DOCTORAL THESIS

Falls Risk in Care Home Residents: a Novel Approach to Exploring the Roles of Chronic Health Conditions, and Multi-morbidity

Author:
Samuel John Watchorn

Supervisors:
Prof. Peter Bath,
Prof. Suzanne Mason,
Dr Jen Lewis

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

School of Medicine and Population Health

March 11, 2024

UNIVERSITY OF SHEFFIELD

Abstract

School of Medicine and Population Health

Doctor of Philosophy

Falls Risk in Care Home Residents: a Novel Approach to Exploring the Roles of Chronic Health Conditions, and Multi-morbidity

by Samuel John Watchorn

Background: Falls in older adults represent a major cause of distress, injury, and mortality. The UK is experiencing concurrent population ageing, increases in the prevalence of multi-morbidity, and a growing care home population. Therefore, an understanding of how multi-morbidity impacts falls risk in care home residents is needed such that effective prediction models can be developed for this population.
Data: Data from the Health Data Research UK learning care homes project specifying care home resident interactions with emergency care and in-hospital wards in the County Durham and Darlington NHS trust were analysed to answer the research question.
Methods: In the sample of 4002 care home residents chronic disease records were grouped using dimensionality reduction and K-means cluster analysis. The resulting clusters were used as an explanatory variable during negative binomial regression analysis with the number of fall presentations per care home resident to the emergency department as the outcome variable. Additional models explored associations between individual chronic conditions, frailty measures, and interactions between the chronic conditions with the fall presentations outcome.
Results: The combined cluster and regression analysis indicated a gradient of effect sizes relating to the type of multi-morbidity present. The Non-Specific-High-Burden (117.8%, 76.1%-169.2%), Cardiovascular-Metabolic (63%, 30.1%-103.7%), Neurological-Psychiatric (41.4%, 17.3%-70.2%), Cardiovascular (23.8%, 3.4%-47.9%) clusters, were all associated with increases in fall presentations when compared to the cluster indicating the absence of chronic disease burden. Of the individual chronic health conditions, hypotension (61.1%, 38.0%-87.9%), dementia (32.3%, 19.3%-46.7%), and peripheral neuropathy (33.7%, 7.7%-65.7%) exhibited the largest impact on falls risk, with smaller effects observed for cerebrovascular disease (18.6%, 2.6%-37.1%), atrial fibrillation (21.8%, 8.4%-36.8%), and osteoarthritis and degenerative joint diseases (15.6%, 1.3%-31.7%). Further models indicated the relationship of frailty with falls is dependent on the index used.
Conclusions: The findings indicate that multi-morbidity impacts falls risk differently depending on the combination of chronic health conditions experienced. However, the role of multi-morbidity in falls risk is complex and in need of further research. Improvements in standardised reporting of fall events at the care home level and linking of this information with electronic health records is the next step for the development of effective falls risk prediction models.

List of Contents

Acknowledgements

The completion of this thesis would not have been possible without the help, support, and advice from a range of sources.

I must first thank the National Institute of Health Research (NIHR), who's funding made this research project possible. This project would have also not been feasible without the data provided by Health Data Research UK (HDRUK), and compilation work by Alex Garner at Lancaster University.

I am also indebted to my supervisory team, who's support, patience and advice were invaluable to both the project and my development as a researcher. You gave me space when I wanted it, support when I asked for it, and sympathy when I needed it. Thank you for not losing faith.

I am grateful to the reviewers for my Confirmation Review; Prof. Steve Goodacre and Dr Suvodeep Mazumdar, whose advice and comments helped shape both the direction taken in the project, and my confidence going forwards. Additionally, members of the Applied Research Collaboration (ARC) and Centre for Urgent and Emergency Care (CURE) for their welcome, kindness and advice.

To my family for their unwavering support and belief, without which I would not be here today. Mum for the phone calls, Dad for the wisdom, Hazel for the dog walks, Rachael for getting me to grow up, and Dave for forever leading the way. I cannot thank you all enough.

Finally to Emma Morris, who's kindness, patience, and enthusiasm helped me embrace the oddness of life. I owe you baked goods in perpetuity.

"If you're going to face a real challenge it has to be a real challenge. You can't accomplish anything without the possibility of failure. "

Lazarus Lake

Declaration of Authorship

I, Samuel John Watchorn, declare that this thesis titled, "Falls Risk in Care Home Residents: a Novel Approach to Exploring the Roles of Chronic Health Conditions, and Multi-morbidity" and the work presented in it are my own. I confirm that:

I am aware of the University's Guidance on the use of Unfair Means. This work has not been previously been presented for an award at this, or any other, university.

Signed: S J Watchorn

Date: March 11, 2024

Contents

List of Figures

List of Tables

List of Abbreviations

AIC	**A**kaike **I**nformation **C**riterion
ANN	**A**rtificial **N**eural **N**etwork
AUC	**A**rea **U**nder the Receiver Operator **C**urve
CDDFT	**C**ounty **D**urham and **D**arlington NHS **F**oundation **T**rust
CI	**C**onfidence **I**nterval
DHARMa	**D**iagnostics for **H**ier**A**rchical **R**egression **M**odels
ED	**E**mergency **D**epartment
EHCH	**E**nhanced **H**ealth in **C**are **H**omes
EHR	**E**lectronic **H**ealth **R**ecord
GCS	**G**lasgow **C**oma **S**cale
HDRUK	**H**ealth **D**ata **R**esearch **UK**
HIC	**H**igh **I**ncome **C**ountry
IRR	**I**ncident **R**ate **R**atio
LR	**L**ogistic **R**egression
MCA	**M**ultiple **C**orrespondence **A**nalysis
MDS	**M**inimum **D**ata **S**et
ML	**M**achine **L**earning
MOI	**M**echanism **O**f **I**njury
NB	**N**aive **B**ayes Classifier
NPV	**N**egative **P**redictive **V**alue
NHS	**N**ational **H**ealth **S**ervice
NICE	**N**ational **I**nstitute of Health and **C**are **E**xcellence
O/E ratio	**O**bserved-**E**xpected ratio
PPV	**P**ositive **P**redictive **V**alue
PRISMA	**P**referred **R**eporting **I**tems for **S**ystematic **R**eviews and **M**eta-**A**nalyses
PROBAST	**P**rediction model **R**isk **O**f **B**ias **AS**sessment **T**ool
RF	**R**andom **F**orest
ROC	**R**eceiver **O**perator **C**urve
ROBIS	**R**isk **O**f **B**ias **I**n **S**ystematic reviews
SMOTE	**S**ynthetic **M**inority **O**versampling **TE**chnique
SVM	**S**upport **V**ector **M**achine
TBI	**T**raumatic **B**rain **I**njury
WHO	**W**orld **H**ealth **O**rganisation

Glossary of Terms

Term	Description
Accuracy:	The proportion of correct classifications in all classifications made by a model. This can be misleading in data with a large class imbalance as predictions of all classes are weighted equally, meaning an algorithm can obtain high accuracy by returning the majority class.
Area under the Curve (AUC):	Summary measure of model discrimination. Patients are ranked based on the model predicted probability of the outcome occurring. Then for each threshold value the sensitivity and 1-specificity are calculated. These values are plotted on a curve, the area under which summarises the ability of a classifier algorithm to separate the classes of an outcome. An AUC of 1 shows perfect discrimination, with 0.5 showing the model is no better than random chance.
Calibration:	Agreement between risk of an outcome predicted by an ML model and the proportion of the outcome in the data. Generally, for individuals with the same characteristics and therefore the same predicted risk of the outcome (X%) there should be on average X in 100 cases in the data with those characteristics for a model to be well calibrated (Calster et al. 2016).
Care Home:	Care homes provide accommodation and personal care for people who need extra support in their daily lives. Also called 'Residential Care'.
Classification:	When the output variable has a finite set of values the learning problem is referred to as classification. Binary classification refers to when the outcome variable only has two possible values.
Cost Function:	Average loss over entire training dataset, optimisation strategies aim to minimise this average loss over multiple iterations.
Cross-Validation:	Randomly splitting the training data into equal sized folds such that a separate model is evaluated on each fold having been trained on the others. Data can be stratified to match distributions in the sample in each fold.
Decision Boundary:	In binary classification, an ML model will classify a point as having the outcome of interest above this threshold. At the decision boundary, the class of a point is ambiguous.
External Validation:	Testing of a model in data not used during the training phase.
Ensemble Learning:	Modelling approaches, which build multiple weaker models and combine their predictions for improved overall performance.

Fall:	Sudden, involuntary transfer of body to the ground and at a lower level than the previous one (Rubenstein et al. 1990).
Hyperparameter:	Parameter whose value determines the learning process of an ML algorithm. This is different to normal parameters, which are derived through model training.
Hyperplane:	Hyperplanes are used as decision boundaries in ML algorithms for classification of data points.
ICD10 Code:	International Statistical Classification of Diseases and Related Health Problems (ICD) standardised codes.
Machine Learning:	An automated computer based process, which extracts patterns from data.
Model Discrimination:	Ability of a model to distinguish between patients with and without a particular outcome.
Negative Predictive Value (NPV):	Proportion of true negative cases in all cases classified as negative by an ML algorithm.
Older Adult:	Individual aged 65 or over.
Overfit:	Situation when an ML model is fit too closely to the training data and therefore cannot classify new instances effectively. Overfit models will have high performance on training data but a large drop in performance in testing.
Penalties:	A penalty (or regularisation term) is included to avoid overfitting a model to the training data such that the resulting algorithm is generalizable to new instances.
Positive Predictive Value (PPV):	Proportion of true positive cases in all cases classified as positive by an ML algorithm.
Sensitivity:	The proportion of correctly classified positive cases by an algorithm in all cases that are positive.
Specificity:	The proportion of correctly classified negative cases by an algorithm in all cases that are negative.
Supervised Learning:	Where historical values of the outcome variable are used to train a predictive model.
Training Set:	Data used by ML algorithm to identify a pattern between input and outcome data. This data is used for optimising weights and parameters in an ML model.
Testing Set:	Data not seen by ML model during the training phase, which is used to assess the generalisability of the model. (Also called a holdout set)
Type-1 ED	A consultant led 24 hour service with full resuscitation facilities and designated accommodation for the reception of accident and emergency patients
Loss Function:	Function for determining the difference between a prediction and an actual observation in Supervised Learning.
Underfit:	Situation when an ML model is oversimplified and cannot classify effectively in the training or testing sets. This can be the result of under-training, over regularisation, or missing input variables.
Unsupervised Learning:	Training an ML model where historical values of the outcome variable are not available. Clustering is a common form of unsupervised learning task.

From the young man who was lost,
For those that helped him find his way.

Chapter 1

Introduction

1.1 Background

This thesis is an investigation into the role of multi-morbidity in determining the risk of falls by care home residents aged over 65. The following chapter summarises the essential background of the topic to contextualise the selection of research question and study aims in Section 1.5.

Falls are defined as a sudden event that results in a person coming to rest on the ground or other lower-level surface unintentionally, which includes falling backwards into bed or chairs (Rubenstein et al., 1990). A distinction is often drawn between ground-level falls, and falls from height due to the differences in severity and outcomes associated with these falls (Yokota et al., 2020). Falls from height are more associated with younger people and work related accidents (Yokota et al., 2020). For the remainder of this thesis, any discussion of falls will refer to ground level falls only, because these are the most common fall event in care home residents aged over 65 (Talbot et al., 2005; Rubenstein, 2006). This thesis uses the common definition of multi-morbidity, which is the co-occurrence of two or more chronic health conditions in the same individual (Johnston et al., 2019). The definition for chronic health conditions used in this thesis is that of a long term physical health condition used by the NHS. This is a health problem requiring management over an extended time period that cannot be cured and instead only controlled or managed through medication and/or other therapies (NHS England, 2023b).

The use of the term care home refers to UK residential care homes, which provide sheltered accommodation to residents and support for daily living activities such as

cooking, washing, dressing, medications, and using toilets (NHS England, 2022). Residential care homes in the UK are distinct to nursing homes, where one or more qualified nurses are also present in the care team (NHS England, 2022). The presence of nursing staff reflects the increased complexity of care needs present in the nursing home setting (NHS England, 2022). Henceforth the the term care home is used to indicate the residential care home setting, which is the focus for this research.

Four areas are addressed in separate sections of this chapter as avenues to explore the background of falls risk in older adults. Section 1.2 addresses why predicting falls in older adults is important by presenting their health effects (Section 1.2.1), the effect of the ageing population (Section 1.2.2), how falls impact the urgent and emergency care system (Section 1.2.3), and the benefits of predicting falls in older adults (Section 1.2.4). Section 1.3 introduces the different types of risk factors for falls, with explanations of the how these risk factors are categorised (Section 1.3.1), and the relationship of falls with frailty (Section 1.3.2).

Having explored the context that falls risk prediction models are used in, findings from the falls risk prediction literature are summarised in Section 1.4 based on where the sample is located. Findings from the community dwelling, and care home settings are summarised separately in Sections 1.4.2 and 1.4.3 respectively. Then a joint discussion of how studies incorporated multi-morbidity into their models and the scope for improvement is presented in Section 1.4.4. Following this the research question, aims, and objectives are presented in Section 1.5, which draw from the key themes presented in the preceding sections.

1.2 Why Predicting Falls in Older Adults is Important

1.2.1 How Falls Impact the Health of Older Adults

Falls represent the second most common cause of unintentional injury and death worldwide across all age groups, with people aged over 60 particularly susceptible to the worst outcomes (World Health Organisation, 2021). In the UK, one in three people aged over 65 will experience at least one fall per year, with this proportion rising to one in two people aged over 80 (Health & Care Excellence, 2013; Bergen, Stevens, and Burns, 2016; NHS, 2021). This high prevalence in older patient groups

meant falls in England were the leading cause of injury in adults and the 9th largest cause of Disability adjusted Life Years (DALYs) during 2013 (Public Health England, 2022).

The effects that falls have on older adults are profound and include distress, pain, injury, reduced mobility, loss of confidence, loss of independence, and mortality (Health & Care Excellence, 2013). Falls also contribute to over two thirds of unintentional injuries, which are the fifth most common cause of death among older people (Pasquetti, Apicella, and Mangone, 2014). Traumatic injuries resulting from falls can range from minor abrasions and bruising, fractures, to serious traumatic brain injuries (Rubenstein, 2006). Due to the possible severity of these fall related injuries, falls in people aged over 65 are a major contributor to increased mortality with only half of those admitted to hospital for a fall surviving beyond a year (Rubenstein, 2006). Furthermore, a fall may lead to a hospital admission and period of immobilisation, which in turn leads to an increased risk of subsequent falls due to a reduction in muscle strength (Stalenhoef et al., 2002; Coker et al., 2015; Valenzuela et al., 2018). This means a major impact of a fall in older people is the increased risk of subsequent falls. Furthermore, falls can also elicit non-injurious effects such as the psychological impact of fear of further falls, although this fear of falling can also be present in people who have not yet fallen (Boyd and Stevens, 2009). Falls and the fear of falls can lead to reduced activity, changes in gait, and depression, which have all been found to raise the risk of future falls creating a vicious cycle (Stalenhoef et al., 2002; Pasquetti, Apicella, and Mangone, 2014; Makino et al., 2017). Therefore, the prevention of falls could elicit important benefits for patients.

However, it is important to recognise that not all falls can be prevented in a safe and efficient manner. While the prevention of falls is highly desirable, implementation of zero falls strategies can have unintended consequences, and would have a large opportunity cost due to the reduced efficiency of falls prevention interventions (King et al., 2016; Public Health England, 2018). Additionally, zero falls strategies can also be harmful to residents through use of immobilisation and restraint to reduce falls risk, which are associated with rapid decline and worsened clinical outcomes in older adults (Gastmans and Milisen, 2006). Therefore, a measured response to the complex issue of falls is needed, where the prevention of falls is

balanced against the capacity to benefit from intervention, and care home resident independence is supported rather than curtailed.

A distinction needs to be drawn between falls in the community and the care home setting due to the differences in both the frequency of falls, and the outcomes experienced following a fall. Falls are three times more common in care home residents than those living in the community setting (Department of Health, 2009). Furthermore, care home residents have worse outcomes after a fall with up to one third of falls resulting in injury (Nurmi et al., 2009). Additionally, when considering the vulnerable state of care home residents, the transport, attendance, and admission to hospital can also carry increased risk of harm and distress beyond the presenting condition. In a multi-centre cohort study of 1,418 patients aged over 65 in Australia, an increase of 0.1 on a frailty index derived from a geriatric assessment score significantly predicted worse outcomes following hospital admission (Hubbard et al., 2017). These worsened outcomes included long (>28 days) lengths of stay, increased inpatient mortality, pressure ulcers, delirium, and in-hospital falls (Hubbard et al., 2017). Further evidence from a Brisbane cohort study found that 13% of patients aged over 70 at discharge had developed daily living impairments and 22% developed bladder incontinence during their hospital stay (Lakhan et al., 2011). Therefore, identification of high falls risk individuals for targeted prevention measures is particularly desirable in the situation where falls and fall related injuries are more likely in care home residents, and hospital transfer for this group carries additional consequences.

In summary, falls are highly prevalent in older patient groups, which as discussed further in Sections 1.2.2 and 1.2.3 makes them a key area of focus for the provision of urgent and emergency care in the context of an ageing population. Additionally, the wide ranging health consequences of falls means the prevention of falls can provide substantial benefits including improvements in health related quality of life and reduced hospital transfers in vulnerable patient groups, which can lead to further health related problems (Lakhan et al., 2011; Stenhagen et al., 2014; Hubbard et al., 2017).

1.2.2 The Importance of Falls Related Research in the Context of an Ageing Population

Research on falls prevention and the identification of high falls risk individuals will become increasingly important as a result of the ageing population in the UK and globally. The number of people aged over 85 in the UK is expected to double over the next 20 years, as the generation born between 1946-1964 ages (Kingston et al., 2018). Between 1990 and 2016 life expectancy in the UK at age 70 increased by 39% for men and 21% for women, which has led to the population of people aged over 70 increasing by 25%, from 4 million to 5 million people (Office for National Statistics, 2018). This ageing of the UK population is expected to lead to an increase in the care home population as the number of publicly funded care home residents is projected to grow by 49% and privately funded by 110% between 2015 and 2035 (Wittenberg and Hu, 2015). Additionally, as discussed in Section 1.2.1, the prevalence of falls is increasing with age and care home residents are at an increased risk of falls and fall related injuries compared to the community dwelling population (Department of Health, 2009). Therefore, there will be more people living in care homes at an increased risk of falls, who could benefit from proactive identification and treatment.

A key characteristic of the ageing population is an increase in the prevalence of multi-morbidity in society, defined in Section 1.1 as the presence of two or more chronic health conditions (Johnston et al., 2019). The prevalence of multi-morbidity in UK adults aged over 65 is projected to rise from 54.0% in 2015 to 67.8% in 2035 (Kingston et al., 2018). An ageing population, and longer survival with chronic conditions will result in more people living in a vulnerable condition for longer. As people live longer with multiple conditions, there is an increased risk of falls caused directly by the conditions themselves and multiple medications, or polypharmacy, used to treat them (Bergen, Stevens, and Burns, 2016; Florence et al., 2018). Therefore, a combination of the ageing population and increased prevalence of multi-morbidity will cause increased falls and fall presentations to the Emergency Department (ED) under the current system.

Faced with the challenges of an ageing population, several strategies published by government and health bodies suggest a supportive legislative environment for

the development of intelligent interventions targeted towards older patient groups (Government office for Science, 2019; NHS, 2019; NHS England, 2020; Department of Health and Social Care, 2021). The integration of health and social care, under the 'whole systems integrated care' proposal will have ramifications for how care homes interact with the wider health system (Department of Health and Social Care, 2021). This increased integration of care homes is echoed in the NHS long-term plan where the expansion of community health teams is intended to provide an alternative route to hospitalisation and support care home residents (NHS, 2019). Furthermore, the Enhanced Health in Care Homes (EHCH) model represents a shift away from a reactive approach to older patient care to a more proactive and integrated system with a focus on individual needs (NHS England, 2020). The EHCH provides a framework for implementing a series of evidence-based interventions and changes to service provision in primary care and care homes to achieve this aim. The roll out of the EHCH model to the whole country in 2023/24 will help cultivate an environment where intelligent health interventions can have impacts on resident wellbeing and outcomes. Furthermore, part of the EHCH model is intended to ease data sharing between care homes and NHS providers, which may mean interventions making use of these sources of linked data will also become more feasible in future (NHS, 2019).

There is also support for implementing intelligent health systems and integrating social care for older patients outside the UK. The European Commission suggested that innovations in e-health, mobile health and telecare alongside integrated care has the potential to substantially improve the efficiency of health systems in the long term (Keifer and Effenberger, 2021). This runs in parallel with the WHO recommendation for the use of focused research into the use of new metrics and analytical methods for investigation into issues surrounding ageing (World Health Organisation, 2022).

A key part of the falls prevention infrastructure is standardised, accurate, and regularly updated sources of linked health records in the care home setting. These sources of standardised information are often referred to as minimum data sets. The international resident assessment instrument (InterRAI) and minimum data set (MDS) 3.0 are examples of this standardised reporting of care home data, and are

used extensively outside the UK (Fries et al., 1997; Saliba and Buchanan, 2008a; interRAI, 2024). A similar prototype structured reporting dataset for use in care homes is being trialled in the UK through the developing resources and minimum dataset for care homes' adoption (DACHA) study (Goodman, 2019). The DACHA study data set differs to the InterRAI and MDS 3.0 because it builds upon the existing data infrastructure in the NHS with the final minimum dataset recommendation for wider use yet to be decided upon (Goodman, 2019; British Geriatrics Society, 2023).

As a result of the ageing population and changing legislative environment, the linkage of data sources, and roll out of intelligent interventions that screen the Electronic Health Record (EHR) for patients at increased falls risk may be made easier and more efficient. However, such intelligent interventions need to be able to account for the increases in multi-morbidity as the population ages. Therefore, research into multi-morbidity and the role it plays in falls risk is essential going forwards. This research is intended to add to a movement towards intelligent and integrated health and social care systems with a focus on identifying high falls risk individuals in the care home setting. By researching the impact of multi-morbidity on falls, the research presented in this thesis is addressing a key challenge of the ageing population. However, as shown in Section 1.2.3, falls in older adults are having a major effect on the provision of urgent and emergency care services in the present day, meaning this research also addresses current problems faced by the health sector.

1.2.3 The Benefits of Falls Related Research in Alleviating Current Pressures in Emergency Care

In setting up the justification for the research question and aims presented in Section 1.5.1, Sections 1.2.1 and 1.2.2 provided evidence for the extensive health impacts of falls in older adults, and how the ageing population necessitates changes in the health system. One of these avenues for change in the urgent and emergency care system is in the Emergency Department (ED). The ED has come under pressure through the confluence of long term demand, and hospital occupancy trends with the aftereffects of the COVID-19 pandemic. A summary flow diagram of these pressures is presented in Figure 1.1.

FIGURE 1.1: Summary of Emergency Department (ED) Pressures

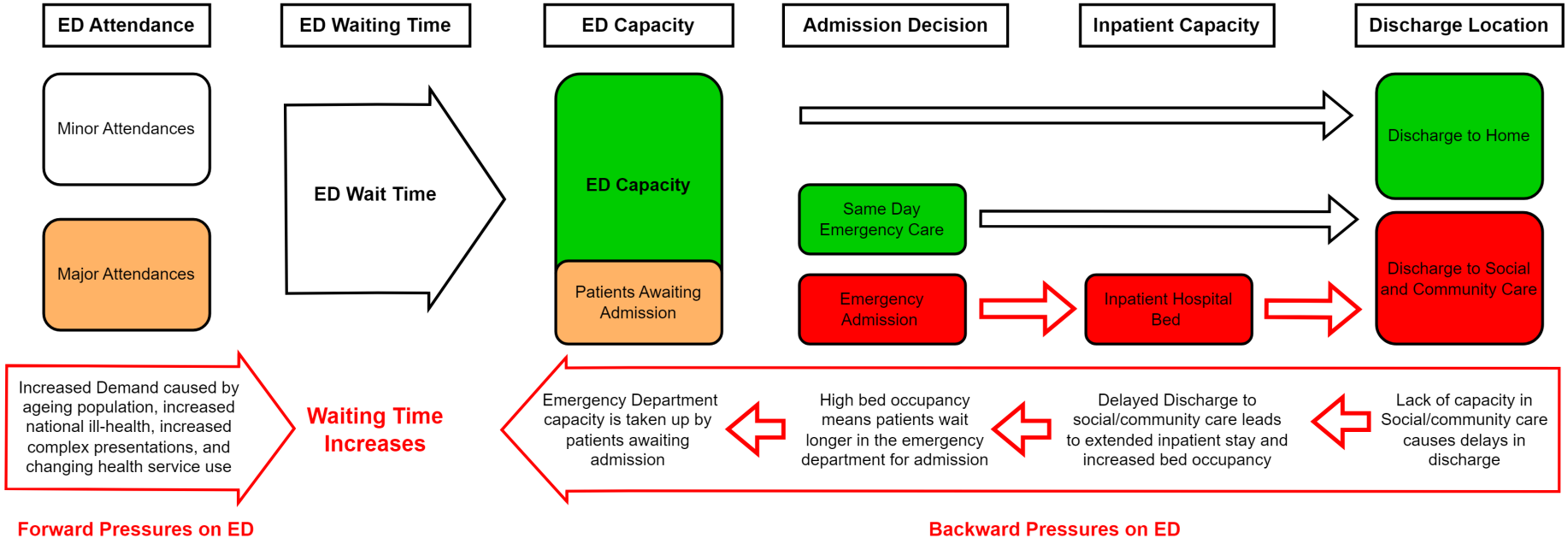


Figure adapted from NHS England (2023a)

Attendances to UK consultant led (Type 1) EDs grew by 10.6% between 2016 and 2019, with the proportion of attendances resulting in 4-hour waits increasing from 8.5% in 2014 to 18.6% in 2018 and 23.8% in 2019 (Baker, 2020). A major cause of worsening ED performance is the high occupancy levels of in-hospital beds, which rose to an average of 95% filled in 2022 (NHS England, 2023a). This issue is further exacerbated by the reduction in the total hospital bed stock, which fell by 8.3% between 2011 and 2020 (British Medical Association, 2022). One factor leading to high hospital occupancy has been difficulty during discharge of patients to social or community care, which means patients are spending longer in hospital beyond the point at which they could be discharged, as seen in Figure 1.1 (Foster, 2023).

High levels of bed occupancy can cause a build up of patients waiting to be admitted from the ED leading to a backlog of demand and longer waiting times (Morley et al., 2018; NHS England, 2023a). In these situations of increasing demand and worsened flow through the department, EDs are increasingly facing problems of overcrowding, which lead to worsened outcomes for patients and hospital staff (Morley et al., 2018).

One cause of ED overcrowding identified in a systematic review was increases in the frequency and complexity of presentations by patients aged over 65 (Morley et al., 2018). Multi-morbidity is present in the majority of older people meaning EDs need to handle increasing numbers of complex presentations as people continue to live longer with multiple chronic health conditions (Barnett et al., 2012; Kehoe et al., 2015; Florence et al., 2018; Head et al., 2021). Additionally, older adults are more likely to be admitted following an ED attendance than those under 65, with multi-morbidity, and residency in care homes both contributing to further increases in this likelihood of admission (Crilly et al., 2008; Bunn et al., 2019; Brewster, O'Keeffe, and Mason, 2019; McParland et al., 2022). This means ED attendances in this group are contributing to the higher levels of bed occupancy described previously. Furthermore, care home residents are also more likely to experience a long length of stay as an inpatient due to their complex needs (Brewster, O'Keeffe, and Mason, 2019). Therefore, preventing falls in care home residents would impact both the forward and backward pressures on the ED shown in Figure 1.1 through reducing attendances by patients who are at an increased risk of both admission and longer

lengths of stay.

In summary, EDs are becoming increasingly pressured meaning research into opportunities for reducing overall ED service demand is worthwhile. A key group identified as contributing to increases in ED service demand are older adults and the leading cause of trauma presentations to EDs in this group are low-level falls. Therefore, research into the identification of older adults at a high risk of a fall, and the prevention of falls could benefit the emergency care system alongside patients.

1.2.4 Summary of the Benefits of Predicting Falls in Older Adults

Section 1.2.1 demonstrated how falls cause a multitude of negative health effects in older adults. This combined with the pressure that fall presentations by older adults place on UK EDs, discussed in Section 1.2.3, means that the prediction and prevention of falls is a highly desirable objective.

Proactive identification of high falls risk individuals is only going to become more important as the population of older adults continues to grow as described in Section 1.2.2. Additionally, a key feature of the ageing population is increases in survival with multi-morbidity. This means that research into developing efficient falls risk flagging systems that are effective in highly multi-morbid samples is worthwhile with the potential to elicit patient and system level benefits.

Having explored the reasoning for why falls risk prediction in older adults is beneficial, Section 1.3 demonstrates the complexity of predicting falls in older adults by exploring the interrelationships between risk factors for falls and how these can be incorporated into models. Following this the key themes in the falls risk prediction literature are explored in Section 1.4 culminating in a discussion of why existing methods for incorporating multi-morbidity into these models need further research.

1.3 Risk Factors for Falls Across the Older Adult Population

1.3.1 Categorising the Risk Factors for Falls

Older adults have different causes of falls and worse outcomes following falls from standing height compared to younger adults (Rubenstein, 2006). The causes of falls in older adults differ from those in younger adults in two ways.

First, a range of internal factors relating to physiological and mental decline change everyday features of the environment into fall hazards (Rubenstein, 2006). Specifically in younger adults, falls are often caused by an identifiable environmental hazard, which would cause anyone to fall (Talbot et al., 2005; Heijnen and Rietdyk, 2016). This contrasts with many falls in the older adult population where there is an interaction between these sources of age related physiological decline and a feature of the environment. For example loose mats, electrical cords, pets, or other items on the floor, which would not have caused a fall in the past but may now contribute to a fall (NHS Inform, 2023). This means there are an increased number of opportunities to fall as a person ages.

Second, the ability and capacity to recover from a loss of balance changes as people age. This change has been presented as a shift from correcting balance using the hips and core, to using multiple rapid steps to regain balance, to then being unable to correct in time to prevent the fall (Rubenstein, 2006).

To add further complexity to this problem, older people often live with multiple risk factors for falls and therefore it can be challenging to determine a singular cause of a fall (Pasquetti, Apicella, and Mangone, 2014). For example alcohol consumption, visual impairment, reduced muscle strength and proprioception in the extremities, and medication side effects can all cause a fall outright, however ascribing a singular cause when they are all present at once is often counterproductive. Instead the causes of falls in older adults are multifactorial and result from interactions between internal and external risk factors.

When considering the falls risk of older adults, it is important to note that not all falls result in injury (Rubenstein, 2006). Additionally, specific features of the fall itself will change the risks of different injuries. For example, falling forwards or sideways increases wrist and hip fracture probability respectively, while falling backwards lowers the fracture risk (Nevitt, Cummings, and Osteoporotic Fractures Research Group, 1993). The focus of the research in this thesis is on the risk that a fall event bad enough to require hospital transfer occurs, rather than the range of possible outcomes following the fall event. Therefore the risk factors discussed in this section refer to features that increase the probability that such a fall will occur. While many of these features will also change the risk of injury following the fall, this is outside

the scope of the research and these effects are not explored further.

The remainder of this section discusses the evidence for different extrinsic (due to the environment) and intrinsic (due to the individual) risk factors for falls in older adults (Pasquetti, Apicella, and Mangone, 2014). Following this, the interrelationship between frailty and falls is discussed in Section 1.3.2. Having investigated the different kinds of risk factors, the discussion then moves to how these factors are incorporated into models in the community and care home settings in Section 1.4.

Extrinsic Risk Factors for Falls

When considering the causes of falls in older adults, a distinction can be drawn between those living in the community and care home residents. Community dwelling older adults living at home are exposed to many extrinsic risk factors for falls, which result from living in a private residence, and moving through the community setting in daily life (Public Health Agency of Canada, 2014).

Identifying the contribution of individual environmental factors on causing a fall is challenging due to their interaction with intrinsic risk factors and falls risk increasing behaviours (Feldman and Chaudhury, 2008). However, environmental risk factors clearly impact the likelihood of a fall, and can be separated into three groups relating to risks in the home, community, and resulting from the weather (Public Health Agency of Canada, 2014). Extrinsic risk factors in the home relate to items such as loose mats, electrical cords, pets, and clutter on the floor but also features of the building such as floor surfaces, stairway width, stair height, door sills, lighting, and bathroom fixtures (Northridge et al., 1995; Pynoos, Steinman, and Nguyen, 2010). Community based extrinsic risk factors relate to features outside the home environment such as raised curbs, and poorly maintained pavements (Gallagher and Scott, 1997). Weather conditions can also increase the risk of a fall through reducing grip and stability while walking on ice, snow, or wet leaves (Beynon et al., 2011).

Older adults living in the community experience a greater variety of extrinsic risk factors than care home dwelling older adults. This is because the care home setting is a more controlled environment, meaning many of the extrinsic risk factors associated with falling are eliminated or mitigated through the design of the care home space (Care Inspectorate, 2016). Additionally, care home staff can conduct

a multi-factorial falls risk assessment for an individual resident to identify specific mitigations that can be put in place (Care Inspectorate, 2016).

However, care home residents are generally more susceptible to falls and fall related injuries than community dwelling older people with three times the risk of falling and ten times the risk of sustaining an injury following a fall (Department of Health, 2009). This difference in fall and fall related injury risk cannot be explained through extrinsic risk factors for the reasons expressed above. However, the degree to which the remaining extrinsic risk factors in the care home, such as a raised step or bathroom floor, would impact an individuals fall risk is dependent on the combination of intrinsic risk factors inherent in that individual (Pasquetti, Apicella, and Mangone, 2014; Public Health Agency of Canada, 2014). It seems reasonable therefore to suggest that intrinsic risk factors have an increased role in determining fall events in the care home setting.

Intrinsic Risk Factors for Falls

Intrinsic risk factors for falls can be generally separated into age-related changes in physiology and the progression of pathological predisposing conditions (Pasquetti, Apicella, and Mangone, 2014). These age related physiological changes can be categorised into the deterioration of sight, hearing, musculoskeletal and central nervous system functioning (Pasquetti, Apicella, and Mangone, 2014).

Walking, standing still, and standing up or sitting down are all complex biomechanical activities, which rely on multiple bodily systems to act in a coordinated manner with sufficient muscle strength, joint mobility, and proprioception in the extremities. The interaction of multiple internal systems needed to maintain and recover balance mean that the disruption or failure of any one of these systems may lead to a fall (Rubenstein, 2006).

Beyond age related physiological changes, these systems can also be disrupted by a wide variety of pathological conditions which also increase the risk of falls. These can be grouped into categories including cardiovascular, internal medicine and endocrine, neurological, musculoskeletal, psychiatric, genitourinary, and iatrogenic (caused by drug side effects or immobilisation during treatment) (Rubenstein, 2006; Pasquetti, Apicella, and Mangone, 2014). These conditions can cause a fall

outright or make the individual more susceptible to extrinsic factors leading to a fall (Public Health Agency of Canada, 2014).

Cardiovascular conditions such as hypotension, and atrial fibrillation can cause dizziness, and fainting (Denfeld et al., 2022). Whereas further conditions such as heart disease can cause reduced proprioception in the extremities as a result of worsened blood flow, which can increase falls risk through reduced foot control and balance (Denfeld et al., 2022).

Internal medical conditions can cause a wide range of fall risk increasing symptoms. For example diabetes can lead to peripheral neuropathy, which reduces feeling in the feet and increases falls risk (Hicks et al., 2023). Furthermore, diabetic retinopathy reduces visual acuity making someone less able to perceive fall hazards (Schwartz et al., 2008). Additionally, if poorly controlled, diabetes can lead to hypoglycemia, which can cause dizziness and fainting (Schwartz et al., 2008).

Neurological conditions can have a range of effects that cause increased falls risk. For example conditions such as dementia, cerebrovascular disease, and epilepsy can cause wandering behaviour, delirium, worsened visual acuity, muscle weakness, balance and gait changes, dizziness, fainting, and seizures all of which have an associated increase in falls risk (Hauer et al., 2003; Lamb et al., 2003; Fernando et al., 2017).

Musculoskeletal conditions impact falls risk primarily through changes in muscle strength, joint pain, impaired balance, changes in gait, and reduced ability to recover from a loss of balance (Sturnieks et al., 2004; Rubenstein, 2006). Genitourinary conditions primarily impact falls risk through causing rushing behaviour as a result of urinary incontinence (Moon et al., 2021).

The most well documented relationship between a psychiatric condition and falls is depression. The mechanisms through which depression impacts falls are cognitive impairment, reduced walking speed, worsened reactions, and muscle weakness (Kvelde et al., 2013).

Finally, iatrogenic mechanisms include the prescription of fall risk increasing drugs, which have side effects including dizziness and fainting (Pasquetti, Apicella, and Mangone, 2014). Examples of medications linked with falls risk are antidepressants, benzodiazipines, and opioids (Bloch et al., 2011). Further iatrogenic

mechanisms include reduced muscle strength as a result of a period of immobilisation in hospital (Growdon, Shorr, and Inouye, 2017).

The literature review in Chapter 3 investigates the relationships between pathological conditions and falls in greater detail. However, there is a further intrinsic risk factor for falls, which cannot be grouped into a single category of pathological conditions. The following section provides a summary of the overlapping role of frailty in the falls risk of older adults.

1.3.2 The Role of Frailty in Falls

There is an intrinsic inter-relationship between falls and frailty (Yang et al., 2023). This means falls can be a partial cause of frailty as well as being a potential outcome of frailty. Additionally many of the causes of frailty are associated with increased falls risk (Clegg et al., 2013). This section will briefly summarise what frailty is, possible causes of frailty, the overlap with multi-morbidity, and evidence for the relationship between frailty and falls in older adults.

Frailty results from the cumulative decline of multiple physiological systems over the course of a lifetime (Clegg et al., 2013). Frailty is a complex health state, which is separate from disabilities and chronic disease although there is much overlap between these categories (Fried et al., 2001; Clegg et al., 2013). The frail state means someone is less able to return to homeostasis following stressor events such as falls (Clegg et al., 2013). The first operationalised definition in Fried et al. (2001) used three or more components of weight loss, weakened muscles, low endurance or energy, slow walking speed, and low physical activity level to identify frailty. The concurrent presence of these components was used to indicate the frailty phenotype, which refers to a cycle of reduced energy, activity, muscle mass and strength, slower walking speed, and chronic under-nutrition (Fried et al., 2001). These features deplete an individual's reserves, meaning they are less able to return to homeostasis following even minor stressor events and more susceptible to a range of worsened health outcomes (Clegg et al., 2013; Roe et al., 2017; Nghiem et al., 2021; Yang et al., 2023).

Frailty is often identified in older adults using clinical indexes such as the electronic frailty index (EFI), and hospital frailty risk score (HFRS) (Clegg et al., 2016;

Gilbert et al., 2018). Further description of these indices is presented in Section 4.2.1.

Most individuals with frailty also have multi-morbidity, although a causal relationship between these factors has not been proven (Vetrano et al., 2019). Using pooled data from nine studies, Vetrano et al. (2019) found 14% of people living with multi-morbidity are also frail, while 68% of people with frailty are also multi-morbid. This suggests that chronic diseases are a major cause of frailty, although only a small proportion of those with multi-morbidity will go on to develop frailty.

A fall can also act as an initial stressor event that leads to hospital admission and a period of immobilisation, which leaves the patient in a frail state even after they have recovered from the initial injury (Brown et al., 2009; Blain et al., 2016; Angulo et al., 2020). Additionally, falls can lead to the fear of further falls, and reductions in activity, which contribute to reduced muscle mass and strength. This reduced muscle mass is one of the components in the Fried et al. (2001) definition for frailty mentioned previously. Therefore, falls can act as a partial cause of frailty, whilst also being an associated outcome of frailty. This can create a spiral effect of a fall leading to future falls through the further development of frailty (Fried et al., 2001; Clegg et al., 2013; Pasquetti, Apicella, and Mangone, 2014).

However, the presence of frailty can also be associated with increased falls risk outside of this spiral effect. The functional decline which leads to frailty, can be seen as a combination of the intrinsic physiological factors discussed previously that mean frail individuals are at a higher risk of future falls (Yang et al., 2023). The mechanism through which frailty impacts falls is through the loss of muscle strength and balance, which were both mentioned in Section 1.3.1 during the discussion of intrinsic risk factors for falls.

The complex interrelated extrinsic and intrinsic risk factors make the prediction of falls in older adults challenging. Additionally, due to the differing risk environment between the community and care home settings, prediction model developed in one may not translate across to the other setting. These issues are discussed further in Section 1.4.

1.4 Falls Prediction in Older Adults

1.4.1 Introduction

Depending on whether a person is living at home in the community, or in a care home, they are likely to be exposed to a differing set of falls risk factors. As a result, models typically examine only one of these populations when modelling falls risk to allow for a better comparison of homogeneous cases (Gade et al., 2021b; Seaman et al., 2022; Shao et al., 2023). In an early systematic review of 38 falls risk and mobility assessment tools, Scott et al. (2007) concluded that due to few tools ever being tested in different populations, or across sub-populations, no single tool could be assumed to be validated across all older patient groups for predicting falls. Due to this lack of generalisability between residence types the discussion of approaches taken and findings are conducted separately for each location.

The differences in risk environment arise because people living in the community are likely to be healthier and more mobile than those in the care home setting, which may increase the number of opportunities to fall however, as discussed in Section 1.4.3, this does not translate into a higher prevalence of falls in the community (Department of Health, 2009). The comparatively higher levels of mobility may also contribute to a difference in the mechanisms leading to the fall, with an increased role for gait, balance and walking related risk factors in the community.

However, as seen in sections 1.4.2 and 1.4.3, even where the location of study is the same, studies are still highly heterogeneous with differences in study design, identification of falls, the statistical methods used in the analysis, and variables included in the models. In relation to the topic of this thesis, studies also differ in how they include multi-morbidity in their models, which is discussed separately for studies in both the community and care home setting in Section 1.4.4. Tables 1.1 and 1.1 provide a cross section of literature in the community and care home setting respectively to aid in this discussion.

TABLE 1.1: Cross Section of Fall Prediction Literature in the Community Setting

Author (Year)	**Country**	**Sample Size**	**Model Type**	**Study Type**	**Exclusions**	**Falls Measure**	**Follow up period**	**Mmorb measure**	**Method**	**Performance**
Kabeshova (2015)	FR	3289	Der	Retro	AI	FH	N/A	None	ANN	Sen 80.4%, Spec 92.54%
Richardson (2015)	IE	6666	Der	Prosp	Severe CI	SR	3 yrs	Main effects	PR	
Howcroft (2017)	CA	100	Der	Prosp	CI, MI	FC	6 mths	None	ROC	84.9% Acc, 50% Sen, 89% Spec
Marques (2018)	DE	102	Der	Prosp	CI, MI	PC	1 yr	None	ROC	Stance time variability (AUC 0.72, Sen 77.8%, and Spec 57.1%), swing time (AUC 0.25, Sen 88.9%, Spec 100%), stride length (AUC 0.97, Sen 77.8%, and Spec 92.9%)
Morin (2019)	SE	49609	Der	Retro CC	None	RD	1 yr	Binary term	LR	each additional drug OR 1.02 (1.01–1.03)
Cella (2020)	IT	96	Der	Prosp	CI, Speech impairment	DR	1 yr	None	LLinR	AUC 0.81 (0.72–0.90)
Bravo (2021)	PT	504	Der	Retro	CI	FH	N/A	Count of diseases	ROC	Predicting recurrent falls in history as the outcome, AUC 0.79 (0.75–0.83)
Gade (2021)	DK	198	Der	Prosp	CI, MI, AI	FC	1 yr	None	PR	MAE 0.88 falls
Lockhart (2021)	US	171	Der	Prosp	Unclear	Unclear		None	RF	81.6% Acc, 86.7% Sen, 80.3% Spec

Dormosh (2022)	NL	39342	Ext Val	Prosp	Died during follow up	Free text in EHR	1 yr	Main effects terms	LR	AUC 0.69
Van de Loo (2022)	NL, DE	5722	Der	Prosp	None	FC	1 yr	None	LR	AUC (any fall) 0.65, AUC (recurrent falls) 0.70
Jacob (2022)	IE	6900	Der	Prosp	Severe CI	SR	3 yrs	Binary term	LR	Mmorb (2+ conds) OR 1.32 (1.06-1.64), Mmorb (4+ conds) OR 1.92 (1.54-2.38)

TABLE 1.2: Cross Section of Fall Prediction Literature in the Care Home Setting

Author (Year)	**Country**	**Sample Size**	**Model Type**	**Study Type**	**Exclusions**	**Falls Measure**	**Follow up period**	**Mmorb measure**	**Method**	**Performance**
Marier (2016)	US	5129	Der	Retro	Missing data	MDS RD	N/A	Main effects terms	LR	AIC comparisson between models
Kuspinar (2019)	CA	88690	Der, Val	Prosp	None	Unclear	6 mths	Main effects terms	DT, LR	Validation across Canadian regions: odds of falling increase consistently across regions with increased risk score
Shaw (2019)	CA	116	Der	Prosp	Less than 2 years in facility	Fall incident reports in CH	3 yrs	None	ROC	Sen 93%, Spec 38%

Vlaeyen (2021)	BE	420	Val	Prosp	CI, severe MI	FC by care home staff	6 mths	Main effects terms	ROC	CaHFRiS: One month AUC 0.65 (0.58-0.72), Three month AUC 0.68 (0.62-0.73), Six months AUC 0.66 (0.61-0.72)
Boyce (2022)	US	3985	Der	Retro	Missing data	MDS RD	3 yrs	None	CART LR	AUC 0.67 (0.64-0.69), Sen (0.57), Spec (0.69)
Thapa (2022)	US	2785	Der	Retro	Missing data	Juniper Com-mu-nities RD	N/A	None	XG-Boost	AUC 0.85 (0.79-0.89), Sen 70.6%, Spec 85.0%
Duprey (2023)	US	733427	Der	Retro	Missing data	Medicare RD	2 yrs	None	LR	AUC 0.67 (0.66-0.67)

Country codes: *BE = Belgium, CA = Canada, DK = Denmark, FR = France, DE = Germany, IE = Ireland, IT = Italy, NL = Netherlands, PT = Portugal, SE = Sweden, US = United States* ***Study Characteristics:*** *Der. = Derivation study, Val. = Validation study, Ext. Val = External validation study, Prosp = Prospective , Retro = Retrospective, CC = Case control* ***Group Exclusions:*** *CI = Cognitive impairments or dementia, MI = Mobility impaired (cannot walk without assistive device), AI = Recent Acute Illness* ***Falls Measures:*** *FC = fall calendar, SR = Self report, RD = Routine Data, PC = Phone call interview, DR = Diary Reporting, FH = Fall history in interview or questionnaire, MDS = Minimum Data Set* ***Multi-morbidity measure:*** *Mmorb = Multi-morbidity* ***Statistical Method:*** *PR = Poisson Regression, LR = Logistic Regression, ANN = Artificial Neural Network, DT = Decision Tree, RF = Random Forest, CART = Classification and Regression Tree, ROC = Receiver Operator Curve threshold analysis, LLinR = Lasso linear regression, XG-Boost = Extreme gradient boosting algorithm* ***Performance Measure:*** *Sen = Sensitivity, Spec = Specificity, Acc = Accuracy*

1.4.2 Predicting Falls in Community Dwelling Older Adults

Table 1.1 presents a cross section of the falls prediction literature based in the community setting. Whilst the studies presented come from a range of countries, all are based in developed economies. However, the features extracted from the studies demonstrate causes of heterogeneity in the literature. The sample sizes in Table 1.1 vary greatly often dependent on whether an in-person assessment was required for the model to be developed versus the use of routinely collected health data. Furthermore, the vast majority of models are developed in derivation studies, and are rarely externally validated, which was also shown by a recent systematic review on the topic where of the 72 models included, only 3 related to the validation of existing models (Gade et al., 2021b). The lack of external validation of existing models is a consistent issue throughout the falls prediction literature, which severely limits the progression of the field because models are developed in heterogeneous samples using inconsistent definitions and sampling criteria, meaning few overarching conclusions can be drawn.

As seen in Table 1.1 community based studies differ in which groups of older adults to exclude based on characteristics such as cognitive impairments (Richardson, Bennett, and Kenny, 2015; Howcroft et al., 2017; Marques et al., 2018; Cella et al., 2020; Bravo et al., 2021; Gade et al., 2021a; Jacob et al., 2022), mobility impairments (Howcroft et al., 2017; Marques et al., 2018; Gade et al., 2021a), acute illnesses (Gade et al., 2021a; Kabeshova et al., 2015), and speech impairments (Cella et al., 2020). Restrictions on cognitively impaired individuals are often introduced to improve the data accuracy of self reported fall measures (Richardson, Bennett, and Kenny, 2015; Howcroft et al., 2017; Marques et al., 2018; Cella et al., 2020; Bravo et al., 2021; Gade et al., 2021a; Jacob et al., 2022). However, these restrictions mean the eventual models developed may not accurately reflect the falls risk experienced by the full older adult population, which reduces their external validity and usability in practice. By excluding cognitively impaired individuals or those at an advanced stage of illness from a study the resulting sample is less representative of the wider population meaning the model findings may be less generalisable to these patients. This is a particular issue when translating findings to care home residents where these

groups are more prevalent.

Within the falls risk prediction in the community literature, a sub-field has developed in the use of gait and balance sensor technology for deriving features to be used in falls risk prediction models (Howcroft et al., 2017; Marques et al., 2018; Cella et al., 2020; Lockhart et al., 2021). As seen in Table 1.1 these studies sometimes place exclusions on those with mobility impairments, which is due to initial assessments being made based on walking without ambulatory devices (Howcroft et al., 2017; Marques et al., 2018). These studies develop prospective models directly from sensor information collected during assessments of gait, balance, and walking after which participants are followed for up to one year to identify falls. While the resulting models often have high reported performance (AUC ≥ 0.7), problems in sample size (≤ 100) due to the need for in person assessment, and repeatability due to prototype or expensive medical devices prevent the results from being widely generalisable or externally validated (Howcroft et al., 2017; Marques et al., 2018; Cella et al., 2020; Lockhart et al., 2021). This difficulty in external validation is not shared by models making use of EHR data, which can be easier to translate across settings, however as seen in Table 1.1 further differences between studies in the identification of falls make the generalisation of results challenging.

The gold standard for prediction models is for training to be conducted using prospectively collected data. This means the assessment of predictors or explanatory variables must happen before a follow up period where falls can be recorded. By conducting the assessment of predictors before the fall event bias introduced through reverse causality, where the fall event changes the explanatory variable values can be avoided. This is particularly important in falls research because, as discussed in Section 1.2.1, falls have a range of health effects and can cause large changes in health outcomes in older adults (Rubenstein, 2006).

Some of the studies in Table 1.1 use a fall outcome based on the history of falls as the outcome for their prediction models and achieve high headline performance (Kabeshova et al., 2015; Bravo et al., 2021). However, by using variables measured after the fall to split their sample there is no proof to support whether the identified signal will be observable before the fall occurs, and instead only refers to splitting a sample into occasional or recurrent fallers in the past (Kabeshova et al., 2015; Bravo

et al., 2021).

Differences in how the studies in Table 1.1 measure falls also contribute to heterogeneity in their findings, performance, and identified relationships. Common approaches include the identification of falls through standard reporting in the EHR (Morin et al., 2019; Dormosh et al., 2022), phone interviews with participants on a regular timescale (Marques et al., 2018), diary or calendar based reporting (Howcroft et al., 2017; Cella et al., 2020; Gade et al., 2021a; Van De Loo et al., 2022), or self reporting falls history on a questionnaire (Kabeshova et al., 2015; Richardson, Bennett, and Kenny, 2015; Bravo et al., 2021; Jacob et al., 2022) with each method having strengths and weaknesses.

Several studies in the community setting have used routine EHR data for identifying falls (Morin et al., 2019; Dormosh et al., 2022). The advantage of using EHR data is that there is a reduced cost meaning larger samples can be used, which can improve the identification of relationships between exposures and outcomes, whilst also improving how representative the sample is within the wider population. Furthermore, the use of information contained in the EHR is already routinely collected in standard practice, which would reduce the barrier to application of the eventual modelling intervention. However, of the studies in Table 1.1 using EHR data, one used fall related admissions to hospital (Morin et al., 2019), and the other used free text in primary care records (Dormosh et al., 2022). Both of these methods for identifying falls introduce bias to the analyses through under-reporting of falls. Using hospital admission information introduces under-reporting because non-injurious falls will not result in hospital transfer and will therefore go unobserved (Morin et al., 2019). Identification in primary care free text may also result in under-reporting for the same reason as hospital admissions, but also because the lack of a standardised field may cause falls to be missed when analysing the free text data (Dormosh et al., 2022). Therefore, while using the EHR to identify falls can allow larger samples there is a trade off in the lack of a reliable and complete fall indicator.

An alternative approach to using standardised sources of information for the identification of fall events is to collect fall information directly from the study participants. Whether this is done prospectively or retrospectively has a direct effect on the quality of information collected with the latter more exposed to recall bias

(Mackenzie, Byles, and D'Este, 2006; Fleming, Matthews, and Brayne, 2008). Additionally, previous research has found older adults significantly under-report falls in retrospective questionnaires (Peel, 2000). These methods also necessitate the exclusion of certain patient groups due to their inability to engage with the collection method (Richardson, Bennett, and Kenny, 2015; Howcroft et al., 2017; Marques et al., 2018; Cella et al., 2020; Gade et al., 2021a; Jacob et al., 2022). This means models based on these approaches to measuring falls may face generalisability problems due to non-representative samples, and inaccurate event rates, which can lead to biases in the associations between exposures and outcomes in the models. These problems are particularly important when considering the application of findings from these community based studies to care home residents as the sample of study due to the increased prevalence of cognitive impairments in this patient group.

There is no single best modelling approach for the prediction of falls in older adults. In previous systematic review of prognostic models for falls prediction in community dwelling older adults, all the included studies were identified as having a high risk of bias due to the differences in statistical methods, study inclusion criteria, and assessment of the falls outcome (Gade et al., 2021b). Studies in Table 1.1 used a range of standard statistical approaches such as logistic regression (Morin et al., 2019; Dormosh et al., 2022; Van De Loo et al., 2022; Jacob et al., 2022), Poisson regression (Richardson, Bennett, and Kenny, 2015; Gade et al., 2021a), lasso linear regression (Cella et al., 2020) and identification of optimal thresholds using the Receiver-Operator-Curve (ROC) (Howcroft et al., 2017; Marques et al., 2018; Bravo et al., 2021). Additionally, machine learning approaches are not uncommon with examples of Artificial Neural Networks (ANN) (Kabeshova et al., 2015), and Random Forest (Lockhart et al., 2021) seen in the Table 1.1 literature cross-section. When considering differences in model performance however, differences caused by the choice of modelling approach are secondary to the sources of heterogeneity between studies discussed previously in this section. Additionally, differences in the way studies report results and performance makes direct comparison challenging. However, there are several overarching themes that can be drawn out of the community dwelling falls prediction models relating to the role of particular variables, which are consistently included in the models.

