clusters were still informative. Middle-aged participants belonging to either the high-decreasing or intermediate-increasing clusters will likely end up in the three non-low-steady clusters at older age, with more patients joining from the low-steady cluster.

9.3 Implications of the research

The findings in this research have great implications for future identification, characterisation, management and treatment of older people with complex health needs and multimorbidity. They highlighted the value of monitoring frailty and polypharmacy over time, and of applying data-driven approaches for identifying the most vulnerable populations using different health markers.

Longitudinal, cross-sectional, and data-driven analyses

This research's first and main implication is the value of monitoring health markers over time offers over traditional cross-sectional measures. Cross-sectional measures -and studies- have the advantage of offering quick and simple population-level observations. They are useful when identifying certain cohorts or calculating the prevalence of drug use and phenomena like frailty and polypharmacy. However, they fail to convey more complete stories on how these measures develop and progress, and whether they interact with each other. The cross-sectional analysis in Chapter 4 reflected these advantages and shortcomings. While the study identified three cohorts with complex health needs, it did not show how the patients developed frailty/polypharmacy and how that affected their short- and long-term prognosis and life expectancy.

On the other hand, the longitudinal studies in later chapters offered insights into the patients' journeys, how they progressed over time, and what were the implications of these progression patterns. Clusters with high frailty/polypharmacy starting values supported previous findings, which reported associations between high levels of these markers and the risk of adverse events[34 69 75]. However, the emergent fast-increasing clusters (e.g., low-rapid and intermediate-increasing) were the most important groups emerging from this work. These clusters included seemingly healthy/average people at the start of the studies who accumulated deficits/prescriptions within a relatively short period of time. These groups require additional

medical attention so that both frailty and polypharmacy can be managed and reversed. Moreover, the longitudinal method employed allows for dynamic predictions, in which every patient visit is a chance to integrate their data into the model, offering better prognosis predictions. Although dynamic predictions were beyond the scope of this thesis, joint latent class models allow for it[100 101]. This facilitates the implementation of the algorithms developed here into clinical practice.

Applying data-driven approaches to identify patients with frailty or polypharmacy is valuable. Both are known to vary in prevalence over time and across geographical locations[21 65], making the use of uniform cut-off values inaccurate in some cases[60]. For example, it is common practice to determine cut-offs for frailty categories using quartiles/quintile values derived from the population being studied[23 30]. However, frailty levels increase over time, and thus, these cut-off values would need to be updated. The comparison study in 3.4 identified similar cohorts of polypharmacy using a data-driven approach, accommodating the different methods for calculating polypharmacy. Then, the data-driven approach in Chapter 4 successfully identified the participants with the highest frailty, hospitalisation, and polypharmacy levels.

Screening and interventions for patients with complex health needs

The published NICE guidelines recommended employing primary care EHR to identify patients with complex health needs. Specifically, they suggested using screening tools like polypharmacy and eFI. The guidelines also advised tailoring individualised treatment packages, taking into consideration the patient's priorities (e.g., maintaining independence, being able to do voluntary work, etc.). This can be achieved through a comprehensive assessment to identify treatments with potential side effects, treatments that can be stopped or non-pharmacological treatment options[12].

Current practice utilises these screening tools based on a single measurement[12]. The models developed in this research present more useful methods for screening using these tools, monitoring variations over time and identifying distinct subgroups that are deteriorating more rapidly and need tailored intervention and treatment packages more urgently.

Multimorbidity, frailty, and polypharmacy can be managed and even reversed. Frailty management and treatment interventions include physical activities, cognitive training, and nutritional supplements. Any combination of these interventions and their individual implementation can preserve and improve the patient's frailty status[32 38 165 166]. Moreover, optimising prescriptions by identifying inappropriate and redundant prescriptions was found to improve the general condition, health status, and functional level in the targeted patients[167].

The earliest interventions are targeted, the better the results for reversing or stabilising frailty and optimising polypharmacy[4 157]. Frailty interventions offer the greatest value when tailored to those at higher levels[166]. This can be extended to polypharmacy and multimorbidity in general.

Despite the calls for including older people with complex health needs in clinical trials[12 168], many feasibility, safety, and framework precautions need to be considered when designing these trials. It could be unethical to include such a vulnerable population in experimental studies[7]. Also, problems like attrition and non-adherence due to polypharmacy are common in older people[7]. Furthermore, it is difficult to design a trial with a number of arms accounting for all possible drug and/or comorbidities combinations[7 12 19]. For example, we would need 32 trial arms for all the possible combinations of only 5 drugs[19]. Nonetheless, clinical trials are still needed and are being increasingly considered for the excluded population. In the meantime, RWE offers a fast, cost-effective alternative for assessing the safety and effectiveness of new treatments in this population.

When designing RWE studies, it is important to account for confounders like polypharmacy and frailty. The heterogeneity of their progression rates and the different association levels of these rates with the risk of adverse health outcomes among older people can pose certain challenges. The analysis and models developed in this work can help to inform pharmaco-epidemiological safety and effectiveness studies on the time-varying effects of polypharmacy and frailty, thus, helping in the production of robust, reliable evidence.

This was mainly a characterisation study to recognise the value of monitoring longitudinal markers in older people and its ability to screen and identify those at higher risks of adverse events. Considerations for practical implementation were beyond the scope of this work. Nonetheless, the studies conducted have confirmed the ability to identify patterns of frailty and polypharmacy trajectories over time, even from as early as middle age. Future guidelines, pharmaco-epidemiological and interventional studies can build on this work to offer recommendations and advice on the identification of patients with complex health needs and the appropriate treatments for them.

9.4 Strengths and limitations

The studies included in this work came with both strengths and limitations. The first study's use of different definitions allowed for the identification of vulnerable patients who may not have been identified using a single definition. Similarly, for Chapters 5-8, longitudinal analysis identified those seemingly healthy but deteriorating rapidly over a few years.

CPRD GOLD, Aurum, IPCI, and the OMOP-CDM

Using UK and Dutch primary care databases has strengths and limitations. In the UK and the Netherlands, GPs are the first point of contact for patients, and the routinely collected comprehensive information is difficult to find elsewhere in both countries. CPRD GOLD, Aurum, and IPCI were updated frequently, representative of their respective general population, and under regular scrutiny to ensure the data's quality. They also recorded electronic health information about the population over extended periods. Consequently, they were frequently used in pharmaco-epidemiology and longitudinal studies, providing good-quality records for research[87 95 169]. Moreover, the representativeness of these databases enables the use of results from this study to inform local public health strategies for older people.

While primary care records include all prescriptions issued by GPs, they do not include hospitalisation information. When they do, there is usually a lag between the actual event and the record updating in the primary care system, which relies on correspondence with the GP[87]. That, in addition to the absence of "over-the-counter" prescription information, could have led to a potential underestimation of polypharmacy burden. Additionally, the lack of linkages with hospital episode information in the data used for longitudinal analysis limited the possibility of assessing the associations between frailty/polypharmacy change over time and the risk of hospitalisation. On the other hand, the prescription data I used do not include dispensation information, which may lead to an overestimation of polypharmacy in some cases. Lastly,

while I identified patients with multiple prescribed drugs, no information on the adequateness of polypharmacy in the individual patients was available.

In the primary care EHR used, the accuracy of the information depended on the individual systems and coding/data-entry practices, as the studies were making secondary use of the data collected for practice management and not for clinical research, leading to potential missing information. Collecting the data from different GP practices with different data entry systems and procedures can lead to data heterogeneity and may affect the validity of certain condition records.

The use of OMOP-CDM was of great value. It facilitated the use of multiple databases with high levels of transparency and efficiency. It enabled direct comparisons between databases, clinical practices, and population characteristics. Moreover, the presence of verified analytical packages using uniform phenotype to identify prescription and condition information allowed for better clinical characterisations of the clusters identified. Thus, the results produced were robust, reliable, and generalisable.

Joint latent class models, external validity and applicability

Joint latent class modelling was a particular strength in this work. The model allowed for the description of longitudinal patterns of health markers, accounted for the heterogeneity of the populations studied, allocated the participants to different groups accordingly, and provided survival probabilities for the different clusters/groups identified. Although linearity assumption may not have always been met in these models -leading to not very accurate estimates (intercepts and slopes)-the overall patterns (e.g., decreasing, fast-increasing) were clinically plausible. Lastly, despite the apparent complexity of the models, the results are intuitive and simple to grasp, especially since the simpler and more parsimonious models were preferred as a trade-off for the more complex models.

Using multiple databases for training and validating the models ensured the generalisability of the models and patterns identified. It also highlighted the differences between different systems, the coding practice and the characteristics of individual populations, even if within the same country. Furthermore, good model validations in different databases and the yielded clusters having similar health marker patterns, survival probabilities, and clinical characteristics proved the success of the models and the generalisability of the results.

There may be a challenge in implementing the algorithms built in this thesis into actual practice. For example, Chapter 3 describes a difference in polypharmacy calculation between source (in-practice) data and mapped data. Bearing these differences in mind, I tried to quantify these differences in the comparison study I conducted (3.4). Moreover, the patterns identified in the thesis are simple enough for the health provider to monitor throughout the patient's history.

9.5 Future work

The work in this thesis covered many topics, methods and populations. It highlighted the need for further investigation, accounting for more factors and associations. Future studies should also consider longer periods and post-pandemic effects. Lastly, the generalisability of identified clusters and associations should be assessed across different data settings and geographical locations.

Research considering different definitions and interactions

A general measure of polypharmacy as the absolute count of ingredients was sufficient to assess its effects and characterise the multimorbidity population in this thesis. However, research on the individual medication classes, multiple drug regimens, and duration of treatment episodes could provide an even better characterisation of the polypharmacy population[170]. Many studies have defined polypharmacy based on these measures but rarely looked at it with the variation over time perspective.

Optimising treatment in older patients with multimorbidity and polypharmacy is a priority since they are the ones at higher risk of adverse health effects[14]. Inappropriate polypharmacy was beyond the scope of this work. Future work can explore further characterisation of the identified clusters to identify the most prevalent inappropriate medications prescribed.

As highlighted in 9.2 and in existing literature, polypharmacy and frailty are highly correlated with each other, cross-sectional and longitudinal[4 16 108 109 111 171]. Modelling polypharmacy and frailty longitudinally and side by side can produce evidence of how they interact and evolve alongside each other. Studies like the one by Chen et al. confirmed the dynamic association between frailty and polypharmacy over time in different groups (e.g., fit without polypharmacy or severe frailty with excess polypharmacy) and found they have different mortality risk levels, but their groups were based on categorical values[111]. The studies conducted here have shown

the value of monitoring these variations continuously, accounting for the unit change in frailty and polypharmacy. Interventions like polypharmacy reduction to manage frailty can then be more optimally designed.

Longer period studies and trends after COVID-19

The longitudinal studies described frailty and polypharmacy progression and their short-term mortality risk over 4-5 years. They modelled these trajectories in the middle and older age populations, separately. Although both frailty and polypharmacy have shown to be more valuable for short-term prediction[55], investigating them over a longer period would provide a more wholesome picture of their progression over time. Longer period studies can be useful in identifying inflection points in the identified clusters, e.g., steeper frailty trajectory beyond the age of 65[51].

Several factors need to be accounted for when designing and interpreting the results of longer-term studies. They include the changes in population-level trends of these health markers, public policies and clinical guidelines. For example, polypharmacy and frailty levels increased for the same age groups over longer periods[17 164 172]. Higher levels can be attributed to increases in early-onset, non-fatal chronic diseases like hypertension and diabetes. They can also be due to better disease recording, and behavioural and socioeconomic status changes[172].

Considerations need to be made for pandemic effects, which may have altered the dynamic associations, population characteristics, and patient behaviour towards clinical care during the pandemic years. For example, research has shown an increase in frailty after the pandemic. The reasons for this increase were divided into lockdown and virus effects. Lockdown effects included healthcare disruptions, and worse lifestyle, nutrition and physical activity behaviours. Virus effects, on the other hand, were incidence of short- and long-term symptoms after a COVID-19 diagnosis, including inflammations, fatigue and accelerated ageing[173].

At the time of this work, there were not enough data points to assess the progression of health markers after the pandemic. It would be beneficial to build similar models for both the middle-aged and older populations in more recent years to understand whether the same patterns and association levels still exist and whether the post-pandemic population would have the same distribution of the different patterns and associations.

Generalisability and external validity

Despite the efforts made in this work to assess the generalisability and external validity of the clusters identified, there is still room for further external validation analysis.

Frailty trajectories performed well in two UK primary databases, and polypharmacy trajectories in older people were validated in UK and Dutch primary care records. Polypharmacy trajectories in middle-aged people were clinically meaningful and showed good internal validation. However, external validation would further support the generalisability of these clusters. Previous studies have shown great heterogeneity in frailty and polypharmacy prevalence across different countries[21 65]. It is important to assess whether the same trajectories and association levels with mortality exist in different nations, even if proportions varied across the clusters compared to their original distributions.

As discussed in 9.4, the study's use of primary care data was a strength due to their representativeness of the population and the evidence supporting their reliability in pharmaco-epidemiological studies. Using other data sources like hospitalisation and medical claims in future studies can add value as they typically contain more hospital-related information. However, the use of these data should be done with caution, as other problems, such as fraud and incomprehensive information, can pose challenges, and adjustments must be made to address these limitations.

9.6 Conclusions

The research in this thesis successfully identified and characterised older and middle-aged people with complex health needs. It described how different health markers overlap and complement each other to identify older people with multimorbidity and complex health needs. It also showed the value of monitoring markers like frailty and polypharmacy over time, how their progression can be different for subgroups of the population, how they can develop from middle age, and how these trajectories are predictive of mortality risks. The work done here can lay ground for better guidelines and treatment advice for the management of patients with multimorbidity, frailty and polypharmacy. Future work should consider modelling frailty and polypharmacy together, monitoring them over longer periods, and further generalisability of the identified patterns.

Reference

1. Organization WH. Ageing and health [Available from: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health accessed 07/06/2024.
2. ONS. Profile of the older population living in England and Wales in 2021 and changes since 2011.
3. ONS. Living longer: is age 70 the new age 65? 2019 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/ageing/articles/livinglongerisage70thenewage65/2019-11-19 accessed 15/07/2024.
4. Woolford SJ, Aggarwal P, Sheikh CJ, et al. Frailty, multimorbidity and polypharmacy. *Medicine* 2021;49(3):166-72. doi: https://doi.org/10.1016/j.mpmed.2020.12.010
5. Zulman DM, Sussman JB, Chen X, et al. Examining the evidence: a systematic review of the inclusion and analysis of older adults in randomized controlled trials. *Journal of general internal medicine* 2011;26(7):783-90. doi: 10.1007/s11606-010-1629-x [published Online First: 2011/02/03]
6. He J, Morales DR, Guthrie B. Exclusion rates in randomized controlled trials of treatments for physical conditions: a systematic review. *Trials* 2020;21(1):228. doi: 10.1186/s13063-020-4139-0 [published Online First: 2020/02/28]
7. Tan YY, Papez V, Chang WH, et al. Comparing clinical trial population representativeness to real-world populations: an external validity analysis encompassing 43 895 trials and 5 685 738 individuals across 989 unique drugs and 286 conditions in England. *The lancet Healthy longevity* 2022;3(10):e674-e89. doi: 10.1016/s2666-7568(22)00186-6 [published Online First: 2022/09/24]
8. Lau SWJ, Huang Y, Hsieh J, et al. Participation of Older Adults in Clinical Trials for New Drug Applications and Biologics License Applications From 2010 Through 2019. *JAMA network open* 2022;5(10):e2236149. doi: 10.1001/jamanetworkopen.2022.36149 [published Online First: 2022/10/15]
9. GOV.UK. COVID-19 vaccines rolled out to people aged 70 years and over from today 2021 [Available from: https://www.gov.uk/government/news/covid-19-vaccines-rolled-out-to-people-aged-70-years-and-over-from-today accessed 15/07/2024.
10. Wong MK, Brooks DJ, Ikejezie J, et al. COVID-19 Mortality and Progress Toward Vaccinating Older Adults - World Health Organization, Worldwide, 2020-2022. *MMWR Morbidity and mortality weekly report* 2023;72(5):113-18. doi: 10.15585/mmwr.mm7205a1 [published Online First: 2023/02/03]
11. Helfand BKI, Webb M, Gartaganis SL, et al. The Exclusion of Older Persons From Vaccine and Treatment Trials for Coronavirus Disease 2019-Missing the Target. *JAMA internal medicine* 2020;180(11):1546-49. doi: 10.1001/jamainternmed.2020.5084 [published Online First: 2020/09/29]
12. National Institute for Health and Care Excellence. NICE guideline [NG56]: Multimorbidity: clinical assessment and management. 2016 [Available from: https://www.nice.org.uk/guidance/NG56/chapter/Recommendations#taking-account-of-multimorbidity-in-tailoring-the-approach-to-care accessed 03/06/2020.
13. Hughes LD, McMurdo MET, Guthrie B. Guidelines for people not for diseases: the challenges of applying UK clinical guidelines to people with multimorbidity. *Age and Ageing* 2012;42(1):62-69. doi: 10.1093/ageing/afs100 %J Age and Ageing
14. Sinnott C, Bradley CP. Multimorbidity or polypharmacy: two sides of the same coin? *Journal of comorbidity* 2015;5:29-31. doi: 10.15256/joc.2015.5.51 [published Online First: 2015/04/24]
15. Nicholson K, Liu W, Fitzpatrick D, et al. Prevalence of multimorbidity and polypharmacy among adults and older adults: a systematic review. *The Lancet Healthy Longevity* 2024;5(4):e287-e96. doi: https://doi.org/10.1016/S2666-7568(24)00007-2