Several variables are consistently found to significantly impact falls risk in the community setting. In their systematic review of prognostic models in the community setting Gade et al. (2021b) identified Falls history as the most commonly included predictor across the 72 included models. In the prospective studies presented in Table 1.1, a history of falls was consistently found to significantly contribute to model predictions (Cella et al., 2020; Gade et al., 2021a; Dormosh et al., 2022; Van De Loo et al., 2022). Further variables relating to functional assessments such as the timed up and go (TUG) test, and activities of daily living (ADL), are also often included by authors (Kabeshova et al., 2015; Cella et al., 2020; Bravo et al., 2021; Van De Loo et al., 2022; Jacob et al., 2022). These functional assessments are consistently significantly associated with falls risk, however in the Kabeshova et al. (2015) study, the use of an ANN model meant the contribution of any variable to the model was unclear (Bravo et al., 2021; Cella et al., 2020; Van De Loo et al., 2022; Jacob et al., 2022). The inclusion of other variables in models beyond demographic characteristics such as alcohol consumption (Kabeshova et al., 2015; Morin et al., 2019; Gade et al., 2021a), or hazards in the environment (Morin et al., 2019; Bravo et al., 2021) are more inconsistent and their inclusion likely relates to data availability.

Further sets of variables which are consistently included in community based models are medications and polypharmacy, chronic diseases, and multi-morbidity. However, community based studies are inconsistent with how these variables are included in models, which when combined with the complexity of the mechanisms being modelled means their role in falls risk is less clearly identified than the more direct features such as falls history or functional impairments.

While several of the studies in Table 1.1 consider the role of polypharmacy (Kabeshova et al., 2015; Cella et al., 2020), four studies provide an opportunity for inference due to the size of their samples and interpretable statistical methods used (Richardson, Bennett, and Kenny, 2015; Morin et al., 2019; Van De Loo et al., 2022; Jacob et al., 2022). Two of these studies used consecutive waves of the Irish longitudinal study of ageing to prospectively analyse the role of polypharmacy on falls. Richardson, Bennett, and Kenny (2015) found polypharmacy only impacted falls risk in the presence of anti-depressants and benzodiazepines, and Jacob et al. (2022) identified polypharmacy accounted for 13% of the identified relationship between

falls and multi-morbidity (measured using a binary indicator). Medications relating to anti-Parkinson's, anti-epileptics, urinary frequency and incontinence, and antihistamines were all found to exhibit significant risk increasing effects for falls in models based in German and Dutch cohorts (Van De Loo et al., 2022). Additionally, a further study using a case-control design found a small significant risk increasing effect for each drug prescribed on a patient record (Morin et al., 2019). These findings all indicate a role for medications and polypharmacy in determining falls risk; however as seen in the Jacob et al. (2022) study there is an inter-relationship present between these medications, chronic diseases, and multi-morbidity.

The role of chronic diseases in falls risk is an extensive topic, which has motivated a wide range of studies, systematic reviews, and meta-analyses. The question of which chronic diseases contribute to increased falls risk motivated the review of reviews presented in Chapter 3. Commonly included chronic diseases in community based studies are depression, urinary incontinence, and cognitive impairment or dementia (Richardson, Bennett, and Kenny, 2015; Kabeshova et al., 2015; Gade et al., 2021a; Dormosh et al., 2022; Van De Loo et al., 2022; Jacob et al., 2022; Bravo et al., 2021). How studies incorporate multi-morbidity into predictive models and a discussion of evidence for a relationship with falls risk is addressed in Section 1.4.4.

In conclusion, studies of falls risk prediction in the community setting are highly heterogeneous due to differences in study design, exclusions of groups, methods of identifying falls, statistical methods, variables included, and reporting of results. Further external validation studies are needed to ensure lessons can be learned regarding which models generalise well to new samples. Additionally, as discussed in Section 1.4.4, studies in the community and care home settings take similar approaches to identifying and incorporating multi-morbidity into their prediction models, which could be improved by applying methods from the multi-morbidity clustering literature. However, while there are many similarities with the care home setting, models developed in the community are unlikely to be generalisable to the care home setting, which is discussed further in Section 1.4.3.

1.4.3 Predicting Falls in Care Home Dwelling Older Adults

When compared to the community setting there are fewer predictive models for falls risk developed in the care home setting. This is likely due to a combination of factors such as the availability of high quality data, and complexity of mechanisms and relationships due to the prevalence of multi-morbidity and frailty in the care home setting (Burton et al., 2021). The care home environment is different to the community setting both in terms of the extrinsic and intrinsic risk environment. Care homes represent a more controlled environment than the community setting, where proactive steps are taken to mitigate or eliminate extrinsic risk factors for falls (Care Inspectorate, 2016; Cooper, 2017). This means there are likely to be fewer opportunities for falls driven by environmental causes alone in a care home compared to the community setting. However, as discussed in Section 1.2.1 falls are three times more common in care home residents, meaning increases in the role of intrinsic risk factors must override any reductions in extrinsic risk (Nurmi et al., 2009). As summarised in Section 1.3.1 intrinsic risk factors relate to changes in physiology caused by ageing, and the progression of pathological conditions over time. Due to the reductions in extrinsic risk, and changing role of intrinsic risk factors, models developed for use in the community setting may not generalise well to the care home setting. The cross section of studies summarised in Table 1.2 are used as a basis for the discussion in this section.

The remainder of this section summarises the key themes in the falls risk prediction literature for older adults living in care homes and contrasts these with the discussion in Section 1.4.2. First decisions in the derivation of samples are discussed, alongside the datasets used. Following this methods of identifying falls are contrasted to those used in the community setting. Next the major variables included in models, and their relative importance to the predictions are discussed, which provides the basis for the discussion of methods for identifying multi-morbidity in Section 1.4.4.

One of the major differences to studies based in the community is the relative lack of data sets available to study falls in care home residents (Burton et al., 2021). This is shown by several studies based in the USA and Canada in Table 1.2, which

all made use of the Minimum Data Set (MDS) for Nursing home residents (Saliba and Buchanan, 2008b; Marier et al., 2016; Kuspinar et al., 2019; Shaw et al., 2019; Duprey et al., 2023; Boyce et al., 2022). Further studies used bespoke databases developed by the Juniper communities chain of care homes in the USA (Thapa et al., 2022), and across 15 Belgian nursing homes (Vlaeyen et al., 2020). The lack of a similar database to the MDS in the UK for care home residents has hindered the development of domestic falls risk prediction models, a point which is returned to in Section 6.7 (Burton et al., 2021).

An additional difference between care home and community based studies is the sample inclusion constraints imposed by authors. Exclusion criteria in care home based studies are often restricted to data availability and completeness rather than eliminating groups from the analysis (Marier et al., 2016; Shaw et al., 2019; Boyce et al., 2022; Thapa et al., 2022; Duprey et al., 2023). The only study in Table 1.2 to restrict their sample due to cognitive or mobility impairments was Vlaeyen et al. (2020), compared to seven studies in Table 1.2 (Richardson, Bennett, and Kenny, 2015; Howcroft et al., 2017; Marques et al., 2018; Cella et al., 2020; Bravo et al., 2021; Gade et al., 2021a; Jacob et al., 2022). This difference in the number of exclusion criteria results from how the explanatory and outcome variables are assessed in studies based in the care home. In community based studies, data on exposures and outcomes are often collected directly from the individual through questionnaires, and surveys (Kabeshova et al., 2015; Richardson, Bennett, and Kenny, 2015; Howcroft et al., 2017; Marques et al., 2018; Cella et al., 2020; Bravo et al., 2021; Gade et al., 2021a; Van De Loo et al., 2022; Jacob et al., 2022). However, this is in contrast to care home based studies, which often use data collected from external observations or routine data sources rather than responses from the individual (Marier et al., 2016; Shaw et al., 2019; Vlaeyen et al., 2020; Boyce et al., 2022; Thapa et al., 2022; Duprey et al., 2023). An example of this is how falls are identified between the two settings, with fall calendars (Howcroft et al., 2017; Gade et al., 2021a; Van De Loo et al., 2022), phone interviews (Marques et al., 2018), and self reporting (Richardson, Bennett, and Kenny, 2015; Jacob et al., 2022) common in the community setting. In contrast, the majority of studies in Table 1.2 used external measures such as routine data fields (Marier et al., 2016; Boyce et al., 2022; Thapa et al., 2022; Duprey et al., 2023), and

care home staff reports (Shaw et al., 2019; Vlaeyen et al., 2020). This means the accuracy of the data in these care home studies is not necessarily dependent on the health state of the individual, which generally allows them to incorporate a wider range of health states than studies in the community. The differences in the types of information available further contributes to difficulty in generalising models developed in the community setting to care homes.

Some of the variables included in falls risk prediction models in the care home setting can be separated into similar groupings to those seen in the community setting in Section 1.4.2. Similar to the community setting, a measure of falls history is consistently included in care home based models with a similar fall risk increasing effect identified (Marier et al., 2016; Shaw et al., 2019; Vlaeyen et al., 2020; Boyce et al., 2022; Thapa et al., 2022; Duprey et al., 2023). A second grouping of explanatory variables consistently seen in the community setting were functional assessments of mobility such as the TUG test and ADL score (Kabeshova et al., 2015; Cella et al., 2020; Bravo et al., 2021; Van De Loo et al., 2022; Jacob et al., 2022). Results from community based studies using interpretable models found that functional impairments measured through these scales were associated with an increase in the risk of falls (Jacob et al., 2022; Van De Loo et al., 2022). However, a single care home based study found that increased mobility and independence was associated with higher falls risk in their care home sample (Duprey et al., 2023). This gives an indication of how relationships between explanatory variables and falls can change between settings, which may reduce the generalisability of models developed in the community setting when applied to the care home. Despite the increased prevalence of disease in care homes compared to the community setting, indicators for the presence of chronic diseases that are known to increase falls risk are not consistently included in care home based models, with only three of the seven studies in Table 1.2 including them (Marier et al., 2016; Kuspinar et al., 2019; Thapa et al., 2022). The role of chronic diseases in falls risk is addressed further in the Chapter 3 review of reviews. Inclusion of multi-morbidity information in models was also lacking in the studies shown in Table 1.2, which is discussed further in Section 1.4.4.

This section has shown that differences in the availability of data contributes to fewer falls risk prediction models being developed for use in care homes compared

to the community setting (Burton et al., 2021). Furthermore, the generalisability of models developed in the community comes into question due to the types of information incorporated into the models, and changes to the relationships identified as a person's health state worsens. However, models developed in both the community and care home settings do not effectively incorporate multi-morbidity into their prediction models, which is discussed further in Section 1.4.4. This presents the opportunity for the research in this thesis and forms the final piece of background before the research aim and objectives are stated in Section 1.5.

1.4.4 Incorporating Multi-morbidity into Falls Prediction

As discussed in Section 1.2.2 people are living longer with multiple chronic health conditions, which means the relationships between chronic health conditions and how they effect falls risk is an important area for study (Bergen, Stevens, and Burns, 2016; Kingston et al., 2018; Florence et al., 2018). Multi-morbidity was defined in Section 1.1 and refers to the presence of two or more chronic health conditions in the same individual (Johnston et al., 2019). In a situation where the size of the older adult population and longevity with multi-morbidity are both increasing, understanding the complex relationship between different types of multi-morbidity and falls will be increasingly important. However, there is a lack of extensive research into the links between multi-morbidity and falls in older adults. Furthermore, when multi-morbidity has been included as a predictor in falls risk models there is often over simplification present. The discussion in this section explores the different methods of incorporating multi-morbidity seen in studies based in the community and care home settings and how these methods fail to capture variation in multi-morbidity presentations, which will have consequences for the relationships identified between multi-morbidity and falls. Following this, alternative approaches to identifying multi-morbidity patterns are summarised, to provide the basis for the research question, aim and objectives in Section 1.5.1.

Similar methods are used to identify multi-morbidity in studies identifying falls risk, fall related injury, and fall related hospitalisation in the community setting. These methods include counts of chronic conditions on record (Teixeira et al., 2019;

Bravo et al., 2021), binary indicators for thresholds of the number of chronic conditions on record (Afrin et al., 2016; Morin et al., 2019; Gade et al., 2021a; Barik et al., 2022; Jacob et al., 2022; You et al., 2023), and using index values such as the age-adjusted Charlson comorbidity index (McCoy et al., 2017; Garu et al., 2021). As seen in Table 1.2, care home based studies predicting falls do not effectively account for multi-morbidity beyond identifying the presence of the individual chronic conditions (Marier et al., 2016; Kuspinar et al., 2019; Vlaeyen et al., 2020).

Identifying the effect of including sub-optimal measures of multi-morbidity into models is challenging due to the heterogeneity between studies discussed in Sections 1.4.2 and 1.4.3. However, the presence of multi-morbidity is consistently identified as significantly increasing the risk of falls across settings and outcome measures (Afrin et al., 2016; Teixeira et al., 2019; Jacob et al., 2022; You et al., 2023; Barik et al., 2022). Additionally, during ROC-threshold analysis, Bravo et al. (2021) found the count of chronic conditions on record achieved a maximum AUC of 0.65 (Sensitivity = 58.1%, Specificity = 63.8%) when discriminating between recurrent fallers ($\geq$2 falls) from occasional fallers ($\leq$ 1 falls).

When identified through a binary threshold, the presence of multi-morbidity ($\geq$ 3 chronic conditions in this study) was associated with increases in all ($\geq$ 1) and recurrent ($\geq$ 2) falls risk in post-menopausal women, with odds ratios of 1.82 (1.56-2.13) and 1.41 (1.24-1.60) respectively when adjusted for age, medications, and smoking status (Afrin et al., 2016). Further evidence from a cross sectional analysis of a Chinese cohort of community dwelling older adults found multi-morbidity, measured through a binary threshold, significantly increased the risk of repeated falls, with a reported odds ratio of 3.45 (1.47-6.97), when adjusting for demographic, geographical, and activity variables (You et al., 2023).

When measured through the count of chronic health conditions, Jacob et al. (2022) identified a prospective association between multi-morbidity and falls in Irish community dwelling older adults, with an increase in falls risk for two conditions on record (OR = 1.32, CI = 1.06-1.64). Although the largest increase in this study was observed for four or more chronic conditions (OR = 1.92, CI = 1.54-2.38) when controlling for demographic variables, alcohol consumption, and falls history (Jacob

et al., 2022). In a systematic review and meta analyses of falls risk factors in community dwelling older adults, Deandrea et al. (2010) derived a pooled odds ratio of 1.23 (1.16-1.30) for the association of incremental co-morbidity with falls based on 10 studies. While providing a greater indication of the gradient of falls risk effects as a result of overall condition burden compared to the binary threshold approach, measuring the number of chronic conditions on record does not differentiate further between the types of multi-morbidity present, which may change the strength of association with falls. This evidence still does not demonstrate whether multi-morbidity effects persist when controlling for the presence of key chronic diseases. Additionally, by using a binary threshold or a count of the chronic conditions present as an indicator for multi-morbidity, authors are inadvertently grouping heterogeneous types of multi-morbidity together, which may have a range of effects on falls. This means the overall multi-morbidity effects identified may be overestimating the importance of certain multi-morbidity presentations, and underestimating the importance of others.

Another alternative approach is to use an index of co-morbidity, where the presence of specific chronic conditions are scored differently based on various criteria. McCoy et al. (2017) found each additional point on the age adjusted Charlson co-morbidity index (AAC) was associated with an odds ratio of 1.05 (1.04-1.05) when predicting hospital admission for fall-related injury. Additionally, Garu et al. (2021) developed a model to predict fragility fractures following falls, finding AAC values of six or higher were associated with an odds ratio of 1.77 (1.14-2.73). The age adjustment in the AAC refers to adding one point for each decade of age over 40 (Charlson et al., 1994). However, the core Charlson co-morbidity index was initially intended to act as a predictor of 10 year survival, meaning the presence of several falls risk increasing chronic conditions are not accounted for on this scale (Charlson et al., 1987). Additionally, using an index scale does not aid in understanding the mechanisms through which multi-morbidity causes changes in falls risk.

In order to understand the relationships between different multi-morbidity patterns and falls risk, first groups of commonly occuring multi-morbidity presentations need to be identified. If reproducible groups of multi-morbidity can be identified and incorporated into falls risk prediction models, this could improve the way

multi-morbidity is handled in these models in future. However, whilst multiple studies have addressed how multi-morbidity presentations differ throughout the older adult population, links between these different patterns and falls have not be investigated.

In the multi-morbidity pattern literature, the focus is on understanding groups of chronic health conditions that regularly co-occur in older adults in an effort to identify disease progressions, possible susceptibilities, and interactions between conditions. Recent advances in computing capacity has motivated authors to apply approaches from the data mining and dimensionality reduction spheres to this problem (Vu, Finch, and Day, 2011; Islam et al., 2014; Violán et al., 2018; Machón et al., 2020).

A common approach is to use cluster analyses to group chronic health conditions together (Prados-Torres et al., 2014). For example, using multiple correspondence analyses (MCA) to reduce the dimensionality in the multi-morbidity data followed by a non-hierarchical clustering solution has been applied to analyse multi-morbidity patterns in community dwelling older adults in Spain to identify whether multi-morbidity patterns can be identified (Violán et al., 2018; Machón et al., 2020). Through these studies common presentations of multi-morbidity were identified relating to cardiovascular-metabolic-musculoskeletal-tobacco consumption cluster, a cluster of musculoskeletal system conditions and connective tissue, and co-occuring diabetes and retinopothy (Violán et al., 2018; Machón et al., 2020). In a systematic review of 63 multi-morbidity patterns from 14 articles, Prados-Torres et al. (2014) identified that a cardiovascular-metabolic pattern, a mental health pattern, and a pattern of musculoskeletal conditions were consistently identified throughout the literature. However, no studies in this review were conducted in the UK, meaning further research is needed to asses whether the multi-morbidity patterns identified in previous studies in other countries are applicable to the UK context (Prados-Torres et al., 2014).

As discussed at the beginning of this section, the current approach of including numbered or threshold based variables does not reflect the complexity present in multi-morbidity information. Additionally, previous research into multi-morbidity patterns has focused on community dwelling older adults outside the UK (Vu, Finch, and Day, 2011; Islam et al., 2014; Prados-Torres et al., 2014; Violán et al., 2018;

Machón et al., 2020). Therefore, there is an opportunity for novel research to be conducted, which simultaneously addresses the lack of multi-morbidity studies into care home residents based in the UK, and the over-simplification of multi-morbidity information in falls risk prediction modelling. These gaps motivate the research question, aims, and objectives of the analyses, which are presented in Section 1.5.1.

1.5 Aims, Objectives, and Outline of the Thesis

1.5.1 Research Question, Aims, and Objectives

As described in this chapter, falls represent a major source of injury, morbidity, mortality, and healthcare resource use in older patient groups worldwide. In the context of an ageing population, an automated falls risk flagging system based on information routinely stored in the EHR is very desirable to aid practitioner decision making and enable timely referral of patients at a high risk of falls to prevention interventions.

A key feature of the ageing population is that patients are living longer with multiple chronic health conditions, or multi-morbidity. This means that any eventual falls risk flagging system must be able to account for the variation in falls risk that different multi-morbidity combinations may cause. Existing models of falls risk in older adults fail to effectively represent the complexity present in multi-morbidity data, meaning they may be unsuitable for wide-scale use. However, useful techniques to categorise and group multiple chronic health conditions used in the wider multi-morbidity literature provide an opportunity to explore the effect of multi-morbidity on falls risk. Therefore, this thesis investigates patterns of chronic diseases that regularly co-occur in UK care home residents, the associations between these patterns and falls risk, and how specific risks can be quantified using statistical models based on data contained in routine EHRs.

The overarching research question for this thesis is "*what is the contribution of multi-morbidity in determining the risk of falls by care home residents aged over 65 and how can these effects be identified in statistical models?*" This question is addressed via the following research aim and objectives.

Research Aim:

Multi-morbidity has not been satisfactorily included in falls risk prediction models intended for use in the community or care home settings. Therefore the differential effects of different multi-morbidity presentations on falls risk is not well understood. The overall aim of the research in this thesis is to develop an understanding for the role of multi-morbidity in determining the falls risk of older adults residing in UK care homes. Once patterns of multi-morbidity were identified using established techniques, links between these patterns and falls risk were thoroughly investigated using statistical models.

Research Objectives:

More specifically, the objectives of the research are to:

1. Review previous research to find evidence for which chronic health conditions contribute to increased falls risk in community dwelling older adults.
2. Identify patterns of chronic health conditions that regularly co-occur in UK care home residents.
3. Establish whether any patterns of multi-morbidity change the risk of falls in care home residents.
4. Determine the association of individual chronic conditions with falls in the sample of UK care home residents.
5. Investigate how specific combinations of chronic health conditions contribute to the changes in falls risk identified for each multi-morbidity pattern.
6. Examine whether the contributions of individual chronic health conditions can explain the effects on falls risk observed for patterns of multi-morbidity.

1.5.2 Thesis Structure

The initial thesis plan was to develop a prediction model for use in the pre-hospital setting to stratify the risks of different outcomes in trauma patients. Such a model was intended to aid in easing the demand pressures faced at the emergency department through the identification of low-acuity cases in the pre-hospital setting. For

this reason, a systematic review of mortality prediction models following traumatic brain injuries in adults was undertaken and is presented in Chapter 2.

During the completion of the Chapter 2 systematic review, the research focus shifted to presentations by older adults to the emergency department. This shift occurred due to several converging themes identified in the wider literature and Chapter 2 review. These themes were the challenges the health system will face from an ageing population, the role of older adult presentations as a cause of emergency department crowding, and the consistent significance of age in determining TBI outcomes identified during the 2 review.

This new focus on older adults meant a statistical model to predict low acuity presentations would be less appropriate due to the increased clinical need and complexity of older adult presentations. Alongside this shift in focus, the opportunity arose to work with a unique dataset, which identified care home residents and their interactions with hospital and community care resources. Therefore, the Chapter 2 represents a key turning point for the project, and the experience gained when conducting this review directly impacted the approach taken in the subsequent Chapter 3 review.

Following the shift in focus to older adults, a gap in the literature was identified surrounding the role of multi-morbidity in falls and fall presentations to emergency departments by the care home residents. A Review of Systematic Reviews was conducted in Chapter 3 to distill the findings from studies that assessed the role of chronic diseases in causing falls in community dwelling older adults. Findings from the community were used due to the lack of studies assessing the relationship between chronic diseases and falls in care home residents. As a result of this review, a shortlist of the chronic health conditions that were consistently associated with increased falls risk was developed and included in later regression models.

Chapter 4 presents the methodology for the analyses in this thesis, drawing a line between the falls risk prediction literature, and multi-morbidity pattern analysis literature to present a novel approach to addressing the research question. Additionally, an overview of the data to be used in the analyses is presented in this chapter, alongside a description of the steps taken to process the chronic health condition information identified from inpatient records. This chronic health condition

information was then used in conjunction with the shortlist derived in the Chapter 3 review of reviews to differentiate the effects of individual chronic health conditions on falls risk, and compare these findings to the patterns of multi-morbidity identified. By demonstrating how no single chronic health condition was responsible for the changes in falls risk associated with the patterns of multi-morbidity, it was then possible to suggest a gradient of effects on falls risk based on the type of multi-morbidity present.

The results of the cluster analysis and regression models are presented in the Results (Chapter 5). Finally, in Chapter 6 the contextualisation of these results in the wider literature, discussion of study limitations, recommendations for future research, and the overall reflection and conclusions of the thesis are provided. While there are still many questions remaining for the role of multi-morbidity in falls risk, this is the first research which has identified a gradient of falls risk effects caused by different patterns of multi-morbidity in UK care home residents.

Chapter 2

Comparing Mortality Prediction Models Following Traumatic Brain Injury: A Systematic Review

2.1 Chapter Introduction

This chapter investigates the effectiveness of existing ML models for the prediction of outcomes following Traumatic Brain Injuries (TBI) in adults. The topic for the systematic review was chosen to investigate the effectiveness of innovative methods for predicting ongoing care needs in patients accessing emergency care. As discussed in Section 1.5.2, the focus of the thesis research shifted during the completion of this systematic review from avoidable attendances to fall attendances to emergency departments by care home dwelling older adults. The shift towards focussing on care home residents was partially motivated by an underlying theme identified in this systematic review, where the differences in outcomes in older patients identified in multiple studies motivated the further investigation of this patient group. This further investigation highlighted how low falls are a common cause of TBI in older patients, which led to the development of the wider project surrounding falls in care home residents aged over 65 (Lawrence et al., 2016).

2.2 Review Introduction

Traumatic Brain Injury (TBI) is a major global cause of death and disability (James, 2019). Rapid determination of the appropriate treatment path for patients with TBI is crucial to maximise the chances of optimum outcomes. Statistical models predicting mortality risk in patients with TBI are intended to aid with this determination by providing information on likely outcomes in the early stages of patient management at the Emergency Department (ED) or Intensive Care Unit (ICU). This information can be used by clinicians to better manage patient needs whilst also giving families justified indications of likely prognosis to help with counselling and managing expectations. In this review I compare the performance of mortality prediction models in adult patients following TBI. The findings from this review were intended to motivate decisions in model development and evaluation, based on the original analysis plan. However, limitations in the data available at the care home level motivated the change in direction for the analyses, which is presented in Chapter 4.

TBI is a form of acquired brain injury where sudden trauma causes damage to the brain. This can be caused by blunt force or when an object penetrates the skull and enters brain tissue (National Institute for Neurological Disorders and Stroke, 2019). A widely used measure of TBI severity is the Glasgow Coma Scale (GCS), where symptoms are classified with a score of 3-15 and a lower score indicates more severe TBI (Teasdale and Jennett, 1974). TBI severity can be grouped into three categories on the GCS: Mild (GCS 13-15), Moderate (GCS 9-12) and Severe (GCS < 9). The focus of this review is moderate-severe TBI because the mortality rate is highest in these categories meaning the potential benefit from early identification and aggressive treatment is higher in this group.

Two individual level prediction models have been developed and validated in large multinational cohorts of patients. The International Mission for Prognosis and Clinical Trial (IMPACT) and Corticosteroid Randomization after Significant Head Injury (CRASH) scores incorporate demographic characteristics, lab results and CT-imaging findings to make their predictions (Maas et al., 2007; Perel et al., 2008). However, these scores are unable to take account of treatment protocols and only make a prediction at admission, rather than updating them dynamically over time.

In addition, they were established over a decade ago meaning incorporation of new evidence and the effect of improved CT-scanning techniques makes updated models necessary.

Whether to use standard statistical approaches or Machine Learning (ML) algorithms for the prediction of binary outcomes such as mortality is an open debate in the academic literature. While the methods may seem similar on the surface, they differ in their fundamental approach taken to identifying relationships in data. At the most basic level standard statistical techniques rely on models where the outcome is related to the predictors based on some function of variable values, stochastic noise, and estimable parameters based on an assumed distribution of observations. In this way, standard statistical approaches attempt to estimate the natural process, that generated the data in the first place. ML algorithms take a different approach however, where the natural link between the predictors and outcomes is seen as complex and unknown. The focus of ML is instead to develop algorithms that best reproduce the output values given the predictor information without making assumptions about the form this relationship will take. In this way ML algorithms can model complex non-linear relationships between variables, although this is often achieved at the cost of interpretability of the findings.

Logistic regression remains the most popular algorithm for binary outcome prediction. However, recent developments in computing capacity have raised the question of whether more complex algorithms can be applied to this setting for improved predictive performance. ML is an automated process, which extracts patterns from data. Supervised ML methods use this automated process to identify a pattern between input variable and outcome values based on historical examples such that predictions can be made in new instances (Kelleher, Mac Namee, and D'Arcy, 2015). The two main phases in ML are model training and evaluation. Training refers to the development of a model based on past examples, where parameters within the model are fine tuned for improved performance. Evaluation is the phase where the predictive performance of a trained model is assessed, ideally in unseen data. Key terms for understanding the ML literature and methods of training and evaluation are provided in the Glossary of terms. Additionally, a description of the methods encountered during this review is provided in Appendix Table B.

The clinical usefulness of a prognostic model is dependent on its clinical and methodological validity. Therefore, interrelationships between features identified in the data must be reflective of clinically meaningful relationships rather than resulting from coincident features of the sample population. While the selection of ML algorithm and corresponding hyperparameters (see the Glossary of Terms) is a key part of this, the quality of the input data is of paramount importance for valid and generalisable models to be developed.

Previous systematic reviews on the application of ML algorithms to head trauma and emergency care settings found methodological quality and reporting is often poor with many studies suffering from an elevated risk of bias (Perel et al., 2006; Miles et al., 2020). This review aims to investigate the methodological validity, performance, and generalisability of ML models for mortality prediction in adult patients with moderate to severe TBI.

2.3 Systematic Review Methodology

The study population for this systematic review is adult patients (16+) in acute care settings with moderate to severe Traumatic Brain Injury (TBI), defined as a score of 12 or below on the Glasgow Coma Scale (GCS). The method of interest is supervised or unsupervised prediction modelling methods where one of the outcomes is mortality or, conversely survival. The outcomes for review are the reported discrimination (C-statistic/AUC), sensitivity, specificity, accuracy, negative predictive value, positive predictive value, and calibration statistics.

Four bibliographic databases (Medline, Web of Science, Embase and CINAHL) were searched using headings, thesaurus terms and key words related to the review title on 12/02/2021. These key words were based around the concepts: 'Traumatic Brain Injury', 'Prediction modelling', 'machine learning and regression', 'outcome of mortality or survival' and 'acute care settings'. To reduce the risk of missing papers relevant to the review, exploded headings and truncations were used in the searches. The full search strategy is available the Appendix Section A. The search results were filtered to studies published between 2015 to 2021 prior to download. Studies were

limited to publication from 2015 onwards to reflect recent advances in computing capacity, which have made ML models more feasible.

Full citations and abstracts of included papers were downloaded into MendeleyTM Version 1.19.8 and duplicates were removed. A single reviewer (SW) screened the titles and abstracts against the inclusion criteria below, recording the results in an Microsoft Excel^{TM} spreadsheet.

Inclusion Criteria:

1. Study based in Acute Care Setting (Emergency department, hospital ward, Intensive Care Unit)
2. Study population is adult patients with moderate to severe traumatic brain injury (defined as 12 or below on the Glasgow Coma Scale (GCS) or other validated measure)
3. Study presents machine learning, deep learning or regression model for outcome prediction
4. One model outcome is either patient mortality or survival
5. Accuracy, discrimination or calibration of predictions reported
6. Study is written in English language

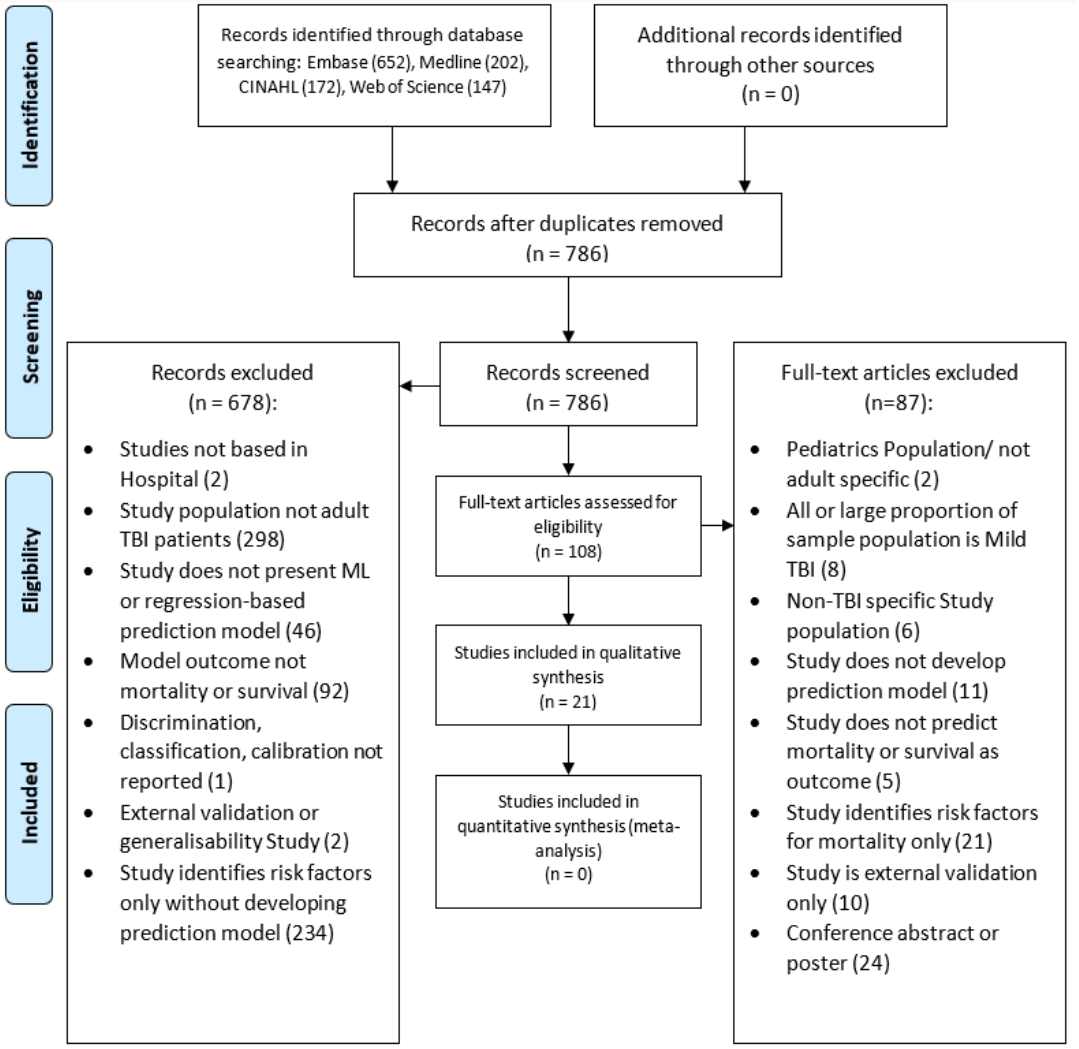
Exclusion Criteria

1. External validation or generalisability study of existing prediction model where no additions or refinements are made
2. Study only identifies risk factors of mortality in TBI patients without developing a prediction model
3. Reviews of previous literature

(Specific age cohorts, for example, studies only including patients aged over 65, were not considered grounds for exclusion in this review.)

Following the initial inclusion judgement based on title and abstract, full texts of included papers were downloaded into Mendeley. A sift of the full texts was performed to ensure studies met the inclusion criteria. The reasons for exclusion at this stage are presented using a PRISMA flowchart in Figure 2.1 alongside the search results. Eighty-seven articles were excluded at the full text stage for the following reasons. 16 studies used populations outside those defined in the inclusion criteria, of these two were based in paediatric populations, eight had a high proportion of mild TBI cases and six were based in non-TBI specific populations such as general trauma. A further 37 studies did not meet the inclusion criteria relating to the models used. This group comprised five studies that did not predict mortality or survival, 11 studies which did not develop prediction models at all, and 21 that only investigated risk factors for the outcome. Ten studies involved externally validating an existing prediction model or score and the remaining 24 were conference abstracts and posters. Following the full-text screening stage, 21 studies were included for data extraction and evidence synthesis.

FIGURE 2.1: PRISMA Flowchart of Search Results



Data were extracted from included studies in the following areas: data source used, country of sample data, descriptive statistics of study population, handling of missing data, prediction model type (and number, if more than one was present), outcomes predicted by the model, explanatory variables included, discrimination, calibration, and classification statistics. This information was organised in a Microsoft Excel^{TM} spreadsheet. The included articles were assessed using the Prediction model study Risk of Bias Assessment Tool (PROBAST) with the results shown in Tables 2.2 and 2.3.

2.4 Systematic Review Results

2.4.1 Study Sample Characteristics

Characteristics of the included studies are presented in Table 2.1. These 21 included studies developed 127 models, of which 47 (from 7 Studies) used prospectively collected data with a mean sample size of 439 (ranging from 185-3496). Of the models using prospective data 35 were based in multiple treatment centres with a mean sample size of 400 (ranging 193-3496, corresponds to 3 studies). The remaining 12 models used prospective data from a single centre with a mean sample size of 553 (range 185-1275, corresponds to 4 studies).

The remaining 80 models (from 15 studies) retrospectively used data from existing datasets for model development with a mean sample size of 31,840 (ranging from 54-212,666). Most models using retrospective data were based in a single treatment centre (54 models) with a mean sample size of 384 patients (ranging 54-1620, corresponds to 7 studies). The remaining 26 models used retrospective data from multiple centres with a mean sample size of 97,212 (range 355-212,666).

Included study populations were often based in high income countries (HIC). Of the 21 studies included 17 were based in a HIC. This corresponds to 7 based in the USA, 3 in the UK, 2 in Taiwan and 1 in Sweden, Finland, Australia and New Zealand, EU based, and Qatar. There were also four studies using populations based in Middle income countries with 3 in Brazil and 1 in Iran. All data included in studies selected for this review was collected between 2000 and 2019.

There was large variety in the criteria used to select study participants between studies. It was common to set a threshold level of TBI severity accepted using the GCS. Common levels for this threshold correspond to the general classifications of TBI using the GCS, ie. GCS $\leq$ 13 for moderate-severe TBI and GCS $\leq$ 9 for only severe TBI cases. Use of the GCS to identify TBI was not uniform across studies however, head Abbreviated Injury Scale (AIS) $\geq$ 2 (Wan-Ting et al., 2020) and CT-scan results (Muehlschlegel et al., 2016; Prosser et al., 2020) were also used to confirm TBI presence. Another common criterion across studies was a threshold age or range of ages for inclusion. Studies would also impose specific criteria depending on the

TABLE 2.1: Included Study Characteristics

Lead Author	Data Collection Period	Country	Study Design	Methods	Study Sample Size	Inclusion Criteria	Outcomes Predicted
Dawes et al. 2015	2009-2010	USA	Multicentre	LR	822	GCS $\leq$ 8, caused by blunt trauma, abnormal IC findings on head CT	Inpatient Mortality
Lu et al. 2015	2009-2012	Taiwan	Single Centre	LR, ANN, NB, DT	115	Admitted to nICU following m-sTBI, age $\geq$ 18 years, did not die within 14 days, no missing data in record	6-month Mortality
Kelly et al. 2015	2008-2012	USA	Multicentre	LR	3496	sTBI pats, age $\geq$ 14 years, Head AIS $\geq$ 3, could be matched to death index	30-day, 6-month Mortality
Thelin et al. 2016	2005-2013	Sweden	Single Centre	LR	417	Age $\geq$ 18 years, minimum of three measures of S100B and NSE (first in 48h and third within 72h after trauma), admission CT data available, LT functional outcome evaluated $\geq$ 3 months after trauma	Long Term Functional Outcome
Muehlschlegel et al. 2016	2000-2013	USA	Multicentre	LR	413	Age $\geq$ 18 years, confirmed pTBI with perforation of dura on head CT, clinical evidence of brain injury following examination, not dead on arrival, medical data available	Inpatient Survival
Alsulaim et al. 2017	2006-2011	USA	Multicentre	LR	93,397	Consciousness data available, all AIS scores for any non-head body regions = 0	Inpatient Mortality
Han et al. et al. 2017	2006-2009	Singapore	Single Centre	LR	300	GCS $\leq$ 8, consecutively admitted to nICU at NNI Singapore	14-day, 6 month Mortality
Junior et al. 2017	2003-2009	Brazil	Single Centre	LR	1275	Not sedated before Neurological assessment, no hemodynamic instability, abnormalities identified on CT scan	In hospital Mortality (unclear)
Zeiler et al. 2018	2005-2016	UK	Single Centre	LR	358	Minimum 6 hours of recorded ICP signals, adult patients, did not receive decompressive craniectomy	6-month Mortality
Winans et al. 2020	2011-2018	USA	Single Centre	LR	402	Age $\geq$ 18 years, Trauma patients, GCS $\leq$ 8, GCS motor response sub-scores of $\leq$ 5 (not following commands), blunt trauma patients, not penetrating trauma patients	In hospital Mortality
O'Briain et al. 2018	2000-2016	Australia and New Zealand	Multicentre	LR	24,148	Age $\geq$ 17 years, Head Trauma patients, no multi trauma, patient is ventilated, no inter ICU transfer, no readmission, not admitted for palliative care/organ donation, not missing outcome data, not missing PaO2 data	In hospital Mortality
Najafi et al. 2018	2016	Iran	Single Centre	LR	185	TBI patient, MOI is traffic accident, transported directly from the scene by EMS ambulance, ISS $\geq$ 9, aged 18-85, at least one vital sign parameter $\geq$ 0 at scene, no incomplete data in either the prehospital or hospital patient records, no pregnancy or comorbidities present on record, transfer time from the scene to hospital $\leq$ 1 hour, patient not transferred in first 24hours	24-hour Mortality

Table 2.1: Included Study Characteristics (continued)

Lead Author	Data Collection Period	Country	Study Design	Methods	Study Sample Size	Inclusion Criteria	Outcomes Predicted
Raj et al. 2019	2003-2017	Finland	Multicentre	LR	472	Age $\geq$ 16, ICP monitoring for more than 24hr, admitted to ICU within 24hr of trauma	30-day Mortality
Abujaber et al. 2020	2014-2019	Qatar	Single Centre	ANN, SVM	1620	Age $\geq$ 17 years, sustained TBI	In hospital Mortality
Wan-Ting et al. 2020	2014-2017	Taiwan	Multicentre	LR	438	Adult patients, ISS score $\geq$ 16 points and head AIS $\geq$ 2, sufficient data on record, no inter hosp transfer, no out of hospital cardiac arrest, patient not DNR	In hospital Mortality
Prosser et al. 2020	2015-2016	UK	Multicentre	LR	355	Age $\geq$ 16 years, AIS severity of $\geq$ 3 for a confirmed TBI/intracranial injury on head CT scan, transported to hospital by land ambulance, not injured nearest to SNC	30-day Survival
Fontoura Solla et al. 2020	2012-2015	Brazil	Single Centre	LR	517	Age $\geq$ 14, not penetrating TBI, patient not transferred from other ICU, no chronic sub-Dural hematoma present	14-day Mortality
Wu et al. 2020	2012-2016	USA	Multicentre	LR, SVM, K-NN, DT, GNB, LDA	212,666	tICH present on record (epidural, subdural, subarachnoid or intraparenchymal haemorrhage)	In hospital Mortality
Zeiler et al. 2020	2015-2017	EU Based	Multicentre	LR	193	Not patients with EVD based ICP data	6-month Mortality
Kim et al. 2020	2003-2018	USA	Single Centre	DT	54	MOI is firearm injury, Patient not DOA, Clinical Imaging results available, no missing brain imaging data, patient has injury with intracranial involvement, no isolated spinal injury, no missing clinical notes data, not duplicated entry	Mortality (time period unspecified)
Amorim et al. 2020	2012-2015	Brazil	Single Centre	NB, RF, BGLM, PDA, GPLS, SGB	517	Age $\geq$ 14, intracranial abnormality on CT, not penetrating TBI, GCS $\leq$ 15, no intracranial lesions on CT not caused by MOI	14-day, In hospital mortality

Method Abbreviations: *LR = Logistic Regression, SVM = Support Vector Machine, KNN = K-Nearest Neighbours Classifier, DT = Decision Tree, GNB = Gaussian Naïve Bayes, NB = Naïve Bayes, LDA = Linear Discriminate Analysis, ANN = Artificial Neural Network, BGLM = Bayesian Generalised Linear Model, PDA = Penalised Discriminant Analysis, RF = Random Forrest, GPLS = Generalised Partial Least Squares, SGB = Stochastic Gradient Boosting,* ***Other Abbreviations:*** *IC = Intracranial, CT scan = computerised tomography, nICU = Neurological ICU, m-sTBI = moderate to severe Traumatic Brain Injury, sTBI = Severe TBI, LT = Long Term, pTBI = penetrating TBI, ICU = Intensive Care Unit, DNR = Do not rescue, SNC = specialised neurological centre, tICH = Traumatic Intracranial Haemorrhage, ICP = intracranial Pressure, MOI = Mechanism of Injury*

injury under investigation, for example including only penetrating (Muehlschlegel et al., 2016) or blunt force (Dawes et al., 2015; Amorim et al., 2020; Fontoura Solla et al., 2020) brain injuries.

2.4.2 Predictors and Outcomes

The included studies incorporated a wide range of predictor variables in the developed models. A breakdown of the common predictors used in ML models developed in the included studies is shown in Appendix Table C. This table shows the variables specified as model inputs for each study because assessment of the significance or contributions of different predictors to overall model predictions can be challenging or impossible for some ML algorithms. As shown in Appendix Table C, commonly included predictor variables were: age (15/21 studies), GCS Score (10/21 studies), pupillary reactivity (11/21 studies), biological sex (8/21 studies), and comorbidities (5/21 studies). Many studies also included findings from CT-scans (13/21 studies) in predictor variables, although a range of information was incorporated from these scans.

Outcomes predicted by the included models can be separated into three categories based on prediction horizon: short-term, long-term and mixed. Short term outcomes are defined as being within one month of initial injury, which corresponds to 13 models that predict 24hr (1), 14-day (8) and 30-day (3) mortality, with a final model predicting 30-day survival (1). Long-term outcomes were predicted by 81 models, specifically 6-month mortality (75) and long-term Glasgow Outcome Scale (GOS) (6) collected after 3 months. Finally, an outcome without a pre-defined time of measurement was predicted by 33 models, which used in-hospital mortality (28) and survival (5).

2.4.3 Modelling, Performance and Validation

127 ML prediction models were included in the review and Logistic Regression (105) was by far the most prevalent. The remaining models used the following approaches: Support Vector machine (5), Decision Tree (3), Naïve Bayes (2), Random Forest (2), Artificial Neural Network (2), Penalised Discriminant Analysis (2), K-nearest neighbours (1), Bayesian generalised linear model (1), Gaussian Naïve Bayes classifier (1), Linear discriminate analysis (1), Generalised Partial Least Squares (1), Stochastic Gradient Boosting algorithm (1).

Univariate analysis was used in the selection of predictors in 8 studies, with 9 using other methods and 4 that were unclear about how predictors were chosen. Of the 9 studies using methods other than univariate analysis, 3 used the findings of previous literature, and 2 used situation based selection of variables that would be available at the time the model was intended for use. A further 2 selected variables for the testing of a specific hypotheses, which were the effect of adding a loss of consciousness variable to existing injury scores (Alsulaim et al., 2017) and the usefulness of ICP monitor derived signal variables (Zeiler et al., 2018). A single study used an iterative variable selection method called recursive feature elimination and the final study inputted all available variables into the ML algorithms. The general methods used for feature selection are reported in Appendix Table C.

Appendix Table D provides a full breakdown of the reported predictive performance of included models and the metrics reported for evaluation. In Appendix Table D, model discrimination was consistently reported using the concordance or c-statistic, which is equivalent to the Area Under the Receiver Operator Curve (AUC) reported for 112 models. This statistic gives the probability a random patient from the sample that experienced the outcome had a higher predicted probability than a random patient that did not experience the outcome. Reporting of sensitivity (14), specificity (13), accuracy (12), negative predictive value (7) (NPV) and positive predictive value (9) (PPV) was far less prevalent. For studies that reported calibration of models (6), the Hosmer-Lemeshow statistic and associated p-Value were used most frequently (5) with two of these studies reporting Brier scores alongside. The final study reporting calibration used the correction rate.

Several approaches were taken to validate model predictions, although 105 models reported no validation method. Of the validated models, 5 used 10-fold cross validation (2 studies), 10 used 5-fold cross validation (2 studies), 2 used Random sampling (1 study), and 4 used bootstrapping for internal validation (1 Study). Only 11 models (2 studies) were reported as having used a training and testing set for evaluation of model performance, of these 9 used an 80/20 and 2 a 70/30 training and test split.

Due to the inconsistent reporting of evaluation metrics, differences in outcomes predicted, and heterogeneity of sample populations meta-analysis was not attempted

TABLE 2.2: PROBAST Risk of Bias

Lead Author Surname	Year	Participant selection	Predictors and Assessment	Outcomes and Determination	Analysis	Overall Risk of Bias
Dawes	2015	Low	Low	Low	Unclear	High
Lu	2015	Unclear	Unclear	Low	Low	Unclear
Kelly	2015	Unclear	Low	Low	High	High
Muehlschlegel	2016	Low	Low	Low	High	High
Alsulaim	2017	Low	Low	Unclear	High	High
Han	2017	Unclear	Low	Unclear	Low	High
Junior	2017	Low	Low	High	High	High
Zeiler	2018	Low	Low	Unclear	High	High
Winans	2020	Unclear	Low	Unclear	Low	High
O Briain	2018	Low	Low	Unclear	High	High
Najafi	2018	Unclear	Low	Unclear	High	High
Raj	2019	Unclear	Low	Low	Low	Unclear
Abujaber	2020	Low	Low	Low	Unclear	Unclear
Wan-Ting	2020	Low	Low	Unclear	Unclear	High
Fontoura Solla	2020	Low	Low	Unclear	High	High
Wu	2020	Low	Low	Low	Low	Low
Zeiler	2020	High	Low	Unclear	High	High
Kim	2020	High	Unclear	Unclear	High	High
Amorim	2020	Low	Low	Unclear	Unclear	Unclear
Prosser	2020	Low	High	Low	High	High
Thelin	2016	Low	Low	Low	High	High

in this review. However, of the 112 models, which reported an AUC value 109 of these were above 0.7, and 80 were above 0.8, meaning a high proportion of models were able to discriminate between target outcomes effectively. However, due to the underuse of appropriate testing data these AUC values were often calculated on performance in training data making them unreliable.

2.4.4 Study Risk of Bias and Applicability

The study PROBAST results are shown in Table 2.2 with the applicability results in Table 2.3. 16 studies were found to have a high risk of bias, mainly resulting from a lack of external validation of model predictions and the poor reporting of evaluation statistics mentioned in the previous section. Several studies also did not adequately address the approach taken to handle missing data or report the degree of missing data in their sample. Another common issue identified was not controlling for the confounding influence of withdrawal of treatment based on the values of predictors used in the model. Despite these issues in reporting, the majority (12/21) of

TABLE 2.3: PROBAST Applicability

Lead Author Surname	**Year**	**Participants and Setting**	**Predictors and Assessment**	**Outcome and Determination**	**Overall Applicability Concern**
Dawes	2015	Low	Low	Low	Low
Lu	2015	Low	Low	Low	Low
Kelly	2015	Unclear	Low	Low	Low
Muehlschlegel	2016	Low	Low	Low	Low
Alsulaim	2017	Low	Low	Low	Low
Han	2017	Low	Low	Low	Unclear
Junior	2017	Unclear	Low	Unclear	High
Zeiler	2018	Low	Low	Low	Low
Winans	2020	Low	Low	Low	Low
O'Briain	2018	Unclear	Low	Low	High
Najafi	2018	Low	Low	Low	Low
Raj	2019	Low	Low	Low	Low
Abujaber	2020	Unclear	Low	Low	Unclear
Wan-Ting	2020	Low	Low	Low	Low
Fontoura Solla	2020	Unclear	Low	Low	Unclear
Wu	2020	Low	Low	Low	Low
Zeiler	2020	Low	Low	Low	Low
Kim	2020	Low	Low	Low	High
Amorim	2020	Unclear	Low	Low	Unclear
Prosser	2020	Low	High	Low	High
Thelin	2016	Low	Low	High	High

studies were found to be highly applicable to the review, as seen in Table 2.3. This is mainly the result of the review placing few constraints on accepted models and study populations.

2.5 Discussion

This review searched four electronic databases and identified 21 studies for inclusion, which corresponded to 127 mortality prediction models in moderate to severe TBI patient populations. Most of these models used logistic regression for prediction and had an elevated PROBAST score for risk of bias. The quality of results reporting was variable, with many papers only reporting discrimination statistics for model performance. Following the aims highlighted in Section 2.2, this discussion is separated into three sections covering methodological validity, model performance, and the generalisability of model findings.

2.5.1 Methodological Validity

While Logistic Regression (LR) was the most popular modelling approach used in the included studies, several other ML approaches were used and warrant further investigation. A short discussion of the key assumptions and limitations of each approach is presented below alongside the characteristics of the studies that used them.

Logistic Regression (LR):

LR is an example of a standard statistical approach. This means the primary aim of studies presenting LR models was often statistical inference rather than prediction accuracy. This difference in primary objective likely informed decisions during the methodology and evaluation of results, which may explain the sparse reporting of key performance metrics and under use of validation approaches in these studies.

Predictors included in LR models need to be pre-defined and 7 of the 17 studies presenting LR models used univariate analysis to inform predictor choice. This approach raises the risk of bias and may lead to the inclusion of non-clinically useful variables being included due to correlation with meaningful variables.

While LR offers a simple and interpretable method for predicting outcomes in TBI patient populations, the inability to handle fuzzy decision boundaries means models often have good predictive performance on average, whilst being poor at the individual level. This makes the use of validation and testing methods alongside the reporting of performance metrics such as NPV and PPV essential for studies making use of LR. Unfortunately, this was often not the case for models included in this review with only 2 LR studies reporting the NPV and 3 reporting PPV. Studies using LR often had small samples with insufficient data for minority classes, meaning predictions made in this group run the risk of small sample bias.

Support Vector Machine (SVM):

Rather than attempting to emulate an underlying natural process through combining weighted predictor variables, SVM simply attempts to split the data in a way that maximises the margin around a decision boundary (hyperplane). Because of this lack of underlying probabilistic logic, variable importance in the model predictions cannot be interpreted as they can be in LR. This lack of interpretability is a key reason why SVM is not a popular model in medical statistics, where model decisions need to be clinically justifiable. This is one possible reason for why only 5 of the 127 included prediction models used SVM (Abujaber et al., 2020; Wu, Marthi, and Asaad, 2020) despite the high predictive performance achieved.

As covered in Appendix Table B, during an application of the SVM algorithm, the 'C' hyperparameter needs to be manually set. This refers to the penalty to be placed on misclassified points during training. Neither study that developed SVM models reported the value this hyperparameter took, nor was any sensitivity analysis of the effect changing this parameter had on model performance.

Artificial Neural Network (ANN):

The two models using ANN identified in the review were examples of supervised learning, which refers to a situation where the actual value of the outcome variable being predicted during training is known, and used to update internal model weights and parameters.

In the two studies that presented ANN models reporting of the methodology used for development was poor (Abujaber et al., 2020; Lu et al., 2015). Neither study

defined what structure the final ANN took nor whether stochastic gradient descent was used in training. One study (Lu et al., 2015) used cross validation to reduce overfitting risk while the other used a standard training and test data split (Abujaber et al., 2020). Both studies reported key performance measures, which is discussed further in Section 2.5.3. The fully connected multi-layer structure and iterative updating of weights in an ANN model means it is often not clear how it arrives at a prediction. This limits their usability as prediction models in clinical practice. Only one study (Lu et al., 2015) assessed variable importance within the ANN model through leave out comparisons of AUC values with t-tests to assess the significance of these changes. While the attempted assessment of variable importance is commendable, because of the way an ANN will combine variables together, individual assessment of variables provides only a limited picture of the contribution a variable makes to the model.

Naïve Bayes Classifier:

As covered in the Glossary of Terms Naïve Bayes models work through applying Bayes Theorem and the simple assumption of conditional independence (CI). Even though the CI assumption is not expected to hold, NB still achieves good performance as a prediction algorithm for categorical outcomes. In the two studies that developed Naïve Bayes (NB) Classifier models (Lu et al., 2015; Amorim et al., 2020), both found the NB models outperformed all other models tested when predicting mortality at different time points. One model had an AUC of 0.901 for predicting 6-month mortality whilst also maintaining high sensitivity, specificity and calibration values (Lu et al., 2015). The other NB model had a similarly high AUC of 0.906 for predicting 14-day mortality. Both models found age and different GCS values were the most important variables contributing to model predictions. Additionally, one of these models incorporated multiple GCS measurements and the changes between these measurements to add a degree of memory in the predictions (Lu et al., 2015).

Tree Based Models:

In the three decision trees included in the review, the C4.5 Algorithm was used in one (Lu et al., 2015) and two did not specify the algorithm used to determine how to split the data at each node (Kim et al., 2020; Wu, Marthi, and Asaad, 2020). None

of the included decision tree models reported how overfitting was prevented in the models or specified any pruning methods used. As seen in Appendix Table B, a key weakness of decision trees is the classification of minority classes in unbalanced data samples, which was often the case in included studies, as shown by Appendix Table D. Unbalanced samples can lead to performance metrics such as accuracy being misleading as discussed in the Glossary of Terms. Of the studies presenting decision tree models, one addressed this problem using the synthetic minority oversampling technique (SMOTE) (Wu, Marthi, and Asaad, 2020). SMOTE creates new minority class instances using a nearest neighbour-based approach of original minority class instances. While the SMOTE approach helps to address class imbalance, the usability of the synthetic instances for training and testing is questionable. This is because even though the new instances are similar to existing minority class examples in the dataset, whether these instances occur in reality is unknown, meaning the end model may be less generalisable to future data (Chawla et al., 2002). The effect of using SMOTE should be addressed with sensitivity analysis and investigation of the feasibility of synthetic instances. In the single study where SMOTE was used, no analysis of the effect of synthetic instances was reported. The major advantages of decision tree models is their ease of interpretation and clarity of how predictions are made. Also, once trained these models require very little computing power to use, as new instances only need to be run through a series of threshold or categorical tests. These strengths make decision trees feasible for use in clinical practice, provided the overfitting risks are handled appropriately and sufficient data is available for training, testing and validation.

Random Forest (RF):

As seen in Appendix Table B, RF models are an example of ensemble learning, which use resampled data to train multiple DT models and aggregate the predictions of these weaker models. Two RF models were identified from the same study in the review, one predicting in-hospital mortality and the other 14-day mortality (Amorim et al., 2020). There was no reporting of how the two Random Forest models included for review were trained or the specific algorithm used. This under-reporting of decisions made during training and development is a common theme among papers

included for review and was especially prevalent in studies, such as the one presenting the Random Forest models that took a scattergun approach to developing many different algorithms.

2.5.2 Variables of Interest

A discussion of variable importance in mortality prediction following moderate to severe TBI is possible here because of the prevalence of LR use in the included studies. A wide range of input variables were included in the prediction models, although the following sections focus on the most widely used between studies. A list of the common predictors is presented in Appendix Table C. The variables are grouped for convenience into demographic, clinical, injury characteristics, physiological, radiology results, and other. This discussion will cover key issues such as measurement, clinical validity, and variable importance in the models.

Demographic:

Age was the most commonly used demographic variable in the included studies and was often identified through patient electronic health records. Studies using ML methods where variable importance could be ascertained found age to be an important predictor of survival outcome (Lu et al., 2015; Dawes et al., 2015; Kelly et al., 2015; Thelin et al., 2016; Han et al., 2017; Junior et al., 2017; Zeiler et al., 2018; Winans et al., 2020; Raj et al., 2019; Abujaber et al., 2020; Wu, Marthi, and Asaad, 2020; Amorim et al., 2020). One study found multiple trauma cases were more common in younger patients, suggesting there may be some interaction between age and mechanism of injury (Thelin et al., 2016). The impact of age on outcomes following TBI was particularly pronounced in the oldest age categories (Han et al., 2017), which is to be expected as patients with advanced age are often experience worse outcomes following traumatic injury, as discussed in Section 1.2.1. Specifically, falls are the leading cause of traumatic injuries in adults aged over 65. Additionally, falls are a major mechanism of injury for TBI in this patient group (Lawrence et al., 2016).

Biological sex was included in models by 8 studies (Dawes et al., 2015; Lu et al., 2015; Kelly et al., 2015; Zeiler et al., 2018; Winans et al., 2020; Wan-Ting et al., 2020; Wu, Marthi, and Asaad, 2020; Amorim et al., 2020). The impact of biological sex on

TBI outcomes is an open area for research where empirical studies have suggested oestrogen and progesterone may have protective effects following TBI (Brotfain et al., 2016). This means biological sex may have a direct effect on outcomes beyond being a proxy for unobserved characteristics (Brotfain et al., 2016). However, in the studies that included biological sex in models only one reported male sex as having a significant association with mortality following severe TBI (Kelly et al., 2015). The remainder either found biological sex to not be significantly associated with mortality (Dawes et al., 2015), did not report individual variable significance (Zeiler et al., 2018; Wan-Ting et al., 2020; Wu, Marthi, and Asaad, 2020), or did not find biological sex to be among the most important attributes in the ML models (Lu et al., 2015; Wu, Marthi, and Asaad, 2020).

Clinical:

The use and importance of GCS scores and components was not consistent between studies. 10 studies included the full GCS score as an input variable for their models (Dawes et al., 2015; Lu et al., 2015; Thelin et al., 2016; Han et al., 2017; Zeiler et al., 2018; Najafi, Zakeri, and Mirhaghi, 2018; Abujaber et al., 2020; Wan-Ting et al., 2020; Amorim et al., 2020). The majority of these studies found GCS score to be significantly associated with mortality outcomes (Dawes et al., 2015; Lu et al., 2015; Thelin et al., 2016; Han et al., 2017; Zeiler et al., 2018; Najafi, Zakeri, and Mirhaghi, 2018; Wan-Ting et al., 2020; Amorim et al., 2020). Abujaber et al., 2020 however found GCS was not an important predictor in their SVM model. A possible explanation of the lack of importance of GCS found in some studies was the distribution of severity of TBI cases in the sample population. In studies with a high proportion of very severe TBI cases, GCS was found to be an important and significant predictor of mortality (Dawes et al., 2015; Thelin et al., 2016; Han et al., 2017; Amorim et al., 2020). However, when much of the sample had moderate to mild TBI, GCS was found to be a less important predictor (Abujaber et al., 2020). The method of including the GCS into prediction models was also inconsistent between studies. While some included GCS as a continuous variable (Dawes et al., 2015; Thelin et al., 2016; Junior et al., 2017) others opted for a three-category approach, which corresponded to the three categories of TBI mentioned previously (Han et al., 2017; Abujaber et al., 2020). It is

possible that, by including GCS as a three-category variable, some of the inconsistency caused by lacking inter-rater reliability may be reduced leading to more robust information being inputted into models. One study measured GCS at multiple time points and incorporated the changes in these values into the model predictions (Lu et al., 2015). By including the initial level and the changes from this, the authors introduced a degree of memory into the model, where the trajectory of a patients GCS status is considered during predictions. This means the models developed are using less of a snapshot of information, possibly allowing better predictions to be made.

Five studies did not include the combined GCS score in their models, rather opting for the constituent elements. Four of these included only the Motor score (Muehlschlegel et al., 2016; Fontoura Solla et al., 2020; Zeiler et al., 2020; Kim et al., 2020) while the last included the eye, verbal and motor scores as separate elements in models (Wu, Marthi, and Asaad, 2020). Only including the motor score may improve the validity of model predictions in severe TBI when patients are intubated, sedated or intoxicated as there is evidence the full GCS score is less reliable in these groups (Meredith et al., 1998; Marmarou et al., 2007).

Information on comorbidities was included in models of 4 studies (Dawes et al., 2015; Lu et al., 2015; Ó Briain et al., 2018; Abujaber et al., 2020). Of those that reported the significance of predictors, comorbidities were found to be not significant predictors in logistic regression (Dawes et al., 2015) or not among the most important predictors in less interpretable models (Abujaber et al., 2020; Lu et al., 2015). These findings may result from small sample bias however as very few patients with each specific comorbidity included had the outcome of interest. This could therefore lead to insignificant findings due to the nature of the sample population rather than the lack of clinical significance.

Injury characteristics and CT findings:

Mechanism of Injury (MOI) was incorporated into models by 3 of the included studies (Dawes et al., 2015; Winans et al., 2020; Abujaber et al., 2020). One of the major causes of TBI identified in these studies was falls from standing height accounting for 26-34% of patients in the study populations. Falls from standing height were found to be predictive of mortality in Dawes et al., 2015, although this was not

supported by Winans et al., 2020. Also, Abujaber et al., 2020 did not find MOI to be in the top ten most important predictors in their SVM model. During analysis of the time taken to follow commands after TBI, a binary variable representing high velocity trauma or fall from standing was statistically significant (Winans et al., 2020). The authors suggest this means the velocity of the TBI-causing trauma has a role to play in the return to consciousness following TBI.

8 studies ascertained the type of TBI present in patients using findings from Computed Tomography (CT) scans (Dawes et al., 2015; Lu et al., 2015; Muehlschlegel et al., 2016; Han et al., 2017; Abujaber et al., 2020; Wu, Marthi, and Asaad, 2020; Zeiler et al., 2020; Amorim et al., 2020). One of the main things to be confirmed through a CT scan is whether a haematoma or haemorrhage is present and where it is located. At a simple level, the major difference between these conditions is haematoma describes bleeding that has clotted, whilst haemorrhage refers to ongoing bleeding. This discussion of findings is split based on the area of the brain where the bleed is identified because the associated effects on mortality is expected to be different. Studies were most concerned with identifying intracerebral rather than intracranial brain bleeds. Only 1 study included the presence of intracranial haematoma in their prediction models, finding it did not significantly add to the prediction of inpatient mortality (Dawes et al., 2015).

Presence of epidural haemorrhage on head CT was included as an input variable in 2 studies (Wu, Marthi, and Asaad, 2020; Amorim et al., 2020) and epidural haematoma in 1 study (Wu, Marthi, and Asaad, 2020). A key difference between epidural and subdural haematomas is that the former does not cross the Suture Line with the bleed located between the dura and skull rather than between the dura and brain tissue. Presence of epidural haemorrhage was not among the most important predictors reported in either study (Wu, Marthi, and Asaad, 2020; Amorim et al., 2020). The reason behind this lack of importance is difficult to quantify due to the methods used and unclear reporting of the studies. However, because prognosis following epidural haemorrhage is generally good following early identification, its importance in predicting mortality may be reduced compared to variables indicating more serious brain bleeds.

One of the more common locations for brain bleeds identified in the included

studies was in the subdural space. Subdural haematomas were present in 65.7-78.3% and haemorrhage present in 32.7-57.7% of patients in study populations. 5 studies included indicators for bleeds in the subdural space, with presence of subdural haemorrhage on head-CT included by 2 studies (Wu, Marthi, and Asaad, 2020; Amorim et al., 2020) and subdural haematoma by 3 studies (Dawes et al., 2015; Lu et al., 2015; Han et al., 2017). All these studies introduced the presence of subdural haemorrhage or haematoma with binary variables. The effect of including variables that indicate subdural bleeds in mortality prediction models was mixed, two studies found presence of a subdural bleed significantly added to the prediction model (Han et al., 2017; Wu, Marthi, and Asaad, 2020) whereas another found it to be insignificant (Dawes et al., 2015). The other two studies that included presence of subdural bleeds either used uninterpretable ML methods (Lu et al., 2015) or presented too many models to effectively show what went into each (scattergun style modelling) (Amorim et al., 2020).

Subarachnoid haemorrhage on head CT was included by 4 studies (Dawes et al., 2015; Wu, Marthi, and Asaad, 2020; Zeiler et al., 2020; Amorim et al., 2020). Subarachnoid haemorrhage refers to a bleed in the subarachnoid space, between the arachnoid and pia layers. Prognosis following this kind of haemorrhage is similar to subdural and generally much worse than epidural bleeds. Subarachnoid haemorrhage is understood to cause mortality through vasospasm and ischemia (Armin, Colohan, and Zhang, 2006). The importance of subarachnoid haemorrhage in prediction models was mixed, with one finding it significantly added to the model predictions (Wu, Marthi, and Asaad, 2020), while another found it insignificant (Dawes et al., 2015), and the final two did not report their findings (Amorim et al., 2020; Zeiler et al., 2020). A possible reason for the difference in findings resides in the study population characteristics. While the study with significant findings had a high proportion of moderate-mild TBI cases (Wu, Marthi, and Asaad, 2020), the other was very severe with over 50% of the sample having a GCS of 3. This difference meant subarachnoid haemorrhage was far less prevalent as a proportion of the study population in Wu, Marthi, and Asaad (2020) and could act as an indicator between the mild-moderate TBI cases and the few severe cases on the LR model. In the study with a high proportion of very severe TBI patients however, presence

of subarachnoid haemorrhage was widespread (59.5% of sample) and other indicators for very severe outcomes such as loss of basal cisterns offered better separation between the survivors and non-survivors in the sample (Dawes et al., 2015).

2.5.3 Model Performance

As highlighted in the results section, performance of prediction models was reported using discrimination statistics in almost all the included studies. As seen in the Glossary of Terms, discrimination refers to the ability of a model to separate between patients with and without the outcome. In other words, a model that discriminates well would predict a higher probability of the outcome occurring for a randomly selected positive case than a negative case. Although AUC is a useful summary metric to capture model discrimination, it must be combined with other statistics to ensure a model is effectively predicting outcomes. Unfortunately, the reporting of sensitivity, specificity, negative predictive value was infrequent in the included studies, as seen in Appendix Table D.1. This means that, while models may exhibit a high AUC value, it is not possible to evaluate whether the model is effectively predicting mortality rather than just returning the majority class.

The generalisability of a prediction model to new settings is reliant on several features of the data, methodology, and testing in a study. Generalisability of findings starts at the beginning of the modelling process with data collection. Collecting data from multiple centres was used by 8 included studies to avoid the raised risk of bias introduced by only using a single centre (Alsulaim et al., 2017; Dawes et al., 2015; Kelly et al., 2015; Muehlschlegel et al., 2016; Ó Briain et al., 2018; Prosser et al., 2020; Wu, Marthi, and Asaad, 2020; Zeiler et al., 2020). By using multiple centres, models are made more robust to differences in protocols and other hospital level fixed effects.

For model findings to be generalisable, care must be taken to ensure overfitting risk is avoided or mitigated during training and appropriately tested for during evaluation. To confirm a model is not overfit to the training data, the use of external validation in new data is needed. A minority of included studies used validation and testing strategies to test the generalisability of model findings in new settings.

The external validation of model findings in unseen data is a key stage of developing a prediction model for use in clinical practice. This is because ML models are designed to closely fit themselves to the training data, meaning the reported performance on this data is not a reliable test for the performance of a model in new data. The methods used for externally validating findings in included studies were n-fold cross validation (Amorim et al., 2020; Lu et al., 2015; Raj et al., 2019; Wu, Marthi, and Asaad, 2020) and splitting the data into a training and testing set (Abujaber et al., 2020; Wu, Marthi, and Asaad, 2020). Bootstrapping was also used for internal validation of model findings in one study (Han et al., 2017). The remaining 15 of the included studies did not report using any validation methods to confirm the model performance.

As highlighted in one study, by rigorously testing performance in a true holdout data set, the comparative performance of a model will be reduced (Wu, Marthi, and Asaad, 2020). Rigorous testing of findings will always represent good practice however, it is possible authors are eschewing this in the pursuit of a high headline AUC value. The idea that high AUC guarantees a good model is flawed, especially when this AUC is based on performance in training data. While not having the sample size available for true holdout testing sets is not necessarily a problem, the limitations this imposes on model generalisability and the conclusions that can be drawn must be made clear.

2.5.4 Future Research Recommendations

Several recommendations for future research into prediction modelling for TBI outcomes are proposed following this review. The first relates to the use of training and testing sets or cross validation to evaluate model performance, which was lacking in the majority of included studies despite this being standard practice in much of the ML literature. The second recommendation surrounds improvements in reporting of model development, performance, and evaluation metrics. While discrimination statistics were often reported the coverage of sensitivity, specificity, NPV and PPV alongside calibration measures were missing from most included studies. Without these figures a comparison between models and evaluation of prediction quality becomes much more difficult. Also, the intended time period and hospital setting for

the prediction model to be set in was rarely reported. This is important to allow comparison of models based in the same setting.

Finally, only a single included study explored the concept of dynamic modelling. Development of support tools that can update predictions to account for new information at different time periods have a clear clinical benefit and warrant further investigation.

2.5.5 Review Limitations

The review had several limitations. First a single reviewer (SW) was responsible for the searching and screening of papers as well as any exclusion judgements. Although any reasons for exclusion were reported, having a single reviewer elevates the risk of bias inherent in the review. Second the inclusion of key terms in the search strategy may have led to the omission of newer papers that have not had key terms assigned. Third the review was limited to English language searches only.

2.6 Review Conclusion

There was no objectively 'best' model identified in the review. Due to the complexity of modelling TBI outcomes and the wide range of factors that affect outcome, sample populations need to be heterogenous and data needs to be collected for wide range of variables for models to be generalisable. Future research needs to ensure appropriate validation of findings and performance metrics are reported clearly to allow direct comparisons between models to take place.

Chapter 3

Association Between Chronic Health Conditions and Falls in Older Adults: A Review of Reviews

3.1 Introduction

As discussed in Section 1.2.1, falls and fall-related injuries are a major source of healthcare use in patients aged over 65 (Office for National Statistics, 2023). As concluded in Section 1.2.4, research into falls is essential due to the ageing population and associated increase in longevity among patients living with chronic health conditions (National Institute of Health & Care Excellence, 2023). Understanding the factors that lead to falls is important in addressing the prevention of falls in future.

Before attempting to model individual fall risk using the chronic condition record in Section 4.4, it is important to identify the conditions to test in regression models. There is a large body of research addressing fall pre-disposing chronic health conditions and a number of systematic reviews and meta-analyses on the topic. Therefore, a review of reviews format was selected to provide an evidence-based approach in identifying a list of conditions to include in the analyses.

The questions this review aimed to answer are as follows:

- In people aged over 65, what chronic health conditions increase the risk of falls?

- What co-morbidity and multi-morbidity chronic health condition combinations increase the risk of falls in people aged over 65?

The remainder of the review is structured as follows. Section 3.2 presents the search strategy, approach to data extraction, and assessment of risk of bias for this review. Section 3.3 describes the search results, tables of extracted data, and results of the risk of bias assessment followed by a written summary of the findings. Section 3.4 discusses these findings, answers to the two highlighted research questions, presents recommendations to future research, and the review limitations. Finally, Section 3.5 presents the review conclusions and indicates where the findings are applied in the wider thesis.

3.2 Review Methodology

3.2.1 Search Strategy

Due to the breadth of literature needed to answer the review questions stated in Section 3.1 a review of reviews format was selected (Smith et al., 2011). This format allows the distillation of findings from a large field of research, and minimises revisiting subjects that have already been covered in detail by previous authors. A review of reviews written by Preston et al. (2021) was used as a framework upon which the methodology taken in this review was based.

The population of interest in this review is adults aged over 65; however, to avoid the unnecessary exclusion of reviews that used an over 60 age criteria where the majority of the sample was aged over 65, the inclusion criteria for age was expanded to capture these reviews. The effect of this expansion is discussed further in Section 3.4.3. The event of interest for the review was ground level falls in the population. The exposure was the chronic health conditions an individual lived with at the time of their fall.

The outcome for the review was the measure of effect a single chronic health condition, or combination of chronic health conditions, had on the risk of a fall. This can be reported as an odds ratio, risk ratio, hazard ratio, or similar, together with associated confidence intervals.

Four bibliographic databases (Medline, Embase, Web of Science, CINAHL) and the PROSPERO review repository were searched on 07/03/2023 using exploded headings, and key words related to the following components:

- Systematic Review
- Accidental falls
- Aged 65 and over
- Chronic health conditions (including co-morbidity and multi-morbidity)

A publication date filter of after 2000 was applied the research results, and only papers available in English were considered for this review. A copy of the full list of search terms is provided in Appendix Section E. The results of these searches are presented using a Page et al. (2021) PRISMA Flowchart in Figure 3.1.

A single reviewer (SW) sifted titles and abstracts from the searches against the following pre-defined inclusion criteria:

- Study is a systematic review, which presents a meta-analysis.
- Published after 2000.
- Review inclusion criteria included participants aged 60 or where no age-related criteria were stated, included studies had a mean age of over 60.
- Review outcome focus is the risk of accidental Falls (not outcomes following a fall).
- At least one chronic health condition is considered as an exposure in relation to accidental falls in the review.

Following the initial sift on the basis of the titles and abstracts, full texts were downloaded, and duplicates were removed. A sift of full texts was performed to ensure the reviews met the inclusion criteria stated above. Reasons for exclusion at this stage are reported in Figure 3.1 alongside the resulting number of papers included in the review.

3.2.2 Data Extraction and Synthesis

A single reviewer (SW) performed the data extraction on included texts using a standardised form for the accurate identification of key features. General review information was extracted at this stage (author, publication year, number of articles in review, combined review sample size, included article publication date range) with the findings reported in Table 3.1. Further information on the review inclusion criteria, measurement of falls, chronic condition requirements, and exclusion criteria were also extracted at this stage, with the findings summarised in Table 3.2.

Information from the meta-analysis in the included reviews was extracted using a standardised form in an MS Excel spreadsheet. The data extracted at this stage were the chronic conditions of interest, type of meta-analysis, number of study estimates pooled, combined sample size in estimate, reported heterogeneity measure, estimate type, pooled estimate, and confidence interval with the findings reported in Table 3.3. No meta-analysis or quantitative synthesis of results were planned for this review of reviews due to time and resource constraints.

3.2.3 Risk of Bias Assessment

The risk of bias for each included systematic review was identified using the ROBIS checklist (Whiting et al., 2016). This checklist provides a structured assessment of four areas of the review (eligibility criteria used, study selection, data extraction and study appraisal, evidence synthesis), which aids with the judgement of risk of bias in the review as High, Low, or Unclear. The ROBIS assessment was conducted in a standardised Microsoft Excel^{TM} spreadsheet by a single reviewer (SW) with Table 3.4 summarising the results.

3.3 Results

3.3.1 Search Results

Figure 3.1 presents a PRISMA flowchart of the search results. There were 506 records returned by the bibliographic searches, of which 475 were excluded through assessment of the title and abstract against the inclusion criteria. The remaining 31 records were downloaded into Mendeley reference manager.

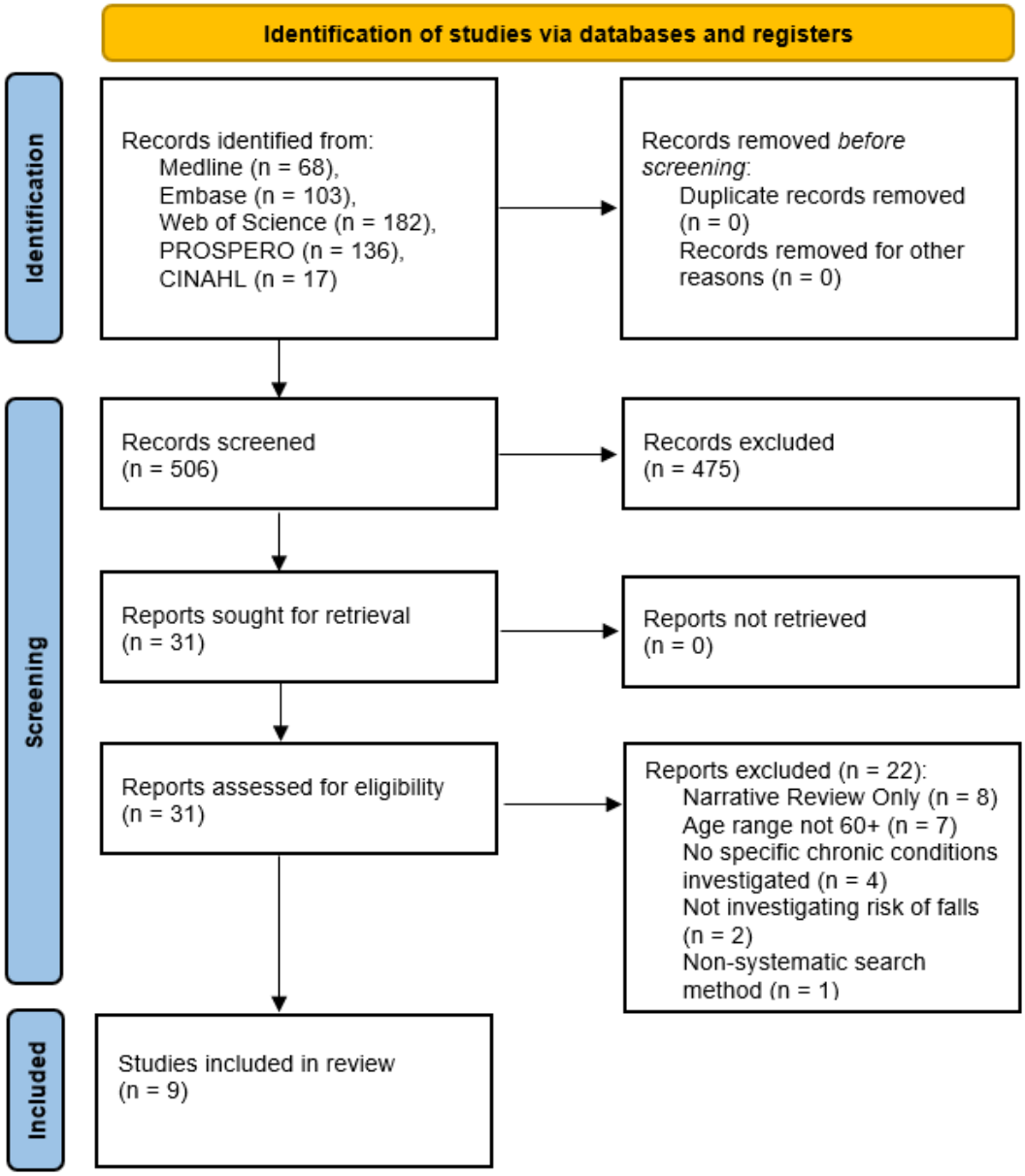


FIGURE 3.1: PRISMA Flowchart of Search Results

A full text sift of these records against the inclusion criteria identified a further 22 papers to be excluded. Reasons for exclusion at this stage, shown in Figure 3.1,

were presenting a narrative review only (n = 8), age criteria or mean age of included papers below 60 (n = 7), no specific chronic conditions investigated (n = 4), investigating outcomes following falls or outcomes not related to the risk of falls (n = 2), and non-systematic search methods (n = 1). Data extracted from the included reviews is presented in Tables 3.1-3.4 alongside a written description of the results in Sections 3.3.2-3.3.4.

3.3.2 Characteristics of Included Reviews

Nine systematic reviews were included for review. Table 3.1 presents the characteristics of the included reviews presenting quantitative meta analyses of risk estimates for chronic health conditions on falls risk. The nine included reviews, which were published between 2013 and 2023, included studies published between 1981 and 2022. The number of studies included in the reviews ranged between 5 and 220 based on the scope of the objectives, resources available to the reviewers, and the paucity of existing research on the relationships of interest.

TABLE 3.1: Included Systematic Review Characteristics

Lead Author	Year of Publication	Number of papers in review	Review Total Sample Size	Study Publication Dates
Bloch	2013	220	Not Reported	1981-2011
Stubbs	2014	21	5367	1999-2013
Yeung	2019	36	45926	2012-2018
Mol	2019	63	49,164	1987-2017
Liu	2020	14	1,284,456	2012-2018
Malik	2020	10	36444	2000-2018
Oliviera	2021	23	46569	2009-2019
Xu	2022	34	76008	2013-2021
Veronese	2023	5	87,554	2016-2022

TABLE 3.2: Inclusion Criteria of Each Systematic Review

Author (Year)	**Age Criteria**	**Study Design Criteria**	**Falls Outcome Measured**	**Chronic Condition(s) requirements**	**Exclusion Criteria**
Bloch (2013)	60+	Observation or intervention studies	Risk of falls	None	Sample is entirely stroke or neurological patients, investigated non-standard falls
Stubbs (2014)	60+	Prospective or retrospective studies with 6 month minimum observation period	Report falls as an outcome	Sample includes patients living with and without chronic pain	Dementia patients, studies where pain caused by a previous fall, Neurological conditions, recent history of trauma or orthopaedic surgery within 6 months
Yeung (2019)	Mean age 60+	None	Falls or fracture outcome reported	Sarcopenia present by any definition and control group	No primary data, no comparison group
Mol (2019)	65+	Cross-sectional and longitudinal designs	Falls reported at follow up	Orthostatic (Postural) Hypotension assessment made at baseline	Non-primary research studies
Liu (2020)	60+	Cases and controls identified as fall and non-fall	Ordinary Falls identified	Total knee or hip arthroplasty patients present in sample	None
Malik (2020)	Mean age 60+	Prospective and retrospective designs	Falls reported as outcome	Sample contains patients with and without Atrial Fibrillation or Syncope	Investigation of falls or syncope on the development of AF
Oliviera (2021)	Mean age 60+	None	Report a recognised fall outcome	Study sample entirely patients with COPD	Various non-observational study types, studies using postural balance as an outcome
Xu (2022)	65+	Case Control Studies	Falls identified in sample	Data on demographics and comorbidities available	Non-primary research articles
Veronese (2023)	Mean age 60+	Meta Analysis of observational studies	Association with falls reported	Knee Osteoarthritis assessed in Meta Analysis	Meta-analyses of cross-sectional studies

Table 3.2 presents a summary of the inclusion criteria used in the included reviews. Five of these reviews imposed an explicit age restriction on the samples of the included studies of over 60 (n = 3), or over 65 (n = 2). The remaining four reviews did not explicitly state an age restriction; however, because accidental falls are most common in older populations, the studies included in these reviews had a mean age of over 60.

The included reviews all imposed eligibility criteria on study design to reduce the included study heterogeneity for the eventual meta-analysis. As shown in Table 3.2, these criteria were either explicitly specified in the inclusion criteria (n = 7), or covered in a general list of exclusion criteria (n = 2). There were few differences in the study design criteria accepted in the included reviews. Seven reviews accepted a wide range of study designs encompassing observational and international approaches (Bloch et al., 2013; Stubbs et al., 2014; Yeung et al., 2019; Mol et al., 2019; Malik et al., 2020; Oliveira et al., 2021; Xu, Ou, and Li, 2022). One review included only meta-analysis of case-control or cohort studies (Veronese et al., 2023). The reporting of study design criteria for the final review was unclear, with the only requirement specified that cases and controls were present in the included study sample (Liu et al., 2020).

3.3.3 Meta-Analyses Results

Table 3.3 presents the 35 pooled meta analyses risk estimates extracted from the included reviews. Of these estimates, four were non-significant. The approach to these meta-analyses was either fixed effects (n = 25) or random effects (n = 9) with the decision to pursue random effects due to either assumed or identified between-study heterogeneity, which made fixed-effects meta-analysis inappropriate. The threshold for deciding to use random effects differed between studies, with different threshold values of various heterogeneity tests used or reviewer assessment based on included study characteristics. The two methods used consistently for measuring heterogeneity were I^2 or the Chi-squared test p-values. To aid comparison between estimates, where both statistics were reported only the I^2 is reported in Table 3.3. The pooled meta-analysis estimates were presented as either Odds Ratios (n = 25), or Risk Ratios (n = 10).

The median sample size in the meta-analysis estimates was 10,649 with a range from 502 to 87,554. The number of studies included in the meta-analysis estimates also varied with a mean of 18, and range of 3 to 63 studies. However, the majority of meta-analysis estimates were drawn from a wide range of studies, with only seven estimates pooled from five studies or less. This figure does not include the estimate by Veronese et al., 2023 because this was based on four previous meta-analyses meaning it drew from a wide range of studies.

There were a range of chronic health conditions investigated in the pooled meta-analysis estimates. The discussion of these results is grouped, based on the chronic health condition of interest below. Of the 22 chronic health conditions addressed by the included reviews, all were found to be significant by at least one meta-analysis. There were four cases of disagreement between meta-analyses relating to estimates for dementia, stroke, diabetes, and visual impairment. Each of these disagreements is described below.

The effect of dementia on the risk of falls was addressed by two of the included meta analyses (Bloch et al., 2013; Xu, Ou, and Li, 2022). A significant increase in falls associated with dementia was identified in Bloch et al. (2013) (Random Effects OR = 1.96, 1.8-2.14, I^2 = 88%) based on 35 studies with a combined sample size of 59,363. However, the estimate presented in Xu, Ou, and Li (2022) found an non-significant effect for dementia (Fixed Effects RR = 1.11, 0.88-1.39, I^2 = 59%) based on 9 studies with a combined sample size of 2277. The cause of this difference in findings cannot be attributed to the meta-analysis method used, because even in the presence of high heterogeneity between included studies, the fixed effects estimate would have narrower confidence intervals than the random effects estimate. However, it is possible that the smaller sample size in Xu, Ou, and Li (2022) both in terms of included studies, and number of participants in those studies could have led to a wider confidence interval surrounding the meta-analysis estimate.

Two of the included meta-analyses addressed the effect of strokes or cerebrovascular disease on the risk of falls (Bloch et al., 2013; Xu, Ou, and Li, 2022). Bloch et al. (2013) identified a significant effect for their pooled estimate, (Fixed effects OR = 1.44, CI = 1.34-1.56, I^2 = 20%) based on 49 studies with a combined sample size of 54,336. Contrasting findings were presented by Xu, Ou, and Li (2022) (Random

TABLE 3.3: Meta Analyses Risk Estimates

Condition	Author (Year)	Studies in estimate	Sample Size	I2 Estimate	Estimate Type	Pooled Estimate	Lower CI	Upper CI
Dementia	Bloch (2013)	35	59363	88%	OR	1.96	1.8	2.14
Dementia	Xu (2022)	9	2277	59%	RR	1.11	0.88	1.39
Stroke	Bloch (2013)	49	54336	20%	OR	1.44	1.34	1.56
Stroke	Xu (2022)	4	8158	87%	RR	1.55	0.72	3.35
Parkinson's	Bloch (2013)	29	39477	50%	OR	2.19	1.68	2.84
Parkinson's	Xu (2022)	3	2293	0%	RR	3.05	1.84	5.05
ND	Bloch (2013)	17	20281	0%	OR	2.18	1.69	2.82
UI	Bloch (2013)	34	59458	49%	OR	1.73	1.6	1.88
UI functional sign	Bloch (2013)	4	1826	19%	OR	1.64	1.16	2.33
Diabetes	Bloch (2013)	40	61028	11%	OR	1.27	1.19	1.36
Diabetes	Xu (2022)	7	10026	84%	RR	1.08	0.87	1.34
Anaemia	Bloch (2013)	5	502	0%	OR	1.47	1.15	1.88
Visual Imp.	Bloch (2013)	39	38671	47%	OR	1.49	1.39	1.59
Visual Imp.	Xu (2022)	4	4661	96%	RR	1.24	0.91	1.69
Visual Imp.	Bloch (2013)	16	31443	24%	OR	1.29	1.18	1.4
Sensory Imp.	Bloch (2013)	9	3125	38%	OR	2.2	1.56	3.11
Hearing Imp.	Bloch (2013)	17	21878	11%	OR	1.37	1.27	1.48
Hypotension	Bloch (2013)	20	11939	9%	OR	1.27	1.09	1.47
Hypotension	Mol (2019)	63	49164	-	OR	1.73	1.5	1.99
Cardiac and Vascular	Bloch (2013)	14	24367	0%	OR	1.6	1.45	1.75
Heart Disease	Xu (2022)	6	11078	0%	RR	1.14	1.09	1.19
Hypertension	Xu (2022)	7	9624	0%	RR	1.08	1.03	1.12
Hypertension	Bloch (2013)	35	45115	42%	OR	1.28	1.19	1.37
Atrial Fibrillation	Bloch (2013)	9	3402	0%	OR	1.42	1.14	1.75
Atrial Fibrillation	Malik (2020)	7	36444	37%	OR	1.19	1.07	1.33
Depression	Bloch (2013)	39	67858	40%	OR	1.64	1.52	1.76
Depression	Xu (2022)	6	9364	98%	RR	4.34	4.02	4.68
Behavioural Disorder	Bloch (2013)	16	35858	25%	OR	1.27	1.14	1.42
Sarcopenia	Yeung (2019)	16	23061	7%	OR	1.75	1.55	1.97
Osteoarthritis	Bloch (2013)	37	5284	88%	OR	1.24	1.2	1.28
Osteoarthritis	Veronese (2023)	4	87554	-	RR	1.34	1.1	1.64
Cancer	Bloch (2013)	17	26642	44%	OR	1.22	1.09	1.35
Chronic Pain	Stubbs (2014)	3	5367	0%	OR	1.8	1.56	2.09
Chronic Pain	Xu (2022)	3	2340	78%	RR	1.22	1.11	1.34
Digestive Disease	Bloch (2013)	8	10649	0%	OR	2.2	1.65	2.93

CI = Confidence Interval, ND = Neurological Disease, Imp = Impairment, UI = Urinary Incontinence, OR = Odds Ratio, RR = Relative Risk

Effects Risk Ratio = 1.55, CI = 0.72 – 3.35, I^2 = 87%) based on four studies with a combined sample of 8158 participants. In this case of disagreement, it seems clear the high level of heterogeneity present in the Xu, Ou, and Li (2022) review, when combined with the smaller sample size, led to a very wide confidence interval and a non-significant finding.

The effect of diabetes on falls risk was assessed in two pooled meta-analysis estimates from two of the included reviews (Bloch et al., 2013; Xu, Ou, and Li, 2022). While Xu, Ou, and Li (2022) found a non-significant effect of diabetes on falls (Random Effects RR = 1.08, CI = 0.87 – 1.34, I^2 = 84%) this is likely a result of the high level of heterogeneity in the seven pooled studies, which would cause the confidence interval surrounding the pooled estimate to be wider. In contrast, the estimate derived in Bloch et al. (2013) from 40 pooled studies with a combined sample size of 61028 found a significant effect when heterogeneity between studies was low (Fixed Effects OR = 1.27, CI = 1.19 – 1.36, I^2 = 11%).

Visual impairment was addressed by two of the included meta-analyses, with seemingly contrasting findings (Bloch et al., 2013; Xu, Ou, and Li, 2022). A significant risk increasing effect of visual impairment on the risk of falls based on a pooled estimate from 39 studies with a combined sample size of 38,671 was identified by Bloch et al. (Random Effects OR = 1.49, CI = 1.39 – 1.59, I^2 = 47%). This result is in contrast to the finding in Xu, Ou, and Li (2022) where four study estimates were pooled with a combined sample size of 4661, resulting in an non-significant effect (Random Effects RR = 1.24, CI = 0.91 - 1.69, I^2 = 96%). The difference in these findings may well have arisen from from the high heterogeneity between studies presented in Xu, Ou, and Li (2022) leading to wider confidence limits surrounding the pooled estimate when compared to the Bloch et al. (2013) pooled estimate.

The disparity of findings between Xu, Ou, and Li (2022) and Bloch et al. (2013) surrounding the significance of dementia, cerebrovascular disease, diabetes, and visual impairment may be caused by differences in the sample sizes, heterogeneity, and method used during the meta-analyses with the true effect of each condition significantly associated with increased falls risk. However, an alternative explanation is that unmeasured confounding, differences in the covariates included in the base-study estimates, or complex interactions could be leading to this inconsistency

TABLE 3.4: ROBIS Assessment Results

Author (Year)	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Risk of bias in the review
Bloch (2013)	LOW	LOW	LOW	HIGH	YES	YES	YES	LOW
Stubbs (2014)	LOW	LOW	LOW	LOW	YES	YES	YES	LOW
Yeung (2019)	LOW	LOW	UNCLEAR	LOW	YES	YES	YES	LOW
Mol (2019)	LOW	LOW	LOW	UNCLEAR	YES	YES	YES	LOW
Liu (2020)	LOW	LOW	HIGH	LOW	NO	YES	NO	HIGH
Malik (2020)	LOW	LOW	HIGH	HIGH	YES	YES	YES	LOW
Oliviera (2021)	LOW	LOW	LOW	HIGH	YES	YES	YES	LOW
Xu (2022)	UNCLEAR	HIGH	LOW	HIGH	NO	YES	YES	HIGH
Veronese (2023)	LOW	LOW	LOW	HIGH	YES	YES	YES	LOW

Q1: *Concerns regarding specification of study eligibility criteria,* ***Q2:*** *Concerns regarding methods used to identify and/or select studies,* ***Q3:*** *Concerns regarding methods used to collect data and appraise studies,* ***Q4:*** *Concerns regarding the synthesis and findings,* ***Q5:*** *Did the interpretation of findings address all of the concerns identified in Domains 1 to 4,* ***Q6:*** *Was the relevance of identified studies to the review's research question appropriately considered?,* ***Q7:*** *Did the reviewers avoid emphasizing results on the basis of their statistical significance?*

between meta-analyses findings.

Of the nine included meta analyses, none attempted meta-analysis of specific co-morbid combinations of chronic health conditions.

3.3.4 Review Risk of Bias (ROBIS) Results

A summary of the findings from the ROBIS assessment is presented in 3.4. Seven of the included reviews were rated as having a low risk of bias following ROBIS assessment. The included reviews largely took effective steps to reduce the risk of bias and error during study selection, and data extraction. Use of study quality assessment tools such as the Newcastle-Ottowa Score was widespread in the included reviews.

Two reviews were labelled as having a high risk of bias following the ROBIS assessment (Liu et al., 2020; Xu, Ou, and Li, 2022). This was due to a lack of a pre-defined study protocol in one (Xu, Ou, and Li, 2022), and the over-reporting of significant study results without sufficient additional information for a reader to make a judgement of the base study quality (Liu et al., 2020).

TABLE 3.5: Chronic Condition Shortlist Following Review of Reviews

Pasquetti (2014) Category	**Chronic Health Condition**	**MA Supporting Evidence:**	**MA Conflicting Evidence:**
Neurological	Cognitive Impairment or Dementia	Bloch et al. 2013	Xu et al. 2022
	Stroke or Cerebrovascular Disease	Bloch et al. 2013	Xu et al. 2022
	Parkinsons Disease	Bloch et al. 2013, Xu et al. 2022	
Urinary Incontinence	Urinary Incontinence	Bloch et al. 2013	
Sensory Impairment	Vision Impairment	Bloch et al. 2013	Xu et al. 2022
	Hearing Impairment	Bloch et al. 2013	
Cardiovascular	Hypotension (postural)	Bloch et al. 2013, Mol et al. 2019	
	Atrial Fibrillation	Bloch et al. 2013, Malik et al. 2020	
	Heart Disease	Xu et al. 2022	
	Hypertension	Bloch et al. 2013, Xu et al. 2022	
Psychiatric	Depression	Bloch et al. 2013, Xu et al. 2022	
Musculoskeletal	Sarcopenia	Yeung et al. 2019	
	Osteoarthritis	Bloch et al. 2013, Veronese et al. 2023	
Other	Cancer	Bloch et al. 2013	
	Diabetes	Bloch et al. 2013	Xu et al. 2022
	Anaemia	Bloch et al. 2013	
	Chronic Pain	Stubbs et al. 2014	
	Digestive Disease	Bloch et al. 2013	

MA: Meta-analysis

3.4 Discussion

3.4.1 Evidence Synthesis

This review of reviews set out to answer in people aged over 65, which chronic health conditions increase the risk of falls in meta-analyses. In the 9 meta-analyses included within the review, significant risk increasing effects in pooled estimates were identified for 18 chronic health conditions summarised in Table 3.5.

The second question of interest to the review was which combinations of chronic co-morbidity and multi-morbidity have been previously identified as increasing falls risk beyond a single fall pre-disposing chronic health condition? However, none of

the included meta analyses addressed any specific chronic condition combinations directly in pooled estimates, so this could not be considered further within the review.

The range of chronic conditions in Table 3.5 demonstrates the complex relationships present when attempting to understand the effect of multi-morbidity and accidental falls in older adults. While it is likely that any two of these conditions, when occurring concurrently, would increase fall risk, it is unclear whether there would be an additional multiplicative component introduced by the interaction between conditions.

Identifying the presence or magnitude of multiplicative components introduced by multimorbid combinations of chronic conditions is important for the development of effective falls risk screening tools intended for use in populations with high levels of multi-morbidity.

A possible avenue for understanding the mechanism of multi-morbidity effects would be through examining the fall pre-disposing symptoms that a chronic health condition causes. The conditions in Table 3.5 lead to a range of symptoms known to increase falls risk such as muscle weakness, confusion/disorientation, worsened balance, reduced proprioception in extremities, urgency or pressure, and reduced ability to perceive hazards (Pasquetti, Apicella, and Mangone, 2014). While these symptoms increase falls risk in isolation, it is possible that combinations of these symptoms may increase fall risk to a larger degree than the individual components would suggest as multiple systems are affected. However, identifying and separating out these combination effects from the role of the individual conditions is a challenging endeavour, especially when considering the inconsistency of presentation and relationship with falls each chronic condition has.

An additional layer of complexity in understanding falls risk in older adults is the mitigating effects presented by treatment and location. Community-dwelling older adults are more exposed to extrinsic factors than those in institutionalised settings, meaning that fall pre-disposing symptoms arising from certain chronic health conditions may have a reduced effect in the care home setting due to proactive removal of fall hazards in the environment. However, as discussed in Section 1.3.2 care home residents have a higher prevalence of frailty and intrinsic risk factors

for falls which may counteract this environmental change. Furthermore, while a chronic health condition may have been identified as increasing the risk of falls in the wider population, the individual level effect will depend on individual treatment and severity of that condition, meaning any multiplicative effects identified may not be consistent across individuals.

Finally, previous authors have suggested a U-shaped relationship between mobility and the risk of falls. In this theoretical framework fall risk is maximised due to the confluence of extrinsic risk factors, and intrinsic factors, while an individual is still mobile enough to be able to fall (Bath and Morgan, 1999). However, beyond a certain level of condition burden, falls risk will reduce as a person becomes less mobile with fewer opportunities to fall until eventually being confined to a bed. Therefore, while multi-morbidity is likely to have a role in determining falls risk, beyond a level of condition severity and disease burden, falls risk may decrease for the individual rather than increase.

3.4.2 Future Research Recommendations

Several recommendations for future research were identified during this review. Reporting of adjustment for covariates present in the estimates taken from base studies during the meta-analyses is important, especially when addressing falls risk, where many factors have been identified as significant. Additionally, it is clear from this review that the effect of single chronic conditions on falls risk is well researched, whilst the effect of combinations of conditions is comparatively understudied. To allow a more holistic view of falls risk resulting from chronic health conditions, further research into the relationships between conditions, and how their co-occurrence impacts the risk of falls is required. This topic is addressed further using the cluster analysis described in Section 4.3.

3.4.3 Review Limitations

This review had several limitations. The first is that a single reviewer was involved in the study search, inclusion decision, data extraction, and risk of bias judgements during the review. While unavoidable due to resource constraints, having a single

reviewer introduces a risk of bias to the review. To limit the effect of this, the reviewer followed pre-defined inclusion criteria during searches and used standardised forms for data extraction and risk of bias assessment. Additionally, the methods, findings, and conclusions of the review were discussed with a multi-disciplinary supervision team.

An additional limitation is the review of reviews format. Having a systematic review requirement necessitates a reasonable number of previously published articles to exist on a relationship to be identified in the review. This means more niche, or understudied, relationships between falls and chronic health conditions would not be identified using this review approach. Such conditions are discussed further in Section 4.4.3 during the application of findings from this review in regression models.

This review is also exposed to publication bias, where significant results are more often accepted for publication, and identified in the included reviews. The risk of exposure to publication bias was assessed during the ROBIS assessment in Section 3.3.4. Additionally, greater confidence that can be placed in a meta-analysis result compared to individual study findings means the relationships identified in this review may be seen as being more robust to the bias inherent in a single study estimate rising from features of the sample data.

3.5 Conclusion

This review of reviews identified a list of 18 chronic health conditions, that have been identified in previous meta-analyses as increasing the risk of falls in adults aged over 65, presented in Table 3.5. This list of conditions will be used in Chapter 4 to derive a list of conditions to include in the models of fall count data. Additionally, this review attempted to identify reviews of combinations of chronic conditions, although no reviews were identified meaning an effective synthesis in this area was not possible.

Chapter 4

Methodology

4.1 Chapter Introduction

Following the literature reviews presented in Chapters 2 and 3, this chapter presents the methodology used for the analyses in this thesis, with the corresponding results presented in Chapter 5. Section 4.2 of this Chapter explores the data to be used in the project, and how these data were processed before the analyses. Sections 4.3 to 4.4 detail the steps taken during the analyses presented in this thesis. The reasoning and theoretical underpinnings supporting the choice of particular approaches and methods of interest are also presented in sections 4.3 and 4.4, together with the steps performed during the analyses.

This chapter also introduces two aspects of novelty into the thesis. First, the thesis uses a combination of methods that has not been previously used for the investigation of associations between multi-morbidity and falls in older adults. The second example is the sample data described in Section 4.2, which identifies care home residents in routinely collected data across an NHS trust and has not been used to investigate association between multi-morbidity and falls in previous research.

4.2 Data Processing

4.2.1 Derivation of Analysis Dataset

Dataset Accessed

The dataset used for this thesis was collected as part of the Health Data Research UK (HDRUK) learning care homes project (Saliba and Buchanan, 2020). This study

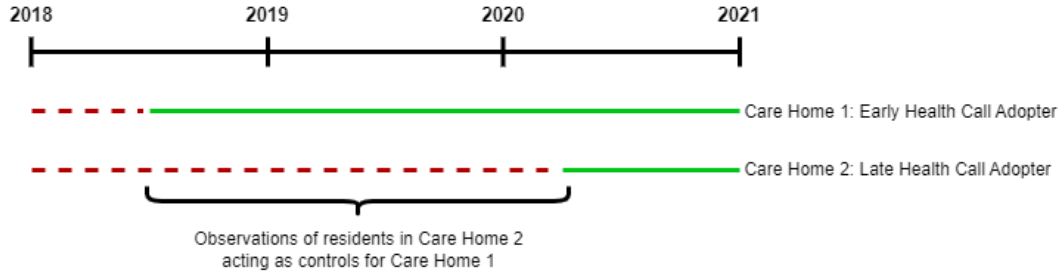


FIGURE 4.1: HealthCall Study Intervention and Controls Approach

linked electronic health records (EHRs) from the County Durham and Darlington NHS Foundation Trust using pseudonymised NHS numbers for patients who interacted with the trust hospitals and community services between 01/04/2018 and 30/09/2021.

The HDRUK study cohort was defined using registration data from the HealthCall application, which included activation date (date resident was activated on the HealthCall application), deactivation date (date of death or relocation away from care home), and care home name for each individual in the dataset. We can be sure the members of the study cohort were care home residents because the HealthCall application was being trialled exclusively in care homes (Garner et al., 2024).

The wider HDRUK study pertained to an evaluation of a digital technology for structured referrals of care home residents to hospital and community healthcare services, which was called HealthCall (HealthCall, 2023). Care home residents were identified from the HealthCall activation dataset by researchers in the HDRUK study (Garner et al., 2024).

Within the HealthCall study, controls were residents in care homes that had not yet adopted the HealthCall application, but had done so by the end of the study period as summarised in Figure 4.1.

A care home resident in the data was identified when the individual was activated using the HealthCall application being trialled exclusively in care homes. Once a care home resident was included in the study cohort, their record of healthcare resource use was identified retrospectively from their first observation in any dataset that placed them in the home (Garner et al., 2024). The HealthCall intervention was delivered through care homes adopting the technology and activating their

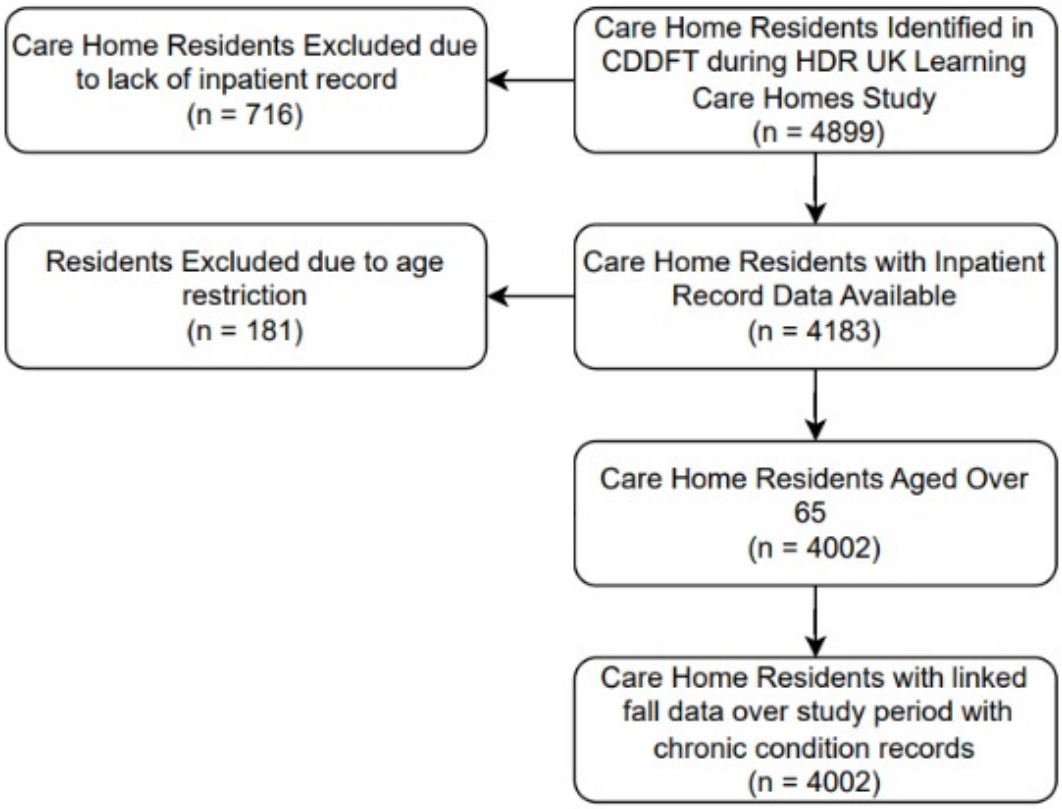


FIGURE 4.2: Sample size during data pre-processing

residents on the system.

Using the data in the HDRUK study as a starting point, the population of interest for the analyses in this thesis was adults aged over 65 residing in care homes in the County Durham and Darlington NHS Foundation Trust. The outcome of interest was fall events that led to an emergency department attendance.

Based on these criteria, a cohort of 4899 care home residents was identified during the study period mentioned previously by researchers in the HDRUK study (Garner et al., 2022; HealthCall, 2023). The pseudonomised NHS Numbers for the 4899 care home residents identified in the HDRUK study represent the starting point of the analysis for this project. EHRs detailing Emergency Department (ED) attendances, and inpatient ward episodes were identified from the available data and linked using these pseudonymised unique NHS numbers. These linked ED and inpatient records were then used for the remainder of the analyses described in this chapter.

Approach to Case Identification

The approach to case identification before the analyses is shown in Figure 4.2. Of the 4899 Care Home residents 4183 had inpatient records relating to an inpatient stay during the study period. In the record of inpatient stays, the presence of acute and chronic medical conditions was recorded using International Classification of Diseases 10th Revision (ICD-10) codes. Using these codes, a list of chronic health conditions present was derived, henceforth referred to as the chronic condition profile, for each individual. A full description for how the chronic condition profile was derived for each individual is provided later in this section. Residents without any inpatient record were excluded from the analysis ($n = 716$). The absence of inpatient records for these residents could indicate that these patients were not hospitalised during the study period. However, it is also possible that care homes on the border of the trust may have sent residents to hospitals in other trusts, meaning their records would not have been available for study in this project. Additionally, it is possible records could be missing or omitted due to problems inherent in structured data reporting (Zahabi, Kaber, and Swangnetr, 2015; Carayon et al., 2017). The impacts of these issues are discussed further in Section 6.5 together with other limitations in the analyses.