16. Gutierrez-Valencia M, Izquierdo M, Cesari M, et al. The relationship between frailty and polypharmacy in older people: A systematic review. *British journal of clinical pharmacology* 2018;84(7):1432-44. doi: 10.1111/bcp.13590 [published Online First: 2018/03/27]
17. van den Akker M, Vaes B, Goderis G, et al. Trends in multimorbidity and polypharmacy in the Flemish-Belgian population between 2000 and 2015. *PLoS One* 2019;14(2):e0212046. doi: 10.1371/journal.pone.0212046 [published Online First: 2019/02/13]
18. Vos R, Boesten J, van den Akker M. Fifteen-year trajectories of multimorbidity and polypharmacy in Dutch primary care—A longitudinal analysis of age and sex patterns. *PLOS ONE* 2022;17(2):e0264343. doi: 10.1371/journal.pone.0264343
19. Mahajan R. Real world data: Additional source for making clinical decisions. *International journal of applied & basic medical research* 2015;5(2):82. doi: 10.4103/2229-516x.157148 [published Online First: 2015/06/23]
20. Liu F, Panagiotakos D. Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC medical research methodology* 2022;22(1):287. doi: 10.1186/s12874-022-01768-6 [published Online First: 2022/11/06]
21. Collard RM, Boter H, Schoevers RA, et al. Prevalence of frailty in community-dwelling older persons: a systematic review. *Journal of the American Geriatrics Society* 2012;60(8):1487-92. doi: 10.1111/j.1532-5415.2012.04054.x [published Online First: 2012/08/14]
22. Clegg A, Young J, Iliffe S, et al. Frailty in elderly people. *Lancet (London, England)* 2013;381(9868):752-62. doi: 10.1016/s0140-6736(12)62167-9 [published Online First: 2013/02/12]
23. Fried LP, Tangen CM, Walston J, et al. Frailty in Older Adults: Evidence for a Phenotype. *The Journals of Gerontology: Series A* 2001;56(3):M146-M57. doi: 10.1093/gerona/56.3.M146 %J The Journals of Gerontology: Series A
24. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. *ScientificWorldJournal* 2001;1:323-36. doi: 10.1100/tsw.2001.58
25. Rockwood K, Mitnitski A. Frailty in Relation to the Accumulation of Deficits. *The Journals of Gerontology: Series A* 2007;62(7):722-27. doi: 10.1093/gerona/62.7.722 %J The Journals of Gerontology: Series A
26. Searle SD, Mitnitski A, Gahbauer EA, et al. A standard procedure for creating a frailty index. *BMC Geriatrics* 2008;8 doi: 10.1186/1471-2318-8-24
27. Rockwood K, Rockwood MR, Mitnitski A. Physiological redundancy in older adults in relation to the change with age in the slope of a frailty index. *Journal of the American Geriatrics Society* 2010;58(2):318-23. doi: 10.1111/j.1532-5415.2009.02667.x [published Online First: 2010/04/08]
28. Pasquet R, Xu M, Sylvestre MP, et al. Comparison of three frailty measures for predicting hospitalization and mortality in the Canadian Longitudinal Study on Aging. *Aging clinical and experimental research* 2024;36(1):48. doi: 10.1007/s40520-024-02706-w [published Online First: 2024/02/29]
29. Dent E, Kowal P, Hoogendijk EO. Frailty measurement in research and clinical practice: A review. *European Journal of Internal Medicine* 2016;31:3-10. doi: https://doi.org/10.1016/j.ejim.2016.03.007
30. Clegg A, Bates C, Young J, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age and Ageing* 2016;45(3):353-60. doi: 10.1093/ageing/afw039
31. NHS England (2017). General Medical Services contract: Supporting routine frailty identification and frailty through the GP Contract 2017/2018. [Available from: https://www.england.nhs.uk/publication/supporting-routine-frailty-identification-and-frailty-through-the-gp-contract-20172018/.
32. Hoogendijk EO, Afilalo J, Ensrud KE, et al. Frailty: implications for clinical practice and public health. *The Lancet* 2019;394(10206):1365-75. doi: https://doi.org/10.1016/S0140-6736(19)31786-6

33. Kojima G. Frailty as a predictor of hospitalisation among community-dwelling older people: a systematic review and meta-analysis. 2016;70(7):722-29. doi: 10.1136/jech-2015-206978 %J Journal of Epidemiology and Community Health
34. Shamliyan T, Talley KM, Ramakrishnan R, et al. Association of frailty with survival: a systematic literature review. *Ageing research reviews* 2013;12(2):719-36. doi: 10.1016/j.arr.2012.03.001 [published Online First: 2012/03/20]
35. Kojima G. Frailty as a Predictor of Future Falls Among Community-Dwelling Older People: A Systematic Review and Meta-Analysis. *Journal of the American Medical Directors Association* 2015;16(12):1027-33. doi: 10.1016/j.jamda.2015.06.018 [published Online First: 2015/08/10]
36. Kojima G. Frailty as a Predictor of Nursing Home Placement Among Community-Dwelling Older Adults: A Systematic Review and Meta-analysis. *Journal of geriatric physical therapy (2001)* 2018;41(1):42-48. doi: 10.1519/jpt.0000000000000097 [published Online First: 2016/06/25]
37. Kojima G. Frailty as a predictor of fractures among community-dwelling older people: A systematic review and meta-analysis. *Bone* 2016;90:116-22. doi: 10.1016/j.bone.2016.06.009 [published Online First: 2016/06/21]
38. Dent E, Martin FC, Bergman H, et al. Management of frailty: opportunities, challenges, and future directions. *The Lancet* 2019;394(10206):1376-86. doi: https://doi.org/10.1016/S0140-6736(19)31785-4
39. Hoogendijk EO, Dent E. Trajectories, Transitions, and Trends in Frailty among Older Adults: A Review. *Ann Geriatr Med Res* 2022;26(4):289-95. doi: 10.4235/agmr.22.0148
40. Kojima G, Taniguchi Y, Iliffe S, et al. Transitions between frailty states among community-dwelling older people: A systematic review and meta-analysis. *Ageing research reviews* 2019;50:81-88. doi: https://doi.org/10.1016/j.arr.2019.01.010
41. Welstead M, Jenkins ND, Russ TC, et al. A Systematic Review of Frailty Trajectories: Their Shape and Influencing Factors. *The Gerontologist* 2021;61(8):e463-e75. doi: 10.1093/geront/gnaa061 [published Online First: 2020/06/03]
42. Gill TM, Gahbauer EA, Allore HG, et al. Transitions Between Frailty States Among Community-Living Older Persons. *Archives of Internal Medicine* 2006;166(4):418-23. doi: 10.1001/archinte.166.4.418 %J Archives of Internal Medicine
43. Pollack LR, Litwack-Harrison S, Cawthon PM, et al. Patterns and Predictors of Frailty Transitions in Older Men: The Osteoporotic Fractures in Men Study. *Journal of the American Geriatrics Society* 2017;65(11):2473-79. doi: 10.1111/jgs.15003 [published Online First: 2017/09/05]
44. Trevisan C, Veronese N, Maggi S, et al. Factors Influencing Transitions Between Frailty States in Elderly Adults: The Progetto Veneto Anziani Longitudinal Study. *Journal of the American Geriatrics Society* 2017;65(1):179-84. doi: 10.1111/jgs.14515 [published Online First: 2016/11/20]
45. Hoogendijk EO, Rockwood K, Theou O, et al. Tracking changes in frailty throughout later life: results from a 17-year longitudinal study in the Netherlands. *Age and Ageing* 2018;47(5):727-33. doi: 10.1093/ageing/afy081
46. Marshall A, Nazroo J, Tampubolon G, et al. Cohort differences in the levels and trajectories of frailty among older people in England. *Journal of epidemiology and community health* 2015;69(4):316. doi: 10.1136/jech-2014-204655
47. Stolz E, Mayerl H, Waxenegger A, et al. Impact of socioeconomic position on frailty trajectories in 10 European countries: evidence from the Survey of Health, Ageing and Retirement in Europe (2004-2013). *Journal of epidemiology and community health* 2017;71(1):73-80. doi: 10.1136/jech-2016-207712 [published Online First: 2016/07/17]
48. Jenkins ND, Hoogendijk EO, Armstrong JJ, et al. Trajectories of Frailty With Aging: Coordinated Analysis of Five Longitudinal Studies. *Innovation in aging* 2022;6(2):igab059. doi: 10.1093/geroni/igab059 [published Online First: 2022/03/03]

49. Stow D, Matthews FE, Hanratty B. Frailty trajectories to identify end of life: a longitudinal population-based study. *BMC medicine* 2018;16(1):171. doi: 10.1186/s12916-018-1148-x [published Online First: 2018/09/22]
50. Verghese J, Ayers E, Sathyan S, et al. Trajectories of frailty in aging: Prospective cohort study. *PLOS ONE* 2021;16(7):e0253976. doi: 10.1371/journal.pone.0253976
51. Raymond E, Reynolds CA, Dahl Aslan AK, et al. Drivers of Frailty from Adulthood into Old Age: Results from a 27-Year Longitudinal Population-Based Study in Sweden. *J Gerontol A Biol Sci Med Sci* 2020;75(10):1943-50. doi: 10.1093/gerona/glaa106 [published Online First: 2020/04/30]
52. Stolz E, Mayerl H, Hoogendijk EO, et al. Acceleration of health deficit accumulation in late-life: evidence of terminal decline in frailty index three years before death in the US Health and Retirement Study. *Annals of epidemiology* 2021;58:156-61. doi: 10.1016/j.annepidem.2021.03.008 [published Online First: 2021/04/05]
53. Chamberlain AM, Finney Rutten LJ, Manemann SM, et al. Frailty Trajectories in an Elderly Population-Based Cohort. *Journal of the American Geriatrics Society* 2016;64(2):285-92. doi: 10.1111/jgs.13944 [published Online First: 2016/02/20]
54. Ward RE, Orkaby AR, Dumontier C, et al. Trajectories of Frailty in the 5 Years Prior to Death Among U.S. Veterans Born 1927-1934. *J Gerontol A Biol Sci Med Sci* 2021;76(11):e347-e53. doi: 10.1093/gerona/glab196 [published Online First: 2021/07/11]
55. Stolz E, Mayerl H, Hoogendijk EO. Frailty in the oldest old: is the current level or the rate of change more predictive of mortality? *Age Ageing* 2022;51(2) doi: 10.1093/ageing/afac020 [published Online First: 2022/02/16]
56. Stolz E, Hoogendijk EO, Mayerl H, et al. Frailty Changes Predict Mortality in 4 Longitudinal Studies of Aging. *J Gerontol A Biol Sci Med Sci* 2021;76(9):1619-26. doi: 10.1093/gerona/glaa266 [published Online First: 2020/10/27]
57. Genolini C, Falissard B. KmL: k-means for longitudinal data. *Computational Statistics* 2010;25(2):317-28. doi: 10.1007/s00180-009-0178-4
58. Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data: With Applications in R: CRC Press 2012.
59. Masnoon N, Shakib S, Kalisch-Ellett L, et al. What is polypharmacy? A systematic review of definitions. *BMC Geriatr* 2017;17(1):230. doi: 10.1186/s12877-017-0621-2 [published Online First: 2017/10/12]
60. Payne RA. The epidemiology of polypharmacy. *Clinical medicine (London, England)* 2016;16(5):465-69. doi: 10.7861/clinmedicine.16-5-465 [published Online First: 2016/10/05]
61. Chiapella LC, Montemarani Menna J, Mamprin ME. Assessment of polypharmacy in elderly patients by using data from dispensed medications in community pharmacies: analysis of results by using different methods of estimation. *Int J Clin Pharm* 2018;40(5):987-90. doi: 10.1007/s11096-018-0663-3 [published Online First: 2018/06/09]
62. Fincke BG, Snyder K, Cantillon C, et al. Three complementary definitions of polypharmacy: methods, application and comparison of findings in a large prescription database. *Pharmacoepidemiology and drug safety* 2005;14(2):121-8. doi: 10.1002/pds.966 [published Online First: 2004/09/24]
63. American Geriatrics Society 2023 updated AGS Beers Criteria® for potentially inappropriate medication use in older adults. *Journal of the American Geriatrics Society* 2023;71(7):2052-81. doi: 10.1111/jgs.18372 [published Online First: 2023/05/04]
64. O'Mahony D, Cherubini A, Guiteras AR, et al. STOPP/START criteria for potentially inappropriate prescribing in older people: version 3. *European geriatric medicine* 2023;14(4):625-32. doi: 10.1007/s41999-023-00777-y [published Online First: 2023/05/31]
65. Midão L, Giardini A, Menditto E, et al. Polypharmacy prevalence among older adults based on the survey of health, ageing and retirement in Europe. *Archives of gerontology and geriatrics* 2018;78:213-20. doi: 10.1016/j.archger.2018.06.018 [published Online First: 2018/07/18]

66. Wastesson JW, Morin L, Tan ECK, et al. An update on the clinical consequences of polypharmacy in older adults: a narrative review. *Expert opinion on drug safety* 2018;17(12):1185-96. doi: 10.1080/14740338.2018.1546841 [published Online First: 2018/12/13]
67. Dhalwani NN, Fahami R, Sathanapally H, et al. Association between polypharmacy and falls in older adults: a longitudinal study from England. *BMJ Open* 2017;7(10):e016358-e58. doi: 10.1136/bmjopen-2017-016358
68. Richardson K, Ananou A, Lafortune L, et al. Variation Over Time in the Association between Polypharmacy and Mortality in the Older Population. *Drugs & aging* 2011;28(7):547-60. doi: 10.2165/11592000-000000000-00000
69. Huang YT, Steptoe A, Wei L, et al. The impact of high-risk medications on mortality risk among older adults with polypharmacy: evidence from the English Longitudinal Study of Ageing. *BMC medicine* 2021;19(1):321. doi: 10.1186/s12916-021-02192-1 [published Online First: 2021/12/17]
70. Zaninotto P, Huang YT, Di Gessa G, et al. Polypharmacy is a risk factor for hospital admission due to a fall: evidence from the English Longitudinal Study of Ageing. *BMC public health* 2020;20(1):1804. doi: 10.1186/s12889-020-09920-x [published Online First: 2020/11/28]
71. Arai S, Ishikawa T, Kato H, et al. Multidrug use positively correlates with high-risk prescriptions in the Japanese elderly: a longitudinal study. *Journal of pharmaceutical health care and sciences* 2019;5:20. doi: 10.1186/s40780-019-0150-6 [published Online First: 2019/09/10]
72. Guthrie B, Makubate B, Hernandez-Santiago V, et al. The rising tide of polypharmacy and drug-drug interactions: population database analysis 1995-2010. *BMC medicine* 2015;13:74. doi: 10.1186/s12916-015-0322-7 [published Online First: 2015/04/19]
73. Pazan F, Wehling M. Polypharmacy in older adults: a narrative review of definitions, epidemiology and consequences. *European geriatric medicine* 2021 doi: 10.1007/s41999-021-00479-3 [published Online First: 2021/03/12]
74. Rodrigues MC, Oliveira C. Drug-drug interactions and adverse drug reactions in polypharmacy among older adults: an integrative review. *Revista latino-americana de enfermagem* 2016;24:e2800. doi: 10.1590/1518-8345.1316.2800 [published Online First: 2016/09/07]
75. Leelakanok N, Holcombe AL, Lund BC, et al. Association between polypharmacy and death: A systematic review and meta-analysis. *Journal of the American Pharmacists Association : JAPhA* 2017;57(6):729-38.e10. doi: 10.1016/j.japh.2017.06.002 [published Online First: 2017/08/09]
76. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical research ed)* 2021;372:n71. doi: 10.1136/bmj.n71 [published Online First: 2021/03/31]
77. Morin L, Vetrano DL, Rizzuto D, et al. Choosing Wisely? Measuring the Burden of Medications in Older Adults near the End of Life: Nationwide, Longitudinal Cohort Study. 2017;130(8):927-36.e9. doi: doi:
78. Denholm R, Morris R, Purdy S, et al. Impact of emergency hospital admissions on patterns of primary care prescribing: a retrospective cohort analysis of electronic records in England. 2020;70(695):e399-e405. doi: doi:
79. Kilpatrick K, Tchouaket E, Jabbour M, et al. A mixed methods quality improvement study to implement nurse practitioner roles and improve care for residents in long-term care facilities. 2020;19(1) doi: doi:
80. Veehof LJG, Stewart RE, Haaijer-Ruskamp FM, et al. The development of polypharmacy. A longitudinal study. 2000;17(3):261-67. doi: doi:
81. Agogo GO, Ramsey CM, Gnjidic D, et al. Longitudinal associations between different dementia diagnoses and medication use jointly accounting for dropout. 2018;30(10):1477-87. doi: doi:
82. Denholm R, Morris R, Payne R. Polypharmacy patterns in the last year of life in patients with dementia. 2019;75(11):1583-91. doi: doi:

83. Wastesson JW, Oksuzyan A, von Bornemann Hjelmborg J, et al. Changes in Drug Use and Polypharmacy After the Age of 90: A Longitudinal Study of the Danish 1905 Cohort. 2017;65(1):160-64. doi: doi:
84. Wastesson JW, Rasmussen L, Oksuzyan A, et al. Drug use among complete responders, partial responders and non-responders in a longitudinal survey of nonagenarians: analysis of prescription register data. 2017;26(2):152-61. doi: doi:
85. Lavikainen P, Leskinen E, Hartikainen S, et al. Impact of missing data mechanism on the estimate of change: A case study on cognitive function and polypharmacy among older persons. 2015;7:169-80. doi: doi:
86. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *International Journal of Epidemiology* 2015;44(3):827-36. doi: 10.1093/ije/dyv098
87. Wolf A, Dedman D, Campbell J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *International Journal of Epidemiology* 2019;48(6):1740-40g. doi: 10.1093/ije/dyz034 %J International Journal of Epidemiology
88. CPRD. Release Notes: CPRD GOLD September 2019 2019 [Available from: https://cprdcw.cprd.com/_docs/Release_Notes_September2019.pdf accessed 13 September 2019.
89. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015;44(3):827-36. doi: 10.1093/ije/dyv098 [published Online First: 2015/06/08]
90. Code of Practice for Statistics [Available from: https://code.statisticsauthority.gov.uk/ accessed 16/10/2024.
91. ONS. User guide to mortality statistics [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/methodologies/userguidetomortalitystatisticsjuly2017#:~:text=Detailed%20annual%20mortality%20statistics%20are%20available%20in,(MSOAs)%20*%20age%2Dstandardised%20and%20age%2Dspecific%20mortality%20rates accessed 16/10/2024.
92. Delmestri A, Prieto-Alhambra D. CPRD GOLD and linked ONS mortality records: Reconciling guidelines. *Int J Med Inform* 2020;136:104038. doi: 10.1016/j.ijmedinf.2019.104038 [published Online First: 2020/02/23]
93. ONS. Impact of registration delays on mortality statistics in England and Wales: 2022 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/impactofregistrationdelaysonmortalitystatisticsinenglandandwales/2022 accessed 16/10/2024.
94. ONS. Impact of registration delays on mortality statistics in England and Wales: 2021 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/impactofregistrationdelaysonmortalitystatisticsinenglandandwales/2021#registration-delays-in-england-and-wales accessed 16/10/2024.
95. de Ridder MAJ, de Wilde M, de Ben C, et al. Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands. *International Journal of Epidemiology* 2022;51(6):e314-e23. doi: 10.1093/ije/dyac026 %J International Journal of Epidemiology
96. OHDSI. The Book of OHDSI: OHDSI, 2019.
97. INFORMATICS OHDSA. Our Journey - Where the OHDSI Community Has Been And Where We Are Going [2022:[Available from: https://www.ohdsi.org/wp-content/uploads/2022/10/OHDSI-OurJourney-2022.pdf accessed 06/07/2024.
98. Clegg A, Bates C, Young J, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age Ageing* 2016;45(3):353-60. doi: 10.1093/ageing/afw039 [published Online First: 2016/03/06]

99. Gallagher AM, Dedman D, Padmanabhan S, et al. The accuracy of date of death recording in the Clinical Practice Research Datalink GOLD database in England compared with the Office for National Statistics death registrations. *Pharmacoepidemiology and drug safety* 2019;28(5):563-69. doi: 10.1002/pds.4747 [published Online First: 2019/03/26]
100. Proust-Lima C, Séne M, Taylor JM, et al. Joint latent class models for longitudinal and time-to-event data: a review. *Statistical methods in medical research* 2014;23(1):74-90. doi: 10.1177/0962280212445839 [published Online First: 2012/04/21]
101. Kyheng M, Babykina G, Ternynck C, et al. Joint latent class model: Simulation study of model properties and application to amyotrophic lateral sclerosis disease. *BMC medical research methodology* 2021;21(1):198. doi: 10.1186/s12874-021-01377-9 [published Online First: 2021/10/02]
102. Proust-Lima C, Philipps V, Liquet B. Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm. *Journal of Statistical Software* 2017;78(2):1 - 56. doi: 10.18637/jss.v078.i02
103. Nagin D, Odgers C. Group-Based Trajectory Modeling in Clinical Research. *Annual review of clinical psychology* 2010;6:109-38. doi: 10.1146/annurev.clinpsy.121208.131413
104. Barnett K, Mercer SW, Norbury M, et al. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *The Lancet* 2012;380(9836):37-43. doi: 10.1016/S0140-6736(12)60240-2
105. MacRae C, McMinn M, Mercer SW, et al. The impact of varying the number and selection of conditions on estimated multimorbidity prevalence: A cross-sectional study using a large, primary care population dataset. *PLoS Med* 2023;20(4):e1004208. doi: 10.1371/journal.pmed.1004208 [published Online First: 2023/04/05]
106. Brilleman SL, Salisbury C. Comparing measures of multimorbidity to predict outcomes in primary care: a cross sectional study. *Fam Pract* 2013;30(2):172-8. doi: 10.1093/fampra/cms060 [published Online First: 2012/10/10]
107. Gnjidic D, Hilmer SN, Blyth FM, et al. High-risk prescribing and incidence of frailty among older community-dwelling men. *Clin Pharmacol Ther* 2012;91(3):521-8. doi: 10.1038/clpt.2011.258 [published Online First: 2012/02/03]
108. Herr M, Robine J-M, Pinot J, et al. Polypharmacy and frailty: prevalence, relationship, and impact on mortality in a French sample of 2350 old people. *Pharmacoepidemiology and drug safety* 2015;24:637–46. doi: 10.1002/pds
109. Saum KU, Schottker B, Meid AD, et al. Is Polypharmacy Associated with Frailty in Older People? Results From the ESTHER Cohort Study. *Journal of the American Geriatrics Society* 2017;65(2):e27-e32. doi: 10.1111/jgs.14718 [published Online First: 2016/12/27]
110. Crocker TF, Brown L, Clegg A, et al. Quality of life is substantially worse for community-dwelling older people living with frailty: systematic review and meta-analysis. *Qual Life Res* 2019;28(8):2041-56. doi: 10.1007/s11136-019-02149-1 [published Online First: 2019/03/16]
111. Chen Y-Z, Huang S-T, Wen Y-W, et al. Combined Effects of Frailty and Polypharmacy on Health Outcomes in Older Adults: Frailty Outweighs Polypharmacy *Journal of the American Medical Directors Association* 2020
112. Kim DH, Schneeweiss S. Measuring frailty using claims data for pharmacoepidemiologic studies of mortality in older adults: evidence and recommendations. *Pharmacoepidemiology and drug safety* 2014;23(9):891-901. doi: 10.1002/pds.3674 [published Online First: 2014/06/26]
113. Davidoff AJ, Zuckerman IH, Pandya N, et al. A novel approach to improve health status measurement in observational claims-based studies of cancer treatment and outcomes. *J Geriatr Oncol* 2013;4(2):157-65. doi: 10.1016/j.jgo.2012.12.005 [published Online First: 2013/06/25]
114. Faurot KR, Jonsson Funk M, Pate V, et al. Using claims data to predict dependency in activities of daily living as a proxy for frailty. *Pharmacoepidemiology and drug safety* 2015;24(1):59-66. doi: 10.1002/pds.3719 [published Online First: 2014/10/23]

115. Segal JB, Chang HY, Du Y, et al. Development of a Claims-based Frailty Indicator Anchored to a Well-established Frailty Phenotype. *Med Care* 2017;55(7):716-22. doi: 10.1097/MLR.0000000000000729 [published Online First: 2017/04/25]
116. Kim DH, Schneeweiss S, Glynn RJ, et al. Measuring Frailty in Medicare Data: Development and Validation of a Claims-Based Frailty Index. *J Gerontol A Biol Sci Med Sci* 2018;73(7):980-87. doi: 10.1093/gerona/glx229 [published Online First: 2017/12/16]
117. NHS England. Supporting routine frailty identification and frailty through the GP Contract 2017/2018 2019 [Available from: https://www.england.nhs.uk/publication/supporting-routine-frailty-identification-and-frailty-through-the-gp-contract-20172018/ accessed 19 Sept 2021 2021.
118. Han L, Clegg A, Doran T, et al. The impact of frailty on healthcare resource use: a longitudinal analysis using the Clinical Practice Research Datalink in England. *Age Ageing* 2019;48(5):665-71. doi: 10.1093/ageing/afz088 [published Online First: 2019/07/13]
119. Ho IS-S, Azcoaga-Lorenzo A, Akbari A, et al. Examining variation in the measurement of multimorbidity in research: a systematic review of 566 studies. *The Lancet Public Health* 2021 doi: 10.1016/s2468-2667(21)00107-9
120. O'Keeffe AG, Nazareth I, Petersen I. Time trends in the prescription of statins for the primary prevention of cardiovascular disease in the United Kingdom: a cohort study using The Health Improvement Network primary care data. *Clinical epidemiology* 2016;8:123-32. doi: 10.2147/clep.S104258 [published Online First: 2016/06/18]
121. Geleedst-De Vooght M, Maitland-van der Zee A, Schalekamp T, et al. Statin Prescribing in the Elderly in the Netherlands: A Pharmacy Database Time Trend Study. *Drugs & aging* 2010;27(7):589-96.
122. Ruokoniemi P, Helin-Salmivaara A, Klaukka T, et al. Shift of statin use towards the elderly in 1995-2005: a nation-wide register study in Finland. *British journal of clinical pharmacology* 2008;66(3):405-10. doi: 10.1111/j.1365-2125.2008.03258.x [published Online First: 2008/09/11]
123. van der Velde RY, Wyers CE, Teesselink E, et al. Trends in oral anti-osteoporosis drug prescription in the United Kingdom between 1990 and 2012: Variation by age, sex, geographic location and ethnicity. *Bone* 2017;94:50-55. doi: 10.1016/j.bone.2016.10.013 [published Online First: 2016/11/03]
124. MacDonald TM, Morant SV, Mozaffari E. Treatment patterns of hypertension and dyslipidaemia in hypertensive patients at higher and lower risk of cardiovascular disease in primary care in the United Kingdom. *J Hum Hypertens* 2007;21(12):925-33. doi: 10.1038/sj.jhh.1002249 [published Online First: 2007/07/06]
125. Dedeyne L, Deschodt M, Verschueren S, et al. Effects of multi-domain interventions in (pre)frail elderly on frailty, functional, and cognitive status: a systematic review. *Clinical interventions in aging* 2017;12:873-96. doi: 10.2147/cia.S130794 [published Online First: 2017/06/06]
126. Stow D, Matthews FE, Barclay S, et al. Evaluating frailty scores to predict mortality in older adults using data from population based electronic health records: case control study. *Age Ageing* 2018;47(4):564-69. doi: 10.1093/ageing/afy022 [published Online First: 2018/03/17]
127. Mitnitski A, Song X, Rockwood K. Trajectories of changes over twelve years in the health status of Canadians from late middle age. *Experimental Gerontology* 2012;47(12):893-99. doi: https://doi.org/10.1016/j.exger.2012.06.015
128. Hollinghurst J, Fry R, Akbari A, et al. External validation of the electronic Frailty Index using the population of Wales within the Secure Anonymised Information Linkage Databank. *Age and Ageing* 2019;48(6):922-26. doi: 10.1093/ageing/afz110
129. Romero-Ortuno R, Hartley P, Knight SP, et al. Frailty index transitions over eight years were frequent in The Irish Longitudinal Study on Ageing. *HRB open research* 2021;4:63. doi: 10.12688/hrbopenres.13286.1 [published Online First: 2021/09/16]

130. Chen EY, Bell JS, Ilomaki J, et al. Medication Regimen Complexity In 8 Australian Residential Aged Care Facilities: Impact Of Age, Length Of Stay, Comorbidity, Frailty, And Dependence In Activities Of Daily Living. *Clinical interventions in aging* 2019;14:1783-95. doi: 10.2147/cia.S216705 [published Online First: 2019/11/07]
131. Turk A, Wong G, Mahtani KR, et al. Optimising a person-centred approach to stopping medicines in older people with multimorbidity and polypharmacy using the DExTruS framework: a realist review. *BMC medicine* 2022;20(1):297. doi: 10.1186/s12916-022-02475-1 [published Online First: 2022/08/31]
132. Eriksen CU, Kyriakidis S, Christensen LD, et al. Medication-related experiences of patients with polypharmacy: a systematic review of qualitative studies. *BMJ Open* 2020;10(9):e036158. doi: 10.1136/bmjopen-2019-036158 [published Online First: 2020/09/09]
133. Burgess CL, Holman CD, Satti AG. Adverse drug reactions in older Australians, 1981-2002. *The Medical journal of Australia* 2005;182(6):267-70. doi: 10.5694/j.1326-5377.2005.tb06698.x [published Online First: 2005/03/22]
134. Milton JC, Hill-Smith I, Jackson SHD. Prescribing for older people. 2008;336(7644):606-09. doi: 10.1136/bmj.39503.424653.80 %J BMJ
135. Palmer K, Villani ER, Vetrano DL, et al. Association of polypharmacy and hyperpolypharmacy with frailty states: a systematic review and meta-analysis. *European geriatric medicine* 2019;10(1):9-36. doi: 10.1007/s41999-018-0124-5 [published Online First: 2019/02/01]
136. Villén N, Guisado-Clavero M, Fernández-Bertolín S, et al. Multimorbidity patterns, polypharmacy and their association with liver and kidney abnormalities in people over 65 years of age: a longitudinal study. *BMC Geriatrics* 2020;20(1):206. doi: 10.1186/s12877-020-01580-1
137. Stensrud MJ, Hernán MA. Why Test for Proportional Hazards? *JAMA* 2020;323(14):1401-02. doi: 10.1001/jama.2020.1267 %J JAMA
138. Holvast F, van Hattem BA, Sinnige J, et al. Late-life depression and the association with multimorbidity and polypharmacy: a cross-sectional study. *Family Practice* 2017;34(5):539-45. doi: 10.1093/fampra/cmx018 %J Family Practice
139. Magnuson A, Sattar S, Nightingale G, et al. A Practical Guide to Geriatric Syndromes in Older Adults With Cancer: A Focus on Falls, Cognition, Polypharmacy, and Depression. 2019(39):e96-e109. doi: 10.1200/edbk_237641
140. Nightingale G, Hajjar E, Swartz K, et al. Evaluation of a pharmacist-led medication assessment used to identify prevalence of and associations with polypharmacy and potentially inappropriate medication use among ambulatory senior adults with cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2015;33(13):1453-9. doi: 10.1200/jco.2014.58.7550 [published Online First: 2015/03/25]
141. Bandidwattanawong C, Rattanaserikulchai P, Jetsadavanit N. Polypharmacy and potentially-inappropriate medications are prevalent in the elderly cancer patients receiving systemic cancer therapy and they co-relate with adverse outcomes. *BMC Geriatr* 2023;23(1):775. doi: 10.1186/s12877-023-04471-3 [published Online First: 2023/11/28]
142. Schmittdiel JA, Raebel MA, Dyer W, et al. Prescription medication burden in patients with newly diagnosed diabetes: A SUrveillance, PREvention, and ManagEment of Diabetes Mellitus (SUPREME-DM) study. *Journal of the American Pharmacists Association* 2014;54(4):374-82. doi: https://doi.org/10.1331/JAPhA.2014.13195
143. Excellence NIfHaC. NICE guideline [NG136]. Hypertension in adults: diagnosis and management 2019 [Available from: https://www.nice.org.uk/guidance/ng136/chapter/Recommendations#choosing-antihypertensive-drug-treatment-for-people-with-or-without-type-2-diabetes.
144. Excellence NIfHaC. NICE guideline [NG28]. Type 2 diabetes in adults: management 2015 [Available from: https://www.nice.org.uk/guidance/ng28 accessed 07/06/2024.
145. Reeve J, Maden M, Hill R, et al. Deprescribing medicines in older people living with multimorbidity and polypharmacy: the TAILOR evidence synthesis. *Health technology assessment*