Age was identified through the minimum age on record when an individual was first observed in the available data. Information on age was available for all members of the cohort. Residents with an age below 65 were excluded from the analysis (n = 181) leaving 4002 residents in the study cohort, as seen in Figure 4.2. Older people living in care homes were the sample in this analysis because, as described in Section 1.2.1, people in care homes are highly susceptible to falls, while the majority of falls-related research is conducted in the community dwelling setting. Therefore, the analyses in this thesis focuses on an under researched population, which together with the novel combination of methods, contributes to the novelty of the research.

Time in the Cohort

Time in the cohort was calculated as the study start date until either the end of the observation period or, where applicable a date of death. Time in the cohort

(measured in days) was used to derive the time offset value described further in Section 4.4 for the count data regression analysis in Sections 5.4 and 5.5.

Identification of Falls

Falls were identified as any ED presentation with treatments and diagnoses indicating a traumatic injury in the EHR. This definition was applied because falls are the leading cause of trauma-related injuries in people over 65 (Samaras et al., 2010; Atinga et al., 2018). A list of the specific treatments and diagnoses used to form the fall presentation definition is available in Appendix Section F.

This approach to identifying falls was taken as a compromise due to no information being available in routine ED or care home data for the identification of falls. An alternative to this approach of identifying falls at ED would have been to use fall-corresponding codes in the ICD-10 codes in the inpatient ward episode data (World Health Organisation, 2004). However, this would have would have led to two major limitations. First, it would have been unclear whether the fall had occurred in the care home or in hospital. This is problematic because in-hospital falls are different in their aetiology, and risk factors to falls occurring in care homes meaning they are outside the scope of the project. Second, individuals with a more serious condition burden may be more likely to be admitted to hospital, leading to a systematic bias in the outcome variable for the regressions. The effect of this would be to artificially strengthen the significance and magnitude of relationships between fall admissions and chronic conditions that cause a more serious health state. More serious condition burden may also increase the likelihood a resident was transferred to ED following a fall, meaning this bias may still persist in the regression outcome, however this bias would be smaller than the alternative identification of falls in the in-hospital setting. This limitation in the analysis is discussed further in Section 6.5.

The presence of a fall event was indicated on each ED presentation and linked to the NHS number. The sum of these fall presentations to ED over the full study period was calculated for each individual NHS number, henceforth referred to as the count of falls. The count of falls during the study period was then used as the primary outcome in the regression models specified in Section 4.4.

Grouping Chronic Health Conditions

Chronic health condition profiles were derived by grouping all ward episodes across all hospital spells during the study period for each individual together, and identifying the list of unique ICD-10 codes recorded for each individual in the inpatient data. This list of unique codes for each individual was then matched against the groups of ICD-10 codes identified in Calderón-Larrañaga et al. (2017). In this study, Calderón-Larrañaga et al. (2017) proposed a classification system for collating ICD-10 codes to identify chronic multi-morbidity in older adults. Distilling definitions of chronic disease from a range of recognised health organisations, the multidisciplinary team of authors defined chronic disease through a series of key features (Calderón-Larrañaga et al., 2017). These features were the duration of the time with the condition, prognosis or trajectory of the condition, the reversibility of symptoms, treatments required, and consequences in terms of disability and changes in quality of life. Calderón-Larrañaga et al. (2017) then considered each ICD-10 code as being chronic against these criteria through a process of selection by two teams followed by review and final judgement by a third independent team of geriatricians. Once codes were identified as being chronic, the authors then grouped these codes into larger overarching categories. This grouping was performed based on clinical criteria, and the relevance of the conditions to one another based on features of the condition, treatment, prognosis, and prevalence. A summary table of the resulting groups of ICD-10 codes is provided in Appendix Section H.

The matching of ICD-10 codes in the sample data to the groups derived in Calderón-Larrañaga et al. (2017) was undertaken using the 2-digit over-arching ICD codes available in the sample data rather than the more specific 3-digit format used by the authors which was not available. This limitation in the data is likely to have introduced a level of misclassification when identifying the chronic condition groups used in the analysis, although this was unavoidable due to constraints in the available data. The effects of this issue are discussed further, together with other limitations, in Section 6.5.

Presence or absence of these groups of codes from Calderón-Larrañaga et al. (2017) was then recorded in the analysis dataset with a single binary indicator for each group of chronic health condition ICD-10 codes, reported alongside each NHS number. For example, in this system an individual with any of the five ICD-10 codes

(E10, E11, E13, E14, E891) relating to the diabetes group in Calderón-Larrañaga et al. (2017) present on their inpatient record would be assigned a positive value for the diabetes group in the analysis.

Chronic condition groups with a prevalence of less than 1% were not included in the analysis. This step was taken to reduce the dimensionality in the data. Of the 60 Calderón-Larrañaga et al. (2017) groups of chronic conditions, 12 were eliminated from the analyses at this stage. These groups were asthma, chromosomal abnormalities, chronic infectious diseases, hematological neoplasms, inflammatory bowel diseases, migraine and facial pain syndromes, multiple sclerosis, neurotic, stress-related and somatoform diseases, obesity, Other skin diseases, peripheral vascular disease, solid neoplasms. The impact of excluding these groups is discussed further in Section 6.5.

Identifying Multi-morbidity

Multi-morbidity was defined within this study as the presence of two or more Calderón-Larrañaga et al. (2017) chronic condition groups on the inpatient record of a study participant. Chronic conditions were used for this project rather than acute because it is reasonable to assume these conditions were present over most of, if not the entire, study period.

Identifying Frailty

As discussed in Section 1.3.2, there is overlap between multi-morbidity and frailty, and frailty can exist outside of multi-morbidity. Additionally, the relationship between frailty and falls risk in care home residents warrants further examination. For this three index scores attempting to capture frailty burden were calculated for each individual in the data set based on their inpatient record data and available demographic information. These were the Charlson Comorbidity Index (CCI) (Charlson et al., 1987), Electronic Frailty Index (EFI) (Clegg et al., 2016), and Hospital Frailty Risk Score (HFRS) (Gilbert et al., 2018).

The CCI is a weighted index score derived from a model of long term survival with co-morbid chronic health conditions in a sample of New York patients in the in-hospital setting. The CCI is calculated through assigning scores for each chronic

health condition present (0, 1, 2, 3, or 6), adding one point for each decade in age over 50, then calculating the sum of points for the individual (Charlson et al., 1987). In the original paper, scores were grouped into 0 (none), 1-2 (mild), 3-4 (moderate), $\geq$ 5 (severe) (Charlson et al., 1987). The CCI has been extensively validated in the in-hospital setting and is often included as a variable in prediction models as a proxy of co-morbidity burden in older adults (Quan et al., 2011; Radovanovic et al., 2014).

The EFI is based on the cumulative deficit frailty model, which uses a range of indicators (deficits) to identify frailty (Mitnitski, Mogilner, and Rockwood, 2001; Clegg et al., 2016). These include chronic diseases, specific symptoms or disabilities, and behaviours (Mitnitski, Mogilner, and Rockwood, 2001). The EFI includes the 36 deficits listed in Table 4.1 for the identification of individuals with frailty. To calculate the EFI a score of 1 is applied to each deficit, with the total for an individual divided by the total number of deficits (36). Categories of frailty can then be applied as follows: fit (EFI $\leq$ 0.12), mild frailty (0.12-0.24), moderate frailty (0.24-0.36), severe frailty (EFI $\geq$ 0.36). The cut-offs for these categories were derived using the quartiles of the EFI value with the 99th centile as the upper limit (Clegg et al., 2016). Following specification of the scale, the authors internally and externally validated the EFI using primary care records of over 900,000 older adults in England (Clegg et al., 2016). This validation identified the EFI effectively differentiated different risk groupings for mortality, hospitalisation, and nursing home admission.

The HFRS is a frailty risk score calculated from ICD-10 code information relating to chronic health conditions, symptoms and behaviours, and care procedures in the in-hospital setting (Gilbert et al., 2018). The HFRS is calculated using 109 ICD-10 codes with point values assigned to each code between 0.1 (fever of unknown origin) and 7.1 (dementia in alzheimers disease). These point values are based on the odds ratios of each ICD-10 code calculated in a logistic regression using membership of a cluster of frail patients as the outcome variable (Gilbert et al., 2018). HFRS can be separated into three categories of frailty risk; low risk (HFRS $\leq$ 5), intermediate risk (5–15), and high risk (HFRS $\geq$ 15). These cut-offs were selected by the authors based on the HFRS scores that best differentiated between proportion of patients experiencing 30-day mortality, long length of hospital stay, and emergency re-admission. Validation for the HFRS was performed in two cohorts of over 75s

using hospital episode statistics and NHS data. During validation in a cohort of 569 patients, the HFRS was compared to the Fried phenotype definition of frailty, and Rockwood frailty scores with positive correlation between the scales (Fried et al., 2001; Rockwood and Mitnitski, 2007; Gilbert et al., 2018).

These three indices were identified in the analyses using two-digit ICD10 codes available in the sample data, a full list of which are available in Appendix Section G for each index. Table 4.1 provides a comparison of the information contained in the three indices for ease of reference.

TABLE 4.1: Frailty Indices Summary Table

Index	Chronic Disease States	Symptoms and Signs	Disabilities and Other Frailty Indicators
EFI	Athritis, atrial fibrillation, chronic kidney disease, coronary heart disease, diabetes, foot problems, fragility fracture, heart failure, heart valve disease, hypertension, hypotension/syncope, osteoporosis, Parkinson's disease, peptic ulcer, peripheral vascular disease, respiratory disease, skin ulcer, stroke and transient ischemic attack, thyroid disorders, urinary system disease, anaemia and haematinic deficiency	Dizziness, dyspnoea, falls, memory & cognitive problems, polypharmacy, sleep disturbance, urinary incontinence, weight loss and anorexia	Activity limitations, hearing loss, housebound, mobility & transfer problems, requirement for care, social vulnerability, vision problems
HFRS*	Alzheimer's disease, arrhythmias, chronic kidney disease, dementia, depression and psychiatric indicators, epilepsy, fractures, gastrointestinal conditions, genitourinary conditions (including urinary incontinence), hearing and visual impairments, hypotension, musculoskeletal conditions (including arthritis and osteoporosis), Parkinson's disease, presence of infectious diseases, stroke	Falls, hospital acquired conditions, speech disturbances, symptoms and signs indicators (including food, fluid, and emotional state), traumatic injuries	Care involving use of rehabilitation procedures, dependence on enabling machines or devices, social environment, care dependency
CCI	Moderate to severe chronic kidney disease, diabetes (separated into uncomplicated/end-organ damage), congestive heart failure, peptic ulcer, peripheral vascular disease, stroke and transient ischemic attack, myocardial infarction, dementia, chronic obstructive pulmonary disease, connective tissue disease, liver disease (separated into mild/moderate to severe), hemiplegia, localized solid tumour, leukemia, lymphoma, metastatic solid tumour, AIDS		

**This is a summary of the 109 ICD-10 codes included in the HFRS Calculation, the full list is provided in Appendix Section G*

Influence of COVID-19

It is important to acknowledge the potential confounding influence of the COVID-19 pandemic on the analyses. This influence means the relationships identified may not generalise to situations where the acute care system is functioning in more standard circumstances. During the COVID-19 pandemic more emphasis was placed on keeping care home residents out of hospital wherever possible (British Geriatrics Society, 2020). This change reduced the likelihood that residents who had fallen would be transported to hospital in an effort to reduce their exposure to the virus

(British Geriatrics Society, 2020). Additionally, COVID-19 may have caused the premature death of care home residents with particular chronic health conditions or advanced frailty meaning these groups may have been under represented in the analyses (Rachas et al., 2023). Finally, COVID-19 changed care home behaviour with an increased emphasis on limiting infection spread, which lead to increased isolation of residents and reductions in physical activity (Mahmood et al., 2021). This had the unintended effect of increased risk of deconditioning in residents, and an associated increase in falls risk as a result (Mahmood et al., 2021).

The effect of COVID-19 on the analyses will likely have led to a dampening of relationships between chronic health conditions and transfer to hospital as a result of a fall. Therefore the estimates of effect identified in models will likely be underestimates of the true values. The confounding influence of COVID-19 during the data collection period described in Section 4.2 means the analyses should be repeated using more complete data sets across multiple trusts during more standard periods in future, which is discussed further in Section 6.7.

4.2.2 Dataset Descriptive Analysis

Following the derivation of the analysis dataset, several sets of descriptive statistics were derived. The results of these tests are presented in Section 5.2. First the continuous age variable was tested for normality using the Kolmogorov-Smirnov test. Mean and standard deviation were derived for the continuous age variable, with frequencies and percentages in each category for categorical variables. The count of falls outcome variable was visually inspected using a histogram in Figure 5.3, together with descriptive statistics presented in Table 5.1. Additional histograms were plotted for the age variable, and number of chronic condition groups recorded in Figures 5.1 and 5.2 respectively.

4.3 Cluster Analysis of Multi-Morbidity Data

4.3.1 Cluster Analysis Introduction and Aims

A core aim of the thesis was to investigate the effect of multi-morbidity on falls in care home residents. Cluster analysis is presented in this thesis to identify the

groups of chronic conditions that are commonly co-occurring in the sample, with the methods used described in Section 4.3.3 and the results presented in Section 5.3. However, before presenting the methodology, an explanation of what cluster analysis is, and the key decisions to be made during the analysis, is required.

The core aim of cluster analysis is to place objects, in this case care home residents, into groups or clusters such that objects within a group are similar but those in different groups are dissimilar (Murphy, 2012). These group characteristics can be generally thought of as within-group homogeneity and between-group heterogeneity. A measure of the similarity between objects is required for a clustering algorithm to identify these groups within data. When objects are plotted in geometric space, similarity can be expressed as the distance between objects. However, one of the key difficulties in clustering multi-morbidity data is that the variables indicating the presence or absence of a condition are binary. This is a problem because similarity between individuals cannot be directly expressed through 2D geometric distance measures such as Euclidean distance when the available information is expressed in binary variables (Bishop, 2006). Therefore, a pre-processing stage is needed before clustering can be performed on the data. Previous studies have used a dimensionality reduction approach called Multiple Correspondence Analysis (MCA) for this pre-processing, which produces output suitable for use with clustering algorithms (Violán et al., 2018; Guisado-clavero et al., 2018; Violán et al., 2019; Machón et al., 2020). MCA is described in greater detail in Section 4.3.2.

Another decision to be made in cluster analysis is the choice between hierarchical or non-hierarchical clustering methods. Hierarchical clustering uses partitions of the data set through agglomerative (bottom up) or divisive (top down) techniques to cluster the data (Murphy, 2012). Output from hierarchical clustering can be viewed using a dendrogram (tree diagram), in which clusters are nested within larger clusters. One reason given for taking a hierarchical approach is that co-occurring diseases may result from underlying risk factors or genetics (Vu, Finch, and Day, 2011).

Non-hierarchical approaches are common in the multi-morbidity clustering literature and work through iteratively splitting or merging clusters based on an objective function. Examples that have been used include K-means clustering (Violán et al., 2018; Guisado-clavero et al., 2018), K-medoids (Islam et al., 2014), and Fuzzy

C-means (Violán et al., 2019). These approaches differ in several ways. The K-means algorithm defines k clusters where membership of a data point to a cluster is calculated according to the least-squared Euclidean distance between data points. K-means differs from K-medoids in that the latter uses a data point for the centre of the cluster rather than the mean value of all points in the cluster, which makes K-medoids less susceptible to outliers. Fuzzy C-means allows data points to be members of multiple clusters meaning an association between each data point and each cluster centre is calculated with closer centres having a greater association. This is particularly useful when clustering overlapping groups of data points where membership in multiple clusters is plausible. Generally, non-hierarchical approaches are more robust to outliers, choice of the distance measure, and inclusion of irrelevant variables in the analysis than their hierarchical counterparts (Everitt, Landau, and Leese, 2010).

In both hierarchical and non-hierarchical approaches a key decision to be made is the number of clusters to identify. There is a risk of bias present in non-hierarchical methods, which results from needing to decide the number of clusters *a priori*. However, the end solution produced by a hierarchical approach is often over trained, with data points separated into their own individual clusters. Therefore a decision needs to be made within hierarchical clustering approaches regarding where to stop the algorithm and take the resulting clusters at that stage. This decision regarding where to cut the dendrogram at an intermediate point also carries a risk of bias. Use of the Calinski-Harabasz (CH) index is common in the literature to justify choosing a particular number of clusters in these situations (Vu, Finch, and Day, 2011; Foguet-Boreu et al., 2015; Guisado-clavero et al., 2018; Violán et al., 2018; Machón et al., 2020). The CH-index is a ratio of the between-and within-cluster dispersion, where high values indicate greater separation between clusters (Caliński and Harabasz, 1974). The optimal number of clusters is identified by conducting multiple iterations with different numbers of clusters and taking the iteration with the highest CH index value. However, this represents a data-driven approach, meaning it is possible the clusters identified will not be reflective of real groups of patients.

Decisions made during the cluster analysis, were motivated by overcoming three challenges. These challenges were the clustering of binary chronic health condition

data, the choice of cluster algorithm, and number of clusters to identify in the final solution.

This cluster analysis aimed to answer three questions related to the overarching thesis research question:

1. In care home residents, what chronic health conditions regularly co-occur together?

2. To what extent do these clusters match those identified previously in community dwelling older adults?

3. To what extent are the clusters of chronic health conditions identified associated with changes in the rate of falls in care home residents?

The first of these questions is answered using the methodology presented in Section 4.3.3 and the results are presented in Section 5.3. The second question is addressed using the cluster results presented in Section 5.3.4, which are compared to previous studies in the community setting in Section 6.3.1. The third question is answered using the methodology described in Section 4.4.3, and the results are presented in Section 5.4.

4.3.2 Multiple Correspondence Analysis (MCA)

As mentioned in Section 4.3.1, binary variables are nominative and therefore similarity between individuals cannot be directly expressed through 2D geometric distance measures such as Euclidean distance (Bishop, 2006). Multiple Correspondence Analysis (MCA) can be used as a pre-processing step when identifying multi-morbidity patterns in categorical data (Le Roux and Rouanet, 2010).

MCA is a dimensionality reduction technique, which summarises the information contained in categorical variables using a graphical representation. This graphical representation is based on the axes which maximise the variance in the data (Husson, Le, and Pagès, 2017).

MCA output plots the row and category profiles together, with the distance and angle with the origin between points relating to the strength of the relationship between them. The axes used to construct this output are the two perpendicular axes

that maintain the maximum variance of the point clouds of row and category profiles that are in multi-dimensional space. It is useful to see the two dimensional plane as a slice being cut through a multi-dimensional point cloud, with all points in the cloud then being projected onto this slice at a 90 degree angle. Distance is expressed in MCA through Chi-squared distance, which is the sum of the squared differences of co-ordinates on each new axis. MCA was conducted using the software package FactoMineR (Lê, Josse, and Husson, 2008).

Within the MCA output, the position of an individual's data point is pulled towards the position of the categories they are in because the co-ordinates of their data point on the plot is the weighted sum of the co-ordinates of the categories they are in. Therefore, the position of the data point representing the individual chronic condition profile will be pulled towards the categories it is in and pushed away from the categories it is not in. This interrelated positioning, where the location of each element is dependent on the location of other elements is referred to as barycentric positioning (Husson, Le, and Pagès, 2017). The result of this barycentric positioning is that individuals (rows) that have a similar chronic health condition profile are plotted closer together.

As mentioned previously, the MCA output also provides the positioning of categories on the output plot. The positioning of these category profiles is governed by a similar system of barycentric properties described for the individual row profiles above. This means a category will be close to individuals that are in it and further away from individuals that are not, because the co-ordinates of categories on the plot correspond to the weighted sum of the co-ordinates of individuals that are in the category. Additionally, chronic health condition categories that have similar profiles of individuals with a positive value for that category are also plotted closer together.

This situation, where the individual row profiles and category profiles exhibit gravity-like effects on one another, is referred to as the double barycentric property in MCA (Husson, Le, and Pagès, 2017). This property leads to a spatial representation of similarity on a plot where the distance from the origin and between points can be calculated using euclidean distance. For the cluster analysis described in Section 4.3.3, the position of individuals on the MCA output was of interest, which is reflected in the plots of MCA output presented in Section 5.3.2.

4.3.3 Cluster Analysis Methodology

The analysis described in the Section 4.4 investigates how patterns of multi-morbidity relate to falls in care home populations. The approach to identifying these patterns of multi-morbidity is described in this section. The data were organised with diseases based on ICD-10 codes codified as binary variables indicating the absence or presence of an ICD-10 coded disease. Only diseases with more than one percent prevalence in the population were included for analysis. This is based on work in previous studies on multi-morbidity patterns in which a minimum prevalence threshold was introduced to reduce the likelihood of spurious relationships being identified during the cluster analysis (Violán et al., 2018; Violán et al., 2019).

Before the cluster analysis, multiple correspondence analysis (MCA) was used to transform the binary disease data as described in Section 4.3.2. The proportion of the total variance captured by the MCA analyses was assessed to indicate the quality of the point cloud representation by the MCA axes. Due to the binary nature of the data, a low proportion of the variance was expected to be captured by the axes (Husson, Le, and Pagès, 2017).

A comparison of clustering algorithms was carried out for the analysis of two-dimensional patterns of chronic health conditions in care home residents. This compared the performance of four clustering algorithms, comprising two hierarchical and two non-hierarchical approaches. Hierarchical algorithms tested were the Ward method, and average linkage while the non-hierarchical approaches were K-means, and K-medoids. These methods were chosen due to their use in community dwelling samples in previous research. Additionally, the relative simplicity of interpretation of the resulting clusters when compared with fuzzy cluster approaches was attractive to avoid over-complicating the analysis. A brief discussion of the theory, strengths, and weaknesses of each clustering algorithm is presented below.

Ward's Method

The Ward method is an agglomerative hierarchical clustering algorithm (Großwendt, Röglin, and Schmidt, 2019). Ward's method utilises a function, which relies on the sum of the squared distance (SSD) between individual data points (Murtagh and Legendre, 2014). The SSD refers to the total squared distance between points in each

cluster summed over all clusters. Ward's method calculates the SSD, then merges the two clusters with the lowest between-cluster sum of squared distance, which results in an increase in the within-cluster sum of squared distance. The increase in the SSD describes the distance between the two clusters that have been merged. This cycle is repeated until the data are contained a single cluster. Graphical output, called a dendogram, is produced by this method, which is used to visualise the hierarchy of clustering stages (Provost and Fawcett, 2013). Ward's method can effectively separate clusters of data in noisy sets although it is biased towards forming rounded clusters. Other drawbacks of this approach are it is computationally expensive, and due to the hierarchical nature, points placed in a cluster cannot be reassigned (Großwendt, Röglin, and Schmidt, 2019).

Average linkage Method

The average linkage method is another agglomerative hierarchical clustering approach, meaning that all points start in their own cluster before being merged together in stages. This algorithm takes all pairs of points in two clusters, computes the distance between them and takes the mean of these distances. This process is completed for all possible pairs of clusters and the two clusters with the smallest average distance are merged at each stage (Murphy, 2012). An advantage of this approach is the ability to handle noisy data, although it becomes computationally expensive in large data sets. Average linkage is also biased towards creating small well-spaced clusters (Murphy, 2012).

K-means

K-means clustering, also known as Lloyd's Algorithm, is a non-hierarchical iterative clustering approach. This requires the number (K) of clusters to group the data points into, to be decided *a priori* (Witten, Frank, and Hall, 2011). Once K is chosen, that number of cluster centres (centroids) is randomly assigned within the data to form the centres of different clusters. The distance between each object and each centroid is then calculated with objects being assigned to their closest centroid. The mean attribute values over all objects in each cluster is calculated following this initial iteration and the cluster centroids are then reassigned to these mean values

(Witten, Frank, and Hall, 2011). The distance from this new centroid to all objects is then re-calculated and objects are assigned to their nearest centroid. This process repeats until the centroids stabilise and objects do not change between clusters or until a threshold of iterations is passed (Bishop, 2006). These stages are based on the optimisation criterion in K-means, where the sum of squared distances between data points and the K centroids is minimised. The standard formula for distance in K-means is the Euclidean distance. Weaknesses of K-means include needing to define K, sensitivity to outliers and sensitivity to the initial centroid positioning (Hastie, Friedman, and Tibshirani, 2009). The K-means algorithm has been used on MCA output in a two-step solution to cluster multi-morbidity data in previous literature (Guisado-clavero et al., 2018; Violán et al., 2018; Machón et al., 2020)

K-Medoids

K-medoids is a further development of the K-means algorithm, which follows the same iterative structure but differs in the way centroid positioning is determined. While the K-means algorithm can be susceptible to outliers in data, K-medoids overcomes this by using the most centrally located data point for the centroid (or medoid) of the cluster (Hastie, Friedman, and Tibshirani, 2009). This is calculated as the point, which minimises the sum of distances to all other points in the cluster. By using the most central data point rather than the mean value, K-medoids is more resistant to cluster centres being 'pulled' towards outlier points, as can happen with the K-means algorithm (Hastie, Friedman, and Tibshirani, 2009; Jin and Han, 2010).

Cluster Algorithm Comparison

To compare the four clustering algorithms mentioned above, solutions for each algorithm needed to be derived and tested for suitability. The optimal numbers of clusters were identified for each algorithm using the Calinski-Harabasz Index (Caliński and Harabasz, 1974). This is a popular method to identify the optimal number of clusters in similar research and works effectively with all the clustering methods to be tested in this analysis (Vu, Finch, and Day, 2011; Foguet-Boreu et al., 2015; Guisado-clavero et al., 2018; Violán et al., 2018; Machón et al., 2020).

Evaluating the quality of competing clustering solutions in this context is challenging because the cluster analysis described is an example of unsupervised learning. In this context, the true answer is not known, meaning an algorithm cannot be evaluated in the same way as supervised learning where the degree of misclassification gives an indication of performance (Everitt, Landau, and Leese, 2010). The different clustering solutions were therefore compared using two intrinsic evaluation metrics.

First, cluster stability was assessed using the mean Jaccard coefficient value over 100 iterations (Hennig, 2007). Jaccard coefficient values refer to the proportion of points assigned to the same cluster as the initial solution when non-parametric bootstrapping is applied to the data to create smaller sets to train the clustering algorithm (Hennig, 2007). The Jaccard coefficient values are averaged over 100 iterations of this process to create a single metric, which indicates the stability of the clustering solution to re-sampling of the data. A mean Jaccard coefficient value of 0.85 or more is seen as indicating highly stable clusters while below 0.6 represents unstable clusters (Hennig, 2007). The more stable a cluster the less likely it is to disappear or change upon the incorporation of new data, under the assumption that the training dataset is representative of the population. Greater confidence can be placed in the external validity of stable clusters as they can be expected to be consistent in new samples of the same population.

Second the Calinski-Harabasz (CH) Index will be used to investigate the definition or separation of the clustering solutions (Caliñski and Harabasz, 1974). This index is a ratio of the sum of between and within cluster dispersion for all clusters, where dispersion is defined as the sum of the squared distances between points and higher CH index values indicate more closely-defined clusters.

The algorithm chosen for the final model was based on the highest CH index value, provided the solutions have mean Jaccard coefficients that indicate a stable solution. This ensures the clustering solution being taken forward will have the best possible separation of clusters once a minimum threshold of 0.75 for stability has been passed. The chosen threshold is based on the rule of thumb proposed in Zumel and Mount (2014) which will ensure at least moderate stability in the solution carried forwards. Further analysis of the final clustering solution was performed as follows.

Prevalence in the cluster and Observed-Expected (O/E) ratios were calculated for each chronic health condition in the resulting clusters. The OE ratio is a ratio of the disease prevalence in the cluster compared to the prevalence in the full study sample (Schäfer et al., 2014; Violán et al., 2018). The OE ratios aid interpretation of the resulting clusters by identifying conditions highly associated with a particular cluster.

The individual cluster membership in the final clustering solution was then used during the analysis of fall count data described in Section 4.4 below. This followed the original analysis plan, which used the cluster solution as an avenue to explore the association between different types of multi-morbidity and falls in care home residents.

4.4 Modelling Fall Count in a Care Home Resident Sample

4.4.1 Regression Analysis of Falls: Aim and Objectives

In order to explore the effects of various key variables from the EHR on fall presentations to ED, a regression approach was used. The aims of this regression analysis relates to overall thesis objectives 3, 4, 5, and 6 presented in Section 1.5.1..

The overall aim of the regression analysis was to analyse whether and how chronic disease and multi-morbidity are associated with the count of fall presentations by cohort members to the ED during the study period.

More specifically, the objectives were to:

- Identify the association between chronic health conditions and the count of fall presentations to ED during the study period.
- Utilise the clustering solution from Section 4.3 further to investigate the association of cluster membership with the count of fall presentations to ED during the study period.
- Investigate the effect of interactions between these chronic health conditions and what they can tell us about their relationship with fall count.

4.4.2 Selection of Generalised Linear Models

The outcome to be modelled in the regression analyses is a count variable, which represents a measure of the number of times an event, in this case fall presentations to the ED, occurred within a specified time period (Cameron and Trivedi, 1998).

Regression analysis of count data involves a dependent variable, which measures a count of an event occurring, with the conditional mean of this count variable dependent on a set of independent variables. However, because count data is collected during a window of time, the dependent variable in count regression actually represents a rate, i.e., the number of events over a period of time (Cameron and Trivedi, 1998).

It is typical for the Log-linear or Poisson distribution to be adopted for count data regression (Hilbe, 2011). However, count data is often over-dispersed with the conditional variance exceeding the conditional mean (Cameron and Trivedi, 1998). This can lead Poisson regression models to underestimate standard errors, possibly causing the significance of explanatory variables to be misinterpreted (Hilbe, 2011). Furthermore, in situations where count data encompass zero, zero-inflation may be occurring (Cameron and Trivedi, 1998). Zero-inflation encompasses a situation where more zero counts than would be consistent with the Poisson distribution are present in the count variable (Cameron and Trivedi, 1998). For the count data model results presented in Section 5.4, over-dispersion was identified using the decision rule in Equation 4.1 (Hilbe, 2011). In the situation where over-dispersion occurs, Negative Binomial regression will be used for the analysis as the equidispersion assumption is relaxed due to an additional term in the relationship (Hilbe, 2011).

$$(RD)/(DF) > 1 \tag{4.1}$$

Where RD = Residual Deviance, DF = Degrees of Freedom

The additional term in Negative Binomial regression allows the conditional variance to exceed the mean (Cameron and Trivedi, 1998). This makes Negative Binomial suitable for modelling over-dispersed count data (Hilbe, 2011).

4.4.3 Regression Methodology

Count of falls during the three-year study period was used as the outcome for five regression models. The tables of results for these regression models, as well as the associated tests described below, are presented in results Section 5.4. Initial models were developed using Poisson regression. The presence of over-dispersion in the resulting models was identified using the expression in Equation 4.1. In the situation where over-dispersion was present, Negative Binomial regression was used instead (Hilbe, 2011).

Explanatory variables included in the five regression models are presented in Table 4.2. A time-offset variable was included in all regression models, calculated by taking the natural log of time spent in the cohort, measured as the difference between the start of the study to either the end of the study period, or a date of death. This was included in the regression models to account for the differing lengths of time spent in the cohort by sample members (Hilbe, 2011). The natural log of time is used in this offset because the event rate in Equation 4.2 is rewritten as Equation 4.3 through the quotient rule of logarithms.

$$log(\frac{\mu}{t}) = \beta_0 + \beta_1 x_1 \tag{4.2}$$

$$log(\mu) = log(t) + \beta_0 + \beta_1 x_1 \tag{4.3}$$

In all models, age and biological sex were identified when an individual entered the cohort, as described in Chapter 4.2.1, with CCI, EFI, HFRS summarised for each individual over the entire study period in regressions 1, 2, and 3 respectively. These regressions were carried out to explore whether frailty and co-morbidity scores were associated with variation in care home resident fall patterns.

Regression model 4, in Table 4.2, included the membership of the final clustering solution derived using the approach detailed in Section 4.3.3, with the resulting cluster model presented and evaluated in results Section 5.3.4. By including the cluster membership, the aim of this regression was to investigate whether different groups of chronic conditions were associated with different numbers of fall presentations to

the emergency department.

In the fifth regression model chronic health condition presence was included individually using binary variables, the derivation of which is described in Section 4.2.1.

Chronic health conditions to be included in this model follow from the short-list derived in the review of reviews in Section 3.4.1, provided that an ICD-10 code could be identified for the condition of interest, and the prevalence of the condition was over 1% in the study sample.

Several of the chronic health conditions mentioned in the shortlist derived from the review of reviews in Chapter 3 did not match up directly with the chronic health condition groups being used in this thesis. The conditions in the shortlist affected were heart disease, postural (orthostatic) hypotension, and digestive diseases. The category of 'heart disease' was split into three groups from Calderón-Larrañaga et al. (2017), cardiac valve disease, ischemic heart disease, and heart failure.

In the absence of a group directly reflecting the postural hypotension category, the more general grouping of hypotension was used instead. The use of this more general hypotension grouping likely introduced misclassification bias due to the wider range of codes included in the overarching category. Specifically, chronic hypotension, idiopathic hypotension and hypotension related to drugs would be captured by the hypotension 2 digit ICD-10 code in addition to postural hypotension. Additionally, a wider range of syndromes can cause hypotension than postural hypotension, meaning the influence of these other syndromes was captured by the hypotension indicator in the analyses. As a result, the findings from the analyses cannot be reliably related directly to postural hypotension, and only relate to the more general hypotension category. The lack of specificity in the ICD-10 codes is a limitation of the analyses, which is discussed further in Section 6.5. Finally, the general category of digestive disease recommended in the review of reviews was split into three groups from Calderón-Larrañaga et al. (2017), colitis and related diseases, oesophagus, stomach, and duodenum diseases, and prostate diseases.

Three of the categories recommended in the Chapter 3 review of reviews were not included in the shortlist regression model. For the Sarcopenia category no appropriate 2-digit format ICD-10 code could be identified. Additionally, the ICD-10

code indicating the presence of chronic pain had a prevalence of less than 1% in the sample and were therefore not included in the regression model. Finally, the two Calderón-Larrañaga et al. (2017) groups indicating the presence of cancer had a prevalence of less than 1% in the sample and were excluded from the analysis on this basis.

Finally, three Calderón-Larrañaga et al. (2017) chronic health condition groups were added to the review of reviews shortlist for inclusion in the regression models. These groups indicated the presence of inflammatory arthropathies, peripheral neuropathy, and chronic kidney disease. Peripheral neuropathy was added to the model because the relationship between diabetes and falls acts partially through peripheral neuropathy (Timar et al., 2016; Riandini et al., 2020). Therefore, to separate out these effects an indicator for peripheral neuropathy was included. The Calderón-Larrañaga et al. (2017) group indicating the presence of inflammatory arthropaties was added to the shortlist model due to symptoms including pain and stiffness in joints which, especially when present in the lower limbs, could cause gait and balance issues, with an associated increase in falls risk (Armstrong et al., 2005; Hayashibara et al., 2010; Stanmore et al., 2013). Finally, chronic kidney disease was included in the regression model following a previous systematic review that found mixed results regarding an association with accidental fall rate but a strong positive association with fracture risk (Goto et al., 2020).

Following all regressions, Incident Rate Ratios (IRR) were derived by taking the exponent of the regression coefficient estimates (Hilbe, 2011). The IRR quantifies how the rate of fall presentations to ED during the study period changed when the explanatory variable of interest was present compared to when it was not present.

Once the regression models were derived, Akaike Information Criterion (AIC) values were used to compare the quality of fit between the regression models, with a lower AIC indicating a better fit of the regression line to the data (Hilbe, 2011).

Due to the difficulty in visually inspecting count data residual plots accurately, model diagnostics were performed using the DHARMa package (Hartig, 2022a). This package simulates expected residuals versus the fitted Generalised Linear Model (GLM) residuals (Hartig, 2022a). The output allows the formation of multiple tests and plots to test for model misspecification. Tests for uniformity of the distribution

TABLE 4.2: Main Effects Regression Models Independent Variables

Main Effects Regression Models	**Independent Variables**
1: Charlson Comorbidity Index	Age, Male Sex, CCI Score
2: Electronic Frailty Index	Age, Male Sex, EFI Category
3: Hospital Frailty Risk Score	Age, Male Sex, HFRS Category
4: Cluster Membership	Age, Male Sex, Cluster Membership
5: Shortlist	Age Male Sex Dementia Cerebrovascular Disease Parkinson's Disease Urinary Incontinence Diabetes Anaemia Blindness and Visual Impairment Other Eye Diseases Glaucoma Deafness and Hearing Impairments Hypotension Atrial Fibrillation Bradycardia and Conduction Diseases Cardiac Valve Diseases Ischemic Heart Disease Heart Failure Hypertension Depression and Mood Diseases Inflammatory Arthropathies Osteoarthritis and other Degenerative Joint Diseases Colitis and related diseases Oesophagus, Stomach and Duodenum Diseases Prostate Diseases Peripheral Neuropathy Chronic Kidney Diseases

EFI Categories: Fit, Mild Frailty, Moderate Frailty, Severely Frail, HFRS Categories: Low Risk, Intermediate Risk, High Risk

versus expected, outliers, over-dispersion, and zero inflation were conducted using the DHARMa package, with the results reported for each count data regression model in Sections 5.4 and 5.5.

4.5 Modelling Fall Count: Interaction Analysis

4.5.1 Aims

Following on from the findings in the initial regression models presented in Section 5.4, further analysis into the interactions between chronic health conditions and their effect on fall presentations to ED was conducted.

This analysis sought to identify whether interaction effects between chronic health conditions could be identified in the data, and whether this could reveal multiplicative or mediating relationships between co-morbidities in determining the count of fall presentations to ED by care home residents.

When conducting interaction analysis, there are several concepts which must be explained such that the final results can be understood in their entirety. These features are explored in Section 4.5.2.

4.5.2 Interactions Interpretation

Interactions are used in statistical modelling when the value of one explanatory variable has an effect on the relationship between another explanatory variable and the outcome (Hilbe, 2009). All interactions proposed in the analysis are between binary explanatory variables, therefore this explanation of their effects and interpretation will relate to binary interactions only.

For the regression Equation 4.4 below, the expected count of falls (μ) is the outcome variable, with binary explanatory variables (x_1 and x_2) representing a specific chronic health condition presence or absence.

$$log(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{4.4}$$

$$IRR_{x_1} = \Delta\mu / \Delta x_1 = exp(\beta_1) \tag{4.5}$$

$$IRR_{x_2} = \Delta\mu/\Delta x_2 = exp(\beta_2) \tag{4.6}$$

As seen in Equations 4.4 to 4.6, when no interaction effect is present in the count regression Equation 4.4 the incident rate ratios are equivalent to those seen in the models described in Section 4.4.2. Equations 4.5 and 4.6 show that the measure of effect on the expected count of each explanatory variable is solely dependent on the regression coefficient for that variable. However, this main effect does not account for any mediating effects the co-occurrence of chronic health conditions may have on their relationship with falls.

For example, in Equation 4.7, an interaction term is added to represent the situation where both binary condition variables (x_1 and x_2) are present. By adding this interaction term, as well as interpreting the main effect of each condition seen in Equations 4.5 and 4.6, there is now an additional relationship which is conditional on the presence or absence of the other condition, as seen in Equations 4.8 and 4.9.

$$log(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 \tag{4.7}$$

$$IRR_{x_1} = \Delta\mu/\Delta x_1 = exp(\beta_1 + \beta_3 x_2) \tag{4.8}$$

$$IRR_{x_2} = \Delta\mu/\Delta x_2 = exp(\beta_2 + \beta_3 x_1) \tag{4.9}$$

Therefore when interpreting the Incident Rate Ratio (IRR) estimates on a main effect ($exp(\beta_1)$ or $exp(\beta_2)$) directly this effect represents the independent impact of the condition of interest; however, when the additional condition is present the additional IRR on the interaction term between the conditions must be taken into account. The direct effect of condition x_1 in Equation 4.7 when condition x_2 is not present is shown in Equation 4.10 below.

$$IRR_{x_1|x_2=0} = \Delta\mu/\Delta x_1 = exp(\beta_1 + \beta_3 x_2) = exp(\beta_1 + \beta_3 * (0)) = exp(\beta_1) \tag{4.10}$$

By incorporating interaction effects in the regression models, it may be possible to understand the mediating role that co-morbidity plays in the association between specific chronic diseases and fall count. However, there is a trade-off present with complexity of the model in interaction analysis, with higher order interactions becoming increasingly difficult to interpret. Additionally, increasing the number of terms in the analysis would increase variance in the model estimates, and increase the risk of identifying spurious relationships.

4.5.3 Interaction Analysis Methodology

Due to the high possible number of interactions to test, several approaches were taken to minimise the number of combinations to consider in the interaction analysis. The methods taken here represent a more data-driven than hypothesis-driven approach, which is a limitation discussed further in Section 6.5.

This analysis took several different approaches in order to investigate possible interactions between chronic health conditions and identify any mitigating effects they may have on the count of fall presentations to ED from care home residents in the sample.

Following the count regression in Section 5.4, which incorporated the cluster model findings derived in Section 5.3, interaction analysis was conducted for each cluster (besides the absence cluster) separately. The absence cluster refers to the pattern of in-patient records with little or no chronic condition groups present that were groups together during the MCA, described in Section 4.3.2.

To identify conditions for interaction in these clusters, the O/E ratios were used to identify the conditions that were most associated with a particular cluster. Interacting these conditions allowed the testing of whether the combination of these conditions led to an increase or decrease in falls risk among people in that cluster. The intention behind these interaction models was to explain why the effects on fall count observed in the cluster regression model occurred. Next, the top five chronic conditions by prevalence in the whole sample and each cluster were interacted in separate models.

For each model derived during the interaction analysis, DHARMa plots and tests were produced, with interpretation equivalent to those described in Section 4.4.3

previously. Plots of residuals and QQ-plots were derived and investigated for each model described above.

4.6 Alternative Methodologies Trialled

Several alternative approaches to answering the research question were explored that are not presented in the thesis. Each of these is mentioned in this section, alongside a brief explanation for why the approach was not continued.

The first of these alternative approaches attempted to identify 'avoidable' fall attendances to ED using an established definition to develop a binary outcome variable. This strand of analysis was abandoned due to the clinical need often present in fall presentations from care homes, which made them necessary by definition.

An alternative approach to the regression analyses was trialled, which used whether a presentation to ED resulted from a fall as a binary outcome in logistic regression analyses. This analysis was discontinued due to the effects of Covid-19 lockdowns, which likely led to a change in the relationships between predictors and the outcome at different times in the analysis.

Attempts to use network analysis of the chronic health condition profile data to visualise the structure of relationships in multi-morbidity were attempted and abandoned due to software availability in the TRE framework and time constraints.

Finally, previous authors have split their sample by biological sex when conducting analyses into multi-morbidity (Violán et al., 2018; Guisado-clavero et al., 2018). A similar approach was trialled for this project however it was abandoned due to existing disparity between sexes in terms of healthcare outcomes, which would likely have led to spurious findings.

4.6.1 Ethical Approval

Ethical approval for the project was obtained through the University of Sheffield Research Ethics Committee self-declaration system, which confirmed that, because the data were fully anonymised then further ethics approval was not required. Approval for the project was given on 19/11/2021. A copy of the Ethical approval letter for the project is provided in Appendix Section I.

Additional approval for accessing the research data was also given through an approved data user authorisation agreement. A full copy of this agreement is provided in Appendix Section J. The time window in which the sample data were accessed for analysis in this project began following ethical approval on 19/11/2021 through until 24/03/2023 at which point they were deleted at the end of the wider HDRUK project. The effect of this hard deadline for the analyses is discussed further in Section 6.5.

4.6.2 Software Used

Software used during the analyses, and construction of the thesis were as follows. R-studio version 1.2.5033 with a range of packages was used for the processing and analyses of data (RStudio Team, 2020; Wickham et al., 2019; Wickham and Bryan, 2023; Grolemund and Wickham, 2011; Wickham, 2016; Lê, Josse, and Husson, 2008; Kassambara and Mundt, 2020; Hennig, 2023; Maechler et al., 2022; Venables and Ripley, 2002; Hartig, 2022a).

4.7 Summary

The methodology presented in this chapter represents a novel combination of methods for the investigation of multi-morbidity and falls in older adults. Additionally, the sample data represents a novel dataset for investigating key questions surrounding multi-morbidity and falls in care home residents. While it is clear that the data described in Section 4.2 were limited in terms of the information contained in the EHR, and the time period of collection, when the influence of the Covid-19 pandemic undoubtedly impacted behaviour of staff and managers etc. in care homes and the emergency care services. In addition, the structured referrals application being trialled in the sample care homes may have led to a change in the relationship between care homes and ED, meaning it is possible the relationships identified may not be transferable to other care home resident populations without external validation. These limitations are explored further in Section 6.5. However, despite these limitations, it was still possible to leverage the existing features in the EHR to analyse falls risk in care home residents, and the novel approach taken in this thesis

could be applied to more robust and rich datasets in future research. The results of the analyses are presented in Chapter 5, after which the key themes and differences with existing literature are explored further in the discussion (Chapter 6).

Chapter 5

Results

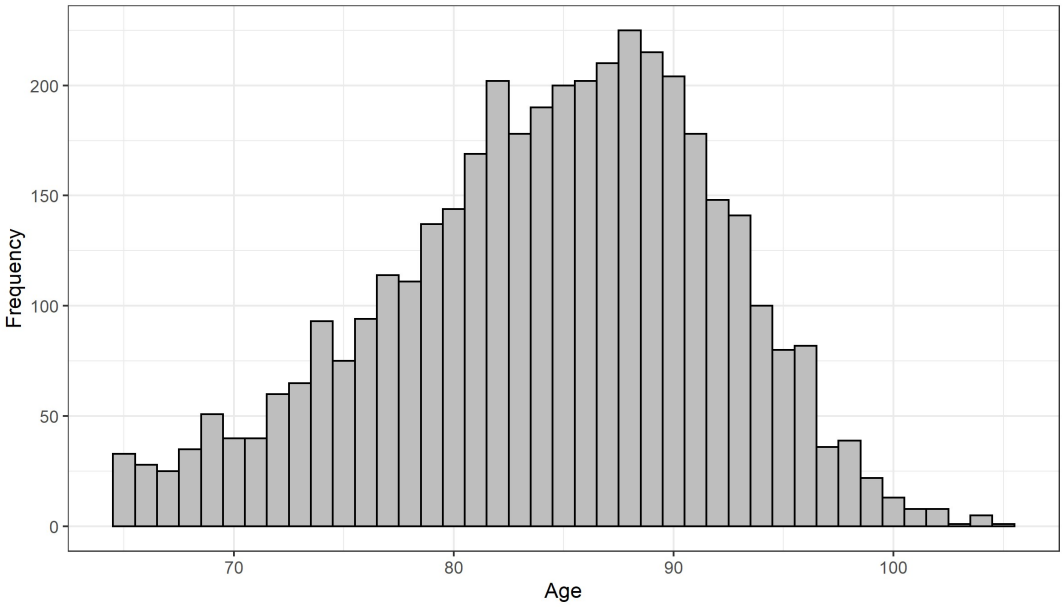
5.1 Introduction

This chapter presents the results of the analyses. Section 5.2 reports the descriptive statistics and tests performed on the sample data. Section 5.3 provides the results of the cluster algorithm comparison and testing. By identifying clusters of sample members based on their multi-morbidity information a picture of the different presentations of multi-morbidity in care home residents was formed. However, these clusters were also leveraged to expose how differences in multi-morbidity are associated with changes in fall presentations by sample members. Section 5.4 specifies the results of the five regression models specified in Section 4.4.3, including a model using the cluster membership from Section 5.3 as an explanatory variable. These regressions also investigated aspects of how chronic disease relates to falls, with three models exploring the role of frailty, and a further model which used the shortlist of chronic health conditions derived in the Chapter 3 review to further explain the differences observed between the clusters. Finally, results from the interaction effect analysis are presented in Section 5.5. These interaction models attempted to identify specific combinations of chronic conditions, which increased falls risk in sample members.

5.2 Descriptive Statistics

The sample size for the analyses was 4002 care home residents. Demographic characteristics of the data were explored using descriptive statistical tests. These results are presented in Table 5.1. 37.4% of the sample was male (n = 1498), with a mean

FIGURE 5.1: Histogram of Age



Age identified as the minimum age on record

age of 84.4 (SD = 7.6). The distribution of age in the sample is shown in Figure 5.1. Use of the Kolmogorov-Smirnov test found age in the sample was significantly ($P \leq 0.05$) non-normally distributed with a negative skew.

TABLE 5.1: Analysis Dataset Descriptive Statistics

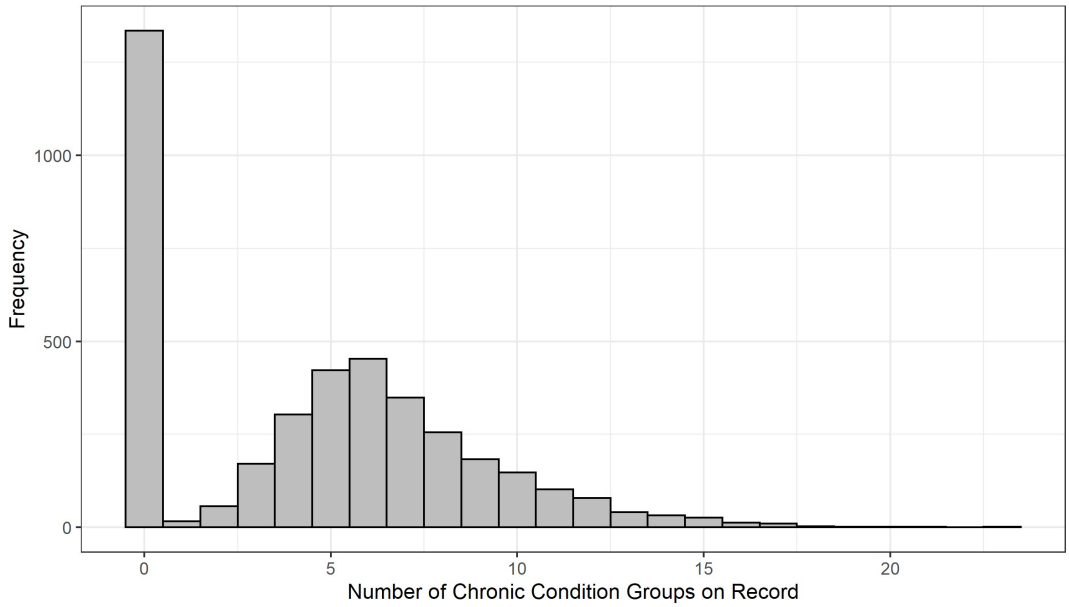
Categorical Variable	**N (%)**
Male Sex	1498 (37.4%)
Multi-morbidity on Record	2651 (66.2%)
Continuous Variable	**Mean (SD)**
Age	84.4 (7.6)
Number of Chronic Condition Groups	4.6 (4)
Days in Cohort	835.6 (398.5)
Number of Fall Presentations on Record	0.7 (1.1)

Multi-morbidity defined as two or more Chronic Condition Groups on inpatient record

The chronic disease burden is high in this sample, with a mean number of Calderón-Larrañaga et al., 2017 chronic condition groups on the EHR of 4.6 (SD = 4). This was further shown by the high proportion of multi-morbidity, defined as two or more chronic condition groups on the Electronic Health Record (EHR), at 66.2% in the sample.

Figure 5.2 shows the number of chronic condition groups on the EHR for each individual. This shows the sample contained a large number of residents with zero

FIGURE 5.2: Number of Chronic Condition Groups on the EHR Histogram



Chronic Condition Groups identified using list derived in Calderón-Larrañaga et al., 2017

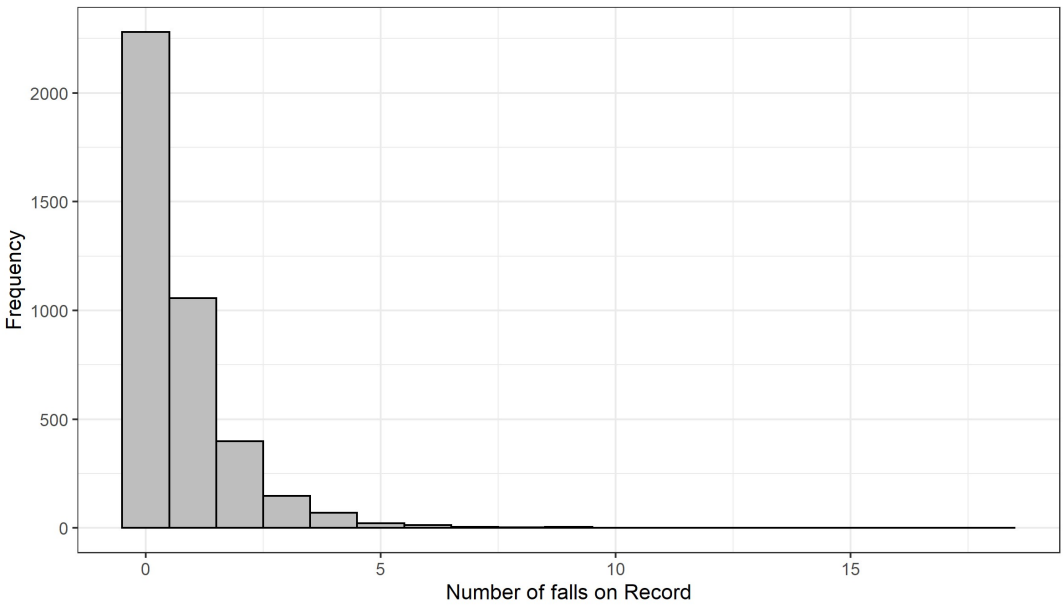
chronic condition groups on their EHR (n = 1335, 33.3%). Whether these residents had no chronic conditions or if this is a result of missingness is unclear.

TABLE 5.2: Analysis Dataset Frailty Index Scores

Index Name	**Mean (SD)**
CCI Score	5.46 (2.7)
Index Name	**N (%)**
EFI	
Fit	1331 (33.3%)
Mild Frailty	556 (13.9%)
Moderate Frailty	1603 (40.1%)
Severely Frail	512 (12.8%)
HFRS	
Low Risk	3127 (78.1%)
Intermediate Risk	773 (19.3%)
High Risk	102 (2.5%)

CCI: Charlson Comorbidity Index, EFI: Electronic Frailty Index, HFRS: Hospital Frailty Risk Score

FIGURE 5.3: Count of Fall Presentations to ED over full study period for each sample member histogram



Fall identified as Emergency Department attendance with investigation or treatment relating to trauma

Further assessment of chronic condition burden in the sample was conducted through calculating three electronic frailty index values based on information in the EHR. A summary of these frailty scores for the whole sample is provided in Table 5.2.

There were differences in the level of frailty identified in the sample depending on the frailty measures used. The Electronic Frailty Index (EFI) identified only 33.3% of the sample as *Fit* with 13.9% of the sample identified as *Mild Frailty* and a combined 52.9% as *Moderate* or *Severely* Frail. This is in contrast to the Hospital Frailty Risk Score (HFRS), which labelled 78.1% of the sample as having a *Low* Frailty risk, and only 21.8% as having an *Intermediate* or *High* Frailty Risk.