(Winchester, England) 2022;26(32):1-148. doi: 10.3310/aafo2475 [published Online First: 2022/07/28]
146. Wang X, Yang C, Jiang J, et al. Polypharmacy, chronic kidney disease, and mortality among older adults: A prospective study of National Health and nutrition examination survey, 1999-2018. *Frontiers in public health* 2023;11:1116583. doi: 10.3389/fpubh.2023.1116583 [published Online First: 2023/04/11]
147. Mangoni AA, Jackson SH. Age-related changes in pharmacokinetics and pharmacodynamics: basic principles and practical applications. *British journal of clinical pharmacology* 2004;57(1):6-14. doi: 10.1046/j.1365-2125.2003.02007.x [published Online First: 2003/12/18]
148. Nightingale G, Schwartz R, Kachur E, et al. Clinical pharmacology of oncology agents in older adults: A comprehensive review of how chronologic and functional age can influence treatment-related effects. *J Geriatr Oncol* 2019;10(1):4-30. doi: 10.1016/j.jgo.2018.06.008 [published Online First: 2018/07/19]
149. Barlow A, Prusak ES, Barlow B, et al. Interventions to reduce polypharmacy and optimize medication use in older adults with cancer. *J Geriatr Oncol* 2021;12(6):863-71. doi: 10.1016/j.jgo.2020.12.007 [published Online First: 2020/12/24]
150. Lash TL, Collin LJ, Van Dyke ME. The replication crisis in epidemiology: snowball, snow job, or winter solstice? *Current epidemiology reports* 2018;5(2):175-83. doi: 10.1007/s40471-018-0148-x [published Online First: 2018/06/01]
151. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature* 2016;533(7604):452-4. doi: 10.1038/533452a [published Online First: 2016/05/27]
152. Mathur MB, Fox MP. Toward Open and Reproducible Epidemiology. *American journal of epidemiology* 2023;192(4):658-64. doi: 10.1093/aje/kwad007 [published Online First: 2023/01/11]
153. Slater N, White S, Venables R, et al. Factors associated with polypharmacy in primary care: a cross-sectional analysis of data from The English Longitudinal Study of Ageing (ELSA). *BMJ Open* 2018;8(3):e020270. doi: 10.1136/bmjopen-2017-020270 [published Online First: 2018/03/16]
154. Guillot J, Rentsch CT, Gordon KS, et al. Potentially inappropriate medication use by level of polypharmacy among US Veterans 49-64 and 65-70 years old. *Pharmacoepidemiology and drug safety* 2022;31(10):1056-74. doi: 10.1002/pds.5506 [published Online First: 2022/07/04]
155. Richardson K, Bennett K, Kenny RA. Polypharmacy including falls risk-increasing medications and subsequent falls in community-dwelling middle-aged and older adults. *Age Ageing* 2015;44(1):90-6. doi: 10.1093/ageing/afu141 [published Online First: 2014/10/15]
156. Katzke VA, Bajracharya R, Nasser MI, et al. Number of medically prescribed pharmaceutical agents as predictor of mortality risk: a longitudinal, time-variable analysis in the EPIC-Heidelberg cohort. *Scientific reports* 2024;14(1):106. doi: 10.1038/s41598-023-50487-5 [published Online First: 2024/01/04]
157. Mitnitski A, Song X, Rockwood K. Improvement and decline in health status from late middle age: modeling age-related changes in deficit accumulation. *Exp Gerontol* 2007;42(11):1109-15. doi: 10.1016/j.exger.2007.08.002 [published Online First: 2007/09/15]
158. Roberts ER, Green D, Kadam UT. Chronic condition comorbidity and multidrug therapy in general practice populations: a cross-sectional linkage study. *BMJ Open* 2014;4(7):e005429. doi: 10.1136/bmjopen-2014-005429 [published Online First: 2014/07/13]
159. Oktora MP, Alfian SD, Bos HJ, et al. Trends in polypharmacy and potentially inappropriate medication (PIM) in older and middle-aged people treated for diabetes. *British journal of clinical pharmacology* 2021;87(7):2807-17. doi: 10.1111/bcp.14685 [published Online First: 2020/12/04]
160. Read JR, Sharpe L, Modini M, et al. Multimorbidity and depression: A systematic review and meta-analysis. *Journal of affective disorders* 2017;221:36-46. doi: 10.1016/j.jad.2017.06.009 [published Online First: 2017/06/20]

161. Mattiuzzi C, Lippi G. Current Cancer Epidemiology. *Journal of epidemiology and global health* 2019;9(4):217-22. doi: 10.2991/jegh.k.191008.001 [published Online First: 2019/12/20]
162. Franchi C, Marcucci M, Mannucci PM, et al. Changes in clinical outcomes for community-dwelling older people exposed to incident chronic polypharmacy: a comparison between 2001 and 2009. *Pharmacoepidemiology and drug safety* 2016;25(2):204-11. doi: 10.1002/pds.3938 [published Online First: 2015/12/22]
163. Lin YH, Chen YC, Tseng YC, et al. Physical activity and successful aging among middle-aged and older adults: a systematic review and meta-analysis of cohort studies. *Aging* 2020;12(9):7704-16. doi: 10.18632/aging.103057 [published Online First: 2020/05/01]
164. Maxwell CJ, Mondor L, Pefoyo Koné AJ, et al. Sex differences in multimorbidity and polypharmacy trends: A repeated cross-sectional study of older adults in Ontario, Canada. *PLoS One* 2021;16(4):e0250567. doi: 10.1371/journal.pone.0250567 [published Online First: 2021/04/27]
165. Walston J, Buta B, Xue QL. Frailty Screening and Interventions: Considerations for Clinical Practice. *Clinics in geriatric medicine* 2018;34(1):25-38. doi: 10.1016/j.cger.2017.09.004 [published Online First: 2017/11/14]
166. Hubbard RE, Fallah N, Searle SD, et al. Impact of exercise in community-dwelling older adults. *PLoS One* 2009;4(7):e6174. doi: 10.1371/journal.pone.0006174 [published Online First: 2009/07/09]
167. Olesen AE, Vaever TJ, Simonsen M, et al. Deprescribing in primary care without deterioration of health-related outcomes: A real-life, quality improvement project. *Basic & clinical pharmacology & toxicology* 2024;134(1):72-82. doi: 10.1111/bcpt.13925 [published Online First: 2023/07/04]
168. Shenoy P, Harugeri A. Elderly patients' participation in clinical trials. *Perspect Clin Res* 2015;6(4):184-89. doi: 10.4103/2229-3485.167099
169. CPRD. Bibliography [Available from: https://cprd.com/bibliography accessed 09/11/2020.
170. Kadam UT, Roberts I, White S, et al. Conceptualizing multiple drug use in patients with comorbidity and multimorbidity: proposal for standard definitions beyond the term polypharmacy. *Journal of clinical epidemiology* 2019;106:98-107. doi: 10.1016/j.jclinepi.2018.10.014 [published Online First: 2018/11/06]
171. Shmuel S, Lund JL, Alvarez C, et al. Polypharmacy and Incident Frailty in a Longitudinal Community-Based Cohort Study. 2019;67(12):2482-89. doi: https://doi.org/10.1111/jgs.16212
172. Mousa A, Savva GM, Mitnitski A, et al. Is frailty a stable predictor of mortality across time? Evidence from the Cognitive Function and Ageing Studies. *Age Ageing* 2018;47(5):721-27. doi: 10.1093/ageing/afy077 [published Online First: 2018/06/16]
173. König M, Gollasch M, Komleva Y. Frailty after COVID-19: The wave after? 2023;6(3):307-16. doi: https://doi.org/10.1002/agm2.12258

Appendix

2. Literature review

Search terms

(“polypharmacy” OR “multiple medication*” OR “multiple medicine*” OR “multiple drug*” OR “Polypharmacy [Mesh]” OR “many medication*” OR “definition of polypharmacy” OR “prevalence of polypharmacy” or “epidemiology of polypharmacy” OR “consequences of polypharmacy” OR “outcomes of polypharmacy”)

AND (“longitudinal” OR “trajector*” OR “over time” OR “over-time” OR “long term” OR “long-term” OR “variation*” OR “trend*” OR “cluster*” OR “chronic” OR “slope*” OR “chang*” OR “growth” OR “progression” OR “pattern*” OR “variabilit*”)

Search Strategies

PubMed

(((((("Polypharmacy"[Mesh]) OR (polypharmac*[Title/Abstract] OR polymedicat*[Title/Abstract] OR "multiple medication*"[Title/Abstract] OR "multiple medicines"[Title/Abstract] OR "multiple drug*"[Title/Abstract] OR "many medication*"[Title/Abstract])) OR (poly-pharmacy[Title/Abstract])) OR (poly-medicat*[Title/Abstract])) AND (("Longitudinal Studies"[Mesh]) OR (longitudinal*[Title/Abstract] OR trajector*[Title/Abstract] OR "over time"[Title/Abstract] OR "long term"[Title/Abstract] OR variation*[Title/Abstract] OR trend*[Title/Abstract] OR cluster*[Title/Abstract] OR chronic[Title/Abstract] OR slope*[Title/Abstract] OR chang*[Title/Abstract] OR growth[Title/Abstract] OR progression[Title/Abstract] OR pattern*[Title/Abstract] OR variabilit*[Title/Abstract]))) NOT ("systematic review"[Title])

Filters applied: English.

Database: Embase 1974 to present

1 *polypharmacy/ (4737)

2 (polypharmac* or polymedicat* or "multiple medication*" or "multiple medicine*" or "multiple drug*" or "many medication*" or poly-pharmacy or poly-medicat* or "multi-drug therap*" or "multidrug therap*" or "multiple pharmacotherap*").ti,ab. (34103)

3 polypragmasia.ti,ab. (75)

4 ("multi-drug treatment*" or "multidrug treatment*").ti,ab. (489)

5 1 or 2 or 3 or 4 (35407)

6 exp *longitudinal study/ (8693)

7 (longitudinal* or trajector* or "over time" or "long term" or variation* or trend* or cluster* or chronic or slope* or chang* or growth or progression or pattern* or variabilit*).ti,ab. (11079970)

8 6 or 7 (11080257)

9 5 and 8 (15630)

10 (systematic* adj2 review*).ti. (196336)

11 9 not 10 (15430)

12 11 (15430)

13 limit 12 to english language (14429)

Cochrane Central Register of Controlled Trials

Issue 7 of 12, July 2021

#1 MeSH descriptor: [Polypharmacy] explode all trees 213

#2 (polypharmac* or polymedicat* or "multiple medication*" or "multiple medicine*" or "multiple drug*" or "many medication*" or poly-pharmacy or poly-medicat* or "multi-drug therap*" or "multidrug therap*" or "multiple pharmacotherap*"):ti,ab,kw 2994

#3 polypragmasia:ti,ab,kw 2

#4 ("multi-drug treatment*" or "multidrug treatment*"):ti,ab,kw 31

#5 #1 or #2 or #3 or #4 3028

#6 MeSH descriptor: [Longitudinal Studies] explode all trees 6263

#7 (longitudinal* or trajector* or "over time" or "long term" or variation* or trend* or cluster* or chronic or slope* or chang* or growth or progression or pattern* or variabilit*):ti,ab,kw 679238

#8 #6 or #7 679238

#9 #5 and #8 1586

Web of Science Core Collection (Editions: All)

#1 polypharmac* or polymedicat* or "multiple medication*" or "multiple medicine*" or "multiple drug*" or "many medication*" or poly-pharmacy or poly-medicat* or "multi-drug therap*" or "multidrug therap*" or "multiple pharmacotherap*" or polypragmasia or "multi-drug treatment*" or "multidrug treatment*" (Topic) and longitudinal* or trajector* or "over time" or "long term" or variation* or trend* or cluster* or chronic or slope* or chang* or growth or progression or pattern* or variabilit* (Topic)

#2 systematic* near/2 review* (Title)

#3 #1 NOT #2

#4 #1 NOT #2 and English (Languages)

3. Data and methods

Table S 3.1 eFI deficits

Activity limitation	Ischaemic heart disease
Anaemia and haematinic deficiency	Memory and cognitive problems
Arthritis	Mobility and transfer problems
Atrial fibrillation	Osteoporosis
Cerebrovascular disease	Parkinsonism and tremor
Chronic kidney disease	Peptic ulcer
Diabetes	Peripheral vascular disease
Dizziness	Polypharmacy
Dyspnoea	Requirement for care
Falls	Respiratory disease
Foot problems	Skin ulcer
Fragility fracture	Sleep disturbance
Hearing impairment	Social vulnerability
Heart failure	Thyroid disease
Heart valve disease	Urinary incontinence
Housebound	Urinary system disease
Hypertension	Visual impairment
Hypotension/syncope	Weight loss and anorexia

4. Characterising complex health needs and the use of preventive therapies in older people

Table S 4.1 Point prevalence of the different antihypertensives classes for the Hospitalisation cohort (N=90597)

Drug name	One-month PP		Three-months PP		One-year PP	
	N prevalent	PP (95% CI)	N prevalent	PP (95% CI)	N prevalent	PP (95% CI)
ACE-I + CCB	11	0 (0, 0)	11	0 (0, 0)	15	0 (0, 0)
ACE-I + Diuretics	281	0.3 (0.3, 0.3)	300	0.3 (0.3, 0.4)	388	0.4 (0.4, 0.5)
ARB + CCB	24	0 (0, 0)	25	0 (0, 0)	26	0 (0, 0)
ARB + Diuretics	141	0.2 (0.1, 0.2)	149	0.2 (0.1, 0.2)	180	0.2 (0.2, 0.2)
Betablockers + CCB	26	0 (0, 0)	27	0 (0, 0)	32	0 (0, 0)
Betablockers + Diuretics	303	0.3 (0.3, 0.4)	323	0.4 (0.3, 0.4)	422	0.5 (0.4, 0.5)
Renin-Inhibitors	60	0.1 (0.1, 0.1)	63	0.1 (0.1, 0.1)	87	0.1 (0.1, 0.1)
ACE-I	27588	30.5 (30.1, 30.8)	28486	31.4 (31.1, 31.8)	31419	34.7 (34.3, 35.1)
ARB	11724	12.9 (12.7, 13.2)	12045	13.3 (13.1, 13.5)	13076	14.4 (14.2, 14.7)
All antihypertensives	62056	68.5 (68.0, 69.0)	63017	69.6 (69.0, 70.1)	65387	72.2 (71.6, 72.7)
Other antihypertensives*	5427	6.0 (5.8, 6.2)	5749	6.3 (6.2, 6.5)	6760	7.5 (7.3, 7.6)
Betablockers	22774	25.1 (24.8, 25.5)	23406	25.8 (25.5, 26.2)	25467	28.1 (27.8, 28.5)
CCB	23376	25.8 (25.5, 26.1)	24378	26.9 (26.6, 27.2)	27540	30.4 (30.0, 30.8)
Diuretics	32672	36.1 (35.7, 36.5)	34173	37.7 (37.3, 38.1)	38126	42.1 (41.7, 42.5)

Note: ACE-I = Angiotensin-converting-enzyme inhibitor, CCB = calcium channel blockers, ARB = Angiotensin-II receptor blockers.
*other antihypertensives comprise: Ambrisentan, Betanidine, Bosentan, Clonidine,Debrisoquine, Doxazosin,Guanethidine, Hydralazine, Indoramin, Ketanserin, Macitentan, Methoserpidine ,Methyldopa, Methyldopate, Metirosine, Minoxidil, Moxonidine, Prazosin, Reserpine, Riociguat, Sitaxentan, Sodium nitroprusside and Trimetaphan camsilate

Table S 4.2 Point prevalence of the different antihypertensives classes for the frailty cohort (N=110225)

Drug name	One-month PP		Three-months PP		One-year PP	
	N prevalent	PP (95% CI)	N prevalent	PP (95% CI)	N prevalent	PP (95% CI)
ACE-I + CCB	29	0 (0, 0)	30	0 (0, 0)	38	0 (0, 0)
ACE-I + Diuretics	473	0.4 (0.4, 0.5)	501	0.5 (0.4, 0.5)	617	0.6 (0.5, 0.6)
ARB + CCB	22	0 (0, 0)	23	0 (0, 0)	23	0 (0, 0)
ARB + Diuretics	228	0.2 (0.2, 0.2)	238	0.2 (0.2, 0.2)	301	0.3 (0.2, 0.3)
Betablockers + CCB	30	0 (0, 0)	31	0 (0, 0)	40	0 (0, 0)
Betablockers + Diuretics	484	0.4 (0.4, 0.5)	501	0.5 (0.4, 0.5)	623	0.6 (0.5, 0.6)
Renin-Inhibitors	123	0.1 (0.1, 0.1)	131	0.1 (0.1, 0.1)	154	0.1 (0.1, 0.2)
ACE-I	42538	38.6 (38.2, 39.0)	43651	39.6 (39.2, 40.0)	47382	43.0 (42.6, 43.4)
ARB	19638	17.8 (17.6, 18.1)	20027	18.2 (17.9, 18.4)	21325	19.3 (19.1, 19.6)
All antihypertensives	88863	80.6 (80.1, 81.2)	89797	81.5 (80.9, 82.0)	92014	83.5 (82.9, 84.0)
Other antihypertensives*	9739	8.8 (8.7, 9.0)	10137	9.2 (9.0, 9.4)	11480	10.4 (10.2, 10.6)
Betablockers	31101	28.2 (27.9, 28.5)	31767	28.8 (28.5, 29.1)	34106	30.9 (30.6, 31.3)
CCB	35338	32.1 (31.7, 32.4)	36522	33.1 (32.8, 33.5)	40453	36.7 (36.3, 37.1)
Diuretics	48364	43.9 (43.5, 44.3)	50160	45.5 (45.1, 45.9)	55022	49.9 (49.5, 50.3)

Note: ACE-I = Angiotensin-converting-enzyme inhibitor, CCB = calcium channel blockers, ARB = Angiotensin-II receptor blockers.
*other antihypertensives comprise: Ambrisentan, Betanidine, Bosentan, Clonidine,Debrisoquine, Doxazosin,Guanethidine, Hydralazine, Indoramin, Ketanserin, Macitentan, Methoserpidine ,Methyldopa, Methyldopate, Metirosine, Minoxidil, Moxonidine, Prazosin, Reserpine, Riociguat, Sitaxentan, Sodium nitroprusside and Trimetaphan camsilate

Table S 4.3 Point prevalence of the different antihypertensives classes for the polypharmacy cohort (N=116076)

Drug name	One-month PP		Three-months PP		One-year PP	
	N prevalent	PP (95% CI)	N prevalent	PP (95% CI)	N prevalent	PP (95% CI)
ACE-I + CCB	18	0 (0, 0)	19	0 (0, 0)	33	0 (0, 0)
ACE-I + Diuretics	503	0.4 (0.4, 0.5)	536	0.5 (0.4, 0.5)	660	0.6 (0.5, 0.6)
ARB + CCB	33	0 (0, 0)	34	0 (0, 0)	36	0 (0, 0)
ARB + Diuretics	308	0.3 (0.2, 0.3)	326	0.3 (0.3, 0.3)	396	0.3 (0.3, 0.4)
Betablockers + CCB	29	0 (0, 0)	32	0 (0, 0)	39	0 (0, 0)
Betablockers + Diuretics	461	0.4 (0.4, 0.4)	490	0.4 (0.4, 0.5)	620	0.5 (0.5, 0.6)
Renin-Inhibitors	173	0.1 (0.1, 0.2)	183	0.2 (0.1, 0.2)	215	0.2 (0.2, 0.2)
ACE-I	45732	39.4 (39.0, 39.8)	46947	40.4 (40.1, 40.8)	50964	43.9 (43.5, 44.3)
ARB	23327	20.1 (19.8, 20.4)	23785	20.5 (20.2, 20.8)	25216	21.7 (21.5, 22.0)
All antihypertensives	98069	84.5 (84.0, 85.0)	98995	85.3 (84.8, 85.8)	101121	87.1 (86.6, 87.7)
Other antihypertensives*	11867	10.2 (10.0, 10.4)	12373	10.7 (10.5, 10.8)	13995	12.1 (11.9, 12.3)
Betablockers	37388	32.2 (31.9, 32.5)	38131	32.9 (32.5, 33.2)	40731	35.1 (34.7, 35.4)
CCB	41720	35.9 (35.6, 36.3)	43092	37.1 (36.8, 37.5)	47568	41.0 (40.6, 41.3)
Diuretics	57259	49.3 (48.9, 49.7)	59339	51.1 (50.7, 51.5)	64681	55.7 (55.3, 56.2)

Note: ACE-I = Angiotensin-converting-enzyme inhibitor, CCB = calcium channel blockers, ARB = Angiotensin-II receptor blockers.
*other antihypertensives comprise: Ambrisentan, Betanidine, Bosentan, Clonidine,Debrisoquine, Doxazosin,Guanethidine, Hydralazine, Indoramin, Ketanserin, Macitentan, Methoserpidine ,Methyldopa, Methyldopate, Metirosine, Minoxidil, Moxonidine, Prazosin, Reserpine, Riociguat, Sitaxentan, Sodium nitroprusside and Trimetaphan camsilate

Table S 4.4 Point prevalence of the different antihypertensives classes for the overlap group (N=28259)

Drug name	One-month PP		Three-months PP		One-year PP	
	N prevalent	PP (95% CI)	N prevalent	PP (95% CI)	N prevalent	PP (95% CI)
ACE-I + CCB	6	0 (0, 0)	6	0 (0, 0)	8	0 (0, 0.001)
ACE-I + Diuretics	94	0.3 (0.3, 0.4)	106	0.4 (0.3, 0.5)	145	0.5 (0.4, 0.6)
ARB + CCB	6	0 (0, 0)	7	0 (0, 0.1)	7	0 (0, 0.1)
ARB + Diuretics	41	0.1 (0.1, 0.2)	45	0.2 (0.1, 0.2)	61	0.2 (0.2, 0.3)
Betablockers + CCB	0	NA	<5	NA	<5	NA
Betablockers + Diuretics	51	0.2 (0.1, 0.2)	57	0.2 (0.2, 0.3)	103	0.4 (0.3, 0.4)
Renin-Inhibitors	31	0.1 (0.1, 0.2)	31	0.1 (0.1, 0.2)	43	0.2 (0.1, 0.2)
ACE-I	11314	40.0 (39.3, 40.8)	11818	41.8 (41.1, 42.6)	13253	46.9 (46.1, 47.7)
ARB	5254	18.6 (18.1, 19.1)	5435	19.2 (18.7, 19.8)	6016	21.3 (20.8, 21.8)
All antihypertensives	23957	84.8 (83.7, 85.9)	24323	86.1 (85.0, 87.2)	25112	88.9 (87.8, 90.0)
Other antihypertensives*	2559	9.1 (8.7, 9.4)	2741	97. (9.3, 10.1)	3300	11.7 (11.3, 12.1)
Betablockers	9314	33.0 (32.3, 33.6)	9626	34.1 (33.4, 34.8)	10562	37.4 (36.7, 38.1)
CCB	9208	32.6 (31.9, 33.3)	9737	34.5 (33.8, 35.1)	11318	40.1 (39.3, 40.8)
Diuretics	14835	52.5 (51.7, 53.3)	15547	55.0 (54.2, 55.9)	17300	61.2 (60.3, 62.1)

Note: ACE-I = Angiotensin-converting-enzyme inhibitor, CCB = calcium channel blockers, ARB = Angiotensin-II receptor blockers.
*other antihypertensives comprise: Ambrisentan, Betanidine, Bosentan, Clonidine,Debrisoquine, Doxazosin,Guanethidine, Hydralazine, Indoramin, Ketanserin, Macitentan, Methoserpidine ,Methyldopa, Methyldopate, Metirosine, Minoxidil, Moxonidine, Prazosin, Reserpine, Riociguat, Sitaxentan, Sodium nitroprusside and Trimetaphan camsilate

Table S 4.5 Point prevalence of the different antihypertensives classes for the background population (N=277332)

Drug name	One-month PP		Three-months PP		One-year PP	
	N prevalent	PP (95% CI)	N prevalent	PP (95% CI)	N prevalent	PP (95% CI)
ACE-I + CCB	50	0 (0, 0)	52	0 (0, 0)	57	0 (0, 0)
ACE-I + Diuretics	1052	0.4 (0.4, 0.4)	1066	0.4 (0.4, 0.4)	1155	0.4 (0.4, 0.4)
ARB + CCB	25	0 (0, 0)	27	0 (0, 0)	30	0 (0, 0)
ARB + Diuretics	426	0.2 (0.1, 0.2)	436	0.2 (0.1, 0.2)	459	0.2 (0.2, 0.2)
Betablockers + CCB	120	0 (0, 0.1)	121	0 (0, 0.1)	124	0 (0, 0.1)
Betablockers + Diuretics	1634	0.6 (0.6, 0.6)	1663	0.6 (0.6, 0.6)	1798	0.6 (0.6, 0.7)
Renin-Inhibitors	83	0 (0, 0)	89	0 (0, 0)	99	0 (0, 0)
ACE-I	54058	19.5 (19.3, 19.7)	54653	19.7 (19.5, 19.9)	57730	20.8 (20.6, 21.0)
ARB	21582	7.8 (7.7, 7.9)	21762	7.8 (7.7, 8.0)	22541	8.1 (8.0, 8.2)
All antihypertensives	130411	47.0 (46.8, 47.3)	131354	47.4 (47.1, 47.6)	134675	48.6 (48.3, 48.8)
Other antihypertensives*	9762	3.5 (3.5, 3.6)	9959	3.6 (3.5, 3.7)	10781	3.9 (3.8, 4.0)
Betablockers	41161	14.8 (14.7, 15.0)	41629	15.0 (14.9, 15.2)	43793	15.8 (15.6, 15.9)
CCB	51332	18.5 (18.3, 18.7)	52055	18.8 (18.6, 18.9)	55234	19.9 (19.8, 20.1)
Diuretics	56577	20.4 (20.2, 20.6)	57695	20.8 (20.6, 21.0)	61844	22.3 (22.1, 22.5)

Note: ACE-I = Angiotensin-converting-enzyme inhibitor, CCB = calcium channel blockers, ARB = Angiotensin-II receptor blockers.
*other antihypertensives comprise: Ambrisentan, Betanidine, Bosentan, Clonidine,Debrisoquine, Doxazosin,Guanethidine, Hydralazine, Indoramin, Ketanserin, Macitentan, Methoserpidine ,Methyldopa, Methyldopate, Metirosine, Minoxidil, Moxonidine, Prazosin, Reserpine, Riociguat, Sitaxentan, Sodium nitroprusside and Trimetaphan camsilate

Table S 4.6 One-year IRs – excluding one-year prevalence of the different antihypertensives classes for the Hospitalisation cohort (N=90597)

Drug name	N incidence	Person-years	IR (95% CI)
ACE-I + CCB	<5	NA	NA
ACE-I + Diuretics	18	81946.7	0.2 (0.1, 0.3)
ARB + CCB	9	82284.4	0.1 (0.0, 0.2)
ARB + Diuretics	13	82145.1	0.2 (0.1, 0.2)
Betablockers + CCB	0	NA	NA
Betablockers + Diuretics	<5	NA	NA
Renin-Inhibitors	25	82217.3	0.3 (0.2, 0.4)
ACE-I	2905	52211.7	55.6 (53.6, 57.7)
ARB	1230	69601.9	17.7 (16.7, 18.7)
All antihypertensives	2793	21610.1	129.3 (124.5, 134.0)
Other antihypertensives*	848	75686.8	11.2 (10.5, 12.0)
Betablockers	2445	57871.6	42.3 (40.6, 43.9)
CCB	2783	55607.2	50.1 (48.2, 51.9)
Diuretics	4141	46198.4	89.6 (86.9, 92.4)

PY=person years

Note: ACE-I = Angiotensin-converting-enzyme inhibitor, CCB = calcium channel blockers, ARB = Angiotensin-II receptor blockers.

*other antihypertensives comprise: Ambrisentan, Betanidine, Bosentan, Clonidine,Debrisoquine, Doxazosin,Guanethidine, Hydralazine, Indoramin, Ketanserin, Macitentan, Methoserpidine ,Methyldopa, Methyldopate, Metirosine, Minoxidil, Moxonidine, Prazosin, Reserpine, Riociguat, Sitaxentan, Sodium nitroprusside and Trimetaphan camsilate

Table S 4.7 One-year IRs – excluding one-year prevalence of the different antihypertensives classes for the frailty cohort (N=110225)

Drug name	N incidence	Person-years	IR (95% CI)
ACE-I + CCB	5	101713.0	0.1 (0.0, 0.1)
ACE-I + Diuretics	29	101161.0	0.3 (0.2, 0.4)
ARB + CCB	19	101717.8	0.2 (0.1, 0.3)
ARB + Diuretics	32	101448.5	0.3 (0.2, 0.3)
Betablockers + CCB	<5	NA	NA
Betablockers + Diuretics	8	101159.6	0.1 (0.0, 0.1)
Renin-Inhibitors	43	101577.2	0.4 (0.3, 0.6)
ACE-I	3400	55998.0	60.7 (58.7, 62.8)
ARB	1716	80840.6	21.2 (20.2, 22.2)
All antihypertensives	2400	15449.0	155.4 (149.1, 161.6)
Other antihypertensives*	1291	90359.1	14.3 (13.5, 15.1)
Betablockers	2933	68671.5	42.7 (41.2, 44.3)
CCB	3334	62256.8	53.6 (51.7, 55.4)
Diuretics	5102	48913.8	104.3 (101.4, 107.2)

PY=person years

Note: ACE-I = Angiotensin-converting-enzyme inhibitor, CCB = calcium channel blockers, ARB = Angiotensin-II receptor blockers.

*other antihypertensives comprise: Ambrisentan, Betanidine, Bosentan, Clonidine,Debrisoquine, Doxazosin,Guanethidine, Hydralazine, Indoramin, Ketanserin, Macitentan, Methoserpidine ,Methyldopa, Methyldopate, Metirosine, Minoxidil, Moxonidine, Prazosin, Reserpine, Riociguat, Sitaxentan, Sodium nitroprusside and Trimetaphan camsilate

Table S 4.8 One-year IRs – excluding one-year prevalence of the different antihypertensives classes for the polypharmacy cohort (N=116076)

Drug name	N incidence	Person-years	IR (95% CI)
ACE-I + CCB	5	107295.7	0.05 (0.01, 0.09)
ACE-I + Diuretics	40	106684.8	0.38 (0.26, 0.49)
ARB + CCB	16	107286.8	0.15 (0.08, 0.22)
ARB + Diuretics	39	106932.9	0.37 (0.25, 0.48)
Betablockers + CCB	<5	NA	NA
Betablockers + Diuretics	12	106732.5	0.11 (0.05, 0.18)
Renin-Inhibitors	53	107091.2	0.50 (0.36, 0.63)
ACE-I	3339	58366.1	57.21 (55.27, 59.15)
ARB	1936	82641.2	23.43 (22.38, 24.47)
All antihypertensives	1972	12759.6	154.55 (147.73, 161.37)
Other antihypertensives*	1454	93430.1	15.56 (14.76, 16.36)
Betablockers	2894	68095.8	42.50 (40.95, 44.05)
CCB	3368	61092.4	55.13 (53.27, 56.99)
Diuretics	4983	45487.4	109.55 (106.51, 112.59)

PY=person years

Note: ACE-I = Angiotensin-converting-enzyme inhibitor, CCB = calcium channel blockers, ARB = Angiotensin-II receptor blockers.

*other antihypertensives comprise: Ambrisentan, Betanidine, Bosentan, Clonidine,Debrisoquine, Doxazosin,Guanethidine, Hydralazine, Indoramin, Ketanserin, Macitentan, Methoserpidine ,Methyldopa, Methyldopate, Metirosine, Minoxidil, Moxonidine, Prazosin, Reserpine, Riociguat, Sitaxentan, Sodium nitroprusside and Trimetaphan camsilate

Table S 4.9 One-year IRs – excluding one-year prevalence of the different antihypertensives classes for the overlap group (N=28259)

Drug name	N incidence	Person-years	IR (95% CI)
ACE-I + CCB	0	NA	NA
ACE-I + Diuretics	5	24729.1	0.20 (0.03, 0.38)
ARB + CCB	<5	NA	NA
ARB + Diuretics	7	24802.3	0.28 (0.07, 0.49)
Betablockers + CCB	0	NA	NA
Betablockers + Diuretics	<5	NA	NA
Renin-Inhibitors	9	24814.1	0.36 (0.13, 0.6)
ACE-I	866	12665.3	68.38 (63.82, 72.9)
ARB	472	19174.3	24.62 (22.40, 26.84)
All antihypertensives	491	2481.9	197.83 (180.33, 215.33)
Other antihypertensives*	315	21729.2	14.50 (12.90, 16.10)
Betablockers	782	15105.2	51.77 (48.14, 55.40)
CCB	828	14269.9	58.03 (54.08, 61.98)
Diuretics	1381	9117.8	151.46 (143.47, 159.45)

PY=person years

Note: ACE-I = Angiotensin-converting-enzyme inhibitor, CCB = calcium channel blockers, ARB = Angiotensin-II receptor blockers.

*other antihypertensives comprise: Ambrisentan, Betanidine, Bosentan, Clonidine,Debrisoquine, Doxazosin,Guanethidine, Hydralazine, Indoramin, Ketanserin, Macitentan, Methoserpidine ,Methyldopa, Methyldopate, Metirosine, Minoxidil, Moxonidine, Prazosin, Reserpine, Riociguat, Sitaxentan, Sodium nitroprusside and Trimetaphan camsilate

Table S 4.10 One-year IRs – excluding one-year prevalence of the different antihypertensives classes for the background population (N=277332)

Drug name	N incidence	Person-years	IR (95% CI)
ACE-I + CCB	5	265656.0	0.02 (0, 0.04)
ACE-I + Diuretics	79	264560.6	0.30 (0.23, 0.36)
ARB + CCB	25	265671.2	0.09 (0.06, 0.13)
ARB + Diuretics	33	265244.8	0.12 (0.08, 0.17)
Betablockers + CCB	<5	NA	NA
Betablockers + Diuretics	24	263954.9	0.09 (0.06, 0.13)
Renin-Inhibitors	45	265589.5	0.17 (0.12, 0.22)
ACE-I	7348	206580.2	35.57 (34.76, 36.38)
ARB	2400	242782.7	9.89 (9.49, 10.28)
All antihypertensives	9622	131709.3	73.06 (71.60, 74.52)
Other antihypertensives*	1553	254571.4	6.10 (5.80, 6.40)
Betablockers	5137	221275.9	23.22 (22.58, 23.85)
CCB	7555	208727.9	36.20 (35.38, 37.01)
Diuretics	8059	202584.7	39.78 (38.91, 40.65)

PY=person years

Note: ACE-I = Angiotensin-converting-enzyme inhibitor, CCB = calcium channel blockers, ARB = Angiotensin-II receptor blockers.