There are several sources that may have introduced differences in the level of frailty identified in the sample by the EFI and HFRS indexes. First, the EFI incorporates several daily living deficits into the calculation of the score, with no corresponding ICD-10 code, this meant several of these deficits were assumed to be present in a care home sample (activity limitations, housebound, mobility and transfer problems, requirement for care, and social vulnerability). By assuming these

deficits were present it is possible the level of frailty in the sample is being over-estimated in the EFI results. However, the HFRS was developed in a cohort of older adults over 75 admitted to hospital, wheras the EFI was derived from primary care records of people age over 65 (Clegg et al., 2016; Gilbert et al., 2018). This may have introduced a difference in the relative definitions of frail between these indicies. Therefore, the HFRS may have a comparatively higher threshold that an individual must pass in order to enter the higher categories of frailty risk than the EFI leading to the differences in the level of frailty observed in Table 5.2.

The mean amount of time spent in the cohort, seen in Table 5.1, measured as time between first contact and either a date of death or the end of the study period was 835.6 days (SD = 398.5) or 2.3 years. The number of fall presentations to the ED per sample member during the study period is the outcome of interest for the regression models presented in Sections 5.4 and 5.5. Descriptive statistics for the fall count outcome are presented in Table 5.3, and a histogram of the distribution is shown in Figure 5.3.

TABLE 5.3: Fall Presentations Descriptive Statistics

Statistic	**Value**
Mean (SD)	0.72 (1.14)
Median	0
Interquartile Range (25th, 75th percentiles)	1 (0, 1)
Minimum	0
Maximum	18
No fall presentations on record	2281 (57%)
Fall presentations on record ≥ 1	1721 (43.0%)

Table 5.3 shows that the prevalence of at least one fall within the sample was 43% (n = 1721) during the study period. As seen in Figure 5.3, the fall count outcome exhibits a positive skew. Testing the distribution of the falls outcome variable using the Kolmogorov-Smirnov test further confirmed the distribution was significantly ($P \leq 0.05$) different to a normal distribution, which is to be expected for count data. Despite a large number of individuals experiencing no fall presentations to the ED during the study period, the prevalence is sufficient for use as an outcome in the regression models, the results of which are presented in Sections 5.4 and 5.5.

Having explored the descriptive characteristics of the sample data, and ensured

the prevalence of falls and multi-morbidity are sufficient for examination through the chosen methods, the Multiple Correspondence Analysis (MCA) processing stage and cluster analysis were performed as described in Section 4.3.3, with the results provided in Section 5.3.

5.3 Cluster Analysis

5.3.1 Introduction

The cluster analysis set out to answer two questions:

- What chronic health conditions commonly co-occur together in care home residents?
- To what extent are these groups similar to clusters identified in community dwelling older adults?

The first of these questions is addressed in Section 5.3.4 with the necessary preliminary data manipulation and analyses presented in Sections 5.3.2 and 5.3.3. The second question is addressed in Section 6.3.1 during the wider contextualisation of the cluster results.

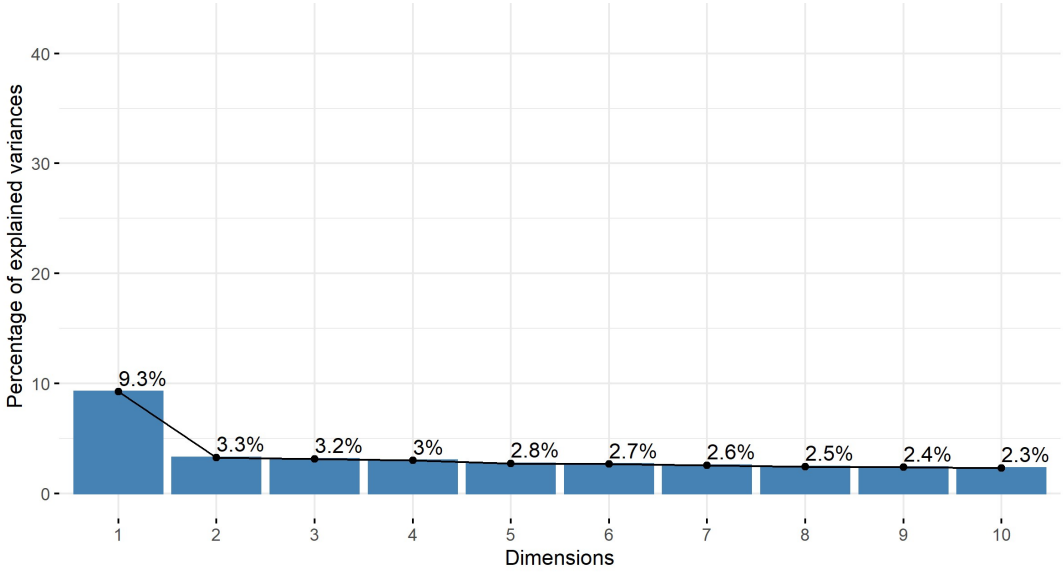
5.3.2 Multiple Correspondence Analysis (MCA)

Following the elimination of chronic conditions below the 1% prevalence threshold, described in Section 4.2.1, there were 48 Calderón-Larrañaga et al., 2017 chronic health condition groups considered for the analyses. The presence or absence of each of these chronic health condition groups was recorded based on sample member inpatient records. This resulted in a binary array of 48 variables, each relating to a chronic condition group from Calderón-Larrañaga et al. (2017), which was used as the input data for the MCA described in Section 4.3.2.

The MCA derived a lower-dimensional set of axes that retained the maximum possible variance in the 48-dimension data. However, Figure 5.4 demonstrates that this was a low proportion of the total variance. As detailed in Section 4.3.2 this low proportion of variance captured is unavoidable due to the binary nature of the input data. Because each subsequent dimension captured only a small amount of variance,

the first two axes were selected for the analyses. As shown in Figure 5.4 these two axes captured 9.3% and 3.3% of the variance respectively.

FIGURE 5.4: Variance captured in Multiple Correspondence Analysis Axes



Following the selection of the first two axes from the MCA output, the co-ordinates of every sample member were identified as the perpendicular projection of their point in the 48-dimension point cloud onto these axes. The co-ordinates of these points on the 2D MCA axes were then used as the input data for the four clustering algorithms described in Section 4.3.3. Results from the training, evaluation, and comparison of these four clustering algorithms is presented in Section 5.3.3 below.

5.3.3 Cluster Solution Comparison

Cluster analysis was used to identify groups of commonly co-occuring chronic health conditions, such that these groups could be assessed for how they impact falls risk during the regression analysis in Section 5.4. Additionally, interpreting these clusters in terms of the chronic conditions highly associated with them also gives a picture of the likelihood of different presentations of multi-morbidty in the sample of care home residents described in Section 5.2.

Four clustering algorithms were trained on the row-profile co-ordinate data from the first two MCA axes, as described in Section 4.3. These algorithms were K-means,

K-medoids, Ward's method, and average linkage to allow comparison between hierarchical and non-hierarchical approaches to clustering. Due to the lack of previous similar clustering research in care home residents, where the chronic condition burden is highest, and people live with many chronic health conditions at once, a data-driven approach was taken to identify the number of clusters determined as optimal by each algorithm. The results of this hyper-parameter optimisation is shown in Table 5.4 where algorithms were trained with differing numbers of clusters and assessed using the Calinski-Harabasz index value. This value is a ratio of the between-and within-cluster dispersion, where high values indicate greater separation between clusters (Caliñski and Harabasz, 1974). A limit of ten clusters was imposed to minimise the risk of over-training in the data and, as seen in Table 5.4, all algorithms saw diminishing returns in terms of Calinski-Harabasz index separation before this limit was reached.

TABLE 5.4: Calinski-Harabasz Index Values for each clustering algorithm

Number of Clusters	K-Means	K-Medoids	Ward's Method	Average Linkage
2	4161.9	4020.1	3429.8	76.5
3	4048.2	3852.6	2986.1	40.9
4	4343.7	4313.6	3192.6	416.1
5	4505.3	4344.8	3394.8	**1056.6**
6	4228.9	**4379.5**	**3462.9**	986.9
7	4510.1	4306.9	3094.9	852.0
8	**4562.1**	4372.0	3016.8	730.9
9	4406.9	4229.2	2860.2	646.2
10	4391.6	4114.8	2826.8	583.9

The number of clusters identified with the highest cluster separation for each algorithm differed, with K-means identifying 8 clusters, K-medoids and Ward's method identifying 6 clusters each, and the average linkage algorithm identifying 5 clusters.

The solution for each algorithm with the highest Calinski-Harabasz index of separation value was further assessed using the Mean Jaccard Coefficient, which evaluates the stability of the cluster solution. These results are shown in Table 5.5.

The mean Jaccard Coefficient value over 100 iterations indicated higher solution stability for the non-hierarchical algorithms (K-means, and K-medoids) when compared with the hierarchical algorithms (Ward's method, average linkage). Furthermore, while no solution was identified as highly stable (Mean Jaccard Coefficient $\geq$

TABLE 5.5: Calinski-Harabasz Index and Jaccard Coefficient of optimised solutions for each clustering algorithm

Clustering Algorithm	Clusters	Calinski-Harabasz Index	Mean Jaccard Coefficient
K-Means	8	4562.1	0.78
K-Medoids	6	4379.5	0.74
Ward's Method	6	3462.9	0.62
Average Linkage	5	1056.6	0.43

Jaccard Coefficient calculated over 100 iterations

0.8), the K-means solution and K-medoids solutions still provided reasonably stable solutions (Mean Jaccard Coefficient ≥ 0.7). However, the clusters should be regarded as having limited generalisability to the wider care home population as a result of this lowered stability.

Following the procedure detailed above, the algorithm selected for final appraisal and use in the fall count regression models in Section 5.4 was the K-means 8 cluster solution. This solution had the largest separation measured through Calinski-Harabasz index (4562.1) and the highest stability measured using the mean Jaccard Coefficient over 100 iterations (0.78). Section 5.3.4 investigates the K-means 8 cluster solution further, and describes each of these clusters in terms of the demographics and chronic health condition patterns identified.

5.3.4 Final Cluster Solution

The 8 cluster K-means solution derived in Section 5.3.3 is plotted on the first two MCA axes in Figure 5.5. The demographic characteristics and descriptive statistics of each cluster are presented in Table 5.6. The position of individual data points on the Figure 5.5 MCA axes represent a linear combination of their profile of chronic health conditions. Therefore, the different positioning of individual points on these axes reflect differences in the underlying chronic health condition profile. Each cluster has been given a name reflecting the major features of that cluster for clarity, which are described in greater detail throughout this section.

In order to understand the clusters in greater detail, Tables 5.7 and 5.8 present the prevalence, and observed expected ratios (O/E ratio) of each chronic health condition. In these ratios the observed prevalence for each chronic health condition in

FIGURE 5.5: K-Means 8 Cluster Solution

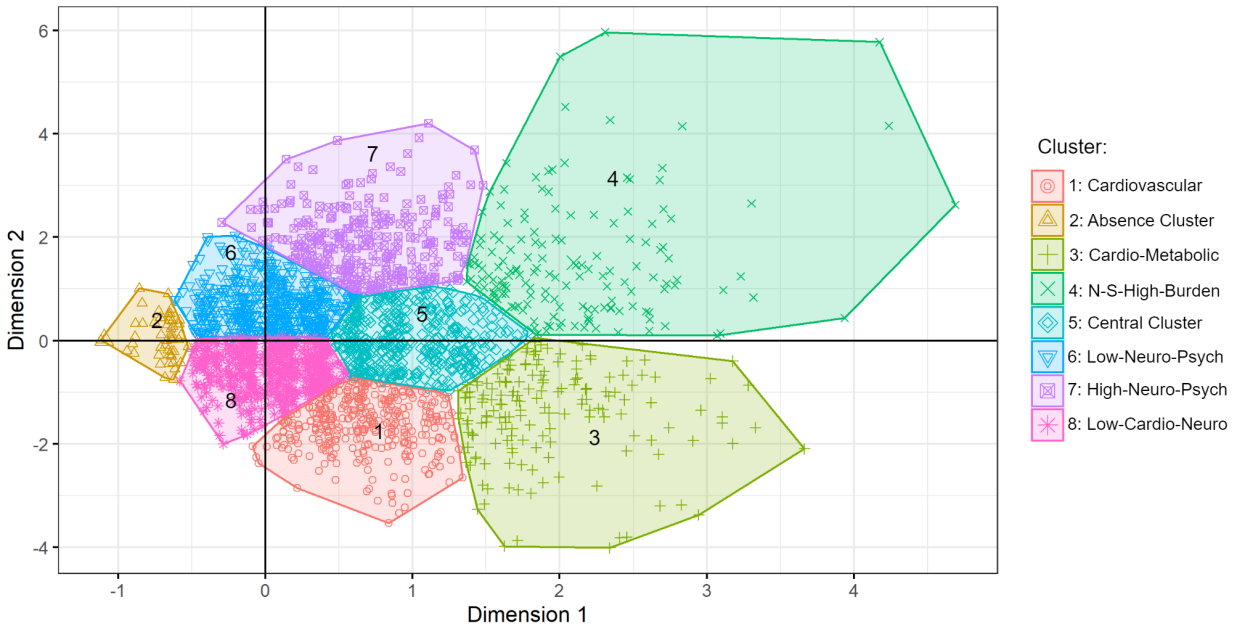


TABLE 5.6: Final Solution Cluster Characteristics

Cluster	Cluster Size	Mean Age (SD)	Mean N Chronic Conditions (SD)	Mean Fall Count (SD)	Mean CCI (SD)
1: Cardiovascular	381	85.5 (6.5)	7.1 (1.2)	0.6 (1.2)	7.5 (2)
2: Absence	1405	84.6 (7.9)	0.1 (0.4)	0.6 (0.9)	3.1 (1)
3: Cardio-Metabolic	168	84 (7)	12.1 (2)	0.9 (1.3)	9 (2)
4: N-S-High-Burden	158	81.6 (7.4)	13 (2.3)	1.3 (1.7)	8.5 (2.2)
5: Central	421	84.4 (7.1)	8.4 (1.3)	0.9 (1.2)	7.6 (2.4)
6: Low-Neuro-Psych	564	83.6 (8)	4.8 (1.1)	0.8 (1.3)	5.4 (1.9)
7: High-Neuro-Psych	289	81.8 (7.9)	7.5 (1.4)	0.8 (1.1)	5.9 (2.2)
8: Low-Cardio-Neuro	616	85.6 (6.6)	4.7 (1)	0.7 (1.1)	6.3 (2.2)
Whole Sample	4002	84.4 (7.6)	4.6 (4)	0.7 (1.1)	5.5 (2.7)

N-S-High-Burden: Non-Specific-High-Burden cluster, Cardio-Metabolic: Cardiovascular-Metabolic cluster, Low/High-Neuro-Psych: Low/High-Neurological-Psychiatric

the cluster is divided by the prevalence in the full sample (the expected prevalence), which highlights the conditions associated with that cluster. An O/E ratio ≥ 2 indicates a high degree of association of a chronic health condition with a particular cluster. An O/E ratio ≥ 2 means the prevalence in the cluster is over double that in the total sample. These tables are split by cluster to illustrate differences in the prevalence of each condition in each cluster.

Interpretation of MCA Axes

The process for identifying a semantic interpretation for the axes was as follows. First, conditions with consistent association across all clusters (diabetes, dementia,

and hypertension) were removed because of their universal association with each cluster besides the Absence cluster. Therefore, these conditions were not what is differentiating between the positions of different clusters.

FIGURE 5.6: K-Means 8 Cluster Solution Pairwise Comparisons

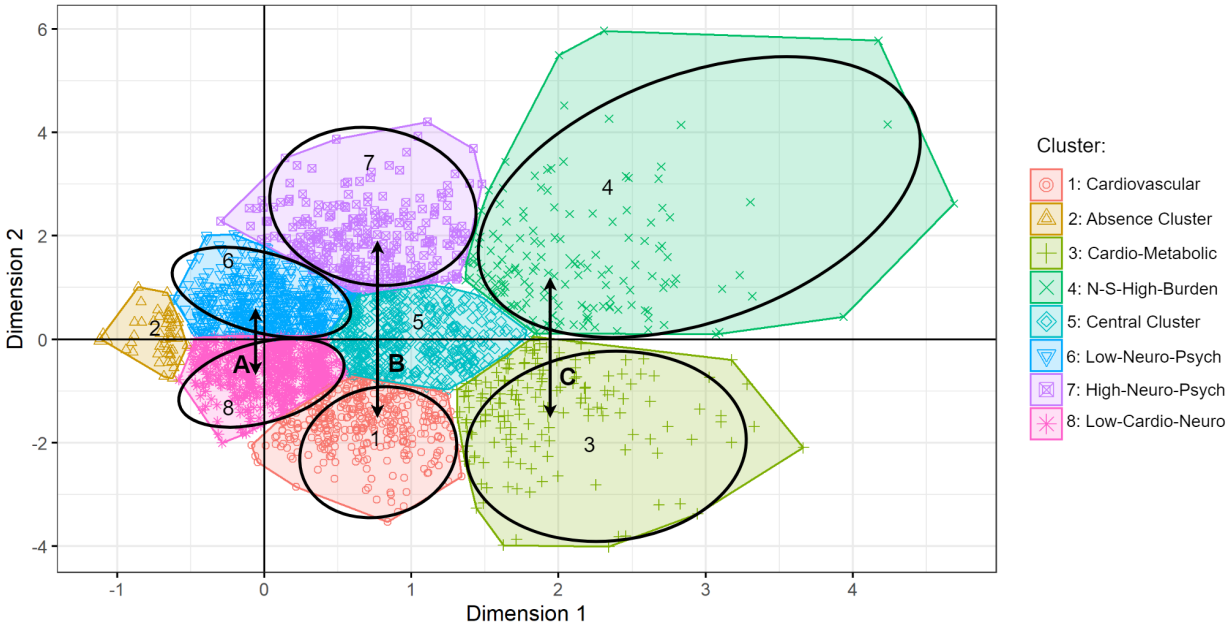
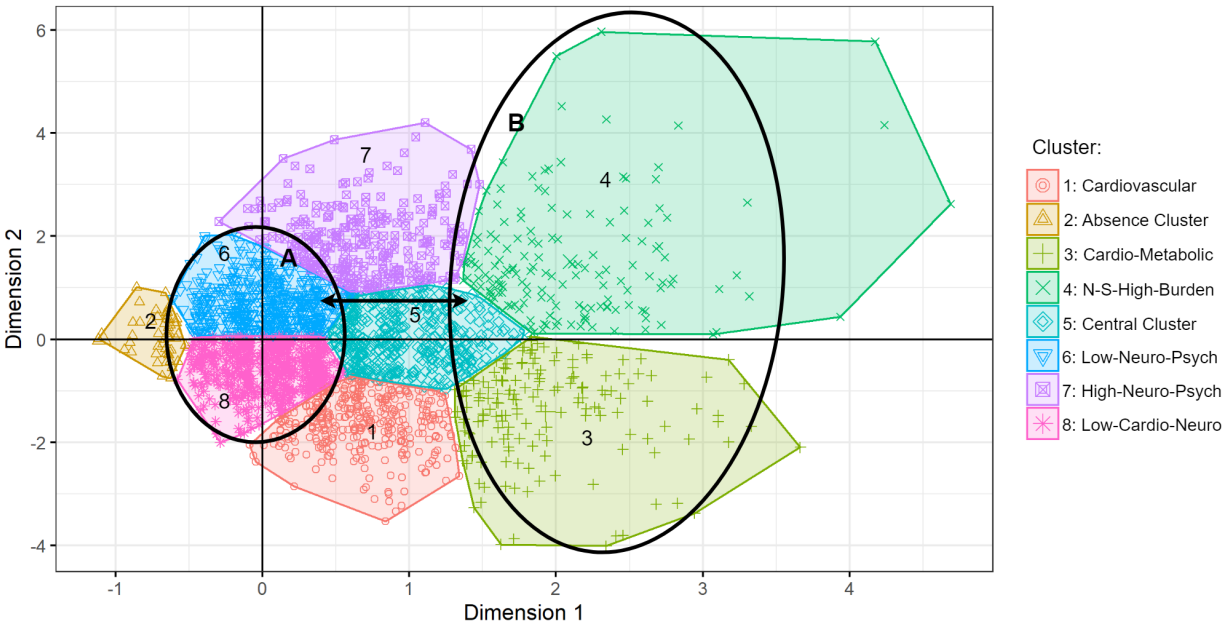


Figure 5.6 indicates the three comparisons made between clusters to identify an explanation of the Y-axis in the MCA output. Differences in O/E ratios for each condition of ≥ 1 between the three pairs of clusters in Figure 5.6 (Comparisons A, B, and C) were recorded as possible explanations of difference as a result of Y-axis positioning. Following this, the consistent themes across the three comparisons were identified for the explanation of the Y-axis. This comparative analysis of O/E ratios identified a difference between the top pane clusters (Low-Neuro-Psychiatric, High-Neuro-Psychiatric, and Non-Specific-High-Burden clusters), which showed stronger associations with neurological and psychiatric conditions, and the bottom pane clusters (Low-Cardio-Neuro, Cardiovascular, and Cardiovascular-Metabolic clusters), which were highly associated with cardiovascular conditions. Further support for this was shown in the Central cluster, which had a moderate degree of association with both sets of conditions whilst also holding a central position in Figure 5.6.

The X-axis in Figure 5.5 represents the primary axes derived during the MCA pre-processing, and this places the Absence cluster at the far left side, whilst the

Cardiovascular-Metabolic and Non-Specific-High-Burden clusters are placed at the far right hand side. Investigating the difference between the clusters positioned at the extremes on this axis in Table 5.6 shows that the Absence cluster represents a large cluster with very low prevalence of chronic condition groups, with a mean of 0.1 (SD = 0.4) chronic condition groups on the EHR. This is in direct contrast with the Cardiovascular-Metabolic and Non-Specific-High-Burden clusters where the mean number of chronic condition groups was 12.1 (SD = 2) and 13 (SD = 2.3) respectively. Therefore it seems reasonable to infer that the X-axis is splitting those sample members that are defined by the absence of most or any of the condition groups studied from those where those groups are present. Furthermore, the Low-Neuro-Psychiatric and Low-Cardio-Neuro clusters have fewer chronic health conditions than the High-Neuro-Psychiatric and Cardiovascular clusters (see Table 5.6). This gives further indication of a reduction in multi-morbidity burden moving from right to left along the X-axis in Figure 5.5.

FIGURE 5.7: K-Means 8 Cluster Solution X-axis Comparisons



To identify the specific condition groups being differentiated by the X-axis, information from Tables 5.7 and 5.8 was used alongside a similar approach for identifying the Y-axis conditions. This involved the comparison of condition prevalence between the Low-Cardio-Neuro and Low-Neuro-Psychiatric clusters with the

Cardiovascular-Metabolic and Non-Specific-High-Burden clusters in a single left-right comparison. On Figure 5.7 this involved looking for the similarities between the Low-Cardio-Neuro and Low-Neuro-Psychiatric clusters in Circle A. Then similarities were identified for the Cardiovascular-Metabolic and Non-Specific-High-Burden clusters in Circle B. The differences between these two groups of clusters were then compared in a single left-right comparison, as seen in Figure 5.7. This comparison identified differences between the clusters based on several conditions, and further confirmation was sought using the Central cluster to identify whether it held a mediating position between these differences.

This comparison identified the changing degree of association with several groups of conditions including musculoskeletal, gastrointestinal, metabolic, and sensory impairment conditions. Generally, the right hand panes have an increasing degree of association with these chronic health conditions, while the left panes do not. This increasing degree of association also reflects an increase in the raw number of co-incident chronic conditions on the EHR as we move between clusters along the X-axis, as shown in Table 5.6.

In summary, the X-axis appears to differentiate between the general number of chronic conditions on the EHR; however this difference is made up by the presence of metabolic, musculoskeletal, gastrointestinal, and sensory impairment chronic conditions on the right hand panes, versus the absence of these conditions on the left hand panes in Figure 5.7. Meanwhile, the Y-axis on Figure 5.6 appears to be differentiating between neurological and psychiatric conditions in the top panes, and cardiovascular conditions in the bottom panes.

It is important to note that these groupings are not absolute, with co-occurrence of conditions on opposite sides of the MCA plot still possible. However, the positioning on the plot means these combinations are a less likely occurrence on aggregate based on this sample.

In the clinical context, the clustering results suggest that multi-morbidity combinations may be differentiated by cardiovascular conditions, and neurological and psychiatric conditions which occur beyond the conditions seen at consistent rates throughout the sample (diabetes, dementia, and hypertension).

TABLE 5.7: K-Means Clustering Prevalence Table

		1: Cardiovascular		2: Absence		3: Cardio-Metabolic		4: N-S-High-Burden	
Chronic Condition Group	**Sample Prev.**	**Cluster Prev.**	**O/E**	**Cluster Prev.**	**O/E**	**Cluster Prev.**	**O/E**	**Cluster Prev.**	**O/E**
Allergy	615 (15.37)	97 (25.46)	1.66	14 (1)	0.06	59 (35.12)	2.29	42 (26.58)	1.73
Anaemia	495 (12.37)	63 (16.54)	1.34	2 (0.14)	0.01	66 (39.29)	3.18	66 (41.77)	3.38
Atrial Fibrillation	979 (24.46)	255 (66.93)	2.74	3 (0.21)	0.01	127 (75.6)	3.09	72 (45.57)	1.86
Autoimmune Diseases	65 (1.62)	7 (1.84)	1.13	0 (0)	0	5 (2.98)	1.83	7 (4.43)	2.73
Blindness and Visual Impairment	58 (1.45)	1 (0.26)	0.18	0 (0)	0	4 (2.38)	1.64	12 (7.59)	5.24
Blood and Blood Forming Organ Diseases	57 (1.42)	11 (2.89)	2.03	0 (0)	0	10 (5.95)	4.18	7 (4.43)	3.11
Bradycardias and Conduction Diseases	156 (3.9)	51 (13.39)	3.43	0 (0)	0	39 (23.21)	5.96	16 (10.13)	2.6
Cardiac Valve Diseases	355 (8.87)	137 (35.96)	4.05	0 (0)	0	84 (50)	5.64	27 (17.09)	1.93
Cataract and Other Lens Diseases	101 (2.52)	2 (0.52)	0.21	0 (0)	0	8 (4.76)	1.89	15 (9.49)	3.76
Cerebrovascular Disease	431 (10.77)	33 (8.66)	0.8	3 (0.21)	0.02	41 (24.4)	2.27	56 (35.44)	3.29
Chronic Kidney Diseases	897 (22.41)	208 (54.59)	2.44	7 (0.5)	0.02	110 (65.48)	2.92	75 (47.47)	2.12
Chronic Liver Diseases	89 (2.22)	6 (1.57)	0.71	0 (0)	0	13 (7.74)	3.48	14 (8.86)	3.98
Chronic Pancreas, Biliary Tract and GBD	158 (3.95)	16 (4.2)	1.06	1 (0.07)	0.02	28 (16.67)	4.22	28 (17.72)	4.49
Chronic Ulcer of the Skin	520 (12.99)	98 (25.72)	1.98	8 (0.57)	0.04	80 (47.62)	3.66	52 (32.91)	2.53
Colitis and Related Diseases	640 (15.99)	40 (10.5)	0.66	2 (0.14)	0.01	68 (40.48)	2.53	89 (56.33)	3.52
COPD, Emphysema, Chronic Bronchitis	594 (14.84)	94 (24.67)	1.66	7 (0.5)	0.03	65 (38.69)	2.61	59 (37.34)	2.52
Deafness and Hearing Impairment	145 (3.62)	7 (1.84)	0.51	0 (0)	0	18 (10.71)	2.96	19 (12.03)	3.32
Dementia	1617 (40.4)	196 (51.44)	1.27	35 (2.49)	0.06	102 (60.71)	1.5	106 (67.09)	1.66
Depression and Mood Diseases	305 (7.62)	9 (2.36)	0.31	1 (0.07)	0.01	9 (5.36)	0.7	63 (39.87)	5.23
Diabetes	840 (20.99)	125 (32.81)	1.56	2 (0.14)	0.01	86 (51.19)	2.44	80 (50.63)	2.41
Dorsopathies	116 (2.9)	3 (0.79)	0.27	1 (0.07)	0.02	2 (1.19)	0.41	35 (22.15)	7.64
Dyslipidemia	205 (5.12)	12 (3.15)	0.61	0 (0)	0	19 (11.31)	2.21	35 (22.15)	4.32
Ear, Nose, and Throat Diseases	59 (1.47)	2 (0.52)	0.36	0 (0)	0	7 (4.17)	2.83	17 (10.76)	7.3
Epilepsy	119 (2.97)	1 (0.26)	0.09	0 (0)	0	6 (3.57)	1.2	14 (8.86)	2.98
Oesophagus Stomach and Duodenum Diseases	195 (4.87)	6 (1.57)	0.32	1 (0.07)	0.01	12 (7.14)	1.47	55 (34.81)	7.14
Glaucoma	110 (2.75)	3 (0.79)	0.29	0 (0)	0	8 (4.76)	1.73	13 (8.23)	2.99
Heart Failure	879 (21.96)	309 (81.1)	3.69	0 (0)	0	156 (92.86)	4.23	78 (49.37)	2.25
Hypertension	1592 (39.78)	237 (62.2)	1.56	7 (0.5)	0.01	152 (90.48)	2.27	128 (81.01)	2.04
Inflammatory Arthropathies	255 (6.37)	37 (9.71)	1.52	0 (0)	0	43 (25.6)	4.02	42 (26.58)	4.17
Ischemic Heart Disease	876 (21.89)	190 (49.87)	2.28	4 (0.28)	0.01	109 (64.88)	2.96	80 (50.63)	2.31
Osteoarthritis and DJD	529 (13.22)	38 (9.97)	0.75	5 (0.36)	0.03	50 (29.76)	2.25	75 (47.47)	3.59
Osteoporosis	287 (7.17)	19 (4.99)	0.7	1 (0.07)	0.01	11 (6.55)	0.91	37 (23.42)	3.27
Other Cardiovascular Diseases	426 (10.64)	52 (13.65)	1.28	2 (0.14)	0.01	66 (39.29)	3.69	54 (34.18)	3.21
Other Digestive Diseases	47 (1.17)	4 (1.05)	0.89	0 (0)	0	4 (2.38)	2.03	8 (5.06)	4.31
Other Eye Diseases	203 (5.07)	6 (1.57)	0.31	3 (0.21)	0.04	16 (9.52)	1.88	37 (23.42)	4.62
Other Genitourinary Diseases	845 (21.11)	90 (23.62)	1.12	8 (0.57)	0.03	87 (51.79)	2.45	97 (61.39)	2.91
Other Metabolic Diseases	208 (5.2)	8 (2.1)	0.4	2 (0.14)	0.03	23 (13.69)	2.63	37 (23.42)	4.51
Other Musculoskeletal and Joint Diseases	326 (8.15)	26 (6.82)	0.84	1 (0.07)	0.01	41 (24.4)	3	70 (44.3)	5.44
Other Neurological Diseases	255 (6.37)	9 (2.36)	0.37	1 (0.07)	0.01	9 (5.36)	0.84	42 (26.58)	4.17
Other Psychiatric and Behavioural Diseases	310 (7.75)	12 (3.15)	0.41	0 (0)	0	29 (17.26)	2.23	44 (27.85)	3.6
Other Respiratory Diseases	372 (9.3)	82 (21.52)	2.32	2 (0.14)	0.02	64 (38.1)	4.1	31 (19.62)	2.11
Parkinsons and Parkinsonism	159 (3.97)	10 (2.62)	0.66	3 (0.21)	0.05	7 (4.17)	1.05	9 (5.7)	1.43
Peripheral Neuropathy	153 (3.82)	9 (2.36)	0.62	0 (0)	0	15 (8.93)	2.34	48 (30.38)	7.95
Prostate Diseases	121 (3.02)	21 (5.51)	1.82	0 (0)	0	17 (10.12)	3.35	6 (3.8)	1.26
Schizophrenia and Delusional Diseases	61 (1.52)	1 (0.26)	0.17	0 (0)	0	1 (0.6)	0.39	9 (5.7)	3.74
Sleep Disorders	40 (1)	8 (2.1)	2.1	0 (0)	0	13 (7.74)	7.74	8 (5.06)	5.07
Thyroid Diseases	338 (8.45)	32 (8.4)	0.99	0 (0)	0	35 (20.83)	2.47	36 (22.78)	2.7
Venous and Lymphatic Diseases	60 (1.5)	17 (4.46)	2.98	0 (0)	0	17 (10.12)	6.75	2 (1.27)	0.84

Prevalence reported as Frequency (%), O/E: Observed-Expected Ratio, DJD: Degenerative Joint Diseases, GBD: Gallbladder Diseases

TABLE 5.8: K-Means Clustering Prevalence Table Continued

		5: Central		6: Low-Neuro-Psych		7: High-Neuro-Psych		8: Low-Cardio-Neuro	
Chronic Condition Group	**Sample Prev.**	**Cluster Prev.**	**O/E**	**Cluster Prev.**	**O/E**	**Cluster Prev.**	**O/E**	**Cluster Prev.**	**O/E**
Allergy	615 (15.37)	106 (25.18)	1.64	89 (15.78)	1.03	41 (14.19)	0.92	167 (27.11)	1.76
Anaemia	495 (12.37)	97 (23.04)	1.86	61 (10.82)	0.87	58 (20.07)	1.62	82 (13.31)	1.08
Atrial Fibrillation	979 (24.46)	190 (45.13)	1.84	73 (12.94)	0.53	44 (15.22)	0.62	215 (34.9)	1.43
Autoimmune Diseases	65 (1.62)	12 (2.85)	1.75	18 (3.19)	1.96	10 (3.46)	2.13	6 (0.97)	0.6
Blindness and Visual Impairment	58 (1.45)	14 (3.33)	2.29	7 (1.24)	0.86	19 (6.57)	4.54	1 (0.16)	0.11
Blood and Blood Forming Organ Diseases	57 (1.42)	8 (1.9)	1.33	9 (1.6)	1.12	8 (2.77)	1.94	4 (0.65)	0.46
Bradycardias and Conduction Diseases	156 (3.9)	22 (5.23)	1.34	6 (1.06)	0.27	2 (0.69)	0.18	20 (3.25)	0.83
Cardiac Valve Diseases	355 (8.87)	56 (13.3)	1.5	6 (1.06)	0.12	3 (1.04)	0.12	42 (6.82)	0.77
Cataract and Other Lens Diseases	101 (2.52)	26 (6.18)	2.45	17 (3.01)	1.19	27 (9.34)	3.7	6 (0.97)	0.39
Cerebrovascular Disease	431 (10.77)	93 (22.09)	2.05	99 (17.55)	1.63	78 (26.99)	2.51	28 (4.55)	0.42
Chronic Kidney Diseases	897 (22.41)	173 (41.09)	1.83	78 (13.83)	0.62	42 (14.53)	0.65	204 (33.12)	1.48
Chronic Liver Diseases	89 (2.22)	19 (4.51)	2.03	14 (2.48)	1.12	21 (7.27)	3.27	2 (0.32)	0.15
Chronic Pancreas, Biliary Tract and GBD	158 (3.95)	26 (6.18)	1.56	24 (4.26)	1.08	16 (5.54)	1.4	19 (3.08)	0.78
Chronic Ulcer of the Skin	520 (12.99)	90 (21.38)	1.65	56 (9.93)	0.76	31 (10.73)	0.83	105 (17.05)	1.31
Colitis and Related Diseases	640 (15.99)	141 (33.49)	2.09	128 (22.7)	1.42	94 (32.53)	2.03	78 (12.66)	0.79
COPD, Emphysema, Chronic Bronchitis	594 (14.84)	115 (27.32)	1.84	86 (15.25)	1.03	57 (19.72)	1.33	111 (18.02)	1.21
Deafness and Hearing Impairment	145 (3.62)	35 (8.31)	2.29	23 (4.08)	1.13	31 (10.73)	2.96	12 (1.95)	0.54
Dementia	1617 (40.4)	302 (71.73)	1.78	335 (59.4)	1.47	173 (59.86)	1.48	368 (59.74)	1.48
Depression and Mood Diseases	305 (7.62)	47 (11.16)	1.46	58 (10.28)	1.35	105 (36.33)	4.77	13 (2.11)	0.28
Diabetes	840 (20.99)	175 (41.57)	1.98	142 (25.18)	1.2	84 (29.07)	1.38	146 (23.7)	1.13
Dorsopathies	116 (2.9)	16 (3.8)	1.31	19 (3.37)	1.16	40 (13.84)	4.78	0 (0)	0
Dyslipidemia	205 (5.12)	49 (11.64)	2.27	36 (6.38)	1.25	37 (12.8)	2.5	17 (2.76)	0.54
Ear, Nose, and Throat Diseases	59 (1.47)	8 (1.9)	1.29	7 (1.24)	0.84	14 (4.84)	3.29	4 (0.65)	0.44
Epilepsy	119 (2.97)	12 (2.85)	0.96	42 (7.45)	2.5	37 (12.8)	4.31	7 (1.14)	0.38
Oesophagus Stomach and Duodenum Diseases	195 (4.87)	40 (9.5)	1.95	30 (5.32)	1.09	46 (15.92)	3.27	5 (0.81)	0.17
Glaucoma	110 (2.75)	22 (5.23)	1.9	26 (4.61)	1.68	35 (12.11)	4.41	3 (0.49)	0.18
Heart Failure	879 (21.96)	165 (39.19)	1.78	9 (1.6)	0.07	12 (4.15)	0.19	150 (24.35)	1.11
Hypertension	1592 (39.78)	331 (78.62)	1.98	265 (46.99)	1.18	169 (58.48)	1.47	303 (49.19)	1.24
Inflammatory Arthropathies	255 (6.37)	54 (12.83)	2.01	25 (4.43)	0.7	22 (7.61)	1.19	32 (5.19)	0.82
Ischemic Heart Disease	876 (21.89)	174 (41.33)	1.89	74 (13.12)	0.6	47 (16.26)	0.74	198 (32.14)	1.47
Osteoarthritis and DJD	529 (13.22)	102 (24.23)	1.83	114 (20.21)	1.53	90 (31.14)	2.36	55 (8.93)	0.68
Osteoporosis	287 (7.17)	42 (9.98)	1.39	75 (13.3)	1.85	72 (24.91)	3.47	30 (4.87)	0.68
Other Cardiovascular Diseases	426 (10.64)	95 (22.57)	2.12	53 (9.4)	0.88	38 (13.15)	1.24	66 (10.71)	1.01
Other Digestive Diseases	47 (1.17)	11 (2.61)	2.22	7 (1.24)	1.06	10 (3.46)	2.95	3 (0.49)	0.41
Other Eye Diseases	203 (5.07)	46 (10.93)	2.15	33 (5.85)	1.15	45 (15.57)	3.07	17 (2.76)	0.54
Other Genitourinary Diseases	845 (21.11)	184 (43.71)	2.07	151 (26.77)	1.27	106 (36.68)	1.74	122 (19.81)	0.94
Other Metabolic Diseases	208 (5.2)	34 (8.08)	1.55	43 (7.62)	1.47	38 (13.15)	2.53	23 (3.73)	0.72
Other Musculoskeletal and Joint Diseases	326 (8.15)	63 (14.96)	1.84	43 (7.62)	0.94	49 (16.96)	2.08	33 (5.36)	0.66
Other Neurological Diseases	255 (6.37)	49 (11.64)	1.83	63 (11.17)	1.75	76 (26.3)	4.13	6 (0.97)	0.15
Other Psychiatric and Behavioural Diseases	310 (7.75)	74 (17.58)	2.27	77 (13.65)	1.76	64 (22.15)	2.86	10 (1.62)	0.21
Other Respiratory Diseases	372 (9.3)	70 (16.63)	1.79	34 (6.03)	0.65	23 (7.96)	0.86	66 (10.71)	1.15
Parkinsons and Parkinsonism	159 (3.97)	31 (7.36)	1.85	44 (7.8)	1.96	20 (6.92)	1.74	35 (5.68)	1.43
Peripheral Neuropathy	153 (3.82)	20 (4.75)	1.24	12 (2.13)	0.56	46 (15.92)	4.16	3 (0.49)	0.13
Prostate Diseases	121 (3.02)	26 (6.18)	2.04	19 (3.37)	1.11	12 (4.15)	1.37	20 (3.25)	1.07
Schizophrenia and Delusional Diseases	61 (1.52)	4 (0.95)	0.62	14 (2.48)	1.63	31 (10.73)	7.04	1 (0.16)	0.11
Sleep Disorders	40 (1)	7 (1.66)	1.66	1 (0.18)	0.18	0 (0)	0	3 (0.49)	0.49
Thyroid Diseases	338 (8.45)	68 (16.15)	1.91	79 (14.01)	1.66	51 (17.65)	2.09	37 (6.01)	0.71
Venous and Lymphatic Diseases	60 (1.5)	8 (1.9)	1.27	1 (0.18)	0.12	3 (1.04)	0.69	12 (1.95)	1.3

Prevalence reported as Frequency (%), O/E: Observed-Expected Ratio, DJD: Degenerative Joint Diseases, GBD: Gallbladder Diseases

Having suggested an interpretation for the MCA axes, the focus now turns to the individual clusters. Each cluster in Figure 5.5 represents a group of sample members with varying prevalence of each chronic health condition. To explore the clusters in greater detail, O/E ratios for each chronic condition were calculated and are presented in Tables 5.7 and 5.8. As mentioned previously, an O/E ratio ≥ 2 indicates conditions with a high degree of association with a particular cluster as the prevalence in the cluster is over double that in the overall sample.

Cardiovascular cluster

The Cardiovascular cluster is an intermediate sized cluster containing 381 (9.5%) sample members. Several chronic health conditions have over 50% prevalence in the cluster with atrial fibrillation (66.93%), chronic kidney diseases (54.59%), heart failure (81.1%), hypertension (62.2%), and dementia (51.44%). Additionally, cardiac valve diseases (35.96%), diabetes (32.81%), and ischemic heart disease (49.87%) all have prevalence over 30%.

The Cardiovascular cluster in Figure 5.5 is strongly associated with a group of cardiovascular conditions. This strong association with cardiovascular conditions is evidenced with atrial fibrillation (O/E = 2.74), bradycardias and conduction diseases (O/E = 3.43), cardiac valve diseases (O/E = 4.05), heart failure (O/E = 3.69) and ischemic heart disease (O/E = 2.28) all showing a higher than expected prevalence in the cluster. Furthermore, four additional chronic condition groups were associated with the cluster. These were blood and blood forming organ diseases (O/E = 2.03), chronic kidney diseases (O/E = 2.44), other respiratory diseases (O/E = 2.32), and venous and lymphatic diseases (O/E = 2.98). The multi-morbidity observed in the Cardiovascular cluster can therefore be seen as the co-occurence of multiple cardiovascular conditions, and several additional internal chronic conditions.

The Cardiovascular cluster has little association with neurological and psychiatric conditions. All included neurological and psychiatric conditions (besides dementia, which is consistent across all clusters besides the Absence cluster) having O/E ratios below 1. These O/E ratios demonstrate that while not entirely absent, compared to other clusters these conditions occur in the Cardiovascular cluster at a lower than expected rate. A similar situation is observed for the series of musculoskeletal and metabolic chronic health conditions.

Absence cluster

The Absence cluster is the largest cluster in the solution presented in Figure 5.5 containing 1405 (35.1%) sample members. The chronic condition profiles in this cluster contained little to no chronic condition presence with all O/E ratios in Table 5.7 at or close to zero. This large 'Absence' cluster is consistent with the large peak of observations with zero chronic health conditions identified in Figure 5.2. The lack of

chronic disease in this group may indicate that 35% of the sample did not suffer from the chronic health conditions included in the analysis. However, this seems unlikely for the majority of the absence cluster considering they are care home residents aged over 65.

There are several possible reasons for this Absence cluster to arise. The first of these is missing data in the electronic health record. As described in Section 4.2, cohort members without an inpatient record during the study period were excluded from the analyses. Therefore, if the chronic condition history of these residents is missing, this is due to issues in the standardised reporting of information in health records rather than a lack of inpatient stays during the study period. This issue is discussed further in Section 6.5.

It is also possible that a small proportion of the members of the Absence cluster do have chronic health conditions, although those conditions were excluded from the analyses due to having a prevalence of below 1% in the sample. However, this is likely to be a small proportion of the 35% of the sample contained in the Absence cluster.

A further possibility is that the Absence cluster is identifying a distinct group of care home residents that are fitter, with fewer chronic health conditions than members of other clusters. For example the absence cluster could represent self-funding care home residents who are less capable of domestic chores, or live with learning difficulties, or mental health diagnoses, which may motivate them to be placed in a care home despite being in a more physically robust state than other residents. For England as a whole, the self funded care home population represented 48.9% of older adult care home residents (Office for National Statistics, 2023). However, the North East of England (where the cohort of study is located) has a lower proportion of self funders at 26.4% in 2022 (Office for National Statistics, 2023). While this explanation may apply to part of the Absence cluster, it is unlikely to hold for the full cluster.

In summary, the Absence cluster likely represents a combination of these different groups of care home residents. There may be some self funded residents in a more robust state, alongside those with chronic health conditions excluded from the

analyses, and a further group of residents with missing information on their inpatient record.

Cardiovascular-Metabolic cluster

The Cardiovascular-Metabolic cluster is highly associated with 34 of the studied chronic health condition groups, as shown by the O/E ratios $\geq$ 2 in Table 5.7. The 168 (4.2%) sample members in this cluster have the second highest chronic condition burden of all clusters with an average of 12.1 chronic condition groups on the EHR. An initial consideration of the condition prevalence in the cluster shows nine conditions have over 50% prevalence. These were atrial fibrillation (75.6%), cardiac valve disease (50%), chronic kidney diseases (65.48%), dementia (60.71%), diabetes (51.19%), heart failure (92.86%), hypertension (90.48%), ischemic heart disease (64.88%), and other genitourinary diseases (51.79%). Heart failure and hypertension were largely ubiquitous in the cluster, meaning that the Cardiovascular-Metabolic cluster typically identifies sample members with these two conditions, alongside a high burden of additional chronic health conditions.

The Cardiovascular-Metabolic cluster is associated not only with cardiovascular conditions, but also multiple musculoskeletal, metabolic, sensory impairments, and gastrointestinal chronic health conditions, as seen in Table 5.7. Furthermore, the few conditions not associated with the Cardiovascular-Metabolic cluster are those expected following the MCA axes description, with both psychiatric condition groups having O/E ratios $\leq$ 1 (schizophrenia and delusional diseases O/E = 0.39, depression and mood diseases O/E = 0.70). Of the neurological conditions included, only cerebrovascular disease showed a strong association with the cluster (O/E = 2.27), with epilepsy (O/E = 1.2) and Parkinson's disease (O/E = 1.05) both occurring at the expected rate in the sample.

Non-Specific-High-Burden cluster

The Non-Specific-High-Burden cluster represents the members with the highest level of multi-morbidity with a mean number of chronic condition groups on the EHR of 13 (SD = 2.3). This translates into 41 of the 47 condition groups included for study having a higher than expected prevalence in the cluster. With only 158 (3.9%) sample members in the cluster, it is the smallest cluster in the solution.

Unlike the Cardiovascular-Metabolic cluster, there are no conditions that are effectively ubiquitous in the cluster, suggesting a breadth of multi-morbidity combinations are present in the Non-Specific-High-Burden cluster. Conditions with over 50% prevalence in the Non-Specific-High-Burden cluster are colitis and related diseases (56.33%), dementia (67.09%), diabetes (50.63%), hypertension (80.01%), ischemic heart disease (50.63%), and other genitourinary diseases (61.39%).

Central cluster

Figure 5.5 shows the Central cluster is placed centrally among the different multi-morbidity clusters. The Central cluster has moderate associations with many conditions including the group of cardiovascular, sensory impairment, and gastrointestinal chronic conditions, with cerebrovascular disease (O/E = 2.05) the only neurological condition.

Low-Neuro-Psychiatric Cluster

The Low-Neuro-Psychiatric cluster contains 564 (14.1%) sample members with a comparatively low prevalence of conditions compared to the other clusters but some association with neurological, and psychiatric condition groups. As seen in Table 5.8 the included psychiatric (schizophrenia and delusional diseases O/E = 1.63, depression and mood diseases O/E = 1.35) and neurological conditions (cerebrovascular disease O/E = 1.63, epilepsy O/E = 2.5, Parkinson's disease O/E = 1.96, and dementia O/E = 1.47) exhibit some association with the cluster, although this is not as strong as other clusters.

Only dementia had a prevalence of over 50% in the Low-Neuro-Psychiatric cluster. The reduced chronic condition burden in this cluster is shown in Table 5.6 where sample members in the Low-Neuro-Psychiatric cluster had a mean number of chronic conditions of 4.8 (SD = 1.1) on the EHR.

High-Neuro-Psychiatric Cluster

The High-Neuro-Psychiatric cluster contains 289 (7.2%) sample members, with a mean number of chronic condition groups on the EHR of 7.5 (SD = 1.4). The High-Neuro-Psychiatric cluster shows a strong association with neurological and

psychiatric conditions in Table 5.8, with very little association to cardiovascular conditions. Psychiatric condition variables showed strong association with the High-Neuro-Psychiatric cluster (schizophrenia and delusional diseases O/E = 7.04, depression and mood diseases O/E = 4.77, other psychiatric and behavioural diseases O/E = 2.86). Several neurological conditions were also strongly associated with the High-Neuro-Psychiatric cluster, with cerebrovascular disease (O/E = 2.51), epilepsy (O/E = 4.31), and other neurological diseases (O/E = 4.13) all showing strong relationships. However, several neurological conditions showed only moderate association with the High-Neuro-Psychiatric cluster, specifically dementia (O/E = 1.48), and Parkinson's disease (O/E = 1.74).

Sensory impairment conditions were also associated with the High-Neuro-Psychiatric cluster, with blindness and visual impairment (O/E = 4.54), cataract and other lens diseases (O/E = 3.7), deafness and hearing impairment (O/E = 2.96), ear nose and throat diseases (O/E = 3.29), glaucoma (O/E = 4.41), and other eye diseases (O/E = 3.07) all presenting strong associations with the cluster. Another groups of conditions associated with the High-Neuro-Psychiatric cluster is musculoskeletal conditions, with dorsopathies (O/E = 4.78), osteoarthritis (O/E = 2.36), osteoporosis (O/E = 3.47), and other musculoskeletal and joint diseases (O/E = 2.08) all showing strong association with the cluster. Therefore, the High-Neuro-Psychiatric cluster can be seen as a cluster representing the confluence of neurological chronic conditions, with sensory impairments, and musculoskeletal chronic conditions.

Low-Cardio-Neuro Cluster

The Low-Cardio-Neuro cluster is the largest non-absence cluster in the solution containing 616 (15.4%) sample members. The Low-Cardio-Neuro cluster exhibits a moderate association with cardiovascular conditions, and the consistent chronic conditions across all clusters, but with comparatively lower condition burden than other clusters. The mean number of condition groups on the EHR of the Low-Cardio-Neuro cluster members is 4.7 (SD = 1) in Table 5.8. The association with cardiovascular conditions in the Low-Cardio-Neuro cluster is weak with only atrial fibrillation (O/E = 1.46), hypertension (O/E = 1.24), heart failure (O/E = 1.11), and ischemic heart disease (O/E = 1.47) having O/E ratios ≥ 1. No O/E ratios calculated

for the Low-Cardio-Neuro cluster pass the ≥ 2 threshold indicating no strong associations, with neurological conditions besides dementia (O/E = 1.48) and Parkinson's disease (O/E = 1.43) exhibiting O/E ratios ≤ 1.

Following the identification of the 8 groups of differing chronic condition prevalence and association, the clusters were incorporated into the regression models to identify whether differences in the makeup of the groups of chronic condition profiles are associated with differences in fall rates in care home residents. The results of this further investigation, are presented in Section 5.4 and 5.5 below.

5.4 Main Effects Regression Analysis

5.4.1 Introduction

The results presented in this section follow from the count data regression methodology covered in Section 4.4. The objective of this regression analyses was to investigate associations between chronic disease, and multi-morbidity with fall presentations to the emergency department by sample members. This overarching aim was separated into four separate objectives in Section 4.4, which are re-iterated below.

1. Explore the role of frailty, and the effectiveness of three frailty index scores (CCI, EFI, HFRS) for accounting for variation in falls in the sample of care home residents

2. Examine whether different groups of chronic health conditions are associated with fall presentations beyond the number of conditions present.

3. Identify the association between specific chronic health conditions and the count of fall presentations to ED during the study period.

4. Investigate the effect of interactions between these chronic health conditions and what they can tell us about their relationship with the number of falls.

Regression models were developed to meet each of these objectives as follows. The first objective is met using three regression models, which used different frailty index scores (Charlson Comorbidity Index, Electronic Frailty Index, and Hospital Frailty Risk Score) as explanatory variables alongside age and sex. Results of these three models are presented in Section 5.4.2 (Table 5.9).

The second aim was answered using the K-means cluster membership, described in Section 5.3.4, as an explanatory variable in a regression model. Cluster membership was used as a proxy for the multi-morbidity combinations because specifying all the combinations would be infeasible. The results of this regression model are presented in Section 5.4.3 (Table 5.10).

The third aim of the regression analysis was investigate associations between specific chronic health conditions and falls. This regression uses a shortlist of the Calderón-Larrañaga et al., 2017 chronic health condition groups, which follows from

the findings of the review of reviews in Chapter 3. The results of this regression model are shown in Section 5.4.4 (Table 5.11).

The fourth aim described above is addressed in the interaction analyses results presented in Section 5.5. All of the regression models described above use the count of Fall presentations to ED during the study period described in Section 5.2, as the outcome for regression. Following preliminary analysis of the fall count outcome variable, overdispersion was identified, and Negative binomial regression models were used in the place of Poisson regression models. The results of this preliminary analysis are summarised in Appendix Section K.

5.4.2 Frailty Regression Analysis

Three negative binomial regression models were developed to investigate the role of frailty in determining fall presentations to the emergency department. The indices used in these models were the Charlson Comorbidity Index (CCI), Electronic Frailty Index (EFI), and Hospital Frailty Risk Score (HFRS) with the results shown in Table 5.9.

In relation to the effect of frailty on fall presentations to the ED, the models suggest a relationship is present; however, there is disagreement present between the scales regarding how increasing frailty impacts the rate of fall presentations to ED.

Based on the CCI (Table 5.9), a one point increase in the CCI results in a 5.4% (3.6%-7.3%) increase in fall presentations during the study period. As seen in Appendix Table G, the CCI point increases range from 1-6 dependent on which health condition is present. Additionally, because the CCI was not developed to reflect falls risk, the conditions associated with the largest point increases on the index are not commonly identified as fall risk increasing conditions. For example AIDs or metastatic solid tumours (6 points on CCI) would be associated with a 32.4% increase in falls presentations based on the model in Table 5.9, whereas dementia or cerebrovascular disease would only be associated with 5.4% increases (1 point on CCI).

The EFI was calculated and split into categories of frailty, based on the author's definitions as described in Section 4.2.1. The EFI category *'Fit'* was used as the reference category, with the other categories effects calculated compared to this category.

TABLE 5.9: Frailty Indices Regression Results

Explanatory Variables	P-Value	IRR	2.5% IRR CL	97.5% IRR CL
(i) CCI Model				
Intercept	0.000	0.000	0.000	0.001
Age*	0.006	1.010	1.003	1.017
Male Sex	0.055	0.898	0.813	0.992
CCI Score*	0.000	1.054	1.036	1.073
(ii) EFI Model				
Intercept	0.000	0.000	0.000	0.000
Age*	0.000	1.015	1.008	1.021
Male Sex*	0.028	0.887	0.804	0.978
EFI Category:				
Fit	Reference	-	-	-
*Mild Frailty**	0.020	1.227	1.047	1.436
*Moderate Frailty**	0.000	1.411	1.263	1.576
*Severely Frail**	0.000	2.185	1.902	2.509
(iii) HFRS Model				
Intercept	0.000	0.000	0.000	0.000
Age*	0.000	1.015	1.008	1.022
Male Sex	0.164	0.925	0.838	1.022
HFRS Category				
Low Risk	Reference	-	-	-
*Intermediate Risk**	0.000	1.482	1.327	1.655
High Risk	0.211	1.211	0.922	1.584

* = $P \leq 0.05$, *CCI: Charlson Comorbidity Index, EF: Electronic Frailty Index, HFRS: Hospital Frailty Risk Score, IRR: Incident Rate Ratio, CL: Confidence Limit*

Significant ($P \leq 0.05$) effects were identified for all frailty categories in comparison to the *'Fit'* category. These categories indicate that an increasing degree of frailty is associated with increasing fall presentations to the ED. Specifically, compared to those in the *Fit* category, *Mild Frailty* increased fall presentations by 22.7% (4.7%-43.6%), wheras *Moderate Frailty* resulted in a 41.1% (26.3%-57.6%) increase. The largest effect was observed for *severely Frail* sample members with an increase in fall presentations of 118.5% (90.2%-150.9%) when compared to those in the *Fit* category. This indicates a clear relationship between both frailty and fall presentations, but also increasing frailty leads to an increasing rate of fall presentations to the ED.

In contrast to the EFI, sample members in the *High Risk* of frailty category of the HFRS did not have a significantly different ($P > 0.05$) count of fall presentations to ED during the study period compared to those with *Low Risk* of frailty, as shown in Table 5.9. However, compared to those in the *Low Risk* category of the HFRS, sample members in the *Intermediate Risk* category experienced 48.2% (32.7%-65.5%) more fall

presentations to ED during the study period.

Based on the results presented above, it is clear the presence of frailty when compared to non-frail individuals increases falls risk in care home residents. However, the HFRS and EFI results contrast over how increasing frailty impacts falls risk. The difference between these indices may possibly arise from the differences in information used to calculate the scores, with the EFI possibly better able to differentiate between the levels of frailty in an information constrained setting. It is possible that the U-shaped relationship indicated in the HFRS model resulted from people in the *High Risk* category having potentially (very) limited activity and therefore fewer opportunities to fall. Furthermore, people in the *Intermediate risk* category may have been more active with a corresponding increase in opportunities to fall. Such a U-shaped relationship has been suggested previously in community dwelling older adults (Bath and Morgan, 1999).

In conclusion, the regression analyses identified that frailty plays a role in determining falls risk. However, the shape this relationship takes is dependent on the index used to differentiate frailty status. Additionally, while there is clear overlap between frailty and multi-morbidity, the contribution of multi-morbidity in determining falls risk may go beyond those effect sizes seen for frailty alone.

5.4.3 Multi-Morbidity Regression Analysis

Following the K-means 8 cluster solution presented in Section 5.3, a negative binomial regression model was developed using cluster membership as an explanatory variable alongside age and sex. As seen in Table 5.10, the Absence cluster was used as the reference category because this appeared to best represent the 'condition absence' cluster in the analysis. The initial aim of this regression model, to identify whether the clusters of chronic health conditions are associated with different levels of falls, is answered by the significant effects of all other clusters in relation to the Absence cluster.

The different Incident Rate Ratio (IRR) values for each cluster in Table 5.10 show that the clusters of chronic health conditions are associated with different rates of fall presentations to the ED by sample members. Specifically, each cluster had a significantly different fall rate when compared to the Absence cluster. Furthermore,

the different combinations of multi-morbidity identified for each cluster, discussed previously in Section 5.3.4, lead to different effects on fall presentations to ED in care home residents.

The results also give some indication of a gradient in effect size based on the types of multi-morbidity present. These different effects can be identified by using the regression results in Table 5.10 in conjunction with the condition prevalence results for each cluster in Tables 5.7 and 5.8. As discussed in Section 5.3, each cluster is identifying mostly distinct groups of multi-morbidity in the sample population, with a different mix of chronic health conditions in each.

TABLE 5.10: K-Means Cluster Negative Binomial Regression Results

Variable	**P-Value**	**IRR**	**2.5% IRR CL**	**97.5% IRR CL**
Intercept	0.000	0.000	0.000	0.000
Age*	0.000	1.016	1.010	1.023
Male Sex	0.073	0.906	0.820	1.000
K-Means Cluster Membership				
2: Absence	Reference	-	-	-
*1: Cardiovascular**	0.033	1.238	1.034	1.479
*3: Cardiovascular-Metabolic**	0.000	1.630	1.301	2.037
*4: Non-Specific-High-Burden**	0.000	2.178	1.761	2.692
*5: Central**	0.000	1.653	1.412	1.934
*6: Low-Neuro-Psychiatric**	0.000	1.430	1.237	1.653
*7: High-Neuro-Psychiatric**	0.001	1.414	1.173	1.702
*8: Low-Cardio-Neuro**	0.004	1.267	1.094	1.465

* = $P \leq 0.05$, *IRR: Incident Rate Ratio, CL: Confidence Limit*

The gradient of effect size provides an opportunity to explore which combinations of chronic health conditions could be driving the increases in falls risk seen in Table 5.10. Based on the IRR values in the table, there are four tiers of effect size. This description starts with the clusters exhibiting the smallest effect sizes and increases to the largest effect size.

The Cardiovascular and Low-Cardio-Neuro clusters have the smallest effect sizes of the clusters in Table 5.10. Cardiovascular cluster membership was associated with a 23.8% (3.4%-47.9%) increase in fall presentations, while the Low-Cardio-Neuro cluster membership led to a 26.7% (9.4%-46.5%) increase. Both effects were found to be significant ($P \leq 0.05$). Observing differences between these clusters and those exhibiting larger effect sizes may give useful insights into the role of multi-morbidity in

determining falls risk. Recalling the cluster descriptions from Section 5.3.4, the Cardiovascular cluster has strong association with range of cardiovascular condition groups, alongside an elevated prevalence of diabetes, and dementia in the cluster. In contrast the Low-Cardio-Neuro cluster is not highly associated with any chronic condition groups in Table 5.8 meaning in isolation few conclusions can be drawn. However, as a comparator to other cluster effect sizes, the Low-Cardio-Neuro cluster may allow useful conclusions to be drawn.

The Low-Neuro-Psychiatric and High-Neuro-Psychiatric clusters exhibited the next level of effect size in Table 5.10, with 43.0% (23.7%-65.3%) and 41.4% (17.3%-70.2%) increases in fall presentations when compared to the Absence cluster. When attempting to explain this increase in effect size several comparisons can be made. First, what are the differences between the Low-Neuro-Psychiatric and High-Neuro-Psychiatric clusters with the Cardiovascular and Low-Cardio-Neuro clusters, how do the clusters compare when condition burden is taken into account (Low-Neuro-Psychiatric cluster vs Low-Cardio-Neuro cluster, and the High-Neuro-Psychiatric cluster vs Cardiovascular cluster), and then how do the Low-Neuro-Psychiatric and High-Neuro-Psychiatric clusters differ?

Examining Tables 5.7 and 5.8, the overarching difference between the Low-Neuro-Psychiatric and High-Neuro-Psychiatric clusters, and the Cardiovascular and Low-Cardio-Neuro clusters is that the former have a much reduced association with cardiovascular conditions, and an increased association with the included psychiatric and neurological conditions. This difference suggests that the confluence of psychiatric and neurological conditions may have a stronger risk increasing effect than the combination of multiple cardiovascular conditions.

However, with the information available we can also make a sub-comparison to identify differences in how the type of multi-morbidity relates to falls risk when the number of chronic conditions is similar between clusters. This involves comparing the Low-Neuro-Psychiatric cluster (mean number of conditions = 4.8) with the Low-Cardio-Neuro cluster (mean number of conditions = 4.7). Additionally a further comparison can be made between the High-Neuro-Psychiatric cluster (mean number of conditions = 7.5) with the Cardiovascular cluster (mean number of conditions = 7.1) to further examine differences in how the type of multi-morbidity may

influence falls risk beyond the number of chronic health conditions present.

When making these comparisons, the Low-Neuro-Psychiatric cluster sees an increase in association with psychiatric and neurological conditions, and decreased association with cardiovascular conditions when compared to the Low-Cardio-Neuro cluster. Examining the effect sizes in Table 5.10 shows the Low-Neuro-Psychiatric cluster (43%) has a larger effect on fall presentations when compared to the Low-Cardio-Neuro cluster (26.7%). This difference in effect sizes when comparing neurological-psychiatric multi-morbidity with cardiovascular multi-morbidity becomes more pronounced when examining the differences between the effect sizes of the High-Neuro-Psychiatric (41.4%) and Cardiovascular (23.8%) clusters.

The increases in effect size observed from the Cardiovascular and Low-Cardio-Neuro clusters to the Low-Neuro-Psychiatric and High-Neuro-Psychiatric clusters cannot be explained through increased condition burden alone. These differences seem to indicate that neurological-psychiatric multi-morbidity exhibits a larger impact on falls risk than isolated cardiovascular multi-morbidity.

Examining the differences between the Low-Neuro-Psychiatric and High-Neuro-Psychiatric clusters, which exhibit a similar effect size, may also provide an opportunity to discount chronic conditions that may not be contributing to increased falls risk. When making this comparison in conjunction with Tables 5.7 and 5.8 a group of sensory and visual impairments is strongly associated with the High-Neuro-Psychiatric cluster and not the Low-Neuro-Psychiatric cluster. The similar effect sizes observed for the Low-Neuro-Psychiatric and High-Neuro-Psychiatric clusters suggest the differences between the clusters are not contributing to a change in falls risk. Therefore, sensory and visual impairments may not be associated with a change in falls risk in this sample. This conclusion will be discussed further when the role of individual conditions is assessed in Section 5.4.4.

The Cardiovascular-Metabolic and Central clusters resulted in 63.0% (30.1%-103.7%) and 65.3% (41.2%-93.4%) increases in fall presentations when compared to the Absence cluster in Table 5.10. The Central cluster presents medium to strong associations with almost all of the chronic conditions included in the analysis: this means that deriving conclusions from this cluster is challenging. However, the cluster exhibits the highest prevalence of dementia (71%) of all clusters, and has the

3rd highest condition burden (mean number of conditions = 8.4, SD = 1.3). In contrast, the Cardiovascular-Metabolic cluster represents the most multi-morbid sample members with cardiovascular conditions. The Cardiovascular-Metabolic cluster has a strong association with all the included cardiovascular conditions, and the second highest condition burden across the clusters (mean number of conditions = 12.1, SD = 2). The increase in effect size seen when comparing the Cardiovascular-Metabolic cluster to the Low-Neuro-Psychiatric and High-Neuro-Psychiatric clusters suggests that concurrent heart failure, hypotension, and atrial fibrillation alongside other chronic conditions have a larger impact on falls rate than the psychiatric and neurological condition combinations seen in the Low-Neuro-Psychiatric and High-Neuro-Psychiatric clusters.

The largest effect size identified in Table 5.10 was identified for sample members in the Non-Specific-High-Burden cluster with a 117.8% (76.1%-169.2%) increase in fall presentations during the study period. This is a small cluster of the members with the highest level of multi-morbidity, with high prevalence in many of the conditions included in the study. However, the level of condition burden seen in the Non-Specific-High-Burden cluster (mean number of conditions = 13, SD = 2.3) is comparable to that seen in the Cardiovascular-Metabolic cluster (mean number of conditions = 12.1, SD = 2) suggesting that the large increase in effect size likely goes beyond the sheer level of multi-morbidity. Investigating Table 5.7, we can see that the Non-Specific-High-Burden cluster has stronger associations with the included psychiatric and neurological conditions than the Cardiovascular-Metabolic cluster. However, the Non-Specific-High-Burden cluster also exhibits strong associations with the cardiovascular conditions that characterise the Cardiovascular-Metabolic cluster. This suggests that the multi-morbidity present in the Non-Specific-High-Burden cluster is the combination of multiple neurological, psychiatric, and cardiovascular conditions which appears to be the most potent combination in leading to increased falls risk in the sample.

The gradient of these effects, differences in chronic disease burden, and specific make up of the clusters leads to several issues for discussion, which are addressed further in Section 6.3.2.

The results in this section suggest there is a gradient of effects observed for different kinds of multi-morbidity on falls risk in care home dwelling older adults. Specifically, neurological-psychiatric multi-morbidity exhibits stronger associations with increased fall presentations than isolated cardiovascular multi-morbidity. However, the cardiovascular-metabolic pattern exhibited a larger effect than neurological-psychiatric multi-morbidity. The largest effect was observed for the Non-Specific-High-Burden cluster suggesting a combination of the cardiovascular-metabolic and neurological-psychiatric patterns may represent the most potent combination for increased falls in care home residents. These themes are discussed further in Section 6.3.

5.4.4 Chronic Health Conditions Regression Analysis

A key aim of the regression analyses was to identify whether the fall pre-disposing chronic health conditions identified in previous literature during the Review of Reviews in Chapter 3 had a consistent impact on fall presentations by care home residents in the sample. The results of the chronic health condition shortlist model are presented in Table 5.11.

The results in Table 5.11 indicate that, while significant risk-increasing effects were identified in meta-analyses during the Chapter 3 review, these effects may not translate to the care home setting directly, or when the outcome is fall presentations to the ED. The summary of results below focuses on the groups of variables included in the chronic condition effects model results reported in Table 5.11. These results are interpreted as the conditions that lead to a significant change in the number of fall presentations during the study period compared to the baseline rate of fall presentations in the sample.

The effect sizes seen in Table 5.11 are small, meaning a statistically significant result may be of limited clinical significance. Additionally, the small effect sizes seen in Table 5.11, when compared to those resulting from the cluster regression model in Table 5.10, indicate a possible effect of multi-morbidity that is not being captured by the individual condition effects.

The results in Table 5.11 are summarised below by discussing conditions impacting similar physiological systems together. Eight informal categories are used below

TABLE 5.11: Fall Pre-disposing Chronic Health Condition Shortlist
Negative Binomial Regression Results

Variable	**P-Value**	**IRR**	**2.5% IRR CL**	**97.5% IRR CL**
Intercept	0.000	0.000	0.000	0.000
Age*	0.000	1.013	1.007	1.020
Male Sex*	0.006	0.856	0.773	0.948
Neurological Conditions:				
*Dementia**	0.000	1.323	1.193	1.467
*Cerebrovascular Disease**	0.032	1.186	1.026	1.371
Parkinson's and Parkinsonism	0.164	1.183	0.950	1.468
Cardiovascular Conditions:				
*Hypotension**	0.000	1.611	1.380	1.879
*Atrial Fibrillation**	0.002	1.218	1.084	1.368
Bradycardia and Conduction Diseases	0.665	0.946	0.748	1.191
Cardiac Valve Diseases	0.652	0.960	0.815	1.129
*Ischemic Heart Disease**	0.004	1.202	1.070	1.350
Heart Failure	0.110	0.891	0.782	1.014
Hypertension	0.876	0.990	0.886	1.107
Sensory Impairment Conditions:				
Blindness and Visual Impairment	0.916	0.980	0.684	1.391
Other Eye Diseases	0.780	1.033	0.839	1.267
Glaucoma	0.122	0.786	0.590	1.038
Deafness and Hearing Impairments	0.126	1.208	0.963	1.512
Musculoskeletal Conditions:				
Inflammatory Arthropathies	0.272	1.115	0.931	1.333
*Osteoarthritis and other DJD**	0.046	1.156	1.013	1.317
Psychiatric Conditions:				
Depression and Mood Diseases	0.502	0.938	0.788	1.113
Other Conditions:				
Urinary Incontinence	0.142	1.099	0.979	1.233
*Diabetes**	0.010	0.838	0.740	0.949
*Peripheral Neuropathy**	0.016	1.337	1.077	1.657
Chronic Kidney Diseases	0.256	0.928	0.823	1.046
Anaemia	0.097	1.136	0.988	1.304
Colitis and Related Diseases	0.188	0.910	0.799	1.035
Oesophagus Stomach, and Duodenum Diseases	0.281	1.131	0.919	1.387
Prostate Diseases	0.415	0.881	0.662	1.163

* = $P \leq 0.05$, *DJD: Degenerative Joint Disease, IRR: Incident Rate Ratio, CL: Confidence Limit*

to aid with reporting whilst also ensuring significant results are not over-emphasised in the reporting compared to non-significant results. These categories are demographic variables, neurological conditions, psychiatric conditions, cardiovascular conditions, musculoskeletal conditions, sensory impairments, metabolic conditions, and other internal medical conditions.

There was a 1.3% (0.7%-2%) increase in fall presentations for each year of age.

Male sex exhibited a protective effect, leading to a significant 14.4% decrease in fall presentations to ED during the study period.

Binary indicators for three neurological conditions were included in the model reported in Table 5.11. These were dementia, cerebrovascular disease, and Parkinson's disease. Dementia was significantly associated with a 32.3% (19.3%-46.7%) increase in fall presentations to ED, while cerebrovascular disease led to an 18.6% (2.6%-37.1%) increase. Parkinson's disease was not significantly ($P > 0.05$) associated with any change in the count of fall presentations to ED. However, because only 3.97% of the sample had Parkinson's this non-significant finding may be due to a Type-2 error rather than lack of association.

There was a single psychiatric condition group included in the regression model, which was Depression and Mood Diseases. This was not significantly ($P > 0.05$) associated with a change in fall presentations to ED.

There were seven cardiovascular chronic health condition groups included in the shortlist regression model, of which only three had a significant association with fall presentations to ED. Hypotension was associated with a large increase of 61.1% (38.0%-87.9%), while more moderate effect sizes were observed for atrial fibrillation and ischemic heart disease, which contributed 21.8% (8.4%-36.8%) and 20.2% (7.0%-35.0%) increases respectively. The remaining included cardiovascular conditions of bradycardia and conduction diseases, cardiac valve disease, heart failure, hypertension had a non-significant association ($P > 0.05$).

Two musculoskeletal chronic health conditions were included in the shortlist regression model. Osteoarthritis and degenerative joint diseases was associated with a 15.6% (1.3%-31.7%) increase in fall presentations whereas inflammatory arthropathies had a non-significant ($P > 0.05$) effect.

Four binary variables were included to represent sensory impairments in the sample members. Blindness and visual impairment, other eye diseases, glaucoma, deafness and hearing impairment were all found to be non-significant ($P > 0.05$) in the model.

Diabetes and peripheral neuropathy both exhibited significant effects on the count of falls. Diabetes contributed to a 16.2% (26.0%-5.1%) decrease in fall presentations

to the ED, while peripheral neuropathy was associated with a 33.7% (7.7%-65.7%) increase in fall presentations.

The remaining conditions in the shortlist model, encompassing a range of internal medical conditions (anaemia, and chronic kidney diseases), gastrointestinal conditions (colitis and related diseases, oesophagus stomach and duodenum diseases, and prostate diseases), and urinary incontinence were all found to be non-significant ($P > 0.05$). Possible explanations for these conditions being non-significant are discussed in Section 6.3.2.

To summarise the results in Table 5.11, each year of age was significantly associated with a small increase in fall presentations to ED over the study period. Additionally, male sex exhibited a moderate protective effect. Of the included chronic health conditions: cerebrovascular disease (18.6%), dementia (32.3%), hypotension (61.1%), atrial fibrillation (21.8%), ischemic heart disease (20.2%), osteoarthritis and degenerative joint diseases (15.6%), and peripheral neuropathy (33.7%) were all found to be associated with significant increases in fall presentations during the study period. Meanwhile, diabetes (-16.2%) exhibited the only significant protective effect of the included chronic health conditions. None of the included sensory impairment or psychiatric conditions were found to significantly impact fall presentations in the model.

These model results further demonstrate that the relationship between multimorbidity and falls will depend on the chronic health conditions present. The presence of hypotension, dementia, or peripheral neuropathy were associated with the largest changes in fall presentations, however these effects were smaller than those observed for the clusters in Table 5.10. How these results contrast with those observed for the gradient of effects seen for different multi-morbidity combinations is discussed further in Section 6.3.2.

Following the regression models several follow-up tests were performed as detailed in Methodology section 4.4.3. The results of these evaluation tests and plots follow in Section 5.4.5 below. The results of further analysis into the possible interaction effects present between variables is presented in Section 5.5.

5.4.5 Model Diagnostics

Several tests and metrics were used to evaluate the model results presented in Section 5.4. The first of these metrics was the Akaike information criterion (AIC), as described in methodology Section 4.4.3. This is a comparative metric, which accounts for model complexity and quality of fit to give an indication of how well different models explain variation in the outcome variable, with a lower AIC value indicating a better comparative fit. AIC values for each model from Section 5.4 are presented in Table 5.12. All models had similar AIC values, with the chronic condition shortlist model having the lowest score of the models from this section.

TABLE 5.12: Single Effects Negative Binomial AIC Comparison

Single Effects Model	**AIC Value**
CCI	9089.667
EFI	9004.621
HFRS	9077.762
K-Means Clustering	9050.46
Chronic Condition Shortlist	8972.798

CCI: Charlson Comorbidity Index, EFI: Electronic Frailty Index, HFRS: Hospital Frailty Risk Score, AIC: Akaike Information Criterion

Expected versus fitted residual plots for negative binomial regression models are not easily interpretable by eye. Therefore, to assist with assessing the appropriateness of the model, the DHARMa package was used. This package generates scaled residuals using the calculated model which can be interpreted graphically, as is done for a linear regression. These scaled residuals are the cumulative density values from the cumulative distribution function for the observed fall count in the sample, and the expected fall count from the model for each sample member. The simulated scaled residuals can then be interpreted via graphical means in the same manner as residuals from a standard linear regression model.

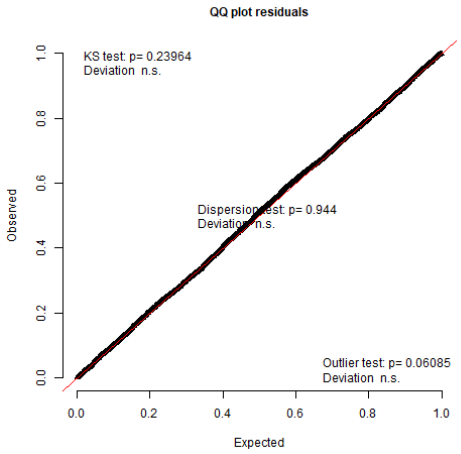
The QQ plots presented in Figure 5.8 show the quantiles of the scaled residuals produced using the DHARMa package. A well fit model is shown in these plots where the quantiles of the scaled residuals are equivalent to the theoretical quantiles. In this situation, the points fall on the 45-degree line in the plot.

The advantage of using the DHARMa package is the interpretation of these rescaled residuals is equivalent irrespective of the underlying distribution used to

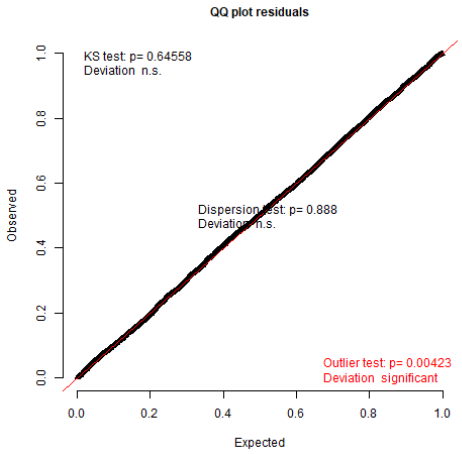
construct the model. As seen in Figure 5.8, the lack of systematic deviation from the 45-degree line in these plots by the five models indicates that there are no major concerns in the fit of the models at different levels of the outcome.

The DHARMa package provides several diagnostic tests for model fit, which are overlayed on the plots in Figure 5.8. These are the Kolmogorov-Smirnov test, Over/Under dispersion test, and outliers test. As noted in the package details, the outlier test is unreliable for use with negative binomial models, and the results are therefore not worth investigating further (Hartig, 2022b). The Kolmogorov-Smirnov test in this instance identifies whether the observed residual quantiles differ significantly from the theoretical quantiles. As seen in Figure 5.8 none of the models showed evidence of deviation from the assumed distributions. Finally, the dispersion test uses a ratio of the observed residuals and simulated residuals to identify whether over or under dispersion are present. None of the models in Figure 5.8 show evidence of over or under dispersion. All of the tests performed suggest the models are well fit, which gives further weight to the conclusions of the analysis.

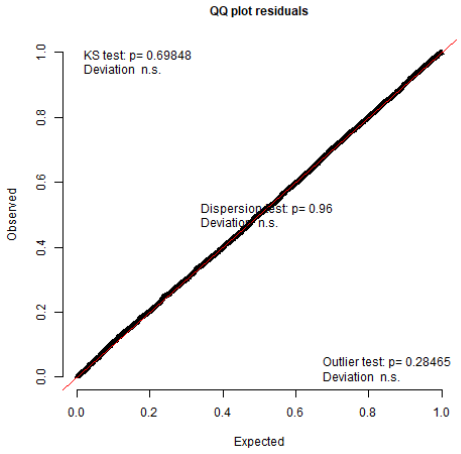
FIGURE 5.8: DHARMa QQ Plots for Single Effects Negative Binomial Models



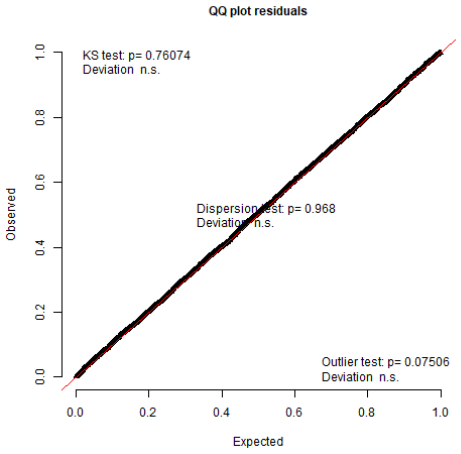
(A) CCI Model DHARMa QQ Plot



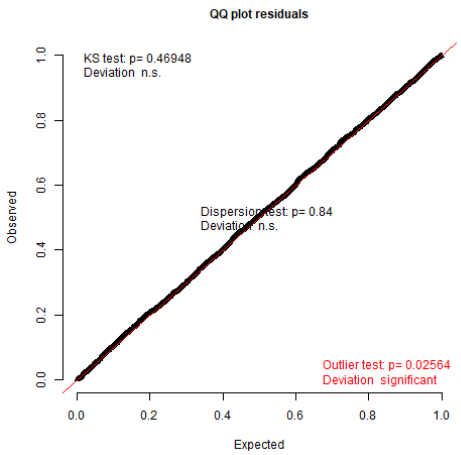
(B) EFI Model DHARMa QQ Plot



(C) HFRS Model DHARMa QQ Plot



(D) K-Means Cluster Model DHARMa QQ Plot



(E) Chronic Condition Shortlist Model DHARMa QQ Plot

5.5 Interaction Regression Analysis

5.5.1 Introduction

I now turn to exploring the effects of condition co-occurrence and how combinations of conditions could present a mitigating or intensifying effect on the relationship between multi-morbidity and falls. The aims of this analysis were to test a systematic range of possible condition interactions to further explore the role of multi-morbidity in determining fall presentations by sample members. This analysis is also used to reveal more about the changes in fall risk described in the previous analyses. Furthermore, the identification of such interactions could have important clinical impacts through the improvement of both the understanding of mechanisms driving the effect of multi-morbidity on falls risk, but also the improvement of falls risk prediction models for use in highly multi-morbid populations. Identifying combinations of chronic health conditions of interest could also lead to proactive intervention and mitigation of increased falls risk in future.

5.5.2 Results

A single model of condition interaction was developed using the full sample. This model explored interactions between the five chronic health conditions with the highest prevalence in the sample, whilst controlling for the shortlist of chronic health conditions derived in the Chapter 3 review of reviews. Interactions between the following conditions were considered in the model: dementia, hypertension, atrial fibrillation, chronic kidney disease, and urinary incontinence. The most prevalent conditions were selected for this model because the interaction of these conditions is the most likely to occur in the care home resident population. Therefore, identifying whether multiplicative effects of multi-morbidity on falls risk exist between these conditions could have clinical implications which, if identified and acted upon, could lead to benefits for patients.

The five most prevalent chronic health condition groups in the sample were combined using pairwise interactions in a negative binomial regression model. The results of this model indicate that, there is no evidence of multiplicative effects of these

chronic conditions on falls risk. This is shown in Table 5.13 where the interaction terms are non-significant ($P > 0.05$) in the model.

Further interaction models were developed for each cluster, using the conditions with over a 15% prevalence threshold in the cluster and an Observed-Expected (O/E) ratio of ≥ 2. These thresholds were used to maximise the possibility that the interaction would be occurring in a sufficiently large sample size for meaningful conclusions to be drawn. Where more than five conditions met the criteria within the cluster, the conditions with the top five O/E ratios were used for the pairwise interactions.

The cluster based interaction models were conducted within the cluster, meaning that it is important to contextualise the results in what is already known about the clusters from previous analyses. Additionally the findings from these sub-sample models cannot be directly compared because the base level of falls differs between the clusters, as seen in Table 5.6. Therefore, these models are making comparisons between individuals within the clusters to identify what conditions were associated with the highest fall rates in the cluster. This means that if all sample members in a cluster exhibited a similar fall count throughout the study period as a result of similar chronic condition influence, the effects of these chronic conditions may be non-significant in these interaction models.

Table 5.14 presents the results of the Cardiovascular cluster interaction analysis model, where five conditions were combined in pairwise interactions alongside the chronic condition shortlist. In this model, all of the included variables and interactions were non-significant ($P > 0.05$). The sample members in the Cardiovascular cluster may all have consistent chronic condition profiles, meaning there is limited variation in explanatory variables versus the outcome. Second, there may be little difference in the fall count between members of the cluster. Also, it is possible that the interactions between conditions took place in a small sample size, leading to wide confidence intervals and non-significant effects. Finally, it is also possible that these conditions do not exhibit multiplicative multi-morbidity effects on falls.

Table 5.15 shows the interaction model derived for the Cardiovascular-Metabolic cluster where five cardiovascular conditions were placed in pairwise interactions

TABLE 5.13: Top 5 in prevalence from Shortlist Regression Interaction Model

Variable	P-Value	IRR	2.5% IRR CL	97.5% IRR CL
Intercept	0.000	0.000	0.000	0.000
Age	0.000	1.013	1.007	1.020
Male Sex	0.006	0.854	0.771	0.946
Interaction Terms				
Dementia - Hypertension	0.454	0.909	0.722	1.144
Dementia - Atrial Fibrillation	0.155	1.207	0.952	1.531
Dementia - Chronic Kidney Diseases	0.978	0.996	0.773	1.285
Dementia - Urinary Incontinence	0.964	1.006	0.790	1.283
Atrial Fibrillation - Hypertension	0.706	0.952	0.754	1.203
Hypertension - Chronic Kidney Diseases	0.274	1.164	0.908	1.494
Urinary Incontinence - Hypertension	0.308	1.143	0.903	1.448
Atrial Fibrillation - Chronic Kidney Diseases	0.978	0.996	0.785	1.265
Urinary Incontinence - Atrial Fibrillation	0.092	0.801	0.632	1.014
Urinary Incontinence - Chronic Kidney Diseases	0.669	0.944	0.740	1.204
Neurological Conditions:				
*Dementia**	0.005	1.312	1.104	1.558
*Cerebrovascular Disease**	0.030	1.190	1.029	1.375
Parkinson's and Parkinsonism	0.188	1.173	0.941	1.457
Cardiovascular Conditions:				
*Hypotension**	0.000	1.611	1.380	1.881
Atrial Fibrillation	0.186	1.200	0.935	1.536
Bradicardia and Conduction Diseases	0.680	0.948	0.749	1.195
Cardiac Valve Disease	0.705	0.966	0.820	1.136
*Ischemic Heart Disease**	0.004	1.203	1.070	1.352
Heart Failure	0.154	0.902	0.791	1.027
Hypertension	0.879	0.983	0.799	1.206
Sensory Impairment Conditions:				
Blindness and Visual Impairment	0.839	0.961	0.671	1.365
Other Eye Diseases	0.893	1.016	0.824	1.248
Glaucoma	0.125	0.787	0.591	1.039
Deafness and Hearing Impairment	0.122	1.211	0.964	1.515
Musculoskeletal Conditions:				
Inflammatory Arthropathies	0.256	1.120	0.934	1.339
*Osteoarthritis and DJD**	0.039	1.162	1.018	1.325
Psychiatric Conditions:				
Depression and Mood Diseases	0.465	0.932	0.783	1.106
Other Conditions:				
Urinary Incontinence	0.432	1.117	0.863	1.440
*Diabetes**	0.014	0.844	0.745	0.956
*Peripheral Neuropathy**	0.019	1.328	1.069	1.647
Chronic Kidney Diseases	0.353	0.862	0.643	1.148
Anaemia	0.088	1.140	0.992	1.309
Colitis and Related Diseases	0.207	0.913	0.801	1.039
Oesophagus, Stomach and Duodenum Diseases	0.276	1.132	0.921	1.389
Prostate Diseases	0.438	0.886	0.665	1.170

* = $P \leq 0.05$, *IRR: Incident Rate Ratio, CL: Confidence Limit, DJD: Degenerative Joint Diseases*

TABLE 5.14: Cardiovascular Cluster: Interaction Regression Model

Variable	**P-Value**	**IRR**	**2.5% IRR CL**	**97.5% IRR CL**
Intercept	0.000	0.000	0.000	0.000
Age*	0.022	1.038	1.008	1.071
Male Sex	0.371	1.215	0.821	1.797
Interaction Terms				
Cardiac Valve Diseases - Heart Failure	0.532	1.518	0.452	5.256
Atrial Fibrillation - Cardiac Valve Diseases	0.147	0.484	0.187	1.225
Cardiac Valve Diseases - Chronic Kidney Diseases	0.328	0.620	0.248	1.524
Cardiac Valve Diseases - Ischemic Heart Disease	0.358	1.546	0.639	3.707
Atrial Fibrillation: - Heart Failure	0.710	1.287	0.355	4.502
Heart Failure - Chronic Kidney Diseases	0.974	0.982	0.341	2.753
Ischemic Heart Disease - Heart Failure	0.966	1.023	0.388	2.674
Atrial Fibrillation - Chronic Kidney Diseases	0.445	0.686	0.273	1.698
Atrial Fibrillation - Ischemic Heart Disease	0.393	1.513	0.606	3.735
Ischemic Heart Disease - Chronic Kidney Diseases	0.399	1.453	0.649	3.251
Neurological Conditions:				
Dementia	0.168	1.330	0.912	1.950
Cerebrovascular Disease	0.177	1.540	0.855	2.756
Parkinson's and Parkinsonism	0.106	2.408	0.844	6.844
Cardiovascular Conditions:				
Hypotension	0.067	1.720	0.982	3.019
Atrial Fibrillation	0.627	1.506	0.326	7.517
Bradicardias and Conduction Diseases	0.749	1.101	0.632	1.895
Cardiac Valve Diseases	0.922	1.084	0.236	5.143
Ischemic Heart Disease	0.427	0.553	0.142	2.207
Heart Failure	0.798	0.792	0.153	4.461
Hypertension	0.323	0.815	0.559	1.190
Sensory Impairment Conditions:				
Other Eye Diseases	0.837	1.163	0.280	4.282
Deafness and Hearing Impairment	0.117	0.227	0.031	1.065
Musculoskeletal Conditions:				
Inflammatory Arthropathies	0.499	1.241	0.683	2.236
Osteoarthritis and DJD	0.418	1.308	0.722	2.354
Psychiatric Conditions:				
Depression and Mood Diseases	0.567	0.678	0.181	2.311
Other Conditions:				
Urinary Incontinence	0.315	1.258	0.827	1.907
Diabetes	0.276	0.772	0.496	1.195
Peripheral Neuropathy	0.899	1.080	0.326	3.333
Chronic Kidney Diseases	0.859	1.139	0.294	4.615
Anaemia	0.885	1.040	0.628	1.707
Colitis and Related Diseases	0.153	0.596	0.297	1.144
Oesophagus, Stomach, and Duodenum Diseases	0.337	1.972	0.474	8.013
Prostate Diseases	0.393	0.689	0.303	1.513

* = $P \leq 0.05$, *IRR: Incident Rate Ratio, CI: Confidence Limit, DJD: Degenerative Joint Diseases*

alongside the shortlisted chronic health conditions. The interaction terms in the Table 5.15 model were all non-significant (P $\geq$ 0.05). The Cardiovascular-Metabolic cluster has only 168 sample members present, meaning these interactions may have been taking place in very small sub-samples with limited statistical power. In this cluster however, heart failure, and hypertension each had prevalence over 90% in the cluster, meaning they are unlikely to explain variation in fall rate within the cluster. In contrast, cerebrovascular disease contributed to a 100.2% (17.8%-242.2%) increase in fall rate in this cluster. These results suggest therefore that when heart failure, and hypertension are present on the inpatient record, the additional presence of cerebrovascular disease can greatly increase the rate of fall presentations to ED from an already elevated level.

The interaction model in Table 5.16 was derived from the Non-Specific-High-Burden cluster sample, which is characterised by a high burden of chronic disease across almost all groups considered in the analyses. In this model five pairwise interaction terms were derived, however these interaction terms were all non-significant ($P > 0.05$) in the Non-Specific-High-Burden cluster model.

The Central cluster interaction model presented in Table 5.17 derived pairwise interactions between four variables associated with the cluster. These interaction terms were all non-significant in the model, with all other terms besides age showing non-significant effects.

In the Low-Neuro-Psychiatric cluster, no condition groups passed the thresholds for interaction, as described previously, and so no interaction model was developed.

The High-Neuro-Psychiatric cluster is associated with neurological and psychiatric condition presence. In the model presented in Table 5.18 five chronic condition groups were included in pairwise interactions alongside the chronic condition shortlist derived following the Chapter 3 review of reviews. The included interactions in this cluster were all non-significant (P $>$ 0.05).

The final Cluster interaction model developed was in the Low-Cardio-Neuro cluster, which as seen previously, is defined by weak association with cardiovascular conditions. For this reason, no chronic condition groups passed the thresholds specified for interaction in this cluster.

The cluster based interaction models described in this section did not identify

TABLE 5.15: Cardiovascular-Metabolic Cluster: Interaction Regression Model

Variable	P-Value	IRR	2.5% IRR CL	97.5% IRR CL
Intercept	0.002	0.001	0.000	0.034
Age	0.866	1.004	0.965	1.044
Male Sex	0.826	0.930	0.524	1.640
Interaction Terms:				
Bradicardias and Conduction Diseases - Cardiac Valve Diseases	0.772	1.240	0.368	4.224
Bradicardias and Conduction Diseases - Heart Failure	0.585	0.340	0.012	9.938
Bradicardias and Conduction Diseases - Inflammatory Arthropathies	0.705	1.357	0.360	5.092
Hypotension - Bradicardias and Conduction Diseases	0.342	0.511	0.164	1.579
Cardiac Valve Diseases - Heart Failure	0.823	1.287	0.199	8.371
Cardiac Valve Diseases - Inflammatory Arthropathies	0.968	1.028	0.329	3.198
Hypotension - Cardiac Valve Diseases	0.593	0.708	0.244	2.046
Heart Failure - Inflammatory Arthropathies	0.756	1.640	0.124	24.909
Hypotension - Heart Failure	0.575	0.523	0.074	3.733
Hypotension - Inflammatory Arthropathies	0.606	1.479	0.396	5.554
Neurological Conditions:				
Dementia	0.329	1.357	0.801	2.317
*Cerebrovascular Disease**	0.028	2.002	1.178	3.422
Parkinson's and Parkinsonism	0.343	1.911	0.602	5.872
Cardiovascular Conditions:				
Hypotension	0.202	4.406	0.610	31.701
Atrial Fibrillation	0.613	1.205	0.656	2.251
Bradicardias and Conduction Diseases	0.659	2.452	0.075	72.149
Cardiac Valve Diseases	0.803	0.750	0.111	5.051
Ischemic Heart Disease	0.562	0.836	0.504	1.385
Heart Failure	0.913	1.122	0.194	6.679
Hypertension	0.183	0.519	0.231	1.179
Sensory Impairment Conditions:				
Other Eye Diseases	0.527	1.360	0.579	3.128
Glaucoma	0.832	0.845	0.204	3.099
Deafness and Hearing Impairment	0.440	1.426	0.662	3.032
Musculoskeletal Conditions:				
Inflammatory Arthropathies	0.760	0.619	0.041	8.121
Osteoarthritis and DJD	0.329	0.712	0.400	1.246
Psychiatric Conditions:				
Depression and Mood Diseases	0.650	0.708	0.186	2.311
Other Conditions:				
Urinary Incontinence	0.580	1.200	0.704	2.059
Diabetes	0.833	1.068	0.643	1.778
Peripheral Neuropathy	0.198	0.382	0.097	1.191
Chronic Kidney Diseases	0.969	0.987	0.579	1.699
Anaemia	0.947	0.980	0.588	1.629
Colitis and Related Diseases	0.564	0.835	0.492	1.410
Oesophagus, Stomach, and Duodenum Diseases	0.353	0.527	0.156	1.577
Prostate Diseases	0.927	0.954	0.408	2.197

* = $P \leq 0.05$, *IRR: Incident Rate Ratio, CL: Confidence Limit, DJD: Degenerative Joint Diseases*

TABLE 5.16: Non-Specific-High-Burden Cluster: Interaction Regression Model

Variable	P-Value	IRR	2.5% IRR CL	97.5% IRR CL
Intercept	0.000	0.000	0.000	0.004
Age	0.815	0.996	0.970	1.023
Male Sex*	0.021	0.560	0.367	0.841
Interaction Terms:				
Espohagus, Stomach, and Duodenum Diseases - Peripheral Neuropathy	0.853	0.923	0.448	1.882
Depression and Mood Diseases- Peripheral Neuropathy	0.134	0.524	0.254	1.064
Other Eye Diseases - Peripheral Neuropathy	0.776	1.195	0.417	3.389
Inflammatory Arthropathies- Peripheral Neuropathy	0.746	1.186	0.484	2.836
Depression and Mood Diseases - Espohagus, Stomach, and Duodenum Diseases	0.114	1.968	0.972	4.030
Other Eye Diseases - Espohagus, Stomach, and Duodenum Diseases	0.212	2.073	0.773	5.422
Inflammatory Arthropathies - Espohagus, Stomach, and Duodenum Diseases	0.910	0.935	0.331	2.409
Other Eye Diseases - Depression and Mood Diseases	0.403	1.685	0.580	4.651
Depression and Mood Diseases - Inflammatory Arthropathies	0.715	1.248	0.444	3.396
Other Eye Diseases - Inflammatory Arthropathies	0.846	1.181	0.278	4.926
Neurological Conditions:				
Dementia	0.899	1.032	0.687	1.571
Cerebrovascular Disease	0.077	1.466	1.024	2.094
Parkinson's and Parkinsonism	0.295	0.594	0.243	1.292
Cardiovascular Conditions:				
*Hypotension**	0.001	2.078	1.428	3.022
*Atrial Fibrillation**	0.003	1.915	1.339	2.769
Bradicardias and Conduction Diseases	0.652	1.170	0.641	2.054
Cardiac Valve Diseases	0.364	1.281	0.807	2.001
Ischemic Heart Disease	0.084	1.466	1.019	2.125
Heart Failure	0.342	0.806	0.551	1.174
Hypertension	0.069	1.699	1.063	2.795
Sensory Impairment Conditions:				
Blindness and Visual Impairment	0.573	0.799	0.399	1.516
Other Eye Diseases	0.599	0.767	0.317	1.721
Glaucoma	0.305	0.636	0.290	1.273
Deafness and Hearing Impairment	0.206	1.488	0.872	2.482
Musculoskeletal Conditions:				
Inflammatory Arthropathies	0.274	0.614	0.284	1.262
Osteoarthritis and DJD	0.500	1.173	0.792	1.742
Psychiatric Conditions:				
Depression and Mood Diseases	0.908	1.041	0.577	1.859
Other Conditions:				
Urinary Incontinence	0.863	0.964	0.678	1.379
Diabetes	0.843	0.960	0.679	1.355
*Peripheral Neuropathy**	0.033	2.108	1.182	3.755
Chronic Kidney Diseases	0.110	0.691	0.469	1.010
*Anaemia**	0.022	1.614	1.144	2.279
Colitis and Related Diseases	0.089	1.479	1.014	2.175
Oesophagus, Stomach, and Duodenum Diseases	0.256	0.648	0.338	1.209
Prostate Diseases	0.099	0.211	0.031	0.820

* = $P \leq 0.05$, *IRR: Incident Rate Ratio, CL: Confidence Limit, DJD: Degenerative Joint Diseases*

TABLE 5.17: Central Cluster: Interaction Regression Model

Variable	P-Value	IRR	2.5% IRR CL	97.5% IRR CL
Intercept	0.000	0.000	0.000	0.000
Age*	0.013	1.030	1.008	1.052
Male Sex	0.325	0.850	0.629	1.144
Interaction Terms:				
Hypotension - Colitis and Related Diseases	0.131	1.860	0.880	3.916
Cerebrovascular Disease - Colitis and Related Diseases	0.242	1.621	0.761	3.387
Urinary Incontinence - Colitis and Related Diseases	0.924	1.032	0.569	1.876
Cerebrovascular Disease - Hypotension	0.628	1.241	0.553	2.754
Urinary Incontinence - Hypotension	0.394	0.722	0.362	1.428
Cerebrovascular Disease - Urinary Incontinence	0.818	0.920	0.476	1.795
Neurological Conditions:				
Dementia	0.244	1.230	0.897	1.697
Cerebrovascular Disease	0.957	0.983	0.548	1.727
Parkinson's and Parkinsonism	0.400	0.784	0.464	1.299
Cardiovascular Conditions:				
Hypotension	0.347	1.333	0.753	2.336
Atrial Fibrillation	0.596	1.085	0.821	1.434
Bradicardias and Conduction Diseases	0.784	1.094	0.596	1.963
Cardiac Valve Diseases	0.721	1.082	0.727	1.598
Ischemic Heart Disease	0.090	1.297	0.987	1.705
Heart Failure	0.650	1.077	0.803	1.444
Hypertension	0.089	0.729	0.522	1.023
Sensory Impairment Conditions:				
Blindness and Visual Impairment	0.085	0.454	0.189	1.000
Other Eye Diseases	0.567	1.151	0.734	1.784
Glaucoma	0.797	0.914	0.482	1.670
Deafness and Hearing Impairment	0.600	1.142	0.717	1.795
Musculoskeletal Conditions:				
Inflammatory Arthropathies	0.180	1.335	0.904	1.958
Osteoarthritis and DJD	0.369	0.850	0.609	1.178
Psychiatric Conditions:				
Depression and Mood Diseases	0.893	1.034	0.650	1.620
Other Conditions:				
Urinary Incontinence	0.278	1.264	0.853	1.874
Diabetes	0.491	0.894	0.667	1.198
Peripheral Neuropathy	0.232	1.455	0.827	2.544
Chronic Kidney Diseases	0.340	1.165	0.872	1.556
Anaemia	0.846	1.036	0.743	1.440
Colitis and Related Diseases	0.062	0.627	0.398	0.978
Oesophagus, Stomach, and Duodenum Diseases	0.304	1.285	0.822	1.986
Prostate Diseases	0.285	1.416	0.776	2.529

* = $P \leq 0.05$, *IRR: Incident Rate Ratio, CL: Confidence Limit, DJD: Degenerative Joint Diseases*

TABLE 5.18: High-Neurological-Psychiatric Cluster: Interaction Regression Model

Variable	P-Value	IRR	2.5% IRR CL	97.5% IRR CL
Intercept	0.000	0.001	0.000	0.004
Age	0.955	0.999	0.975	1.024
Male Sex	0.524	0.871	0.603	1.248
Interaction Terms:				
Depression and Mood Diseases - Peripheral Neuropathy	0.672	0.781	0.287	2.004
Depression and Mood Diseases - Espohagus, Stomach, and Duodenum Diseases	0.554	1.414	0.520	3.650
Other Eye Diseases - Depression and Mood Diseases	0.747	0.794	0.226	2.404
Cerebrovascular Disease - Depression and Mood Diseases	0.706	1.196	0.538	2.589
Espohagus, Stomach, and Duodenum Diseases - Peripheral Neuropathy	0.610	0.602	0.082	2.675
Other Eye Diseases - Peripheral Neuropathy	0.879	0.874	0.166	3.518
Cerebrovascular Disease- Peripheral Neuropathy	0.129	2.298	0.925	5.658
Other Eye Diseases- Espohagus, Stomach, and Duodenum Diseases	0.365	1.981	0.516	6.833
Cerebrovascular Disease- Espohagus, Stomach, and Duodenum Diseases	0.771	1.239	0.347	4.008
Cerebrovascular Disease- Other Eye Diseases	0.403	0.529	0.131	1.727
Neurological Conditions:				
*Dementia**	0.009	1.761	1.236	2.537
Cerebrovascular Disease	0.219	0.695	0.419	1.136
Parkinson's and Parkinsonism	0.788	0.892	0.418	1.750
Cardiovascular Conditions:				
Hypotension	0.635	1.161	0.676	1.930
Atrial Fibrillation	0.115	1.441	0.972	2.114
Ischemic Heart Disease	0.475	0.814	0.498	1.296
Heart Failure	0.066	0.323	0.103	0.827
Hypertension	0.265	1.249	0.897	1.749
Sensory Impairment Conditions:				
Blindness and Visual Impairment	0.243	1.496	0.836	2.611
Other Eye Diseases	0.864	0.937	0.494	1.719
Glaucoma	0.556	0.841	0.507	1.361
Deafness and Hearing Impairment	0.147	1.539	0.929	2.499
Musculoskeletal Conditions:				
Inflammatory Arthropathies	0.871	0.939	0.476	1.734
Osteoarthritis and DJD	0.486	1.153	0.820	1.618
Psychiatric Conditions:				
Depression and Mood Diseases	0.138	0.664	0.418	1.045
Other Conditions:				
*Urinary Incontinence**	0.019	1.588	1.144	2.204
Diabetes	0.699	0.917	0.628	1.326
Peripheral Neuropathy	0.874	1.060	0.557	1.958
Chronic Kidney Diseases	0.349	0.763	0.468	1.210
Anaemia	0.997	0.999	0.677	1.453
Colitis and Related Diseases	0.487	0.864	0.606	1.221
Oesophagus, Stomach, and Duodenum Diseases	0.796	1.096	0.596	1.962
Prostate Diseases	0.910	1.057	0.450	2.305

* = $P \leq 0.05$, *IRR: Incident Rate Ratio, CL: Confidence Limit, DJD: Degenerative Joint Diseases*

any multiplicative multi-morbidity interaction effects between chronic condition variables. However, this analysis is not definitive, and alternative approaches may be better suited for identifying these effects, as discussed in Section 6.5. However, the consistency of these findings indicates that the conditions tested for interactions within this analysis are unlikely to interact in a meaningful way to cause changes in the risk of fall presentations during the study period.

5.5.3 Interaction Analysis Further Plots and Tests

DHARMa residual plots for the interaction models presented in Section 5.5.2 are shown in Appendix Section L. These plots show no strong systematic relationships in the residuals of these interaction models. This suggests that the lack of significant interactions was not a result of systematic bias within the models.

5.6 Conclusion

The results presented in this chapter and Sections 5.3-5.5 in particular, used a novel combination of methods to analyse an understudied population in health data research. The aim of the analyses was to investigate what chronic health conditions have an impact on the rate of fall presentations to the ED by care home residents, and whether their is evidence for a multiplicative role of multi-morbidity in this relationship. While several useful findings were identified, the results of the analyses suggest there is not a multiplicative effect of multi-morbidity on falls. Specifically, Section 5.3 identified clusters of chronic health conditions, which exhibited different relationships with the conditions studied in the analyses. Differences in fall presentations during the study period between these clusters were identified in Section 5.4. However, the analyses in Section 5.5 found no evidence that the effects observed went beyond additive effects of individual chronic conditions. Possible explanations for this lack of finding are explored further in the Section 6.5. The following chapter contextualises the results of the analyses in the wider literature, whilst also addressing study limitations, and opportunities for future research.

Chapter 6

Discussion and Conclusions

6.1 Introduction to Chapter

The empirical research presented in this thesis set out to identify patterns of multi-morbidity in UK based care home residents, and what impact these patterns had on the risk of falls. Answering these questions is essential in the context of an ageing population and resource constrained healthcare setting and the research presented in this thesis is intended to motivate further study into this area. The discussion in this chapter is structured as follows: after an initial comparison of findings with previous literature in Section 6.3, clinical and policy implications of the results are discussed in Section 6.4. Next the limitations of the research are discussed in Section 6.5, following which a summary of the novel contribution made by this thesis is provided in Section 6.6. The recommendations for future research are presented in Section 6.7. Finally, the conclusions of the thesis, and reflections on the PhD process are summarised in Section 6.8.

6.2 Summary of Main Findings

Based on the results in Section 5.4.3 a gradient of multi-morbidity effects can be drawn with relation to falls in care home residents. The High-Neuro-Psychiatric and Low-Neuro-Psychiatric clusters had a stronger association with fall presentations to emergency departments than the Cardiovascular condition cluster, with 41.4%, 43% and 23.8% increases respectively when compared to the Absence cluster. However, the Cardiovascular-Metabolic cluster exhibited a stronger effect than

either of the Neurological-Psychiatric clusters with a 63% increase in fall presentations associated with this cluster compared to the Absence cluster. Finally, the strongest increase of 117.8% was observed for the Non-Specific-High-Burden cluster, which may represent sample members where both the neurological-psychiatric pattern, and cardiovascular-metabolic patterns are present. Therefore, this may represent the most potent combination for increased falls risk in care home dwelling older adults in the sample.

When considering the role of individual chronic health conditions, hypotension (61.1%), dementia (32.3%), and peripheral neuropathy (33.7%) exhibited the largest effects, with smaller risk increasing effects observed for cerebrovascular disease (18.6%), atrial fibrillation (21.8%), and osteoarthritis and degenerative joint diseases (15.6%).

In an investigation of the relationship between frailty and fall presentations in care home dwelling older adults, three indices were examined in Section 5.4.2. The results of these regression models indicate how the measure used to asses frailty will partially determine the relationship identified. Additionally, the results highlight the risk of using measures such as the Charlson Comorbidity Index (CCI) to reflect the role of frailty in falls risk, when this metric was designed for prediction of ten year survival (Charlson et al., 1987).

6.3 Discussion of Findings

6.3.1 Clusters of Chronic Health Conditions in UK Care Home Residents

The cluster analysis presented in this thesis had two aims. The first was to identify groups of chronic health conditions in the sample of care home residents, and compare the resulting patterns to those previously identified in care home resident populations. The second aim of the cluster analyses was to examine whether the fall rate varies between the identified patterns of multi-morbidity. In meeting these two aims the analyses made two novel contributions to knowledge. This is because patterns of multi-morbidity are under-researched in UK care home resident populations, which this analyses addresses. The second source of novelty is that multi-morbidity patterns have not previously been linked to falls risk in either care home

resident or community dwelling populations. Considering how many individual chronic conditions are linked with changes in falls rate in previous literature, this analysis represents an important step towards understanding how groups of these conditions may impact falls (Rubenstein, 2006; Deandrea et al., 2010; Bloch et al., 2013).