*other antihypertensives comprise: Ambrisentan, Betanidine, Bosentan, Clonidine,Debrisoquine, Doxazosin,Guanethidine, Hydralazine, Indoramin, Ketanserin, Macitentan, Methoserpidine ,Methyldopa, Methyldopate, Metirosine, Minoxidil, Moxonidine, Prazosin, Reserpine, Riociguat, Sitaxentan, Sodium nitroprusside and Trimetaphan camsilate

5. Longitudinal trajectories of frailty in older people, and their association with mortality

Table S 5.1 Comparing joint models

Trajectory	Log-likelihood
Linear	4141453
Quadratic	4142843

Table S 5.2 Comparing joint latent class models

Number of clusters	BIC	Entropy	Smallest cluster size (%)	More than 50% of the population has a posterior probability >0.7
1	6370269	1	NA	NA
2	6319479	0.78	12.0	Yes
3	6255902	0.79	7.9	Yes
4	6234100	0.77	2.5	Yes
5	6224469	0.77	1.9	Yes
6*	NA	NA	NA	NA

* Did not converge

Table S 5.3 Joint latent class model results in GOLD (including baseline age and gender)

N eligible=475503	N events=68977	N observations=2074677
Longitudinal model	Parameter	Estimate (95% CI)
	Low-slow	Intercept = 2.04 (2.03, 2.05)
		Slope = 0.18 (0.18, 0.18)
	Low-moderate	Intercept = 2.15 (2.13, 2.16)
		Slope = 0.80 (0.80, 0.81)
	Low-rapid	Intercept = 2.40 (2.37, 2.43)
		Slope = 1.65 (1.65, 1.66)
	High-slow	Intercept = 7.37 (7.35, 7.40)
		Slope = 0.26 (0.26, 0.26)
	High-rapid	Intercept = 7.88 (7.82, 7.95)
		Slope = 0.97 (0.97, 0.98)
	Gender=male	-0.20 (-0.21, -0.20)
	Age-65	0.11 (0.11, 0.11)
Survival model	Parameter	HR (95% CI)
	Low-slow	Ref
	Low-moderate	1.24 (1.13, 1.37)
	Low-rapid	3.73 (3.71, 3.76)
	High-slow	2.80 (2.74, 2.85)
	High-rapid	3.63 (3.57, 3.69)
	Gender=male	1.40 (1.39, 1.41)
	Age-65	1.11 (1.11, 1.11)

Table S 5.4 Summary eFI score and number of deaths during study follow-up for all clusters

		Low-slow N = 338029	Low-moderate N = 69794	Low-rapid N = 12869	High-slow N = 45707	High-rapid N = 9104
Follow-up time (median [IQR])		4.0 [2.5, 4.0]	4.0 [3.4, 4.0]	3.0 [1.5, 4.0]	3.4 [1.5, 4.0]	3.3 [1.8, 4.0]
Death (n (%))	Year 1	11503 (3.4)	117 (0.2)	2110 (16.4)	5329 (11.7)	696 (7.6)
	Year 2	19784 (5.6)	2474 (3.5)	4175 (32.4)	8911 (19.5)	1940 (21.3)
	Year 3	27224 (8.1)	6108 (8.8)	5488 (42.6)	12103 (26.5)	2989 (32.8)
	Year 4	33611 (9.9)	10352 (14.8)	6489 (50.4)	14654 (32.1)	3871 (42.5)
eFI score (mean (SD))	Year 1	3.24 (2.19)	4.03 (2.44)	5.53 (2.80)	9.71 (2.13)	11.01 (2.34)
	Year 2	3.33 (2.23)	4.87 (2.54)	7.16 (2.97)	9.88 (2.16)	11.92 (2.28)
	Year 3	3.56 (2.30)	5.80 (2.48)	8.86 (2.82)	10.13 (2.16)	12.85 (2.20)
	Year 4	3.76 (2.36)	6.51 (2.49)	9.96 (2.81)	10.35 (2.18)	13.55 (2.14)

Table S 5.5 Prevalence of eFI deficits in CPRD GOLD and Aurum

N (%)	Study population – GOLD	Study population – Aurum
eFI 1: Activity Limitation	6422 (1.4)	4613 (1.2)
eFI 2: Anaemic & haematinic deficiency	42879 (9.0)	39138 (10.0)
eFI 3: Arthritis	67117 (14.1)	126585 (32.4)
eFI 4: Atrial fibrillation	34396 (7.2)	34695 (8.9)
eFI 5: Cerebrovascular disease	30477 (6.4)	33794 (8.7)
eFI 6: CKD	90030 (18.9)	76541 (19.6)
eFI 7: Diabetes	66699 (14.0)	58483 (15.0)
eFI 8: Dizziness	67670 (14.2)	61284 (15.7)
eFI 9: Dyspnoea	77555 (16.3)	59128 (15.2)
eFI 10: Falls	14032 (3.0)	37920 (9.7)
eFI 11: Foot problems	20385 (4.3)	15224 (3.9)
eFI 12: Fragility fracture	18463 (3.9)	25076 (6.4)
eFI 13: Hearing impairment	61743 (13.0)	60947 (15.6)
eFI 14: Heart failure	15853 (3.3)	15760 (4.0)
eFI 15: Heart valve disease	2765 (0.6)	4258 (1.1)
eFI 16: Housebound	36138 (7.6)	50722 (13.0)
eFI 17: Hypertension	170820 (35.9)	192522 (49.3)
eFI 18: Hypotension / syncope	33208 (7.0)	31393 (8.0)
eFI 19: Ischaemic heart disease	123014 (25.9)	89168 (22.9)
eFI 20: Memory & cognitive problems	18739 (3.9)	18468 (4.7)
eFI 21: Mobility and transfer problems	18260 (3.8)	26711 (6.8)
eFI 22: Osteoporosis	37759 (7.9)	33192 (8.5)
eFI 23: Parkinsonism & Tremor	7692 (1.6)	8580 (2.2)
eFI 24: Peptic ulcer	7860 (1.7)	12101 (3.1)
eFI 25: Peripheral vascular disease	15364 (3.2)	17638 (4.5)
eFI 26: Polypharmacy	288453 (60.7)	265432 (68.0)
eFI 27: Requirement for care	6880 (1.4)	16204 (4.2)
eFI 28: Respiratory disease	103441 (21.8)	103935 (26.6)
eFI 29: Skin ulcer	22173 (4.7)	20304 (5.2)
eFI 30: Sleep disturbance	33813 (7.1)	25264 (6.5)
eFI 31: Social vulnerability	13488 (2.8)	28014 (7.2)
eFI 32: Thyroid disease	16640 (3.5)	32871 (8.4)
eFI 33: Urinary incontinence	22803 (4.8)	19383 (5.0)
eFI 34: Urinary system disease	128758 (27.1)	114477 (29.3)
eFI 35: Visual impairment	95132 (20.0)	116276 (29.8)
eFI 36: Weight loss and anorexia	13710 (2.9)	14394 (3.7)

Table S 5.6 Proportion of posterior probability above different thresholds in each cluster

Posterior probability	Low-moderate	Low-rapid	Low-slow	High-slow	High-rapid
Posterior probability ≥ 0.7	66.8%	78.0%	91.1%	79.4%	73.4%
Posterior probability ≥ 0.8	60.2%	67.7%	83.5%	69.2%	61.8%
Posterior probability ≥ 0.9	43.0%	56.5%	73.7%	51.4%	45.9%

6. Longitudinal trajectories of polypharmacy in older people, and their association with mortality

CPRD GOLD

Table S 6.1 Number of patients and data points used for each KmL model

Number of longitudinal data points	3 data points	4 data points	5 data points	6 data points
Patients included N (%)	238486 (79.5%)	193762 (64.6%)	167471 (55.8%)	150588 (50.2%)

Figure S 6.1 Trajectories captured by KmL for (a) 3 data points (b) 4 data points (c) 5 data points (d) 6 data points

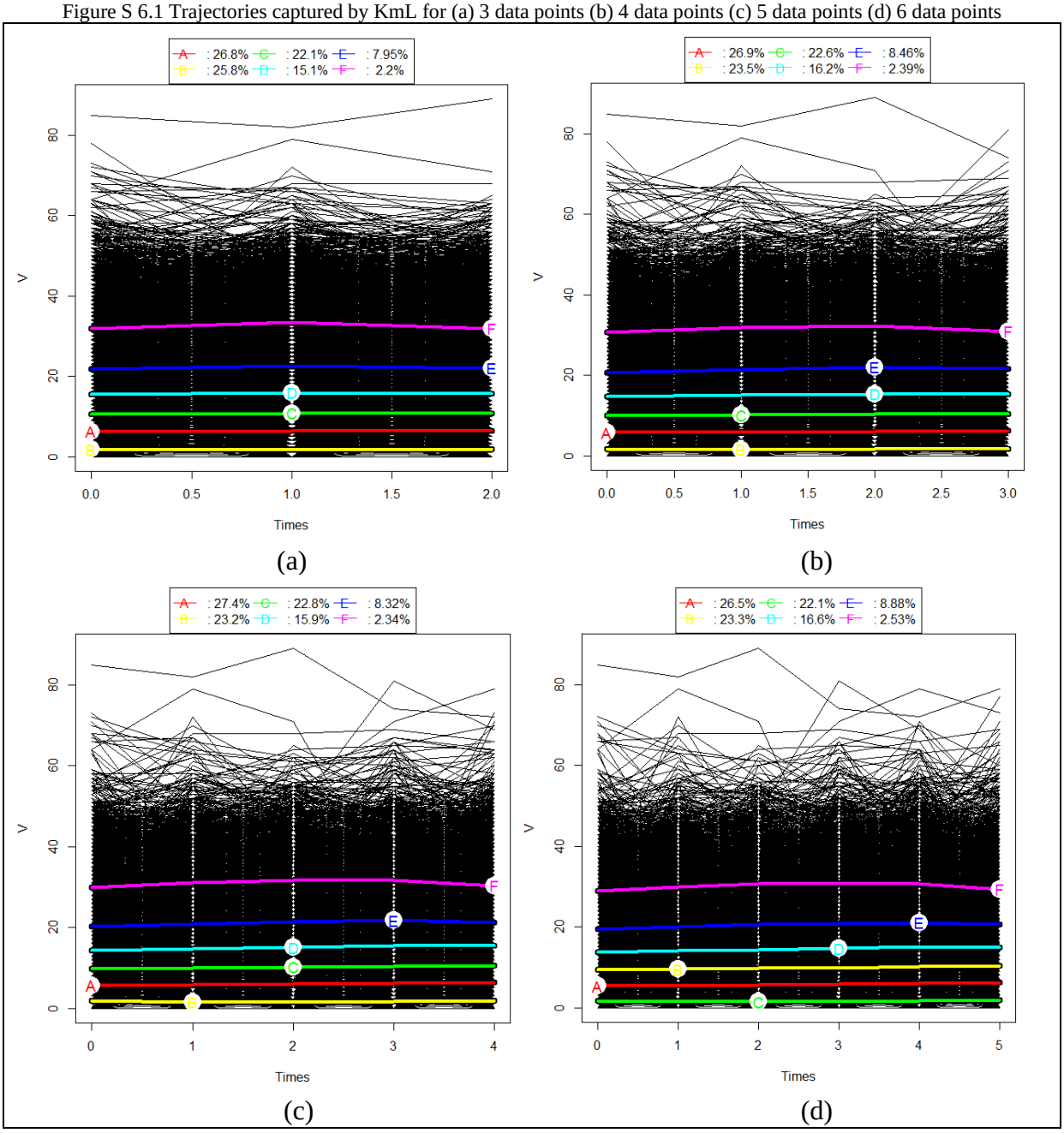


Table S 6.2 Comparison between different link functions for joint latent class models

	Linear model	Spline (3-knots)	Spline (6-knots)	Beta model*
BIC	8446892	7902865	7798489	NA
Variance explained by the model at time=5	80.0%	83.1%	83.0%	NA

* Did not converge

Table S 6.3 Comparison between different hazard functions for joint latent class models

	Weibull (no PH assumption)	Weibull (with PH assumption)	Piecewise*	Splines*
BIC	8399793	8399786	NA	NA

* Did not converge

Table S 6.4 Comparison between different cluster numbers for joint latent class models

Number of clusters	BIC	Entropy	Smallest cluster size (%)	More than 50% of the population has a posterior probability >0.7
1	8446892	1.00	NA	NA
2	8399786	0.84	10.0	Yes
3	8368178	0.84	4.8	Yes
4	8358633	0.82	1.2	Yes
5*	NA	NA	NA	NA

* Did not converge

Table S 6.5 Joint latent class model results in GOLD (including baseline age and gender)

N eligible=299859	N events=38031	N observations=1350561
Longitudinal model	Parameter	Estimate (95% CI)
	Intermediate-fast/increasing	Intercept = 10.70 (10.55, 10.85)
		Slope = 6.40 (6.33, 6.47)
	Low-steady	Intercept = 5.97 (5.95, 6.00)
		Slope = 0.07 (0.06, 0.07)
	Intermediate-slow/increasing	Intercept = 11.22 (11.11, 11.32)
		Slope = 1.98 (1.95, 2.00)
	High-decreasing	Intercept = 23.44 (23.35, 23.54)
		Slope = -1.75 (-1.77, -1.73)
	Gender=male	-0.51 (-0.53, -0.49)
	Age-65	0.16 (0.16, 0.16)
Survival model	Parameter	HR (95% CI)
	Low-steady	Ref
	Intermediate-fast/increasing	20.53 (19.87, 21.21)
	Intermediate-slow/increasing	4.95 (4.84, 5.07)
	High-decreasing	4.64 (4.58, 4.73)
	Gender=male	1.36 (1.34, 1.38)
	Age-65	1.17 (1.17, 1.17)

Table S 6.6 Summary polypharmacy and number of deaths during study follow-up for all clusters in GOLD

GOLD		Intermediate-fast/increasing N=3708	Low-steady N=256923	Intermediate-slow/increasing N=19207	High-decreasing N=20021	Overall N=299859
Follow-up time (median [IQR])		1.7 [0.3, 2.8]	4.3 [1.4, 5.0]	4.6 [2.4, 5.0]	2.5 [1.0, 5.0]	4.1 [1.4, 5.0]
Death (n (%))	Year 1	1049 (28.3)	5274 (2.1)	2005 (10.4)	2224 (11.1)	10552 (3.5)
	Year 2	2108 (56.9)	9497 (3.7)	3492 (18.2)	3748 (18.7)	18845 (6.3)
	Year 3	2718 (73.3)	13130 (5.1)	5188 (27.0)	4900 (24.5)	25936 (8.6)
	Year 4	2988 (80.6)	16496 (6.4)	6938 (36.1)	5882 (29.4)	32304 (10.8)
	Year 5	3107 (83.8)	19648 (7.6)	8612 (44.8)	6664 (33.3)	38031 (12.7)
Polypharmacy (median [IQR])	Year 1	20 [12, 26]	7 [3, 11]	16 [11, 21]	24 [20, 29]	8 [4, 13]
	Year 2	24 [17, 30]	7 [3, 11]	18 [13, 23]	22 [17, 27]	8 [4, 13]
	Year 3	28 [21, 35]	7 [3, 11]	20 15, 25]	20 [15, 25]	8 [4, 13]
	Year 4	33 [27, 40]	7 [3, 11]	22 [18, 27]	18 [13, 24]	8 [4, 14]
	Year 5	37 [30, 44]	7 [3, 11]	23 [19, 28]	17 [12, 23]	8 [4, 14]

Table S 6.7 Difference in percentage between baseline and last observation of individual morbidities in GOLD