Some of the patterns of multi-morbidity in the cluster solution were similar to two major patterns identified in previous multi-morbidity studies based on samples of community dwelling older adults. These patterns were a cardiovascular-metabolic pattern and neurological-psychiatric pattern observed in the Cardiovascular-Metabolic and High-Neuro-Psychiatric clusters respectively. These patterns persist across samples and different analysis methods, suggesting they are valid groupings of chronic disease presentations in older adults. However, while the overarching groupings were the same, there were differences identified in the present study when compared to previous research in the community setting which are discussed for each pattern below.

Cardiovascular-Metabolic Pattern

A cardiovascular-metabolic pattern of multi-morbidity has been consistently identified in the community setting using a range of samples and analysis approaches (Islam et al., 2014; García-Olmos et al., 2012; Machón et al., 2020; Vu, Finch, and Day, 2011; Guisado-clavero et al., 2018; Déruaz-Luyet et al., 2017; Schafer et al., 2010; Violán et al., 2019; Marengoni et al., 2020). The main features of this pattern are the coincidence of diabetes, hypertension, and a range of cardiovascular diseases (cardiac valve disease, ischemic heart disease, atrial fibrillation). Additional related conditions inconsistently observed in the pattern are obesity, dislipidemia, chronic kidney disease, chronic liver disease, and peripheral neuropathy. This cardiovascular-metabolic pattern is explained by metabolic syndrome, where the presence of diabetes, hypertension, and obesity raise the risks of developing cardiovascular conditions (Byrne and Wild, 2011).

In contrast to previous studies in the community setting, the cardiovascular-metabolic pattern identified in the care home resident sample also incorporates further links with neuro-degenerative chronic conditions (dementia, cerebrovascular

disease), and cardiovascular conditions (heart failure). This suggests that when identifying patterns of multi-morbidity in care home residents, we may be observing a further progression of those patterns seen previously in the community setting. In studies based exclusively in community dwelling older adults neurological conditions such as dementia, and cerebrovascular disease have not been associated with the cardiovascular-metabolic pattern (Vu, Finch, and Day, 2011; García-Olmos et al., 2012; Islam et al., 2014; Déruaz-Luyet et al., 2017; Guisado-clavero et al., 2018; Machón et al., 2020). However, three studies based outside the UK with mixed samples of community dwelling and institutionalised older adults identified at least one neurological condition in a cardiovascular-metabolic pattern cluster (Schafer et al., 2010; Violán et al., 2019; Marengoni et al., 2020). By incorporating mixed samples it is possible that these studies were identifying a similar cluster of sample members to that seen in the Cardiovascular-Metabolic pattern in the present study. These findings suggest a further progression of the cardiovascular-metabolic pattern of chronic conditions may be the development of neuro-degenerative conditions in the future, however further research is needed to investigate this relationship further, which is discussed in Section 6.7.

Neurological-Psychiatric Pattern

The second overarching pattern identified in multiple studies of community dwelling older adults is the neurological-psychiatric pattern (Schafer et al., 2010; Guisado-clavero et al., 2018; Machón et al., 2020; Marengoni et al., 2020). This pattern involves a cluster of neurological conditions such as dementia, cerebrovascular disease, epilepsy, or Parkinson's disease with depression and mood diseases the most common psychiatric condition included for study (Schafer et al., 2010; Guisado-clavero et al., 2018; Machón et al., 2020; Marengoni et al., 2020). Additional to the neurological and psychiatric conditions included in this pattern, studies based in the community setting also consistently include a range of gastrointestinal, musculoskeletal, and sensory impairment chronic conditions (Schafer et al., 2010; Guisado-clavero et al., 2018; Machón et al., 2020; Marengoni et al., 2020). In studies where no neurological-psychiatric cluster was identified, this combination of neurological conditions with gastrointestinal, musculoskeletal, and sensory impairment chronic

conditions persisted (García-Olmos et al., 2012; Violán et al., 2019).

The High-Neuro-Psychiatric pattern in Section 5.3 was strongly associated with neurological (dementia, epilepsy, cerebrovascular disease), psychiatric (depression and mood disorders, schizophrenia and delusional disorders), gastrointestinal (oesophagus, stomach, and duodenum diseases, colitis, and other digestive diseases), musculoskeletal (osteoporosis, osteoarthritis, dorsopathies), and sensory impairments (blindness and visual impairments, deafness and hearing impairments, glaucoma).

Whilst some combination of all the conditions associated with the High-Neuro-Psychiatric cluster have been identified in studies based in the community setting, these conditions were all strongly associated with a single cluster in the care home setting. This may suggest that the progression of this pattern of multi-morbidity is a further development of a pattern seen in the community setting (Prados-Torres et al., 2014).

The association of dementia and psychiatric conditions such as depression has been noted in previous research, although the relationship may be bi-directional with each condition contributing to the state of the other (Quinn and Dickinson, 2014). Furthermore, symptoms of dementia and cerebrovascular disease include loss of emotional control and increased delirium, which may also increase the risk of an individual being diagnosed with concurrent psychiatric conditions (NHS England, 2023d; NHS England, 2023c).

Several mechanisms explain the presence of sensory impairment, gastrointestinal, and musculoskeletal conditions in the neurological-psychiatric pattern of multi-morbidity observed in the community setting, and now in the care home setting, as a result of the research presented in this thesis.

In a recent systematic review and meta-analysis, visual impairments in older adults was found to be associated with the presence of dementia (Shang et al., 2021). However, this association may be caused by the high prevalence of visual impairments due to age related changes in the eye leading to coincidental co-occurrence of vision problems with dementia in older adults. Processing of visual information in the brain can also be impaired through damage caused by the symptoms of neurological conditions such as dementia, and cerebrovascular disease, so there may be a causal relationship present for a subset of this group of residents (NHS

England, 2023d; NHS England, 2023c). Recent studies in older adults have suggested a relationship between digestive diseases and psychiatric conditions due to a bi-directional gut-brain relationship, which may explain the occurrence of these groups alongside the neurological-psychiatric pattern (Bi et al., 2021; Zheng et al., 2023).

The consistent inclusion of musculoskeletal conditions in the neurological-psychiatric pattern may indicate a causal link between the conditions in the pattern and bone health, fracture risk, or muscle strength. However, a recent review found little evidence to support this association more than co-incident occurrence of osteoporosis with dementia (Downey et al., 2017). It is possible however that reductions in activity as a result of the neurological and psychiatric conditions in the pattern may increase the risk of developing the musculoskeletal conditions.

Conclusion of Cluster Analysis Discussion:

The cluster analyses identified evidence that the Cardiovascular-Metabolic and Neurological-Psychiatric patterns of chronic disease previously observed in community dwelling older adults also occur in care home residents. The continuation of multi-morbidity patterns from the community to care home setting is to be expected as the core underlying chronic conditions would be unchanged during this transition. However, the multi-morbidity patterns observed in this care home sample may represent a further progression of the major patterns when compared to the community setting, which demonstrate the possible progression of the underlying mechanisms present. How these patterns evolve over time, and lead to individuals developing further chronic conditions in future requires further research, which is discussed further in Section 6.7.

Furthermore, while previous studies focused on highly prevalent multi-morbidity combinations, or functionally independent older adults, this analysis was able to show possible links between several rarer neuro-degenerative conditions, which were often not included in previous studies. Finally, while the inclusion of a wide range of chronic health conditions increased the complexity of the analyses and interpretation it also provided a realistic view of multi-morbidity in care home residents.

In conclusion, while the cluster analyses presented in the thesis represent a promising area of study, further research is needed to establish important multi-morbidity patterns in UK care home residents. By identifying these patterns in care home residents, clinicians may be better able to identify the likely progression of multi-morbidity over time, which may provide opportunities to identify candidates for fall-prevention interventions. To ensure the findings are transferable to the UK-wide care home population, the cohorts used in this further research must be drawn from all regions, rather than the single hospital trust that this analysis was based on.

6.3.2 The Role of Chronic Disease and Multi-Morbidity in Determining Fall Count in Care Home Residents

Negative Binomial regression analysis was used to identify whether the patterns of chronic health conditions described in Section 6.3.1 caused a change in falls risk among care home residents in the sample. This analysis further explored whether different clusters differ in their falls risk and addressed two aims of the research presented in Section 1.5.1. Identifying the falls risk associated with different combinations of conditions would allow for the further identification of possible mechanisms, and improve predictive performance of falls risk prediction models intended for use in highly multi-morbid samples. The results of this analysis, presented in Sections 5.4 and 5.5, found the clusters of chronic health conditions were associated with different counts of fall presentations.

The analysis of multi-morbidity patterns identified several novel findings for discussion. First, a gradient of falls risk effects was identified for different patterns of multi-morbidity. Four tiers of effect were identified, with the smallest increases in fall presentations to the ED associated with the Cardiovascular (23.8%) and Low-Cardio-Neuro (26.7%) clusters when compared with the Absence cluster. The second tier of effect was identified for two patterns of neurological-psychiatric multi-morbidity with the Low-Neuro-Psychiatric and High-Neuro-Psychiatric clusters associated with 43.0% and 41.4% increases respectively.

However, while neurological-psychiatric multi-morbidity was identified as having a larger effect on the rate of fall presentations than isolated cardiovascular multi-morbidity, the Cardiovascular-Metabolic cluster exhibited a still larger increase of

63.0% compared to the Absence cluster.

The final tier of effect was identified for the Non-Specific-High-Burden cluster with an associated 117.8% increase in fall presentations when compared to the Absence cluster. However, when analysing the condition make-up of this cluster it may represent a combination of the Neurological-Psychiatric and Cardiovascular-Metabolic patterns.

The identification of a gradient of associations with fall rate between the patterns of chronic disease discussed in Section 6.3.1 represents a novel finding in the multi-morbidity pattern, and falls literature. This is the first time the overarching patterns of chronic disease have been linked with differing fall rates in the care home setting. However, further analysis attempting to explain this gradient of effects using the shortlist of chronic conditions that increase falls risk, identified during the Chapter 3 literature review, found no evidence of specific multiplicative multi-morbidity effects beyond the main condition effects identified.

The core argument of this thesis is that the combination of conditions that make up multi-morbidity matter, which is not reflected by the measures commonly used in the falls risk literature (Morin et al., 2019; Bravo et al., 2021; Gade et al., 2021a; Barik et al., 2022; Jacob et al., 2022; You et al., 2023). As seen in the review of systematic reviews presented in Chapter 3, there are a wide range of chronic health conditions identified as increasing the risk of fall events in older adults. Due to the variety of mechanisms and strengths of relationships present, it is unreasonable to expect the impact of two chronic conditions co-occurring will be the same as another pair of conditions. When this is scaled up to reflect the variety of multi-morbidity presentations in care home residents, incorporating multi-morbidity using a count of chronic conditions, binary thresholds, or standardised index values into these models may represent an over-generalisation of multi-morbidity presentations. This is because these methods do not allow for the complex range of effects when conditions co-occur and present a single estimate for a 'multi-morbidity effect', when in reality multi-morbidity will exhibit a different effect dependent on the conditions present. Therefore, an approach is needed to identify and incorporate the salient multi-morbidity combinations that impact falls risk beyond the individual conditions themselves. The cluster based negative binomial regression results found that

different types of multi-morbidity exhibited large changes in fall rate, which suggest that different combinations of multi-morbidity exhibit different levels of falls risk. This means that the way multi-morbidity will impact an outcome is dependent on the context, or conditions that make up the specific multi-morbidity presentation. If this is the case, attempts to group these different relationships into a single generalised estimate of a multi-morbidity effect will be met with a subsequent reduction in model performance. Therefore, indicating the type of multi-morbidity present will better reflect the underlying data generating process. However, efforts to reflect the multi-morbidity context will come with a trade off in model complexity and reduction in degrees of freedom.

Furthermore, the effect sizes observed in the main effects regression models, presented in Section 5.4.4, were not large enough to explain the changes in falls risk seen in the cluster regression results, presented in Section 5.4.3. This suggests there is a role for multi-morbidity in determining falls risk beyond the individual effects of the chronic conditions present.

The third aim of the research, presented in Section 1.5.1 was the identification of specific combinations of chronic health conditions that change falls risk. The interaction analyses results identified no evidence of multiplicative multi-morbidity effects in the combinations tested. This is in contrast to the findings of the cluster-based regression model, presented in Section 5.4.3, where different multi-morbidity presentations exhibited very different associations with falls. This difference in findings shows that incorporating multi-morbidity into fall risk models is a challenging task because of the range of possible effects that could arise from the combination of conditions. Additionally, even within a single combination of chronic health conditions, there will be a range of risk effects and mechanisms in operation. The combination and degree of condition severity, response to treatment, and polypharmacy effects will also exhibit moderating effects on the actual impact of multi-morbidity on falls risk. This complex array of competing effects and interactions may be the next barrier for falls risk prediction models to overcome, which may unlock their predictive effectiveness in highly multi-morbid samples. I contend that there exists a middle ground to be pursued between highly complex modelling, and the oversimplified measures being used currently (Morin et al., 2019; Bravo et al., 2021;

Gade et al., 2021a; Barik et al., 2022; Jacob et al., 2022; You et al., 2023). In the initial stages of understanding multi-morbidity, being able to identify consistent groups is the first step. Once these groupings can be identified, decision rules for determining group membership need to be defined, which allow the categorisation of new samples without the need for complex cluster analyses. The results of the cluster analyses made some headway in this area, and demonstrated there a relationships between multi-morbidity and falls risk present, which warrant further study.

Once a sample is stratified into the presence of specific multi-morbidity groupings, it may be possible to derive risk effects of these groupings, which allow a model to reflect some of the individuality presented by multi-morbidity effects. A promising direction for this research is network analysis, where chronic condition information is translated into individual combinations of conditions, with a network visualisation used to facilitate understanding of the complexity in these data. This is a growing area in the multi-morbidity literature, which is discussed further in Section 6.7 (Hernández, Reilly, and Kenny, 2019). However, as of yet, no study has linked the groups identified through network analyses to falls risk in care home residents.

This is the first study to draw a link between specific patterns of multi-morbidity and falls risk in care home residents based in the UK. While the findings point to the existence of multi-morbidity effects, no specific mechanisms were identified for why these effects occur beyond the cluster analyses groupings. This research presented a novel combination of multi-morbidity cluster analysis, and associated analysis of falls risk to try and draw a link between the patterns identified and the risk of falls. However, while the clusters of chronic disease presentations had different falls risk associated with them, the attempt to identify which specific combinations were driving this falls risk found limited evidence for why these effects occurred. Therefore, as is discussed in Section 6.7, future research is needed to identify which combinations were motivating the effects seen, why specific multi-morbidity patterns exhibit different risk effects, and how these patterns can be operationalised into future falls risk prediction models. Reflecting the individuality present in multi-morbidity is a key challenge for falls risk prediction models to overcome in order to predict falls in groups with high levels of multi-morbidity.

6.4 Clinical and Policy Implications

The patterns of chronic disease discussed in Section 6.3.1 were found to increase the risk of falls in care home residents as seen in Section 5.4.3. The specific groups of chronic health conditions identified in this research need to be externally validated in other samples of care home residents before the findings can be considered robust. However, identifying the groups of chronic diseases that most regularly co-occur facilitates the development of interventions that target the specific mechanisms at play in each group to reduce the risk of a fall. Furthermore, identification of these patterns in care home residents may provide an opportunity for practitioners to better understand the likely trajectory of multi-morbidity and put in place preventative measures in the community setting. Such proactive management could lead to a reduction in falls in future cohorts of care home residents. A large effect was identified for the cluster that showed evidence of the cardiovascular-metabolic syndrome pattern of disease. Further patterns that increased falls risk were identified in a cluster of neurological conditions, a further group emulating the neurological-psychiatric condition pattern, and finally a group of cardiovascular conditions. These groups of conditions are likely to impact falls through a variety of mechanisms, meaning non-targeted intervention may not have a consistent effect on the prevention of these falls.

In the context of the ageing population in the UK, where the number of care home residents is expected to grow substantially in a resource constrained setting, improving the efficiency with which healthcare resources are allocated in this patient group is essential (Kingston et al., 2018; Office for National Statistics, 2018; Wittenberg and Hu, 2015). As discussed in Section 1.2.1, falls in older adults are the leading cause of trauma presentations to the emergency department in this group. Falls risk prediction models offer an avenue to identify people at a high risk of falls based on routinely collected information stored in the electronic health record (EHR). However, because people are living longer with multi-morbidity, any fall risk prediction model for use in older adult populations needs to be effective in highly multi-morbid populations.

By improving the accuracy of fall risk prediction models in highly multi-morbid

patients, system level cost savings can be elicited through the better targeting of fall prevention interventions. Furthermore, because the mechanisms through which multi-morbidity impacts falls differ between multi-morbidity presentations, identifying those presentations at the highest risk of falls provides the opportunity for better targeted intervention in these patient groups. Additionally, by understanding the progression multi-morbidity patterns over time, clinicians may be better able to mitigate these increases in falls risk sooner and prevent future falls through review of medications and preventative measures.

If a sufficiently high performing falls risk prediction model can be developed, people at a high risk of falls can be flagged for review by clinicians. This process has two major potential benefits. First, it could lead to the identification of high falls risk people sooner and at a reduced time cost. Second, by leveraging information contained in the EHR through risk prediction models, clinician decisions surrounding treatment and interventions could be based on a range of in-depth information, which may improve the overall the effectiveness of the resulting decisions.

Incorporating the individuality of multi-morbidity presentations into prediction models may aid in improving their predictive performance in care home residents. However, in order to distinguish between different risk groups, a reliable and comparable source of data is required for the care home setting. For this reason, a key policy development would be to emulate the Minimum Data Set (MDS), which is used in the USA and Canada. The MDS is a standardised questionnaire completed for all residents of nursing homes certified by Medicare or Medicaid in the USA (Saliba and Buchanan, 2008b). The information contained in the MDS covers demographic information, functional or activity based impairments, observations of psychological and physical functioning, medications, treatments and therapies, and records of health conditions present. The development of a similar data set in the UK is ongoing through the developing research resources and minimum data set for care homes' adoption and use (DACHA) study (Goodman, 2019; Burton et al., 2022). Such a data set, bringing together multiple sources of information on care home residents, will provide an invaluable opportunity to further explore the relationships identified in this research.

By having a such a repository of information, many of the issues faced in this thesis surrounding the identification of falls at the care home level, and reliable reporting of chronic diseases could be overcome. Furthermore, additional observations such as Activities of Daily Living (ADL), Timed Up and Go (TUG) test observations, and other functional test results could be recorded, which are beneficial for the development of predictive models in UK populations. Finally, the development of a standardised data reporting structure would improve the feasibility of rolling out a falls risk prediction model in practice. This recommendation is discussed further in Section 6.7.

Predictive modelling in patients susceptible to falls is an area of research that will become more important as the population ages. However, the existing challenges of clinical interpretability, trustworthiness of model predictions, and the complexity of the underlying mechanisms that cause falls means there are unavoidable trade offs in developing a tool for use in clinical practice. Bodies such as National Institute for Health and Clinical Excellence (NICE) have taken a moderate approach, demanding that clearly effective models are demonstrated through external validation before such a prediction model would be considered for clinical practice (Health & Care Excellence, 2013). However, research into the comparative predictive performance of standard practice would be informative to indicate where predictive modelling could be useful. Understanding the sensitivity and specificity of practitioner predictions in different situations would indicate where a predictive model would be most useful, and allow more targeted development of such models. Rather than attempting to replace the clinician in the decision, the statistical model must always be a tool that aids the clinician in their decision. However, for this to be done effectively, the role of clinical intuition, and degree of evidence used by practitioners needs to be understood.

6.5 Limitations of the Research

The research presented in this thesis had several limitations. Where possible, steps were taken to mitigate these limitations and their effects on the analyses.

1. **Identification of Falls**

The first limitation related to the identification of falls in the available data. Being unable to identify falls at the care home level would have meant that a proportion of falls by sample members will have gone unobserved. To overcome this, falls were identified at the emergency department level instead. However, no information was available regarding the cause of an attendance. To mitigate this, I applied a definition of trauma presentations to the emergency department by the cohort members, which was used as a proxy for falls. This definition is reasonable because a high proportion of trauma presentations are caused by falls in older adults (Sterling, O'Connor, and Bonadies, 2001; Labib et al., 2011; Benhamed et al., 2023). However, the use of such a proxy definition to identify falls may have led to a degree of mis-classification in the outcome variable used during the regression analyses.

Furthermore, by identifying falls at attendance to the emergency department level, falls that did not require treatment or transport to hospital were not recorded in this dataset. This would have had two effects on the analyses. First, explanatory variables which raise the risk of injurious falls, would have been more likely to be identified in this analysis. Additionally, sample members with a higher condition burden may have been more likely to be transferred to hospital due to the seriousness of their condition, which would raise the likelihood their fall was observed in the data and artificially strengthen the identified relationship. Finally, it is possible that some sample members who fell bypassed the emergency department entirely and were admitted directly, meaning their fall would have been unrecorded. While this measure of falls is not perfect for identifying all falls, it does give an accurate picture of emergency department service use resulting from falls in this patient group.

2. **ICD-10 Code Availability**

The second group of limitations refer to the use of two digit ICD-10 codes collected through standardised data reporting as a result of an inpatient stay in

one of the County Durham and Darlington Foundation Trust (CDDFT) hospitals to identify the chronic disease records. This introduced several limitations into the analysis. First, chronic condition information for sample members without an inpatient attendance during the study period could not be observed. However, the three year study period, and high level of healthcare resource use in care home residents, meant the effect of this is expected to be small. This missing data adds to the mis-classification problem, with the effect on the analysis being to reduce the clarity of effects observable through regression analysis and a subsequent increase in standard errors. This effect would also lead to a widening of confidence intervals, and underestimation of effect sizes of explanatory variables on the outcome. Additionally only the two digit ICD-10 codes were available for use, rather than the more specific 3-digit ICD-10 codes, which are recommended by the original authors of the chronic health condition groups used in this thesis (Calderón-Larrañaga et al., 2017). Using the less specific codes introduced a level of miss-classification into the analysis, which would have impacted the accuracy and precision of the results during the regression analysis.

The final problems with the use of ICD-10 codes for identifying chronic disease, relate to the codes themselves. First, the use of ICD-10 codes collected in inpatient stays means the research was reliant on standardised data reporting. In the case of patients with a high degree of multi-morbidity, it is possible that the only conditions reported on their inpatient record would be those directly related to their stay. Furthermore, using ICD-10 codes in isolation to indicate chronic disease, is that the codes themselves do not give any indication of condition severity, management, or treatment. This means that the effects estimated during the regression analysis may not include the full array of possible presentations, severity, and management of a condition. This is important because multiplicative multi-morbidity effects may only result from more severe, or unmanaged, chronic condition states. The expected effect of this conflation of possible disease states is wider confidence intervals and biased estimates of effect by explanatory variables. Finally, the risk of falls presented by a chronic condition will be directly related to how that condition is being treated and

managed.

3. **Accounting for Frailty in Regression Models**

   The next limitation to be discussed is the non-inclusion of a measure of frailty in the negative binomial regression model, which used the clusters as input data. By not including a measure of frailty in this regression, the effect attributed to the clusters may have been partially due to the effect of frailty rather than multi-morbidity effects. Additionally, because the outcome was fall-related ED attendances, sample members in a frail condition would have been more likely to sustain an injury, which required treatment in hospital, meaning they were more likely to be observed in the outcome. The measure of frailty was not included in this model because much of the input data used in deciding the individual level of frailty was the presence of several chronic conditions and age, which were all captured in the regression already, meaning the score would have been highly inter-correlated with other explanatory variables in the regression.

4. **Inclusion of Severity Information**

   An additional limitation in the analyses is the lack of severity information regarding the chronic health conditions identified through inpatient records. This is important because the severity and presence of specific symptoms related to a single chronic health condition may have large consequences for how that chronic condition relates to falls. For example, dementia was present in 40.4% of the sample, however wandering behaviour would only have been present in a sub group of these sample members with dementia. This means the dementia indicator in the regression models is capturing multiple sub groups, with dementia contributing differently to overall falls risk in each. For future models to effectively identify falls risk in care home residents, they will need to include indications of condition severity, or the presence of specific falls risk increasing symptoms to differentiate between subgroups within the chronic health condition indicators.

5. **Interaction Analyses**

The final limitations to be discussed in this section relate to the method used for the interaction analyses. This approach identified the most prevalent conditions by cluster over an O/E ratio threshold for inclusion in regression models as interaction effects. This approach was taken to reduce down the number of possible combinations needed to test in the analyses; however this also led to several problems.

First, by identifying these effects separately within each cluster, what made a particular cluster have a higher risk effect compared to other clusters was not identified. Additionally, the choice of interactions often did not reflect the unique attributes of the specific cluster, and by using only pairwise combinations the complexity inherent in the multi-morbidity data was not being reflected in the analyses. In addition, O/E ratios are calculated relative to the prevalence of a condition in the whole sample, meaning conditions with lower prevalence were often attributed with very high O/E ratios, while it was much harder for a cluster to be differentiated based on a highly prevalent disease such as dementia, which was not associated with any cluster through O/E ratios, despite having over 50% prevalence in many clusters. This meant the decision rules used to select chronic conditions for interaction did not always identify the most salient combinations for a particular cluster and reduced the overall quality of the findings. Finally, adopting a data-driven systematic approach approach meant the interaction of conditions was not based on predetermined clinically viable mechanisms, which were expected to impact falls.

A data-driven approach was adopted for this stage due to time and resource constraints and the need to develop some way of selecting conditions to interact in a re-producible way. The combination of these problems and the limitations mentioned previously may have contributed to much of the insignificance of many of the interaction effects in the analysis in Section 5.5.

The research conducted in this thesis had to contend with challenging circumstances regarding the availability of data, and fallout from the COVID-19 pandemic. However, as is discussed in Section 6.6 these challenges motivated creative solutions that led to novel contributions to knowledge being derived.

6.6 Novel Contributions to Knowledge

This thesis makes an important contribution to new knowledge in several ways surrounding the methods used, reviews of the literature, and results found. The reviews of literature in Chapters 2 and 3 addressed unanswered questions on topics within Traumatic Brain Injuries, and falls. This led to the identification of key themes and conclusions, which were applied in developing the research methodology and impacted the direction of the analyses.

The Chapter 2 review introduced novelty by comparing the effectiveness of machine learning approaches for mortality prediction following TBI. This review identified that there was no objectively best model developed, due to differences in underlying samples and information available. Furthermore, the review identified a need to improve the transparency and reporting of performance metrics and decisions made in training such that models developed could by compared in future. This review impacted the direction of the research presented in the thesis by prompting a change in direction towards exploring falls in older adults. This was due to falls causing the majority of TBIs in older adults and age was a consistently important factor in the studied ML models regarding mortality following TBI in the Chapter 2 review (Lawrence et al., 2016).

The Chapter 3 review was the first to identify conditions associated with changes in falls risk from the range of reviews published since 2000 using a review of reviews format. While individual reviews have drawn conclusions, there has been no previous attempt to draw their conclusions together in a single summary of the literature surrounding falls risk increasing chronic health conditions. This was a gap that the Chapter 3 review filled, with the resulting shortlist of chronic health conditions used as the basis for the regression models using individual chronic conditions in Sections 5.4 and 5.5.

This is the first empirical research study to identify clusters of chronic health conditions in UK based care home residents, and link these clusters with fall presentations to the emergency department. Additionally, this was undertaken using a novel combination of methods involving the three step analyses of MCA pre-processing of

chronic condition data, cluster analysis to form groups of sample members with similar chronic condition records, and inclusion of cluster membership as an explanatory variable for regression with falls as an outcome.

Throughout these analyses, the resulting models demonstrated the inconsistent effects of multi-morbidity on falls in older adults. This means that there is possible scope for improving falls risk prediction models that are used in highly multi-morbid populations.

As a result of the novel combination of methods, estimates for the effect of membership in patterns of multi-morbidity were derived, which allowed a comparative analysis to identify those patterns that most impacted falls in care home residents. This is a new application of existing approaches to the question of falls risk in older adults, which can be replicated in future studies.

This approach could also be applied to new data sets because it is based on the commonly available two digit ICD-10 codes. Additionally, in comparison with previous studies focused on multi-morbidity in older adults, the research presented in this thesis adds to the body of evidence through the number of conditions included in the analysis, and the full picture these provided surrounding multi-morbidity in UK care home residents.

The research presented in this thesis also involved the use of a unique data set, developed as part of the HDRUK 'Learning Care Homes: Continuous improvement of structured referrals' project (Saliba and Buchanan, 2020). By using a data set based covering an entire NHS trust, robust analyses were developed, which avoided problems arising from single centre, or provider bias.

6.7 Recommendations for Future Research

Many of the limitations encountered in the thesis relate to data quality and usability. Therefore, the primary recommendation for future research into falls in the care home is the development of a standardised data set, which is consistent across care homes. Development of a minimum data set, similar to that seen in the US, for use in care homes would allow research into key geriatric syndromes such as falls through

accurate identification at the care home level without the need for researchers to develop bespoke data sets. Moving away from bespoke data sets into a standardised environment would be hugely beneficial to the falls risk prediction literature, as a common data set could be used to train, test, and apply these models in practice, which will inevitably lead to improvements in performance and usability. This is especially important when researching multi-morbidity patterns in data using cluster algorithms, which are highly susceptible to changes in the underlying data, which makes comparison of results between studies challenging. Furthermore, if a standardised data set is compiled in the community setting, research into the progression of multi-morbidity and health outcomes during the transition between the community and care home settings setting would be made possible. Additionally, this standardised data set could be a source of more accurate records such that the approaches taken in this thesis could be repeated using more reliable base data across multiple regions of the UK, rather than a single NHS trust.

An opportunity for future research also exists in the linking of multi-morbidity patterns in the community setting with fall rates. Using a similar approach to that used in this thesis would also allow for further investigation of how chronic disease patterns progress between the community and care home settings and whether relationships with falls change during this transition.

Additionally, authors should move away from treating 'multi-morbidity' as a single effect that can be captured in a single measure or estimate, rather multi-morbidity is context specific and will have different effects dependent on the conditions, and treatments an individual experiences. Therefore further research is needed into how the individual complexity in multi-morbidity can be incorporated into falls risk models in such a way that the mechanisms are methodologically valid, and understandable to clinicians.

In summarising this individual level complexity, further study is needed to identify combinations of chronic conditions that relate to falls in care homes, such that the findings from this research can be validated in external populations. Identifying consistent combinations of chronic conditions is essential for improving the performance and usability of models with larger prediction horizon. This is especially important in the context of an ageing population in a resource constrained

healthcare setting. Pursuing approaches that allow individuals to be members of multiple multi-morbidity groupings rather than attempting to define them into single groups is desirable to reflect the individual heterogeneity in the presentation of multi-morbidity. Network analysis allows effective visualisation of multi-morbidity patterns, and using this in conjunction with cluster approaches may provide further opportunities to explore the role of multi-morbidity in falls risk (Hernández, Reilly, and Kenny, 2019).

Finally, the movement towards tailored individual risk scores based on specific chronic conditions on record and observations in electronic health records would allow for a greater degree of personalisation in the management of falls risk. However, research evaluating the effectiveness of combinations tying together the output from falls risk models with specific interventions and mitigations to allow for this personalisation will also be required in future.

6.8 Concluding Comments

6.8.1 Reflections on the PhD Process

The PhD process has been a challenging endeavour, but not for the reasons I expected, and I will carry forward several key lessons into my future career. The challenges discussed in this section have demonstrated to me the importance of seeking out advice, finding roles that allow me to work with others, and believing that my abilities and work rate are enough to be proud of. I see now that the PhD process is primarily intended to develop an early career researcher, as well as the pursuit of actionable research findings. While it is difficult to distil the lessons of the last three years into a narrative, the discussion that follows will attempt to draw together the experiences that shaped the PhD experience and make clear the lessons I have drawn from each.

The isolationist tendency brought on during the pandemic had a subtle effect on my desire to ask for help. Starting the PhD during this period meant I did not see another PhD student struggling, having difficulties, or seeking the advice throughout the first year. Instead I saw only progress during infrequent group meetings. I think now that not seeing others asking for help, which they undoubtedly were,

influenced a sub-conscious calculation that asking for help was an admission of failure in some way. Laid on top of this was an assumption that failure was an absolute threshold beyond which lies little prospect of a rewarding career. This unwillingness to reach out for help, and aversion to asking for advice greatly hampered me through the course of the PhD, and I definitely experienced much unnecessary stress around expectations, and personal failings throughout the three years. This ate up much of the bandwidth in my mind at times, however I do not regret these periods, for they helped me see the behaviours that I can change that made me more resilient in the long term. Moving to a place where I see seeking help and advice as a sign of strength through showing the desire to improve is now a major goal in my personal development. Towards this aim I intend to set measurable goals, which will help me break out the habit of isolationism in work. I see this development in mindset as one of the major victories during the PhD process, and can see this change will pay dividends throughout my career.

The second reflection I will make on the last three years is linked to the unwillingness to ask for help discussed previously. I think now that some of the reason I was unwilling to ask for help related to my views at the time of my abilities as a researcher and how hard I was working. Seeing my own view as a single opinion, rather than a measure of absolute fact has been a major development through the PhD. The effects of these negative personal views meant I was very susceptible to focusing on the negative aspects of feedback, decisions taken during the research, and results rather than seeing these things on a continuum. By not seeing the positive attributes of what I had done, my overall view of the projects progress was often completely at odds to those around me. I categorised everything in terms of either success or failure, and failed to see the limitations of this. Reflecting on this there are several lessons to be learned, which can be applied in future. First, it is clear I was caught in the simultaneous grip of imposter syndrome, perceived expectations of others, high personal expectations of myself, and a degree of perfectionism. I believe the root of all these challenges come from a fear of exposing myself to uncertainty. In order to overcome this I intend to have a focus on actively seeking out and reflecting upon feedback until this becomes a habit. Furthermore, within this I need to focus on seeing both the positive and negative aspects of feedback with an equal

degree of weight and recognise my own tendency to focus on the negative. Finally, I need to recognise all outcomes exist on a spectrum and accept that my best effort is enough, rather than tying my view of the effort to the outcome that occurred. I think these lessons will help me develop a healthier approach to work in future if applied effectively.

The final reflection on the PhD process to be discussed is my approach to problem solving, and how I manage this process internally. Throughout the PhD my approach to problem solving was to consume myself with the details, and obsess until a solution was found. However, this often led to situations where only I understood the reasoning for a solution, or where errors occurred, I had to re-approach the problem entirely. I have begun to accept that problems experienced in work do not reflect negatively on my abilities. Furthermore, focusing on why the problem occurred and blaming myself does not solve problems. To overcome these tendencies I have accepted that when a problem occurs I have the responsibility to handle it effectively, then put in place steps to ensure it does not happen again rather than wasting effort and time blaming myself for not forseeing the difficulties. By changing my focus from things I cannot control such as past actions, into how I act in the moment and work towards finding a solution I am more efficient and positive as a worker, whilst also being more sustainable in my approach to work.

While my PhD experience has been far from linear, the lessons learned, challenges overcome, and mindset shift has been worth the time spent irrespective of whether I attain the degree. Divorcing the success of the project from the results of the research was the single poignant change, that allowed me to grow as a researcher, and gave me the space to unpick the lessons discussed in this section, which will support me throughout my career.

6.8.2 Overall Conclusion

The research presented in this thesis is directed towards identifying patterns of multi-morbidity in care home residents, and how these patterns relate to falls. Despite failing to identify specific combinations of multi-morbidity that impact falls risk through interaction analyses, the results still provide evidence that the impact of multi-morbidity on falls is context specific. Four patterns of multi-morbidity were

identified in the solution, which all raised the risk of fall presentations to the emergency department. These were cardiovascular conditions with and without associated metabolic syndrome conditions, a late stage development of the cardiovascular-metabolic syndrome pattern with associated neurological components. Further groupings related to neuro-degenerative conditions were identified without a psychiatric component. Finally, a neurological-psychiatric pattern was identified with sensory, digestive, and musculoskeletal components beyond those seen in the community setting. While further research is needed in care home populations to validate these patterns, their being broadly in accordance with previous findings from the community setting provide support for their validity. Fall risk prediction models used in highly multi-morbid samples need to adopt an approach to incorporate some of the individuality present in multi-morbidity rather than adopting over-generalised approaches. Future research needs to be directed towards identifying the impacts of multi-morbidity on falls due to the expansion of multi-morbidity resulting from the ageing population. Furthermore, the development of minimum standardised data sets across care homes in the UK would allow further development of falls risk prediction models in this population.

Bibliography

Abujaber, Ahmad et al. (2020). "Prediction of in-hospital mortality in patients on mechanical ventilation post traumatic brain injury: machine learning approach". In: *BMC Medical Informatics and Decision Making* 20.336. DOI: 10.1186/s12911-020-01363-z.

Afrin, Nadia et al. (2016). "Multimorbidity predicts falls differentially according to the type of fall in postmenopausal women". In: *Maturitas* 91, pp. 19–24. DOI: 10.1016/j.maturitas.2016.05.004.

Alsulaim, Hatim A et al. (2017). "Conscious status predicts mortality among patients with isolated traumatic brain injury in administrative data". In: *American Journal of Surgery* 214.2, pp. 207–210. DOI: 10.1016/j.amjsurg.2016.07.012.

Amorim, Robson Luis et al. (2020). "Prediction of Early TBI Mortality Using a Machine Learning Approach in a LMIC Population". In: *Frontiers in Neurology* 10.1366, pp. 1–9. DOI: 10.3389/fneur.2019.01366.

Angulo, Javier et al. (2020). "Physical activity and exercise: Strategies to manage frailty". In: *Redox biology* 35.101513. DOI: 10.1016/j.redox.2020.101513.

Armin, Sean S., Austin R. T. Colohan, and John H. Zhang (2006). "Traumatic subarachnoid hemorrhage: our current understanding and its evolution over the past half century". In: *Neurological research* 28.4, pp. 445–452. DOI: 10.1179/016164106X115053.

Armstrong, C et al. (2005). "Occurrence and risk factors for falls in rheumatoid arthritis". In: *Annals of the rheumatic diseases* 64.11, pp. 1602–1604. DOI: 10.1136/ard.2004.031195.

Atinga, Angela et al. (2018). "Trauma in the elderly patient". In: *British Journal of Radiology* 91.1087, pp. 1–10. DOI: 10.1259/bjr.20170739.

Baker, Carl (2020). *NHS Key Statistics, England, February 2020*. Tech. rep. House of Commons Library. URL: https://researchbriefings.files.parliament.uk/documents/CBP-7281/CBP07281-feb2020.pdf.

Barik, Manish et al. (2022). "Is multimorbidity associated with higher risk of falls among older adults in India?" In: *BMC Geriatrics* 22.1, pp. 1–8. DOI: 10.1186/s12877-022-03158-5.

Barnett, Karen et al. (2012). "Epidemiology of multimorbidity and implications for health care, research, and medical education: A cross-sectional study". In: *The Lancet* 380.9836, pp. 37–43. DOI: 10.1016/S0140-6736(12)60240-2.

Bath, P.A. and K. Morgan (1999). "Differential risk factor profiles for indoor and outdoor falls in older people living at home in Nottingham, UK". In: *European Journal of Epidemiology* 15.1, pp. 65–73. DOI: 10.1023/A:1007531101765.

Benhamed, Axel et al. (2023). "Epidemiology, injury pattern and outcome of older trauma patients: A 15-year study of level-I trauma centers". In: *PloS one* 18.1, e0280345. DOI: 10.1371/journal.pone.0280345.

Bergen, Gwen, Mark R. Stevens, and Elizabeth R. Burns (2016). "Falls and Fall Injuries Among Adults Aged $\geq$ 65 Years — United States, 2014". In: *Morbidity and Mortality Weekly Report* 65.37, pp. 993–998. ISSN: 0149-2195. DOI: 10.15585/mmwr.mm6537a2.

Beynon, Caryl et al. (2011). "The cost of emergency hospital admissions for falls on snow and ice in England during winter 2009/10: A cross sectional analysis". In: *Environmental health* 10.1, p. 60. DOI: 10.1186/1476-069X-10-60.

Bi, Yu-Han et al. (2021). "The relationship between chronic diseases and depression in middle-aged and older adults: A 4-year follow-up study from the China Health and Retirement Longitudinal Study". In: *Journal of affective disorders* 289, pp. 160–166. DOI: 10.1016/j.jad.2021.04.032.

Bishop, Christopher M (2006). *Pattern recognition and machine learning*. Information science and statistics. New York: Springer. ISBN: 9780387310732.

Blain, H. et al. (2016). "A comprehensive fracture prevention strategy in older adults: the European Union Geriatric Medicine Society (EUGMS) statement". In: *Aging clinical and experimental research* 28.4, pp. 797–803. DOI: 10.1007/s40520-016-0588-4.

Bloch, Frederic et al. (2013). "Estimation of the risk factors for falls in the elderly: Can meta-analysis provide a valid answer?" In: *Geriatrics and Gerontology International* 13.2, pp. 250–263. DOI: 10.1111/j.1447-0594.2012.00965.x.

Bloch, Frédéric et al. (2011). "Psychotropic Drugs and Falls in the Elderly People: Updated Literature Review and Meta-Analysis". In: *Journal of Aging and Health* 23.2, pp. 329–346. DOI: 10.1177/0898264310381277.

Boyce, Richard D. et al. (2022). "Falls prediction using the nursing home minimum dataset". In: *Journal of the American Medical Informatics Association* 29.9, pp. 1497–1507. DOI: 10.1093/jamia/ocac111.

Boyd, Rebecca and Judy A. Stevens (2009). "Falls and fear of falling: Burden, beliefs and behaviours". In: *Age and Ageing* 38.4, pp. 423–428. DOI: 10.1093/ageing/afp053.

Bravo, Jorge et al. (2021). "Development and validation of a continuous fall risk score in community-dwelling older people: an ecological approach". In: *BMC Public Health* 21, pp. 1–11. DOI: 10.1186/s12889-021-10813-w.

Brewster, M., C. O'Keeffe, and S. Mason (2019). "027 The utilisation of urgent emergency care services by older care home residents in the UK and the subsequent impact on emergency departments". In: *Emergency medicine journal* 36.12. DOI: 10.1136/emermed-2019-RCEM.27.

British Geriatrics Society (2020). *Managing the COVID-19 Pandemic in care homes: Good Practice Guide*. Tech. rep. British Geriatrics Society.

— (2023). *Smarter data, better care: Empowering care homes to use data to transform quality of care*. URL: https://www.bgs.org.uk/CareHomesData.

British Medical Association (2022). *NHS hospital beds data analysis*. URL: https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/pressures/nhs-hospital-beds-data-analysis.

Brotfain, Evgeni et al. (2016). "Neuroprotection by Estrogen and Progesterone in Traumatic Brain Injury and Spinal Cord Injury". In: *Current neuropharmacology* 14.6, pp. 641–653. DOI: 10.2174/1570159x14666160309123554.

Brown, Cynthia J. et al. (2009). "The Underrecognized Epidemic of Low Mobility During Hospitalization of Older Adults". In: *Journal of the American Geriatrics Society* 57.9, pp. 1660–1665. DOI: 10.1111/j.1532-5415.2009.02393.x.

Bunn, Jonathan Graeme et al. (2019). "Urgent care axis for the older adult: where is best to target interventions?" In: *Emergency medicine journal: EMJ* 36.1, pp. 22–26. DOI: 10.1136/emermed-2018-207505.

Burton, Jennifer Kirsty et al. (2021). "Closing the UK care home data gap – methodological challenges and solutions: Closing the UK care home data gap". eng. In: *International journal of population data science* 5.4, p. 1391. DOI: 10.23889/ijpds.v5i4.1391.

Burton, Jennifer Kirsty et al. (2022). "Developing a minimum data set for older adult care homes in the UK: exploring the concept and defining early core principles". In: *The Lancet. Healthy longevity* 3.3, e186–e193. DOI: 10.1016/S2666-7568(22)00010-1.

Byrne, Christopher D and Sarah H Wild (2011). *The metabolic syndrome*. 2nd edition. Wiley-Blackwell. ISBN: 1-4443-4730-6. DOI: 10.1002/9781444347319.

Calderón-Larrañaga, Amaia et al. (2017). "Assessing and Measuring Chronic Multimorbidity in the Older Population: A Proposal for Its Operationalization". In: *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 72.10, pp. 1417–1423. DOI: 10.1093/gerona/glw233.

Caliński, T. and J. Harabasz (1974). "A Dendrite Method For Cluster Analysis". In: *Communications in Statistics* 3.1, pp. 1–27. DOI: 10.1080/03610927408827101.

Cameron, A. C. and P. K. Trivedi (1998). *Regression analysis of count data*. Econometric Society monographs ; 30. Cambridge University Press. ISBN: 1-316-08502-3.

Carayon, Pascale et al. (2017). "EHR-related medication errors in two ICUs". In: *Journal of healthcare risk management* 36.3, pp. 6–15. ISSN: 1074-4797. DOI: 10.1002/jhrm.21259.

Care Inspectorate (2016). *Managing Falls and Fractures in Care Homes for Older People – good practice resource, Revised edition*. Tech. rep. Care Inspectorate and NHS Scotland. URL: https://www.careinspectorate.com/images/documents/2737/2016/Falls-and-fractures-new-resource-low-res.pdf.

Cella, Alberto et al. (2020). "Development and validation of a robotic multifactorial fall-risk predictive model: A one-year prospective study in community-dwelling older adults". In: *PLoS ONE* 15.6, pp. 1–22. DOI: 10.1371/journal.pone.0234904.

Charlson, Mary et al. (1994). "Validation of a combined comorbidity index". In: *Journal of clinical epidemiology* 47.11, pp. 1245–1251. DOI: 10.1016/0895-4356(94)90129-5.

Charlson, Mary E et al. (1987). "A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation". In: *Journal of chronic diseases* 40.5, pp. 373–383. DOI: 10.1016/0021-9681(87)90171-8.

Chawla, N. V. et al. (2002). "SMOTE: Synthetic Minority Over-sampling Technique". In: *Journal of Artificial Intelligence Research* 16. DOI: 10.1613/jair.953.

Clegg, Andrew et al. (2013). "Frailty in elderly people". In: *The Lancet* 381.9868, pp. 752–762. DOI: 10.1016/S0140-6736(12)62167-9.

Clegg, Andrew et al. (2016). "Development and validation of an electronic frailty index using routine primary care electronic health record data". In: *Age and ageing* 45.3, pp. 353–360. DOI: 10.1093/ageing/afw039.

Coker, Robert H et al. (2015). "Bed rest promotes reductions in walking speed, functional parameters, and aerobic fitness in older, healthy adults". In: *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 70.1, pp. 91–96. DOI: 10.1093/gerona/glu123.

Cooper, Rosie (2017). "Reducing falls in a care home". In: *BMJ quality improvement reports* 6.1. DOI: 10.1136/bmjquality.u214186.w5626.

Crilly, Julia et al. (2008). "Predictive outcomes for older people who present to the emergency department". In: *Australasian emergency nursing journal* 11.4, pp. 178–183. DOI: 10.1016/j.aenj.2008.07.002.

Dawes, Aaron J et al. (2015). "Intracranial pressure monitoring and inpatient mortality in severe traumatic brain injury: A propensity score-matched analysis". In: *The journal of trauma and acute care surgery* 78.3, pp. 492–502. DOI: 10.1097/TA.0000000000000559.

Deandrea, Silvia et al. (2010). "Risk Factors for Falls in Community-dwelling Older People: A Systematic Review and Meta-analysis". In: *Epidemiology (Cambridge, Mass.)* 21.5, pp. 658–668. DOI: 10.1097/EDE.0b013e3181e89905.

Denfeld, Quin E. et al. (2022). "Preventing and Managing Falls in Adults With Cardiovascular Disease: A Scientific Statement From the American Heart Association". In: *Circulation: Cardiovascular Quality and Outcomes* 15.6, E000108. DOI: 10.1161/HCQ.0000000000000108.

Department of Health (2009). *Falls and Fractures Effective interventions in health and social care*. Tech. rep. URL: https://laterlifetraining.co.uk/wp-content/uploads/2011/12/FF_Effective-Interventions-in-health-and-social-care.pdf.

Department of Health and Social Care (2021). *Integration and innovation: working together to improve health and social care for all*. URL: https://www.gov.uk/government/publications/working-together-to-improve-health-and-social-care-for-all/integration-and-innovation-working-together-to-improve-health-and-social-care-for-all-html-version.

Déruaz-Luyet, Anouk et al. (2017). "Multimorbidity and patterns of chronic conditions in a primary care population in Switzerland: A cross-sectional study". In: *BMJ Open* 7.6, pp. 1–10. DOI: 10.1136/bmjopen-2016-013664.

Dormosh, Noman et al. (2022). "Development and Internal Validation of a Risk Prediction Model for Falls Among Older People Using Primary Care Electronic Health Records". In: *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 77.7, pp. 1438–1445. DOI: 10.1093/gerona/glab311.

Downey, Candice L. et al. (2017). "Dementia and osteoporosis in a geriatric population: Is there a common link?" In: *World journal of orthopedics* 8.5, pp. 412–423. DOI: 10.5312/wjo.v8.i5.412.

Duprey, Matthew S. et al. (2023). "Development and validation of the fall-related injury risk in nursing homes (INJURE-NH) prediction tool". In: *Journal of the American Geriatrics Society* 71.6, 1851—1860. DOI: 10.1111/jgs.18277.

Everitt, Brian, Sabine Landau, and Morven Leese (2010). *Cluster Analysis*. 5th edition. Chicester: Wiley. ISBN: 1-280-76795-2.

Feldman, Fabio and Habib Chaudhury (2008). "Falls and the physical environment: A review and a new multifactorial falls-risk conceptual framework". In: *Canadian journal of occupational therapy* 75.2, pp. 82–95. DOI: 10.1177/000841740807500204.

Fernando, Eresha et al. (2017). "Risk factors associated with falls in older adults with dementia: A systematic review". In: *Physiotherapy Canada* 69.2, pp. 161–170. DOI: 10.3138/ptc.2016-14.

Fleming, Jane, Fiona E Matthews, and Carol Brayne (2008). "Falls in advanced old age: recalled falls and prospective follow-up of over-90-year-olds in the Cambridge City over-75s Cohort study". In: *BMC Geriatrics* 8.6, pp. 1–11. DOI: 10.1186/1471-2318-8-6.

Florence, Curtis S. et al. (2018). "Medical Costs of Fatal and Nonfatal Falls in Older Adults". In: *Journal of the American Geriatrics Society* 66.4, pp. 693–698. DOI: 10.1111/jgs.15304.

Foguet-Boreu, Quintí et al. (2015). "Multimorbidity patterns in elderly primary health care patients in a South Mediterranean European region: A cluster analysis". In: *PLoS ONE* 10.11, e0141155. DOI: 10.1371/journal.pone.0141155.

Fontoura Solla, Davi J. et al. (2020). "Incremental Prognostic Value of Coagulopathy in Addition to the Crash Score in Traumatic Brain Injury Patients". In: *Neurocritical Care* 34.1, pp. 130–138. DOI: 10.1007/s12028-020-00991-7.

Foster, David (2023). *Delayed hospital discharges and adult social care*. Tech. rep. UK Parliament. URL: https://commonslibrary.parliament.uk/delayed-hospital-discharges-and-adult-social-care/.

Fried, L. P. et al. (2001). "Frailty in older adults: Evidence for a phenotype". In: *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 56.3, pp. 146–156. DOI: 10.1093/gerona/56.3.m146.

Fries, Brant E. et al. (May 1997). "Approaching Cross-National Comparisons of Nursing Home Residents". In: *Age and Ageing* 26.suppl$_2$, pp. 13–18. DOI: 10.1093/ageing/26.suppl_2.13.

Gade, Gustav et al. (2021a). "Development of a multivariable prognostic PREdiction model for 1-year risk of FALLing in a cohort of community-dwelling older adults aged 75 years and above (PREFALL)". In: *BMC Geriatrics* 21.1, pp. 1–12. DOI: 10.1186/s12877-021-02346-z.

Gade, Gustav Valentin et al. (2021b). "Predicting falls in community-dwelling older adults: a systematic review of prognostic models". In: *BMJ open* 11.5, e044170. DOI: 10.1136/bmjopen-2020-044170.

Gallagher, Elaine M. and Vicky J. Scott (1997). "The STEPS Project: Participatory Action Research to Reduce Falls in Public Places Among Seniors and Persons with Disabilities". In: *Canadian journal of public health* 88.2, pp. 129–133. DOI: 10.1007/BF03403876.

García-Olmos, Luis et al. (2012). "Comorbidity patterns in patients with chronic diseases in general practice". In: *PLoS ONE* 7.2, e32141. DOI: 10.1371/journal.pone.0032141.

Garner, Alex et al. (2022). "1583 Impact of digital technology in care homes on Emergency Department attendances". In: *Emergency medicine journal: EMJ* 39, A974–A974. DOI: 10.1136/emermed-2022-RCEM2.22.

Garner, Alex et al. (Feb. 2024). "The impact of digital technology in care homes on unplanned secondary care usage and associated costs". In: *Age and Ageing* 53.2, afae004. DOI: 10.1093/ageing/afae004.

Garu, A. et al. (2021). "Effect of multimorbidity on fragility fractures in community-dwelling older adults: Shimane cohre study". In: *Journal of Clinical Medicine* 10.15, pp. 1–11. DOI: 10.3390/jcm10153225.

Gastmans, C and K Milisen (2006). "Use of physical restraint in nursing homes: clinical-ethical considerations". In: *Journal of Medical Ethics* 32.3, pp. 148–152. DOI: 10.1136/jme.2005.012708.

Gilbert, Thomas et al. (2018). "Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study". In: *The Lancet* 391.10132, pp. 1775–1782. DOI: 10.1016/S0140-6736(18)30668-8.

Goodman Claire, et al. (2019). *The DACHA Study: Developing resources And minimum data set for Care Homes' Adoption study*. URL: https://arc-eoe.nihr.ac.uk/research-implementation/research-themes/ageing-and-multi-morbidity/dacha-study-developing-resources.

Goto, N. A. et al. (2020). "The association between chronic kidney disease, falls, and fractures: a systematic review and meta-analysis". In: *Osteoporosis International* 31.1, pp. 13–29. DOI: 10.1007/s00198-019-05190-5.

Government office for Science (2019). *Future of an Ageing Population*. Tech. rep. URL: https://www.gov.uk/government/publications/future-of-an-ageing-population.

Grolemund, Garrett and Hadley Wickham (2011). "Dates and Times Made Easy with lubridate". In: *Journal of Statistical Software* 40.3, pp. 1–25. DOI: 10.18637/jss.v040.i03.

Großwendt, Anna, Heiko Röglin, and Melanie Schmidt (2019). "Analysis of Ward's Method". In: pp. 1–27. DOI: 10.48550/arXiv.1907.05094.

Growdon, Matthew E, Ronald I Shorr, and Sharon K Inouye (2017). "The Tension Between Promoting Mobility and Preventing Falls in the Hospital". In: *JAMA internal medicine* 177.6, pp. 759–760. DOI: 10.1001/jamainternmed.2017.0840.

Guisado-clavero, Marina et al. (2018). "Multimorbidity patterns in the elderly : a prospective cohort study with cluster analysis". In: *BMC Geriatrics* 18.1, pp. 1–11. DOI: 10.1186/s12877-018-0705-7.

Han, Julian Xinguang et al. (2017). "Models of Mortality and Morbidity in Severe Traumatic Brain Injury: An Analysis of a Singapore Neurotrauma Database." In: *World neurosurgery* 108, pp. 885–894. DOI: 10.1016/j.wneu.2017.08.147.