	% difference				
	Intermediate-fast/increasing N=3708	Low-steady N=256923	Intermediate-slow/increasing N=19207	High-decreasing N=20021	Overall N=299859
Myocardial infarction	3.8	1.0	4.0	1.9	1.3
Congestive heart failure	5.9	1.4	7.2	3.8	2.0
Peripheral vascular disease	1.4	0.5	2.2	1.1	0.7
Cerebrovascular disease	4.7	2.4	5.7	3.3	2.7
Dementia	4.6	2.8	6.2	4.7	3.2
Chronic obstructive pulmonary disease	4.8	2.6	7.4	3.9	3.0
Rheumatologic disease	1.1	0.9	2.8	1.1	1.0
Peptic ulcer disease	1.3	0.5	1.4	0.8	0.5
Mild liver disease	0.8	0.1	0.5	0.4	0.2
Diabetes with chronic complications	2.8	1.6	4.8	3.7	1.9
Hemiplegia or paraplegia	0.2	0	0.1	0	0
Renal disease	7.3	4.1	9.4	6.3	4.6
Any malignancy	27.8	5.2	15.0	5.8	6.2
Moderate to severe liver disease	0.6	0	0.3	0.2	0.1
Metastatic solid tumour	7.5	0.3	2.6	0.7	0.5
AIDS	0	0	0	0	0
Hypertension	2.2	3.7	4.5	2.2	3.6
Heart failure	6.0	1.5	7.5	4.0	2.1
Osteoporosis	3.4	1.7	5.4	2.6	2.0
Gastroesophageal reflux disease	1.1	0.8	1.6	1.1	0.9
Chronic kidney disease	4.9	3.5	7.3	5.0	3.8
Venous thromboembolism	6.2	1.3	4.4	2.5	1.7
Hypothyroidism	1.3	0.9	1.7	1.1	0.9
Stroke	3.1	1.3	3.5	2.0	1.5
Anxiety	3.5	1.7	4.2	2.9	2.0
Asthma	1.3	1.0	2.6	1.8	1.1
Pneumonia	8.8	1.6	7.8	5.3	2.3
Diabetes	3.8	2.3	4.8	3.2	2.5
Inflammatory bowel disease	0.2	0.1	0.3	0.1	0.1
Depressive disorder	2.8	1.1	3.3	2.3	1.4

Table S 6.8 Difference in percentage between baseline and last observation in drug use by class and for common drugs in GOLD

	% difference				
	Intermediate-fast/increasing N=3708	Low-steady N=256923	Intermediate-slow/increasing N=19207	High-decreasing N=20021	Overall N=299859
Drug classes					
Alimentary tract and metabolism	23.4	-0.3	12.8	-4.9	0.5
Blood and blood forming organs	20.9	3.0	14.1	-4.9	3.4
Cardiovascular system	9.8	1.7	5.2	-6.0	1.6
Dermatological	26.6	-4	16	-18.6	-3.4
Genito-urinary system and sex hormones	6.0	0.1	6.0	-7.7	0
Systemic hormonal preparations, excluding sex hormones and insulins	4.1	0.5	3.7	-1.6	0.7
Anti-infective for systemic use	28.5	-7.1	16.5	-21.7	-6.1
Antineoplastic and immunomodulation agents	3.8	0.6	3.5	-1.4	0.6
Musculoskeletal system	13.1	-3.7	8.8	-15.3	-3.5
Nervous system	31.2	-1.2	15.9	-6.0	0
Anti-parasitic products, insecticides and repellents	3.8	0.6	3.5	-1.4	0.6
Respiratory system	30.6	-2.4	16.7	-12.6	-1.4
Sensory organs	24.2	0.3	9.2	-5.8	0.7
Various	1.4	0	0.6	-0.1	0.1
Drugs					
Psychostimulants	NA	0	0	0	0
Opioids	20.9	6.3	12.0	3.3	6.7
Drugs for obstructive airway disease	9.9	6.1	10.8	4.7	6.4
Drugs used in diabetes	3.9	1.8	4.1	2.2	2.0
Psycholeptics	34.5	6.1	18.8	6.8	7.3
Agents renin-angiotensin systemic	5.6	4.3	7.6	3.1	4.4
Anti-thrombotic	19.0	7.3	19.1	8.4	8.3
Lipid modifying agents	5.7	5.9	7.2	2.6	5.8
Calcium channel blockers	5.3	5.7	8.4	3.8	5.8
Diuretics	15.3	4.6	12.8	5.0	5.3
Antibacterial systemic	7.7	4.8	4.9	0.8	4.6
Immunosuppressant	0.7	0.3	1.4	0.7	0.4
Anti-inflammatory and/or anti-rheumatic	13.6	4.8	8.9	3.0	5.1
Hormonal contraceptives systemic	0.7	0.1	0.2	0.1	0.1
Anti-epileptics	18.2	3.7	13.5	7.7	4.7
Drugs for acid related disorder	20.0	8.1	12.6	2.9	8.1
Beta blocking agents	10.0	4.2	10.9	4.5	4.7
Antidepressants	17.5	6.0	14.8	6.5	6.8
Antineoplastic agents	0.8	0.5	1.4	0.8	0.5

CPRD Aurum

Table S 6.9 Posterior probability in each cluster in Aurum

n (%)	Intermediate-fast/increasing	Low-steady	Intermediate-slow/increasing	High-decreasing
Posterior probability ≥ 0.7	1919 (56.5)	210692 (93.5)	11498 (55.7)	11041 (70.4)
Posterior probability ≥ 0.8	1509 (44.4)	201603 (89.4)	8915(43.2)	9243 (58.9)
Posterior probability ≥ 0.9	1084 (31.9)	186354 (82.7)	6085 (29.5)	7036 (44.9)

Table S 6.10 Posterior classification table in Aurum

		True classification			
		Intermediate-fast/increasing	Low-steady	Intermediate-slow/increasing	High-decreasing
Mean of posterior probabilities in each class	Intermediate-fast/increasing	0.7377	0.0034	0.0461	0.0136
	Low-steady	0.0449	0.9419	0.1822	0.1170
	Intermediate-slow/increasing	0.1861	0.0395	0.7301	0.0612
	High-decreasing	0.0314	0.0152	0.0415	0.8082

Table S 6.11 Summary polypharmacy and number of deaths during study follow-up for all clusters in Aurum

		Intermediate-fast/increasing N=3396	Low-steady N=225383	Intermediate-slow/increasing N=20639	High-decreasing N=15683	Overall N=265101
Follow-up time (median [IQR])		1.7 [1.0, 2.6]	5.0 [4.5, 5.0]	4.8 [2.7, 5.0]	4.1 [1.8, 5.0]	5.0 [3.9, 5.0]
Death (n (%))	Year 1	974 (28.7)	5212 (2.3)	1923 (9.3)	1881 (12.0)	9990 (3.8)
	Year 2	2127 (62.6)	9947 (4.4)	3585 (17.4)	3342 (21.3)	19001 (7.2)
	Year 3	2733 (80.5)	14662 (6.5)	5605 (27.2)	4617 (29.4)	27617 (10.4)
	Year 4	2966 (87.3)	19342 (8.6)	7848 (38.0)	5716 (36.4)	35872 (13.5)
	Year 5	3052 (89.9)	23924 (10.6)	9832 (47.6)	6538 (41.7)	43346 (16.4)
Polypharmacy (median [IQR])	Year 1	19 [13, 26]	7 [3, 11]	15 [11, 20]	24 [20, 29]	8 [4, 13]
	Year 2	24 [17, 30]	7 [3, 11]	17 [12, 22]	22 [18, 28]	8 [4, 13]
	Year 3	29 [22, 35]	7 [3, 11]	20 [15, 25]	20 [15, 26]	8 [4, 13]
	Year 4	32 [26, 39]	7 [3, 11]	22 [17, 26]	18 [14, 24]	8 [4, 13]
	Year 5	37 [32, 45]	7 [3, 11]	23 [19, 28]	17 [12, 22]	8 [4, 13]

Table S 6.12 Percentage difference between baseline and last observation in Aurum

	% difference				
	Intermediate-fast/increasing N=3396	Low-steady N=225383	Intermediate-slow/increasing N=20639	High-decreasing N=15683	Overall N=265101
Myocardial infarction	3.2	1.3	4.2	2.3	1.6
Congestive heart failure	7.3	2.4	9.6	5.8	3.3
Peripheral vascular disease	1.5	0.8	2.3	1.6	0.9
Cerebrovascular disease	4.2	3.1	6.3	4.1	3.4
Dementia	4.7	3.8	6.8	6.5	4.3
Chronic obstructive pulmonary disease	4.6	3.6	8.0	4.3	4.0
Rheumatologic disease	0.9	1.3	2.7	1.3	1.4
Peptic ulcer disease	1.8	0.7	1.7	1.1	0.8
Mild liver disease	0.8	0.2	0.6	0.5	0.3
Diabetes with chronic complications	3.1	2.1	5.4	4.5	2.5
Hemiplegia or paraplegia	0.1	0	0	0	0.1
Renal disease	7.6	6.7	11.8	8.9	7.3
Any malignancy	24.4	7.0	14.6	6.2	7.9
Moderate to severe liver disease	0.4	0.1	0.3	0.2	0.2
Metastatic solid tumour	8.1	0.4	3.0	0.8	0.7
AIDS	0	0	0	0	0
Hypertension	2.9	6.6	6.1	3.9	6.4
Heart failure	7.7	2.6	10.0	6.1	3.5
Osteoporosis	3.1	3.0	6.6	3.8	3.3
Gastroesophageal reflux disease	1.4	1.3	2.2	2.1	1.4
Chronic kidney disease	4.6	5.9	9.0	7.0	6.2
Venous thromboembolism	7.3	1.7	5.0	2.9	2.1
Hypothyroidism	1.5	1.6	2.5	1.7	1.6
Stroke	3.4	1.9	4.4	2.8	2.1
Anxiety	3.7	2.6	5.0	3.8	2.9
Asthma	0.9	1.0	2.1	1.5	1.1
Pneumonia	10.9	2.8	11.1	8.4	3.9
Diabetes	3.7	3.5	5.6	3.9	3.7
Inflammatory bowel disease	0.1	0.1	0.3	0.2	0.2
Depressive disorder	2.4	1.5	3.7	3.0	1.8

Table S 6.13 Difference in percentage between baseline and last observation in drug use by class and for common drugs in Aurum

	% difference				
	Intermediate-fast/increasing N=3396	Low-steady N=225383	Intermediate-slow/increasing N=20639	High-decreasing N=15683	Overall N=265101
Drug classes					
Alimentary tract and metabolism	20.5	3.2	13.2	-5.1	3.8
Blood and blood forming organs	19.2	5.5	16.9	-4.5	6.0
Cardiovascular system	7.5	4.5	5.1	-6.1	4.0
Dermatological	19.8	-2.7	13.2	-21.3	-2.3
Genito-urinary system and sex hormones	4.8	1.3	6.1	-8.0	1.2
Systemic hormonal preparations, excluding sex hormones and insulins	3.3	1.3	4.0	-1.5	1.3
Anti-infective for systemic use	27.6	-3.5	17.0	-21.8	-2.6
Antineoplastic and immunomodulation agents	2.7	1.1	3.6	-1.7	1.1
Musculoskeletal system	11.6	-2.9	6.7	-16.3	-2.8
Nervous system	32.3	0.3	17.0	-7.5	1.6
Anti-parasitic products, insecticides and repellents	2.7	1.1	3.6	-1.7	1.1
Respiratory system	32.3	-0.9	16.4	-13.9	0.1
Sensory organs	23.3	1.2	9.2	-6.0	1.7
Various	0.3	0	0.2	-0.2	0
Drugs					
Psychostimulants	0.1	0	0	0	0
Opioids	22.2	7.1	12.1	3.8	7.5
Drugs for obstructive airway disease	9.3	7.6	10.9	5.1	7.7
Drugs used in diabetes	3.7	2.5	4.5	2.7	2.6
Psycholeptics	34.8	6.9	18.4	7.5	8.2
Agents renin-angiotensin systemic	5.2	5.9	7.6	3.5	5.9
Anti-thrombotic	18.7	10.0	20.8	10.4	11.0
Lipid modifying agents	4.9	9.3	8.6	3.5	8.9
Calcium channel blockers	5.2	8.1	8.8	4.8	7.9
Diuretics	14.4	5.6	13.0	5.3	6.3
Antibacterial systemic	5.4	5.5	4.5	1.0	5.1
Immunosuppressant	0.9	0.5	1.8	1.0	0.7
Anti-inflammatory and/or anti-rheumatic	12.0	5.4	7.3	3.1	5.5
Hormonal contraceptives systemic	0.4	0.1	0.1	0.1	0.1
Anti-epileptics	16.4	5.2	15.6	10.5	6.5
Drugs for acid related disorder	19.4	10.6	13.6	3.1	10.5
Beta blocking agents	8.6	5.6	11.5	5.8	6.1
Antidepressants	15.8	6.9	13.4	6.2	7.5
Antineoplastic agents	1.3	0.7	2.0	0.9	0.8

7. Generalisability of longitudinal trajectories of polypharmacy in older people: a validation study in the Dutch population

Table S 7.1 Posterior probability in each cluster in IPCI

n (%)	Intermediate-fast/increasing	Low-steady	Intermediate-slow/increasing	High-decreasing
Posterior probability ≥ 0.7	1130 (79.1)	122519 (97.4)	4778 (57.8)	2917 (77.3)
Posterior probability ≥ 0.8	939 (65.7)	120169 (95.5)	3560 (43.1)	2343 (62.1)
Posterior probability ≥ 0.9	706 (49.4)	115553 (91.8)	2033 (24.6)	1674 (44.3)

Table S 7.2 Summary polypharmacy and number of deaths during study follow-up for all clusters in IPCI

		Intermediate-fast/increasing N=1429	Low-steady N=125835	Intermediate-slow/increasing N=8268	High-decreasing N=3775	Overall N=139307
Follow-up time (median [IQR])		1.6 [0.9, 2.5]	5.0 [3.8, 5.0]	5.0 [3.0, 5.0]	3.8 [1.7, 5.0]	5.0 [3.4, 5.0]
Death (n (%))	Year 1	455 (31.8)	2731 (2.2)	605 (7.3)	396 (10.5)	4187 (3.0)
	Year 2	936 (65.5)	5307 (4.2)	1202 (14.5)	754 (20.0)	8199 (5.9)
	Year 3	1171 (81.9)	7765 (6.2)	1970 (23.8)	1014 (26.9)	11920 (8.6)
	Year 4	1263 (88.4)	10072 (8.0)	2831 (34.2)	1277 (33.8)	15443 (11.1)
	Year 5	1288 (90.1)	12304 (9.8)	3633 (43.9)	1493 (39.5)	18718 (13.4)
Polypharmacy (median [IQR])	Year 1	17 [10, 24]	6 [3, 10]	13 [8, 18]	23 [19, 27]	6 [3, 11]
	Year 2	22 [16, 27]	6 [3, 10]	15 [10, 20]	20 [16, 25]	6 [3, 11]
	Year 3	26 [21, 31]	6 [3, 10]	18 [13, 22]	18 [14, 23]	7 [3, 11]
	Year 4	30 [24, 35]	6 [3, 10]	20 [16, 24]	17 [12, 21]	7 [3, 11]
	Year 5	33 [29, 39]	6 [3, 10]	22 [18, 26]	16 [11, 20]	7 [3, 11]

Table S 7.3 Difference in percentage between baseline and last observation in comorbidities diagnosis of IPCI

	% difference				
	Intermediate-fast/increasing N=1429	Low-steady N=125835	Intermediate-slow/increasing N=8268	High-decreasing N=3775	Overall N=139307
Myocardial infarction	3.0	1.6	4.9	2.2	1.9
Congestive heart failure	11.3	3.4	13.2	8.6	4.2
Peripheral vascular disease	2.8	2.2	5.4	5.2	2.5
Cerebrovascular disease	6.3	4.7	9.0	6.0	5.0
Dementia	3.3	4.1	4.6	5.9	4.2
Chronic obstructive pulmonary disease	3.8	2.0	5.3	3.1	2.2
Rheumatologic disease	1.1	1.4	2.8	1.3	1.5
Peptic ulcer disease	1.2	0.4	0.9	0.8	0.6
Mild liver disease	0.5	0.1	0.4	0.2	0.1
Diabetes with chronic complications	1.3	0.9	2.0	2.6	1.0
Hemiplegia or paraplegia	0.0	0.0	0.0	0.0	0.0
Renal disease	7.3	8.0	14.7	12.8	8.6
Any malignancy	30.6	9.0	18.9	8.5	9.8
Moderate to severe liver disease	0.1	0.0	0.0	0.1	0.1
Metastatic solid tumour	0.0	0.0	0.0	0.0	0.0
AIDS	0.0	0.1	0.0	0.0	0.1
Hypertension	6.4	2.1	4.9	3.6	2.4
Heart failure	12.5	3.9	15.3	10.1	4.8
Osteoporosis	2.7	2.4	4.5	3.9	2.6
Gastroesophageal reflux disease	0.8	0.7	1.0	1.0	0.7
Chronic kidney disease	0.1	0.1	0.3	0.1	0.1
Venous thromboembolism	6.4	2.1	4.9	3.6	2.4
Hypothyroidism	1.9	1.8	3.2	2.6	1.9
Stroke	5.5	3.8	7.3	5.0	4.0
Anxiety	7.4	6.9	11.3	9.1	7.2
Asthma	1.3	1.6	3.8	3.7	1.8
Pneumonia	17.4	7.4	19.9	14.9	8.5
Diabetes	5.1	3.4	7.0	5.2	3.7
Inflammatory bowel disease	0.3	0.2	0.5	0.5	0.3
Depressive disorder	3.5	1.6	4.1	3.8	1.8

Table S 7.4 Difference in percentage between baseline and last observation in drug use by class and for common drugs in IPCI

	% difference				
	Intermediate-fast/increasing N=1429	Low-steady N=125835	Intermediate-slow/increasing N=8268	High-decreasing N=3775	Overall N=139307
Drug classes					
Alimentary tract and metabolism	24.0	4.2	15.3	-6.7	4.8
Blood and blood forming organs	27.6	8.0	26.3	-7.0	8.8
Cardiovascular system	11.5	2.8	8.6	-8.9	2.9
Dermatological	18.4	-1.9	13.8	-20.1	-1.2
Genito-urinary system and sex hormones	9.6	0.3	7.4	-10.3	0.6
Systemic hormonal preparations, excluding sex hormones and insulins	10.1	0.9	6.2	-4.1	1.1
Anti-infective for systemic use	32.4	-0.4	21.5	-23.1	0.7
Antineoplastic and immunomodulation agents	8.2	0.8	5.4	-3.5	1.0
Musculoskeletal system	19.7	-4.1	9.1	-20.5	-3.5
Nervous system	50.4	2.4	29.2	-9.3	4.1
Anti-parasitic products, insecticides and repellents	8.2	0.8	5.4	-3.5	1.0
Respiratory system	19.8	-0.6	13.8	-15.5	0.1
Sensory organs	27.4	1.8	12.1	-6.8	2.5
Various	4.5	0.0	1.3	-1.0	0.1
Drugs					
Psychostimulants	1.8	0.1	0.4	0.5	0.1
Opioids	58.4	14.4	39.1	15.6	16.4
Drugs for obstructive airway disease	16.5	10.9	18.9	8.5	11.4
Drugs used in diabetes	5.5	2.5	5.5	2.6	2.7
Psycholeptics	53.0	10.5	29.9	12.1	12.1
Agents renin-angiotensin systemic	7.5	8.4	12.7	4.9	8.6
Anti-thrombotic	26.2	13.5	30.8	13.3	14.6
Lipid modifying agents	7.6	8.5	12.3	4.3	8.6
Calcium channel blockers	10.0	8.7	15.0	7.5	9.0
Diuretics	21.6	11.0	24.6	11.3	11.9
Antibacterial systemic	29.4	22.2	28.7	9.5	22.3
Immunosuppressant	2.4	0.6	3.1	1.6	0.8
Anti-inflammatory and/or anti-rheumatic	33.0	17.7	26.1	9.9	18.2
Hormonal contraceptives systemic	NA	0.1	0.2	0	0.1
Anti-epileptics	14.9	3.6	13.3	9.6	4.5
Drugs for acid related disorder	29.3	17.3	22.5	4.5	17.4
Beta blocking agents	14.8	7.7	16.6	6.1	8.3
Antidepressants	16.8	5.1	14.6	10.4	5.9
Antineoplastic agents	8.5	2.1	7.2	3.6	2.5

8. Longitudinal trajectories of polypharmacy in middle-aged people, and their association with mortality

Table S 8.1 Comparison between different link functions for joint latent class models

	Linear model	Spline (3-knots)	Spline (6-knots)	Beta model*
BIC	7980261	6744495	6600732	NA
Variance explained by the model at time=5	82.7%	81.4%	81.2%	NA

* Did not converge

Table S 8.2 Comparison between different hazard functions for joint latent class models

	Weibull (no PH assumption)	Weibull (with PH assumption)	Piecewise*	Splines*
BIC	7885977	7886051	NA	NA

* Did not converge

Table S 8.3 Comparing joint latent class models

No. of clusters	BIC
1	7980261
2	7886051
3	7850094
4*	7850144

* Did not converge

Table S 8.4 Joint latent class model results in GOLD (including baseline age and gender)

N eligible=299859	N events=38031	N observations=1350561
Longitudinal model	Parameter	Estimate (95% CI)
	Intermediate-increasing	Intercept = 9.33 (9.26, 9.40)
		Slope = 2.48 (2.46, 2.50)
	Low-steady	Intercept = 3.35 (3.34, 3.37)
		Slope = 0.06 (0.06, 0.06)
	High-decreasing	Intercept = 20.47 (20.42, 20.50)
		Slope = -0.89 (-0.90, -0.88)
	Gender=male	- 0.97 (-0.98, -0.95)
	Age-50	0.13 (0.13, 0.14)
Survival model	Parameter	HR (95% CI)
	Low-steady	Ref
	Intermediate-increasing	24.6 (23.6, 25.7)
	High-decreasing	8.9 (8.5, 9.3)
	Gender=male	1.59 (1.54, 1.63)
	Age-50	1.07 (1.06, 1.07)

Table S 8.5 Summary polypharmacy and number of deaths during study follow-up for all clusters in the training set

		Intermediate-increasing N= 9605	Low-steady N= 273234	High-decreasing N=17097	Overall N=299936
Follow-up time (median [IQR])		5.0 [2.8, 5.0]	5.0 [1.6, 5.0]	5.0 [1.9, 5.0]	5.0 [1.6, 5.0]
Death (n (%))	Year 1	516 (5.4)	341 (0.1)	389 (2.3)	1246 (0.4)
	Year 2	1023 (10.7)	724 (0.2)	661 (3.9)	2408 (0.8)
	Year 3	1510 (15.7)	1173 (0.4)	854 (5.0)	3537 (1.2)
	Year 4	1968 (20.5)	1651 (0.6)	1081 (6.3)	4700 (1.6)
	Year 5	2373 (24.7)	2168 (0.8)	1292 (7.6)	5833 (1.9)
Polypharmacy (median [IQR])	Year 1	12 [8, 17]	3 [0, 6]	19 [16, 24]	3 [0, 7]
	Year 2	15 [10, 19]	3 [0, 6]	18 [15, 23]	3 [1, 8]
	Year 3	17 [12, 22]	3 [0, 6]	17 [13, 22]	3 [1, 8]
	Year 4	19 [15, 24]	3 [0, 7]	17 [13, 22]	4 [1, 8]
	Year 5	21 [17, 26]	3 [1, 7]	16 [12, 21]	4 [1, 8]

Table S 8.6 Difference in percentage between baseline and last observation of individual morbidities in CPRD GOLD training set

	% difference			
	Intermediate-increasing n= 9605	Low-steady N= 273234	High-decreasing N=17097	Overall N=299936
Myocardial infarction	4.1	0.6	1.2	0.7
Congestive heart failure	2.8	0.3	1.2	0.4
Peripheral vascular disease	1.6	0.2	0.9	0.3
Cerebrovascular disease	3.8	0.7	1.9	0.8
Dementia	0.6	0.2	0.5	0.1
Chronic obstructive pulmonary disease	9.4	2.0	5.6	2.4
Rheumatologic disease	1.6	0.3	1.2	0.5
Peptic ulcer disease	1.3	0.3	1.1	0.4
Mild liver disease	1.6	0.2	0.6	0.2
Diabetes with chronic complications	4.6	0.8	4.6	1.2
Hemiplegia or paraplegia	0.1	0.1	0	0
Renal disease	5.5	1.0	4.2	1.3
Any malignancy	15.0	2.3	3.7	2.9
Moderate to severe liver disease	0.9	0	0.3	0.1
Metastatic solid tumour	3.6	0.1	0.4	0.2
AIDS	0	0	0	0
Hypertension	7.5	4.2	4.1	4.3
Heart failure	3.2	0.3	1.3	0.4
Osteoporosis	3.0	0.6	1.8	0.8
Gastroesophageal reflux disease	2.4	0.8	2.0	1
Chronic kidney disease	3.8	0.8	3.1	1.1
Venous thromboembolism	4.2	0.7	1.9	0.9
Hypothyroidism	2.4	0.8	1.6	0.9
Stroke	2.3	0.4	1.1	0.5
Anxiety	6.4	2.6	4.6	2.9
Asthma	4.5	1.1	3.3	1.3
Pneumonia	4.3	0.4	2.4	0.6
Diabetes	7.6	2.1	4.9	2.4
Inflammatory bowel disease	0.6	0.1	0.3	0.1
Depressive disorder	5.1	1.8	3.9	2.0

Table S 8.7 Difference in percentage between baseline and last observation in drug use by class and for common drugs in CPRD GOLD training set

	% difference			
	Intermediate-increasing N= 9605	Low-steady N= 273234	High-decreasing N=17097	Overall N=299936
Drug classes				
Alimentary tract and metabolism	22.1	0	-3.7	0.5
Blood and blood forming organs	25.0	1.7	-0.4	2.3
Cardiovascular system	18.5	4.5	-0.5	4.6
Dermatological	23.9	-2.2	-11.6	-1.9
Genito-urinary system and sex hormones	10.1	-0.1	-5.1	-0.1
Systemic hormonal preparations, excluding sex hormones and insulins	5.2	0.7	0	0.8
Anti-infective for systemic use	20.3	-6.6	-18.8	-6.4
Antineoplastic and immunomodulation agents	4.1	0.2	-1.3	0.1
Musculoskeletal system	12.8	-3.2	-11.0	-3.1
Nervous system	20.4	-2.0	-4.8	-1.5
Anti-parasitic products, insecticides and repellents	4.1	0.2	-1.3	0.1
Respiratory system	21.9	-1.6	-8.8	-1.3
Sensory organs	15.0	0.5	-2.4	0.7
Various	0.4	0	-0.1	0
Drugs				
Psychostimulants	0.1	0	0	0
Opioids	14.6	6.7	3.4	6.7
Drugs for obstructive airway disease	14.7	6.7	7.1	7
Drugs used in diabetes	7.6	1.8	4.2	2.1
Psycholeptics	17.9	5.5	6.4	6
Agents renin-angiotensin systemic	12.4	4.2	4.9	4.5
Anti-thrombotic	18	2.9	7.2	3.7
Lipid modifying agents	14.9	5.6	6.4	5.9
Calcium channel blockers	12.4	5	6.2	5.3
Diuretics	14.5	2.2	6.1	2.8
Antibacterial systemic	6.7	5.5	1.8	5.3
Immunosuppressant	2.7	0.4	1.7	0.5
Anti-inflammatory and/or anti-rheumatic	10.8	6.2	3.7	6.2
Hormonal contraceptives systemic	1.7	0.9	1.3	0.9
Anti-epileptics	19.8	3.1	10.8	4.1
Drugs for acid related disorder	18.7	9.9	4.9	9.9
Beta blocking agents	13	3.4	5.3	3.8
Antidepressants	14.4	5.9	5.7	6.2
Antineoplastic agents	2.5	0.5	1.9	0.6

Table S 8.8 Posterior probability in each cluster in the CPRD GOLD testing set

Posterior probability (n (%))	Intermediate-slow/increasing	Low-steady	High-decreasing
Posterior probability ≥ 0.7	7126 (75.7)	268429 (98.1)	14071 (82.8)
Posterior probability ≥ 0.8	6168 (65.5)	265095 (96.9)	12714 (74.8)
Posterior probability ≥ 0.9	5022 (53.3)	258977 (94.7)	10822 (63.6)

Table S 8.9 Posterior classification table in the CPRD GOLD testing set

		True classification		
		Intermediate-slow/increasing	Low-steady	High-decreasing
Mean of posterior probabilities in each class	Intermediate-slow/increasing	0.8433	0.0110	0.0412
	Low-steady	0.1071	0.9813	0.0794
	High-decreasing	0.0497	0.0077	0.8794

Table S 8.10 Summary polypharmacy and number of deaths during study follow-up for all clusters in the testing set

		Intermediate-increasing N= 9605	Low-steady N= 273234	High-decreasing N=17097	Overall N=299936
Follow-up time (median [IQR])		5.0 [2.8, 5.0]	5.0 [1.6, 5.0]	5.0 [1.9, 5.0]	5.0 [1.6, 5.0]
Death (n (%))	Year 1	534 (5.6)	367 (0.1)	412 (2.4)	1313 (0.4)
	Year 2	1014 (10.6)	753 (0.3)	687 (4.0)	2454 (0.8)
	Year 3	1505 (15.7)	1175 (0.4)	930 (5.4)	3610 (1.2)
	Year 4	1969 (20.5)	1636 (0.6)	1148 (6.7)	4753 (1.6)
	Year 5	2369 (24.7)	2134 (0.8)	1335 (7.8)	5838 (1.9)
Polypharmacy (median [IQR])	Year 1	12 [8, 17]	3 [0, 6]	19 [16, 24]	3 [0, 7]
	Year 2	15 [10, 20]	3 [0, 6]	19 [15, 24]	3 [1, 8]
	Year 3	17 [12, 22]	3 [0, 6]	17 [14, 23]	3 [1, 8]
	Year 4	19 [15, 24]	3 [0, 7]	17 [13, 22]	4 [1, 8]
	Year 5	21 [17, 26]	3 [0, 7]	16 [12, 21]	4 [1, 8]

Table S 8.11 Difference in percentage between baseline and last observation of individual morbidities in CPRD GOLD testing set

	% difference			
	Intermediate-increasing N=9415	Low-steady N=273517	High-decreasing N=17004	Overall N= 299936
Myocardial infarction	4.1	0.6	1.5	0.7
Congestive heart failure	2.7	0.2	1.4	0.4
Peripheral vascular disease	1.8	0.2	0.8	0.3
Cerebrovascular disease	3.7	0.7	2.2	0.9
Dementia	0.8	0.1	0.4	0.1
Chronic obstructive pulmonary disease	9.0	2.1	5.5	2.5
Rheumatologic disease	2.2	0.4	0.9	0.5
Peptic ulcer disease	1.4	0.3	0.9	0.3
Mild liver disease	1.5	0.1	0.6	0.2
Diabetes with chronic complications	4.4	0.8	4.5	1.2
Hemiplegia or paraplegia	0.1	0.1	0.1	0.0
Renal disease	5.1	1.0	4.0	1.3
Any malignancy	15.1	2.4	4.0	2.9
Moderate to severe liver disease	0.9	0.0	0.4	0.1
Metastatic solid tumour	3.4	0.1	0.5	0.2
AIDS	0.1	0	0	0
Hypertension	7.7	4.2	4.3	4.3
Heart failure	3.1	0.3	1.5	0.4
Osteoporosis	3.0	0.6	1.9	0.8
Gastroesophageal reflux disease	2.2	0.8	1.7	0.9
Chronic kidney disease	3.6	0.9	3.0	1.1
Venous thromboembolism	4.3	0.7	1.7	0.8
Hypothyroidism	2.1	0.8	1.5	0.9
Stroke	2.4	0.4	1.3	0.5
Anxiety	6.2	2.6	4.8	2.9
Asthma	4.7	1.2	3.3	1.4
Pneumonia	4.9	0.4	2.7	0.7
Diabetes	7.1	2.1	5	2.4
Inflammatory bowel disease	0.4	0.1	0.3	0.1
Depressive disorder	5.3	1.8	3.8	2.0

Table S 8.12 Difference in percentage between baseline and last observation in drug use by class and for common drugs in CPRD GOLD testing set

	Difference			
	Intermediate-increasing N=9415	Low-steady N=273517	High-decreasing N=17004	Overall N= 299936
Drug classes				
Alimentary tract and metabolism	23.0	0.2	-3.8	0.7
Blood and blood forming organs	24.9	1.8	-0.7	2.4
Cardiovascular system	18.5	4.5	-0.9	4.7
Dermatological	23.9	-2.2	-12.0	-1.9
Genito-urinary system and sex hormones	9.3	-0.1	-5.5	-0.2
Systemic hormonal preparations, excluding sex hormones and insulins	4.4	0.6	0.0	0.6
Anti-infective for systemic use	20.4	-6.5	-18.2	-6.4
Antineoplastic and immunomodulation agents	3.8	0.2	-1.3	0.2
Musculoskeletal system	11.7	-3.2	-11.3	-3.2
Nervous system	19.7	-2.0	-4.8	-1.5
Anti-parasitic products, insecticides and repellents	3.8	0.2	-1.3	0.2
Respiratory system	21.8	-1.5	-8.7	-1.1
Sensory organs	14.7	0.3	-2.8	0.6
Various	0.5	0.0	-0.2	0.0
Drugs				
Psychostimulants	0.1	0	0.1	0
Opioids	14.0	6.7	3.3	6.7
Drugs for obstructive airway disease	15.6	6.8	7.2	7.1
Drugs used in diabetes	7.1	1.8	4.3	2.1
Psycholeptics	17.9	5.5	6.4	6.0
Agents renin-angiotensin systemic	12.1	4.1	5.1	4.4
Anti-thrombotic	17.8	3.0	7.6	3.7
Lipid modifying agents	15.9	5.6	6.5	6.0
Calcium channel blockers	12.6	4.8	6.8	5.2
Diuretics	14.3	2.2	6.5	2.8
Antibacterial systemic	7.1	5.5	1.8	5.3
Immunosuppressant	2.9	0.3	1.4	0.5
Anti-inflammatory and/or anti-rheumatic	10.6	6.3	3.3	6.2
Hormonal contraceptives systemic	1.5	0.9	1.4	0.9
Anti-epileptics	19.3	3.1	10.9	4.1
Drugs for acid related disorder	19.6	9.8	5.3	9.9
Beta blocking agents	13.0	3.4	5.5	3.8
Antidepressants	14.3	5.9	5.3	6.1
Antineoplastic agents	2.8	0.5	1.7	0.6