Hartig, Florian (2022a). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level/Mixed) Regression Models*. R package version 0.4.6. URL: https://cran.r-project.org/package=DHARMa.

— (2022b). *testOutliers: Test for outliers*. URL: https://www.rdocumentation.org/packages/DHARMa/versions/0.4.6/topics/testOutliers.

Hastie, Trevor, Jerome Friedman, and Robert Tibshirani (2009). *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. Springer series in statistics. New York: Springer. DOI: 10.1007/978-0-387-84858-7.

Hauer, Klaus et al. (2003). "Cognitive Impairment Decreases Postural Control During Dual Tasks in Geriatric Patients with a History of Severe Falls". In: *Journal of the American Geriatrics Society (JAGS)* 51.11, pp. 1638–1644. DOI: 10.1046/j.1532-5415.2003.51517.x.

Hayashibara, M et al. (2010). "Incidence and risk factors of falling in ambulatory patients with rheumatoid arthritis: a prospective 1-year study". In: *Osteoporosis international* 21.11, pp. 1825–1833. DOI: 10.1007/s00198-009-1150-4.

Head, Anna et al. (2021). "Inequalities in incident and prevalent multimorbidity in England, 2004–19: a population-based, descriptive study". In: *The Lancet. Healthy longevity* 2.8, e489–e497. DOI: 10.1016/S2666-7568(21)00146-X.

Health & Care Excellence, National Institute of (2013). *Falls In Older People: Assessing Risk and Prevention*. Tech. rep. NICE. URL: https://www.nice.org.uk/guidance/cg161/resources/falls-in-older-people-assessing-risk-and-prevention-35109686728645.

HealthCall (2023). *HealthCall*. URL: https://nhshealthcall.co.uk (visited on 09/05/2023).

Heijnen, Michel Johannes Hubertus and Shirley Rietdyk (2016). "Falls in young adults: Perceived causes and environmental factors assessed with a daily online survey". In: *Human movement science* 46, pp. 86–95. DOI: 10.1016/j.humov.2015.12.007.

Hennig, Christian (2007). "Cluster-wise assessment of cluster stability". In: *Computational Statistics Data Analysis* 52.1, pp. 258–271. DOI: 10.1016/j.csda.2006.11.025.

— (2023). *fpc: Flexible Procedures for Clustering*. https://cran.r-project.org/package=fpc.

Hernández, Belinda, Richard B. Reilly, and Rose Anne Kenny (2019). "Investigation of multimorbidity and prevalent disease combinations in older Irish adults using network analysis and association rules". In: *Scientific Reports* 9.1, pp. 1–12. DOI: 10.1038/s41598-019-51135-7.

Hicks, Caitlin W. et al. (2023). "The association of peripheral neuropathy detected by monofilament testing with risk of falls and fractures in older adults". In: *Journal of the American Geriatrics Society (JAGS)* 71.6, pp. 1902–1909. DOI: 10.1111/jgs.18338.

Hilbe, Joseph (2009). *Logistic regression models*. A Chapman and Hall Book. Boca Raton: Chapman and Hall/CRC. DOI: 10.1201/9781420075779.

— (2011). *Negative binomial regression*. Second edition. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511973420.

Howcroft, Jennifer et al. (2017). "Elderly fall risk prediction using static posturography". In: *PLoS ONE* 12.2, pp. 1–13. DOI: 10.1371/journal.pone.0172398.

Hubbard, R E et al. (2017). "Frailty status at admission to hospital predicts multiple adverse outcomes". In: *Age and Ageing* 46.5, pp. 801–806. DOI: 10.1093/ageing/afx081.

Husson, F, S Le, and J Pagès (2017). *Exploratory Multivariate Analysis by Example Using R*. Ed. by Chapman and Hall/CRC. Second edition, p. 262. DOI: 10.1201/b21874.

interRAI (2024). *Instruments overview*. URL: https://interrai.org/instruments/.

Islam, M Mofizul et al. (2014). "Multimorbidity and comorbidity of chronic diseases among the senior australians: Prevalence and patterns". In: *PLoS ONE* 9.1, pp. 1–11. DOI: 10.1371/journal.pone.0083783.

Jacob, Louis et al. (2022). "Prospective Association between Multimorbidity and Falls and Its Mediators: Findings from the Irish Longitudinal Study on Ageing". In: *Journal of Clinical Medicine* 11.15, pp. 1–11. DOI: 10.3390/jcm11154470.

James S. L., et al. (2019). "Global, regional, and national burden of traumatic brain injury and spinal cord injury, 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016". In: *The Lancet Neurology* 18.1, pp. 56–87. ISSN: 14744465. DOI: 10.1016/S1474-4422(18)30415-0.

Jin, Xin and Jiawei Han (2010). "K-Medoids Clustering". In: *Encyclopedia of Machine Learning*. Ed. by Claude Sammut and Geoffrey I Webb. Boston, MA: Springer US, pp. 564–565. DOI: 10.1007/978-0-387-30164-8_426.

Johnston, Marjorie et al. (2019). "Defining and measuring multimorbidity: a systematic review of systematic reviews". In: *European journal of public health* 29.1, pp. 182–189. DOI: 10.1093/eurpub/cky098.

Junior, Jefferson Rosi et al. (2017). "Prognostic model for patients with traumatic brain injuries and abnormal computed tomography scans". In: *Journal of Clinical Neuroscience* 42, pp. 122–128. DOI: 10.1016/j.jocn.2017.03.012.

Kabeshova, Anastasiia et al. (2015). "Artificial Neural Network and Falls in Community-Dwellers: A New Approach to Identify the Risk of Recurrent Falling?" In: *Journal of the American Medical Directors Association* 16.4, pp. 277–281. DOI: 10.1016/j.jamda.2014.09.013.

Kassambara, Alboukadel and Fabian Mundt (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. https://CRAN.R-project.org/package=factoextra.

Kehoe, A. et al. (2015). "The changing face of major trauma in the UK". In: *Emergency Medicine Journal* 32.12, pp. 911–915. DOI: 10.1136/emermed-2015-205265.

Keifer, G. and F. Effenberger (2021). *Green Paper on Ageing*. Tech. rep. European Comission, pp. 951–952. DOI: 10.2775/785789.

Kelleher, John D., Brian Mac Namee, and Aoife D'Arcy (2015). *Fundamentals of machine learning for predictive data analytics : algorithms, worked examples, and case studies*. Cambridge, Massachusetts: The MIT Press. ISBN: 9780262029445.

Kelly, Michael L. et al. (2015). "Functional and long-term outcomes in severe traumatic brain injury following regionalization of a trauma system". In: *Journal of Trauma and Acute Care Surgery* 79.3, pp. 372–377. DOI: 10.1097/TA.0000000000000762.

Kim, Lily H et al. (2020). "Mortality prediction and long-term outcomes for civilian cerebral gunshot wounds: A decision-tree algorithm based on a single trauma center". In: *Journal of clinical neuroscience* 75, pp. 71–79. DOI: 10.1016/j.jocn.2020.03.027.

King, Barbara et al. (2016). "Impact of Fall Prevention on Nurses and Care of Fall Risk Patients". In: *The Gerontologist* 58.2, pp. 331–340. DOI: 10.1093/geront/gnw156.

Kingston, Andrew et al. (2018). "Projections of multi-morbidity in the older population in England to 2035: Estimates from the Population Ageing and Care Simulation (PACSim) model". In: *Age and Ageing* 47.3, pp. 374–380. DOI: 10.1093/ageing/afx201.

Kuspinar, Ayse et al. (2019). "Development and validation of an algorithm to assess risk of first-time falling among home care clients". In: *BMC Geriatrics* 19.1, pp. 4–11. DOI: 10.1186/s12877-019-1300-2.

Kvelde, Tasha et al. (2013). "Depressive Symptomatology as a Risk Factor for Falls in Older People: Systematic Review and Meta-Analysis". In: *Journal of the American Geriatrics Society (JAGS)* 61.5, pp. 694–706. DOI: 10.1111/jgs.12209.

Labib, Noura et al. (2011). "Severely Injured Geriatric Population: Morbidity, Mortality, and Risk Factors". In: *The Journal of trauma, injury, infection, and critical care* 71.6, pp. 1908–1914. DOI: 10.1097/TA.0b013e31820989ed.

Lakhan, Prabha et al. (2011). "A Prospective Cohort Study of Geriatric Syndromes Among Older Medical Patients Admitted to Acute Care Hospitals". In: *American*

Geriatrics Society 59.11, pp. 2001–2008. DOI: 10.1111/j.1532-5415.2011.03663.x.

Lamb, Sarah E. et al. (2003). "Risk factors for falling in home-dwelling older women with stroke: The women's health and aging study". In: *Stroke* 34.2, pp. 494–501. DOI: 10.1161/01.STR.0000053444.00582.B7.

Lawrence, T et al. (2016). "Traumatic brain injury in England and Wales: prospective audit of epidemiology, complications and standardised mortality". In: *BMJ open* 6.11, e012197–e012197. DOI: 10.1136/bmjopen-2016-012197.

Lê, Sébastien, Julie Josse, and François Husson (2008). "FactoMineR: A Package for Multivariate Analysis". In: *Journal of Statistical Software* 25.1, pp. 1–18. DOI: 10.18637/jss.v025.i01.

Le Roux, Brigitte and Henry Rouanet (2010). *Multiple correspondence analysis*. Quantitative applications in the social sciences ; 163. London: SAGE. DOI: 10.4135/9781412993906.

Liu, Yang et al. (2020). "A systematic review and meta-analysis of fall incidence and risk factors in elderly patients after total joint arthroplasty". In: *Medicine* 99.50, e23664. DOI: 10.1097/MD.0000000000023664.

Lockhart, Thurmon E. et al. (2021). "Prediction of fall risk among community-dwelling older adults using a wearable system". In: *Scientific Reports* 11.1, pp. 1–14. DOI: 10.1038/s41598-021-00458-5.

Lu, Hsueh-Yi et al. (2015). "Predicting Long-Term Outcome After Traumatic Brain Injury Using Repeated Measurements of Glasgow Coma Scale and Data Mining Methods". In: *Journal of Medical Systems* 39.2, pp. 1–10. DOI: 10.1007/s10916-014-0187-x.

Maas, Andrew I.R. et al. (2007). "Prognosis and Clinical Trial Design in Traumatic Brain Injury: The IMPACT Study". In: *Journal of Neurotrauma* 24.2, pp. 232–238. DOI: 10.1089/neu.2006.0024.

Machón, Mónica et al. (2020). "Multimorbidity and functional status in older people: a cluster analysis". In: *European Geriatric Medicine* 11.2, pp. 321–332. DOI: 10.1007/s41999-020-00291-5.

Mackenzie, Lynette, Julie Byles, and Catherine D'Este (2006). "Validation of self-reported fall events in intervention studies". In: *Clinical Rehabilitation* 20.4, pp. 331–339. DOI: 10.1191/0269215506cr947oa.

Maechler, Martin et al. (2022). *cluster: Cluster Analysis Basics and Extensions*. URL: https://cran.r-project.org/package=cluster.

Mahmood, Faizan et al. (2021). *Wider impacts of COVID-19 on physical activity, deconditioning and falls in older adults*. Tech. rep. Public Health England.

Makino, Keitaro et al. (2017). "Fear of falling and gait parameters in older adults with and without fall history". In: *Geriatrics and Gerontology International* 17.12, pp. 2455–2459. DOI: 10.1111/ggi.13102.

Malik, Varun et al. (2020). "Atrial Fibrillation Is Associated With Syncope and Falls in Older Adults: A Systematic Review and Meta-analysis". In: *Mayo Clinic Proceedings* 95.4, pp. 676–687. DOI: 10.1016/j.mayocp.2019.09.029.

Marengoni, Alessandra et al. (2020). "Patterns of multimorbidity in a population-based cohort of older people: Sociodemographic, lifestyle, clinical, and functional differences". In: *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 75.4, pp. 798–805. DOI: 10.1093/gerona/glz137.

Marier, Allison et al. (2016). "Improving prediction of fall risk among nursing home residents using electronic medical records". In: *Journal of the American Medical Informatics Association* 23.2, pp. 276–282. DOI: 10.1093/jamia/ocv061.

Marmarou, Anthony et al. (2007). "Prognostic Value of The Glasgow Coma Scale And Pupil Reactivity in Traumatic Brain Injury Assessed Pre-Hospital And on Enrollment: An IMPACT Analysis". In: *Journal of Neurotrauma* 24.2, pp. 270–280. DOI: 10.1089/neu.2006.0029.

Marques, Nise Ribeiro et al. (2018). "Is it possible to predict falls in older adults using gait kinematics?" In: *Clinical Biomechanics* 59, pp. 15–18. DOI: 10.1016/j.clinbiomech.2018.08.006.

McCoy, Thomas H. et al. (2017). "Validation of a risk stratification tool for fall-related injury in a state-wide cohort". In: *BMJ Open* 7.2, pp. 1–7. DOI: 10.1136/bmjopen-2016-012189.

McParland, Chris et al. (2022). "Multimorbidity, disease count, mortality and emergency care use in persons attending the emergency department: a cross-sectional data-linkage study". In: *Journal of comorbidity* 12, pp. 1–15. DOI: 10.1177/26335565221147417.

Meredith, W et al. (1998). "The conundrum of the Glasgow Coma Scale in intubated patients: a linear regression prediction of the Glasgow verbal score from the Glasgow eye and motor scores." In: *The Journal of trauma* 44.5, pp. 835–839. DOI: 10.1097/00005373-199805000-00016.

Miles, Jamie et al. (2020). "Using machine-learning risk prediction models to triage the acuity of undifferentiated patients entering the emergency care system: a systematic review". In: *Diagnostic and Prognostic Research* 4.1. DOI: 10.1186/s41512-020-00084-1.

Mitnitski, A. B., A. J. Mogilner, and K. Rockwood (2001). "Accumulation of deficits as a proxy measure of aging". In: *The Scientific World* 1, pp. 323–336.

Mol, Arjen et al. (2019). "Orthostatic Hypotension and Falls in Older Adults: A Systematic Review and Meta-analysis". In: *Journal of the American Medical Directors Association* 20.5, pp. 589–597. DOI: 10.1016/j.jamda.2018.11.003.

Moon, Shinje et al. (2021). "The impact of urinary incontinence on falls: A systematic review and meta-analysis". In: *PloS one* 16.5, e0251711–e0251711. DOI: 10.1371/journal.pone.0251711.

Morin, Lucas et al. (2019). "Polypharmacy and injurious falls in older adults: A nationwide nested case-control study". In: *Clinical Epidemiology* 11, pp. 483–493. DOI: 10.2147/CLEP.S201614.

Morley, Claire et al. (2018). "Emergency department crowding: A systematic review of causes, consequences and solutions". In: *PLoS ONE* 13.8, e0203316. DOI: 10.1371/journal.pone.0203316.

Muehlschlegel, Susanne et al. (2016). "Predicting survival after acute civilian penetrating brain injuries: The SPIN score." In: *American Academy of Neurology* 87.21, pp. 2244–2253. DOI: 10.1212/WNL.0000000000003355.

Murphy, Kevin P (2012). *Machine learning: a probabilistic perspective*. Adaptive computation and machine learning series. Cambridge, MA: MIT Press. ISBN: 0-262-30432-5.

Murtagh, Fionn and Pierre Legendre (2014). "Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?" In: *Journal of Classification* 31, pp. 274–295. DOI: 10.1007/s00357-014-9161-z.

Najafi, Zohre, Hossien Zakeri, and Amir Mirhaghi (2018). "The accuracy of acuity scoring tools to predict 24-h mortality in traumatic brain injury patients: A guide to triage criteria". In: *International Emergency Nursing* 36, pp. 27–33. DOI: 10.1016/j.ienj.2017.08.003.

National Institute for Neurological Disorders and Stroke (2019). *Traumatic Brain Injury Information Page*. URL: https://www.ninds.nih.gov/Disorders/All-Disorders/Traumatic-Brain-Injury-Information-Page (visited on 08/10/2021).

National Institute of Health & Care Excellence (2023). *Multimorbidity*. URL: https://cks.nice.org.uk/topics/multimorbidity/background-information/prevalence/ (visited on 09/05/2023).

Nevitt, Michael C., Steven R. Cummings, and Study of Osteoporotic Fractures Research Group (1993). "Type of Fall and Risk of Hip and Wrist Fractures: The Study of Osteoporotic Fractures". In: *Journal of the American Geriatrics Society* 41.11, pp. 1226–1234. DOI: https://doi.org/10.1111/j.1532-5415.1993.tb07307.x.

Nghiem, Son et al. (2021). "Hospital frailty risk score and adverse health outcomes: Evidence from longitudinal record linkage cardiac data". In: *Age and ageing* 50.5, pp. 1778–1784. DOI: 10.1093/ageing/afab073.

NHS (2019). *The NHS Long Term Plan*. Tech. rep. NHS. DOI: 10.38192/12.1.4. URL: https://www.longtermplan.nhs.uk/publication/nhs-long-term-plan/.

— (2021). *Overview Falls*. URL: https://www.nhs.uk/conditions/falls/#:~:text=Around\%201\%20in\%203\%20adults,they\%20have\%20lost\%20their\%20independence (visited on 05/05/2023).

NHS England (2020). *The framework for Enhanced Health in Care Homes*. Tech. rep. NHS England. URL: https://www.england.nhs.uk/publication/enhanced-health-in-care-homes-framework/.

— (2022). *Care homes*. URL: https://www.nhs.uk/conditions/social-care-and-support-guide/care-services-equipment-and-care-homes/care-homes/.

— (2023a). *Delivery Plan for Recovering Urgent and Emergency Care Services*. Tech. rep. Department of Health and Social Care. URL: https://www.england.nhs.uk/wp-content/uploads/2023/01/B2034-delivery-plan-for-recovering-urgent-and-emergency-care-services.pdf.

— (2023b). *Long Term Physical Health Condition*. URL: https://www.datadictionary.nhs.uk/nhs_business_definitions/long_term_physical_health_condition.html.

— (2023c). *Stroke Overview*. URL: https://www.nhs.uk/conditions/stroke/.

— (2023d). *Symptoms of dementia*. URL: https://www.nhs.uk/conditions/dementia/symptoms-and-diagnosis/symptoms/.

NHS Inform (2023). *Prevent falls by identifying hazards at home*. Tech. rep. The Scottish Government. URL: https://www.nhsinform.scot/healthy-living/preventing-falls/home-safety/prevent-falls-by-identifying-hazards-at-home/.

Northridge, M E et al. (1995). "Home hazards and falls in the elderly: the role of health and functional status". In: *American journal of public health (1971)* 85.4, pp. 509–515. DOI: 10.2105/ajph.85.4.509.

Nurmi, Ilona et al. (2009). "Incidence and costs of falls and fall injuries among elderly in institutional care". In: 20, pp. 118–122. DOI: 10.1080/pri.20.2.118.122.

Ó Briain, Diarmuid et al. (2018). "Early Hyperoxia in Patients with Traumatic Brain Injury Admitted to Intensive Care in Australia and New Zealand: A Retrospective Multicenter Cohort Study". In: *Neurocritical Care* 29.3, pp. 443–451. DOI: 10.1007/s12028-018-0553-5.

Office for National Statistics (2018). *How do the post-World War baby boom generations compare?* Tech. rep. Office for National Statistics. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/ageing/articles/howdothepostworldwarbabyboomgenerationscompare/2018-03-06.

Office for National Statistics (2023). *Care homes and estimating the self-funding population, England: 2022 to 2023*. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/socialcare/articles/carehomesandestimatingtheselffundingpopulationen 2022to2023 (visited on 02/27/2024).

Office for National Statistics (2023). *Profile of the older population living in England and Wales in 2021 and changes since 2011*. Tech. rep. Office for National Statistics. URL:

https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/ageing/articles/profileoftheolderpopulationlivinginenglandandwalesin2021andchangessi 2023-04-03.

Oliveira, Cristino C et al. (2021). "Falls prevalence and risk factors in people with chronic obstructive pulmonary disease : A systematic review". In: *Respiratory Medicine* 176, p. 106284. DOI: 10.1016/j.rmed.2020.106284.

Page, Matthew J et al. (2021). "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews". In: *BMJ* 18.3, e1003583–e1003583. DOI: 10.1136/bmj.n71.

Pasquetti, Pietro, Lorenzo Apicella, and Giuseppe Mangone (2014). "Pathogenesis and treatment of falls in elderly". In: *Clinical Cases in Mineral and Bone Metabolism* 11.3, pp. 222–225. DOI: 10.11138/ccmbm/2014.11.3.222.

Peel, N (2000). "Validating recall of falls by older people". In: *Accident; analysis and prevention* 32.3, pp. 371–372. DOI: 10.1016/s0001-4575(99)00066-4.

Perel, Pablo et al. (2006). "Systematic review of prognostic models in traumatic brain injury". In: *BMC medical informatics and decision making* 6.1, p. 38. DOI: 10.1186/1472-6947-6-38.

Perel, Pablo et al. (2008). "Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients". In: *BMJ* 336.7641, pp. 425–429. DOI: 10.1136/bmj.39461.643438.25.

Prados-Torres, Alexandra et al. (2014). "Multimorbidity Patterns: A systematic review". In: *Journal of Clinical Epidemiology* 67.3, pp. 254–266. DOI: 10.1016/j.jclinepi.2013.09.021.

Preston, Louise et al. (2021). "Improving outcomes for older people in the emergency department: a review of reviews". In: *Emergency medicine journal: EMJ* 38.12, pp. 882–888. DOI: 10.1136/emermed-2020-209514.

Prosser, Callum J et al. (2020). "Bypassing the nearest emergency department for a more distant neurosurgical centre in traumatic brain injury patients". In: *British Journal of Neurosurgery* 36.1, pp. 31–37. DOI: 10.1080/02688697.2020.1858026.

Provost, Foster and Tom Fawcett (2013). *Data science for business*. Beijing: O'Reilly. ISBN: 9781449361327.

Public Health Agency of Canada (2014). *Seniors' Falls in Canada: Second Report*. Tech. rep. Government of Canada. URL: https://www.canada.ca/en/public-health/services/health-promotion/aging-seniors/publications/publications-general-public/seniors-falls-canada-second-report.html#s2-3.

Public Health England (2018). *A Return on Investment Tool for the Assessment of Falls Prevention Programmes for Older People Living in the Community*. Tech. rep. Public Health England. URL: https://assets.publishing.service.gov.uk/media/5a7aedafe5274a319e77bb5d/A_return_on_investment_tool_for_falls_prevention_programmes.pdf.

Public Health England (2022). *Falls: applying All Our Health*. URL: https://www.gov.uk/government/publications/falls-applying-all-our-health/falls-applying-all-our-health?_ga=2.43988389.1814564206.1685694498-696290112.1685694498 (visited on 02/06/2023).

Pynoos, J, B Steinman, and A Nguyen (2010). "Environmental Assessment and Modification as Fall-Prevention Strategies for Older Adults". In: *Clinics in geriatric medicine* 26.4, pp. 633–644. DOI: 10.1016/j.cger.2010.07.001.

Quan, Hude et al. (2011). "Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries". In: *American Journal of Epidemiology* 173.6, pp. 676–682. DOI: 10.1093/aje/kwq433.

Quinn, Joseph F. and Sarah Dickinson (2014). *Dementia*. Neurology in practice (Series). Chichester, England: Wiley-Blackwell. DOI: 10.1002/9781118656082.

Rachas, Antoine et al. (Aug. 2023). "Individual and contextual risk factors for mortality in nursing home residents during the first wave of COVID-19 in France: a multilevel analysis of a nationwide cohort study". In: *Age and Ageing* 52.8. DOI: 10.1093/ageing/afad165.

Radovanovic, Dragana et al. (2014). "Validity of Charlson Comorbidity Index in patients hospitalised with acute coronary syndrome. Insights from the nationwide AMIS Plus registry 2002–2012". In: *Heart* 100.4, pp. 288–294. DOI: 10.1136/heartjnl-2013-304588.

Raj, Rahul et al. (2019). "Machine learning-based dynamic mortality prediction after traumatic brain injury." In: *Scientific reports* 9.17672, pp. 1–14. DOI: 10.1038/s41598-019-53889-6.

Riandini, Tessa et al. (2020). "Fall Risk and Balance Confidence in Patients With Diabetic Peripheral Neuropathy: An Observational Study". In: *Frontiers in endocrinology* 11, p. 573804. DOI: 10.3389/fendo.2020.573804.

Richardson, Kathryn, Kathleen Bennett, and Rose Anne Kenny (2015). "Polypharmacy including falls risk-increasing medications and subsequent falls in community-dwelling middle-aged and older adults". In: *Age and Ageing* 44.1, pp. 90–96. DOI: 10.1093/AGEING/AFU141.

Rockwood, Kenneth and Arnold Mitnitski (2007). "Frailty in Relation to the Accumulation of Deficits". In: *The journals of gerontology. Series A, Biological sciences and medical sciences* 62.7, pp. 722–727.

Roe, Lorna et al. (2017). "The impact of frailty on healthcare utilisation in Ireland: Evidence from the Irish longitudinal study on ageing". In: *BMC geriatrics* 17.1, p. 203. DOI: 10.1186/s12877-017-0579-0.

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC. Boston, MA. URL: http://www.rstudio.com/.

Rubenstein, Laurence (2006). "Falls in older people: epidemiology, risk factors and strategies for prevention". In: *Age and Ageing* 35.2, pp. 37–41. DOI: 10.1093/ageing/afl084.

Rubenstein, Laurence et al. (1990). "The Value of Assessing Falls in an Elderly Population". In: *Annals of Internal Medicine* 113.4, pp. 308–316. DOI: 10.7326/0003-4819-113-4-308.

Saliba, Debra and Joan Buchanan (2008a). *DEVELOPMENT VALIDATION OF A REVISED NURSING HOME ASSESSMENT TOOL: MDS 3.0*. Tech. rep. Centers for Medicare Medicaid Services. URL: https://www.cms.gov/medicare/quality/nursing-home-improvement/minimum-data-sets-swing-bed-providers#:~:text=In%202010%20with%20the%20implementation,protocols%20used%20in%20other%20settings..

— (2008b). *Development Validation of a Revised Nursing Home Assessment Tool: MDS 3.0*. Tech. rep. United States Centers for Medicare Medicaid Services. URL: https:

//www.cms.gov/medicare/quality/nursing-home-improvement/resident-assessment-instrument-manual.

— (2020). *Learning Care Homes: Continuous improvement of structured referrals*. URL: https://www.hdruk.ac.uk/projects/better-care-northern-partnership-learning-care-homes-continuous-improvement-of-structured-referrals/.

Samaras, Nikolaos et al. (2010). "Older patients in the emergency department: A review". In: *Annals of Emergency Medicine* 56.3, pp. 261–269. DOI: 10.1016/j.annemergmed.2010.04.015.

Schafer, Ingmar et al. (2010). "Multimorbidity Patterns in the Elderly : A New Approach of Disease Clustering Identifies Complex Interrelations between Chronic Conditions". In: *PloS one* 5.12, e15941. DOI: 10.1371/journal.pone.0015941.

Schwartz, Ann V et al. (2008). "Diabetes-Related Complications, Glycemic Control, and Falls in Older Adults". In: *Diabetes care* 31.3, pp. 391–396. DOI: 10.2337/dc07-1152.

Schäfer, Ingmar et al. (2014). "Reducing complexity: A visualisation of multimorbidity by combining disease clusters and triads". In: *BMC public health* 14.1, p. 1285. DOI: 10.1186/1471-2458-14-1285.

Scott, Vicky et al. (2007). "Multifactorial and functional mobility assessment tools for fall risk among older adults in community, home-support, long-term and acute care settings". In: *Age and Ageing* 36.2, pp. 130–139. DOI: 10.1093/ageing/afl165.

Seaman, Karla et al. (2022). "The use of predictive fall models for older adults receiving aged care, using routinely collected electronic health record data: a systematic review". In: *BMC geriatrics* 22.210, pp. 1–13. DOI: 10.1186/s12877-022-02901-2.

Shang, Xianwen et al. (2021). "The Association between Vision Impairment and Incidence of Dementia and Cognitive Impairment: A Systematic Review and Meta-analysis". In: *Ophthalmology* 128.8, pp. 1135–1149. DOI: 10.1016/j.ophtha.2020.12.029.

Shao, Lu et al. (2023). "Incidence and Risk Factors of Falls Among Older People in Nursing Homes: Systematic Review and Meta-Analysis". In: *Journal of the American Medical Directors Association* 24.11, pp. 1708–1717. DOI: 10.1016/j.jamda.2023.06.002.

Shaw, Brett H et al. (2019). "Relationships between orthostatic hypotension, frailty, falling and mortality in elderly care home residents". In: *BMC Geriatrics* 19.80, pp. 1–14. DOI: 10.1186/s12877-019-1082-6.

Smith, Valerie et al. (2011). "Methodology in conducting a systematic review of systematic reviews of healthcare interventions". In: *BMC medical research methodology* 11.15, pp. 1–6. DOI: 10.1186/1471-2288-11-15.

Stalenhoef, P A et al. (2002). "A risk model for the prediction of recurrent falls in community-dwelling elderly: A prospective cohort study". In: *Journal of Clinical Epidemiology* 55.11, pp. 1088–1094. DOI: 10.1016/S0895-4356(02)00502-4.

Stanmore, Emma K. et al. (2013). "Fall Incidence and Outcomes of Falls in a Prospective Study of Adults With Rheumatoid Arthritis". In: *Arthritis Care and Research* 65.5, pp. 737–744. DOI: 10.1002/acr.21892.

Stenhagen, Magnus et al. (2014). "Accidental falls, health-related quality of life and life satisfaction: A prospective study of the general elderly population". In: *Archives of gerontology and geriatrics* 58.1, pp. 95–100. DOI: 10.1016/j.archger.2013.07.006.

Sterling, Daniel A., Judith A. O'Connor, and John Bonadies (2001). "Geriatric Falls: Injury Severity Is High and Disproportionate to Mechanism". In: *The Journal of Trauma: Injury, Infection, and Critical Care* 50.1, pp. 116–119. DOI: 10.1097/00005373-200101000-00021.

Stubbs, Brendon et al. (2014). "Pain and the Risk for Falls in Community-Dwelling Older Adults : Systematic Review and Meta-Analysis". In: *Archives of Physical Medicine and Rehabilitation* 95.1, pp. 175–187. DOI: 10.1016/j.apmr.2013.08.241.

Sturnieks, Daina L et al. (2004). "Physiological risk factors for falls in older people with lower limb arthritis". In: *Journal of Rheumatology* 31.11, pp. 2272–2279. DOI: 10.1046/j.1532-5415.2001.49107.

Talbot, Laura A. et al. (2005). "Falls in young, middle-aged and older community dwelling adults: Perceived cause, environmental factors and injury". In: *BMC public health* 5.86, pp. 1–9. DOI: 10.1186/1471-2458-5-86.

Teasdale, Graham and Bryan Jennett (1974). "ASSESSMENT OF COMA AND IMPAIRED CONSCIOUSNESS: A Practical Scale". In: *The Lancet* 49.3, pp. 81–84. DOI: 10.1097/00000542-197809000-00023.

Teixeira, Laetitia et al. (2019). "Falls and fear of falling in a sample of centenarians: the role of multimorbidity, pain and anxiety". In: *Psychogeriatrics* 19.5, pp. 457–464. DOI: 10.1111/psyg.12423.

Thapa, Rahul et al. (2022). "Predicting Falls in Long-term Care Facilities: Machine Learning Study". In: *JMIR Aging* 5.2, pp. 1–15. DOI: 10.2196/35373.

Thelin, Eric Peter et al. (2016). "Utility of neuron-specific enolase in traumatic brain injury; relations to S100B levels, outcome, and extracranial injury severity". In: *Critical Care* 20.285, pp. 1–15. DOI: 10.1186/s13054-016-1450-y.

Timar, Bogdan et al. (2016). "The impact of diabetic neuropathy on balance and on the risk of falls in patients with type 2 diabetes mellitus: A cross-sectional study". In: *PloS one* 11.4, e0154654.

Valenzuela, Pedro L et al. (2018). "Physical strategies to prevent disuse-induced functional decline in the elderly". In: *Ageing Research Reviews* 47, pp. 80–88. DOI: 10.1016/j.arr.2018.07.003.

Van De Loo, Bob et al. (2022). "Development of the ADFICE$_{I}$*TModelsforPredictingFallsandRecurrentFallsi*
DwellingOlderAdults : *PooledAnalysesofEuropeanCohortsWithSpecialAttentiontoMedication*".
In: *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 77.7, pp. 1446–1454. DOI: 10.1093/gerona/glac080.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S*. Fourth. ISBN 0-387-95457-0. New York: Springer. URL: https://www.stats.ox.ac.uk/pub/MASS4/.

Veronese, Nicola et al. (2023). "Knee osteoarthritis and adverse health outcomes: an umbrella review of meta-analyses of observational studies". In: *Aging Clinical and Experimental Research* 35.2, pp. 245–252. DOI: 10.1007/s40520-022-02289-4.

Vetrano, Davide L et al. (2019). "Frailty and Multimorbidity: A Systematic Review and Meta-analysis". In: *The journals of gerontology. Series A, Biological sciences and medical sciences* 74.5, pp. 659–666. DOI: 10.1093/gerona/gly110.

Violán, Concepción et al. (2018). "Multimorbidity patterns with K-means nonhierarchical cluster analysis". In: *BMC Family Practice* 19.108, pp. 1–11. DOI: 10.1186/s12875-018-0790-x.

Violán, Concepción et al. (2019). "Soft clustering using real-world data for the identification of multimorbidity patterns in an elderly population: Cross-sectional

study in a Mediterranean population". In: *BMJ Open* 9.8, pp. 1–14. DOI: 10.1136/bmjopen-2019-029594.

Vlaeyen, E et al. (2020). "Predicting falls in nursing homes: A prospective multicenter cohort study comparing fall history, staff clinical judgment, the CaHFRiS and the FRiCA". In: *Journal of the American Medical Directors Association* 22.2, pp. 380–387. DOI: 10.1016/j.jamda.2020.06.037.

Vu, Trang, Caroline F Finch, and Lesley Day (2011). "Patterns of comorbidity in community-dwelling older people hospitalised for fall-related injury: A cluster analysis". In: *BMC Geriatrics* 11.45, pp. 1–10. DOI: 10.1186/1471-2318-11-45.

Wan-Ting, Chu et al. (2020). "Reverse shock index multiplied by Glasgow Coma Scale (rSIG) predicts mortality in severe trauma patients with head injury." In: *Scientific reports* 10.2095, pp. 1–8. DOI: 10.1038/s41598-020-59044-w.

Whiting, Penny et al. (2016). "ROBIS: A new tool to assess risk of bias in systematic reviews was developed". In: *Journal of clinical epidemiology* 69, pp. 225–234. DOI: 10.1016/j.jclinepi.2015.06.005.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

Wickham, Hadley and Jennifer Bryan (2023). *readxl: Read Excel Files*. https://readxl.tidyverse.org, https://github.com/tidyverse/readxl.

Wickham, Hadley et al. (2019). "Welcome to the tidyverse". In: *Journal of Open Source Software* 4.43, p. 1686. DOI: 10.21105/joss.01686.

Winans, Nathan J. et al. (2020). "Modeling the return to consciousness after severe traumatic brain injury at a large academic level 1 trauma center". In: *Journal of Neurosurgery* 133.2, pp. 477–485. DOI: 10.3171/2019.2.JNS183568.

Witten, Ian H., Eibe Frank, and Mark A. Hall (2011). *Data mining [electronic resource] : practical machine learning tools and techniques*. 3rd ed. The Morgan Kaufmann Series in Data Management Systems. Amsterdam: Elsevier/Morgan Kaufmann. ISBN: 1-282-95388-5.

Wittenberg, Raphael and Bo Hu (2015). *Projections of Demand for and Costs of Social Care for Older People and Younger Adults in England, 2015 to 2035*. Tech. rep. URL: www.pssru.ac.uk/pdf/DP2900.pdf.

World Health Organisation (2004). *ICD 10: international statistical classification of diseases and related health problems: tenth revision*. Tech. rep. URL: https://apps.who.int/iris/handle/10665/42980.

— (2021). *Falls*. URL: https://www.who.int/news-room/fact-sheets/detail/falls (visited on 02/06/2023).

— (2022). *Ageing and health*. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health (visited on 02/06/2023).

Wu, Esther, Siddharth Marthi, and Wael F. Asaad (2020). "Predictors of Mortality in Traumatic Intracranial Hemorrhage: A National Trauma Data Bank Study". In: *Frontiers in Neurology* 11, pp. 1–8. DOI: 10.3389/fneur.2020.587587.

Xu, Qingmei, Xuemei Ou, and Jinfeng Li (2022). "The risk of falls among the aging population: A systematic review and meta-analysis". In: *Frontiers in Public Health* 10.4, pp. 1–8. DOI: 10.3389/fpubh.2022.902599.

Yang, Z.-C. et al. (2023). "Frailty Is a Risk Factor for Falls in the Older Adults: A Systematic Review and Meta-Analysis". In: *The Journal of nutrition, health aging* 27.6, pp. 487–495. DOI: 10.1007/s12603-023-1935-8.

Yeung, Suey S Y et al. (2019). "Sarcopenia and its association with falls and fractures in older adults : A systematic review and meta-analysis". In: *Journal of cachexia, sarcopenia and muscle* 10.3, pp. 485–500. DOI: 10.1002/jcsm.12411.

Yokota, M. et al. (2020). "Clarifying differences in injury patterns between ground-level falls and falls from heights among the elderly in Japan". In: *Public Health* 181, pp. 114–118. DOI: 10.1016/j.puhe.2019.11.020.

You, Liuqing et al. (2023). "Association between multimorbidity and falls and fear of falling among older adults in eastern China: a cross-sectional study". In: *Frontiers in public health* 11, p. 1146899. DOI: 10.3389/fpubh.2023.1146899.

Zahabi, Maryam, David B. Kaber, and Manida Swangnetr (2015). "Usability and Safety in Electronic Medical Records Interface Design: A Review of Recent Literature and Guideline Formulation". In: *Human factors* 57.5, pp. 805–834. DOI: 10.1177/0018720815576827.

Zeiler, Frederick A. et al. (2018). "Critical Thresholds of Intracranial Pressure-Derived Continuous Cerebrovascular Reactivity Indices for Outcome Prediction in Noncraniectomized Patients with Traumatic Brain Injury". In: *Journal of Neurotrauma* 35.10, pp. 1107–1115. DOI: 10.1089/neu.2017.5472.

Zeiler, Frederick A et al. (2020). "Association between Cerebrovascular Reactivity Monitoring and Mortality is preserved when adjusting for baseline admission characteristics in Adult TBI: A CENTER-TBI Study". In: *Journal of Neurotrauma* 37.10, pp. 1233–1241. DOI: 10.1089/neu.2019.6808.

Zheng, Jiali et al. (2023). "Bidirectional associations and a causal mediation analysis between depressive symptoms and chronic digestive diseases: A longitudinal investigation". In: *Journal of affective disorders* 333, pp. 278–289. DOI: 10.1016/j.jad.2023.04.025.

Zumel, N and J Mount (2014). *Practical Data Science with R*. 1st. USA: Manning Publications Co. ISBN: 9781617291562.

Appendix A

Traumatic Brain Injury Systematic Review Search Strategy

TABLE A.1: Medline Search Terms: 202 Papers returned

Filters	(2015, English)
Concept	Terms
Prediction	Predict*.ti,ab. or *Prognosis/ or *Risk/ or Prognos*.ti,ab. or risk*.ti,ab. or likelihood.ti,ab. or *Probability/ or *Likelihood Functions/ or probability.ti,ab. or chance.ti,ab. or *Odds Ratio/ or odds.ti,ab.
Machine Learning and statistical approaches	exp Regression Analysis/ or regression.ti,ab. or logistic.ti,ab. or exp Machine Learning/ or machine learning.ti,ab. or *Neural Networks, Computer/ or neural network*.ti,ab. or *Decision Trees/ or Decision Tree*.ti,ab. or Random For?est*.ti,ab. or *Artificial Intelligence/ or AI.ti,ab. or Artificial Intelligence.ti,ab. or *Models, Statistical/ or Transfer Learning.ti,ab. or *Support Vector Machine/ or Support Vector Machine.ti,ab. or *Bayes Theorem/ or na?ve bayes classifier.ti,ab. or exp Deep Learning/ or deep learning.ti,ab.
Mortality	exp Mortality/ or mortality.ti,ab. or exp Death/ or death.ti,ab. or die.ti,ab. or dying.ti,ab. exp Survival Analysis/ surviv*.ti,ab.
Emergency care	exp Emergency Service, Hospital/ or Emergenc*.ti,ab. or exp Critical Care/ or Intensive Care.ti,ab. or Critical Care.ti,ab. or exp Ambulatory Care/ or Urgent.ti,ab. or exp Emergency Medical Services/ or exp Trauma Centers/
Traumatic Brain Injuries	exp Brain Injuries, Traumatic/ or traumatic Brain Injury.ti,ab. or brain injury.ti,ab. or head injury.ti,ab. or TBI.ti,ab. or exp Brain Injuries/ or exp Skull Fractures/
Adults	exp Adult/

TABLE A.2: Embase Search Terms: 652 Papers returned

Filters	2015+, English
Concept	**Terms**
Prediction	*prediction/ or Predict*.ti,ab. or *prognosis/ or *risk/ or Prognos*.ti,ab. or risk*.ti,ab. or likelihood.ti,ab. or *probability/ or probability.ti,ab. or chance.ti,ab. or *odds ratio/ or odds.ti,ab.
Machine Learning and statistical approaches	*multiple linear regression analysis/ or *multivariate logistic regression analysis/ or *linear regression analysis/ or *regression analysis/ or *multiple regression/ or *nonlinear regression analysis/ or *logistic regression analysis/ or regression.ti,ab. or logistic.ti,ab. or exp machine learning/ or machine learning.ti,ab. or *artificial neural network/ or neural network*.ti,ab. or *"decision tree"/ or Decision Tree*.ti,ab. or Random For?est*.ti,ab. or *artificial intelligence/ or Artificial Intelligence.ti,ab. or *statistical model/ or Transfer Learning.ti,ab. or *support vector machine/ or Support Vector Machine.ti,ab. or *Bayesian learning/ or naive bayes classifier.ti,ab. or exp deep learning/ or deep learning.ti,ab.
Mortality	exp mortality/ or mortality.ti,ab. or exp death/ or death.ti,ab. or die.ti,ab. or dying.ti,ab. or surviv*.ti,ab. or exp Survival Analysis/
Emergency care	exp emergency care/ or Emergency*.ti,ab. or exp intensive care/ or Intensive.ti,ab. or Critical.ti,ab. or exp ambulatory care/ or Urgent.ti,ab. or exp emergency health service/
Traumatic Brain Injuries	exp traumatic brain injury/ or traumatic Brain Injury.ti,ab. or brain injury.ti,ab. or exp brain injury/ or head injury.ti,ab. or exp head injury/ or TBI.ti,ab. or exp skull fracture/
Adults	exp Adult/

TABLE A.3: Web of Science Search Terms: 147 Papers returned

Filters	Core collection, 2015-21, English, articles
Concept	Terms
Prediction	TI = (Predict OR Prognosis OR Risk OR Prognostic OR likelihood OR Probability OR chance OR odds) OR AB = (Predict OR Prognosis OR Risk OR Prognostic OR likelihood OR Probability OR chance OR odds)
Machine Learning and statistical approaches	TI = (Regression OR logistic OR Machine Learning OR Neural Network OR Decision Tree OR Random Forest OR Random Forrest OR Artificial Intelligence OR Statistical Model OR Transfer Learning OR Support Vector Machine OR naive bayes classifier OR Deep Learning) OR AB = (Regression OR logistic OR Machine Learning OR Neural Network OR Decision Tree OR Random Forest OR Random Forrest OR Artificial Intelligence OR Statistical Model OR Transfer Learning OR Support Vector Machine OR naive bayes classifier OR Deep Learning)
Mortality	TI = (Mortality OR Death OR die OR dying OR survive OR survival) OR AB = (Mortality OR Death OR die OR dying OR survive OR survival)
Emergency care	TI = (Emergency Care OR Critical OR Ambulatory OR Urgent OR Emergency Medical Service OR Emergency Department OR Emergency) OR AB = (Emergency Care OR Critical OR Ambulatory OR Urgent OR Emergency Medical Service OR Emergency Department OR Emergency)
Traumatic Brain Injuries	TI = (Traumatic Brain Injury OR Brain Injury OR Head Injury OR TBI OR Skull Fracture) OR AB = (Traumatic Brain Injury OR Brain Injury OR Head Injury OR TBI OR Skull Fracture)

TABLE A.4: CINAHL Search Terms

Search	**Query**	**Results**
S1	(MM "Trauma+")	12,636
S2	TI trauma* or AB Trauma*	120,502
S3	TI injur* or AB injur*	220,789
S4	TI wound* or AB wound*	52,290
S5	TI pierc* or AB pierc*	1,414
S6	TI penetrat* or AB penetrat*	12,557
S7	TI broken or AB broken	4,091
S8	TI break or AB break	10,206
S9	(S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8)	358,097
S10	(MH "Prediction Models")	270
S11	TI predict* or AB predict*	372,640
S12	(MM "Prognosis+")	66,949
S13	TI Prognos* or AB Prognos*	111,411
S14	TI risk or AB risk	736,168
S15	TI likelihood or AB likelihood	45,470
S16	(MM "Probability+")	2,980
S17	TI probability or AB probability	39,006
S18	TI chance or AB chance	22,324
S19	(MM "Odds Ratio")	195
S20	TI Odds or AB Odds	127,158
S21	S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20	1,200,859
S22	(MM "Regression+")	1,544
S23	TI regression or AB regression	253,524
S24	TI logistic or AB logistic	122,149
S25	(MM "Machine Learning+")	1,124
S26	TI machine learning or AB machine learning	6,132
S27	(MM "Neural Networks (Computer)")	1,474
S28	TI neural network or AB neural network	4,405
S29	(MM "Decision Trees+") OR (MM "Random Forest")	540
S30	TI Decision Tree or AB Decision Tree	2,382
S31	TI Random Forest or AB Random Forest	1,541
S32	(MM "Artificial Intelligence+")	12,374
S33	TI Artificial Intelligence or AB Artificial Intelligence	3,397
S34	(MM "Models, Statistical+")	9,491
S35	TI Transfer Learning or AB Transfer Learning	802
S36	(MH "Support Vector Machine")	30
S37	TI Support Vector Machine or AB Support Vector Machine	1,853
S38	TI naive bayes or AB naive bayes	328
S39	TI bayesian or AB bayesian	6,698
S40	(MH "Deep Learning")	458

TABLE A.4: CINAHL Search Terms

Search	**Query**	**Results**
S41	TI deep learning or AB deep learning	2,118
S42	(S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41)	299,858
S43	(MM "Mortality+")	31,611
S44	TI mortality or AB mortality	197,529
S45	(MM "Death+")	26,628
S46	TI Death or AB Death	177,026
S47	TI die or AB die	19,251
S48	TI dying or AB dying	16,328
S49	TI surviv* or AB surviv*	210,100
S50	S43 OR S44 OR S45 OR S46 OR S47 OR S48 OR S49	522,711
S51	(MM "Emergency Care+")	27,761
S52	TI Emergency or AB Emergency	129,457
S53	(MM "Critical Care")	15,055
S54	(MM "Intensive Care Units")	15,679
S55	TI Intensive or AB Intensive	90,625
S56	TI Critical or AB Critical	143,675
S57	(MM "Ambulatory Care")	6,724
S58	TI Urgent or AB Urgent	19,121
S59	(MM "Emergency Medical Services")	20,413
S60	S51 OR S52 OR S53 OR S54 OR S55 OR S56 OR S57 OR S58 OR S59	398,356
S61	S9 AND S21 AND S42 AND S50 AND S60	872

Appendix B

Description of ML and Statistical Approaches Seen in Traumatic Brain Injury Systematic Review

TABLE B.1: Description of ML and Statistical Approaches Seen in Systematic Review

Approach	Description and Characteristics
Logistic Regression (LR)	LR is an example of a standard statistical approach, which transforms a linear combination of input variable values into a non-linear sigmoid function (s-shaped) using the logistic function. Following this transformation, model estimates of the outcome value are bound between zero and one and interpreted as the predicted probability the outcome will occur. Setting a threshold for this predicted probability value allows LR to be used for binary classification problems. The individual contributions of different input variables towards model predictions are reported using odds ratios, which makes LR models more interpretable than other more complex ML methods. LR assumes individuals can be separated using a linear decision boundary. In addition, LR requires little to no multicollinearity between input variables for unbiased predictions to be made.
Support Vector Machines (SVM)	SVMs create a linear decision boundary between different classes of the outcome variable in higher dimensional space. Standard SVM involves transforming the original data into a higher dimensional space where the data points are linearly separable by a decision boundary. This boundary is a hyperplane, which separates the data and maximises the margin between the separated classes. The margin being maximised by an SVM refers to the perpendicular Euclidean distance between the closest data points (support vectors) on either side of the hyperplane (Kelleher et al. 2015). Predictions are made by an SVM based on the position of new data points to the decision boundary. Because the predictions made by an SVM only rely on the support vector values, there is a reduced risk of overfitting the data compared to other ML approaches (Kelleher et al. 2015).
Artificial Neural Networks (ANN)	Artificial Neural Networks (ANN) are a collection of iterative error-based learning approaches, which use layers of connected nodes (neurons) to model complex non-linear relationships between outcome and predictor variables. Connections between nodes in an ANN represent the strength of association between features. While unsupervised ANN algorithms exist, those seen in this review were all supervised. Supervised ANN are trained using back-propagation of the weighted connections and stochastic gradient descent. Effectively this means the ANN will make a single pass over the training data, identify where wrong predictions were made on average and iteratively update the weights of the connections such that fewer errors are made in the next pass of the training data. This approach is known as stochastic gradient descent with the intention that an ANN will be trained once the average error stops decreasing, falls below a predefined threshold, or passes a certain number of pre-defined iterations. This learning approach does raise the issue of local minima, where the model settles on a sub-optimal outcome due to the complexity of the underlying error surface. However, local minima are not by-default problematic as a global minima will likely be overfit to the training data.

TABLE B.1: Description of ML and Statistical Approaches Seen in Systematic Review

Approach	**Description and Characteristics**
Naïve Bayes Classifier (NB)	Naïve Bayes models are a relatively simple ML method, which have fast training times and require less computing power to run than other more complex approaches. This is achieved through the application of Bayes theorem and the use of the class conditional independence (CI) assumption to enable high dimensional probabilities to be estimated using one-dimensional conditional probabilities. CI refers to a property where the effect of an attribute value on a given class is independent of the values of the other attributes (Han et al. 2006). The CI assumption is not expected to hold, which is why these models are named naive Bayes. However, NB still achieves comparative predictive performance when applied to real world data for classification. This is because predictions are made based on the relative sizes of the predicted probabilities, rather than the probabilities themselves. This means the model is somewhat robust to errors when calculating the exact probabilities. NB models also benefit from fast training times as relatively few predicted probabilities need to be calculated and these calculations are simplified through the CI assumption.
Decision Tree (DT)	DTs make predictions through a series of ordered tests of the explanatory variable values. Tree based models start with an initial test at the root node, followed by further tests in interior nodes until the tree terminates at a leaf node. When training a DT, each node is chosen in a way that best splits the data such that there is homogeneity of classes in the resulting partitions. This is also referred to as maximising the information gain, by using the explanatory variable that is most informative of the outcome classes to form the node. The aim in a DT is to finish with leaf nodes that have maximum purity, meaning there is maximal separation between the classes. During testing new instances are classified based on the majority class in the resulting leaf node. A major decision to be made in the development of a DT is what stopping criteria to use. Stopping criteria are required because DTs are inherently susceptible to overfitting, as an algorithm can continue until every point occupies a leaf node. Frequently used criteria are stopping splits being made when the number of resulting instances fall under a user defined threshold, minimum information gain, and maximal tree depth. The major advantages of DT models is their ease of interpretation and clarity of how predictions are made. Also, once trained these models require very little computing power to use, as new instances only need to be run through a series of threshold or categorical tests.
Random Forest (RF)	RF is an ensemble learning approach, where the predictions of multiple weak DT models are combined to make predictions with higher performance. In binary outcome prediction, Random Forests work by training many different DT models on resampled data, then classifying a new instance based on the majority vote of all the trees. The reason behind this data resampling is to make the resulting ensemble model more robust to changes in the sample data, with the hope that this translates to better generalisability in new settings.

Appendix C

Traumatic Brain Injury Systematic Review Included Study Predictors Table

TABLE C.1: Included Study predictors

Lead Author (Year)		Dawes (2015)	Lu (2015)	Kelly (2015)	Thelin (2016)	Mueh. (2016)	Alsulaim (2017)	Han (2017)	Junior (2017)	Zeiler (2018)	Winans (2020)	O'Briain (2018)	Najafi (2018)	Raj (2019)	Abujaber (2020)	Wan-Ting (2020)	Prosser (2020)	Fontoura Solla (2020)	Wu (2020)	Zeiler (2020)	Kim (2020	Amorim (2020)
Demographic																						
Age	15	Yes	Yes	Yes	Yes			Yes	Yes	Yes	Yes	Yes		Yes	Yes			Yes	Yes	Yes		Yes
Gender	8	Yes	Yes	Yes						Yes	Yes					Yes			Yes			Yes
Ethnicity/Race	3	Yes		Yes															Yes			
Clinical:																						
GCS Score	10	Yes	Yes		Yes			Yes	Yes	Yes			Yes		Yes	Yes						Yes
GCS Eye	2												Yes						Yes			
GCS Verbal	2												Yes						Yes			
GCS Motor Score	7					Yes							Yes					Yes	Yes	Yes	Yes	Yes
IMPACT	2				Yes															Yes		
ISS	5	Yes		Yes		Yes					Yes								Yes			
Comorbidities	4	Yes	Yes									Yes			Yes							
Pupillary Reactivity	11	Yes			Yes	Yes		Yes			Yes		Yes	Yes				Yes		Yes	Yes	Yes
Pupil Size	5		Yes			Yes			Yes		Yes		Yes									

TABLE C.1: Included Study predictors

Lead Author (Year)		Dawes (2015)	Lu (2015)	Kelly (2015)	Thelin (2016)	Mueh. (2016)	Alsulaim (2017)	Han (2017)	Junior (2017)	Zeiler (2018)	Winans (2020)	O'Briain (2018)	Najafi (2018)	Raj (2019)	Abujaber (2020)	Wan-Ting (2020)	Prosser (2020)	Fontoura Solla (2020)	Wu (2020)	Zeiler (2020)	Kim (2020	Amorim (2020)
In-hospital Complications	2														Yes				Yes			
Injury Characteristics																						
MOI	3	Yes									Yes				Yes							
Physiological:																						
Heart Rate	4	Yes										Yes	Yes		Yes							
Systolic Blood Pressure (SBP)	4	Yes											Yes		Yes				Yes			
Diastolic blood pressure (DBP)	2												Yes									
INR (int norm rat)	4	Yes				Yes												Yes				Yes
Blood Glucose Level	3		Yes		Yes							Yes										

TABLE C.1: Included Study predictors

Lead Author (Year)		Dawes (2015)	Lu (2015)	Kelly (2015)	Thelin (2016)	Mueh. (2016)	Alsulaim (2017)	Han (2017)	Junior (2017)	Zeiler (2018)	Winans (2020)	O'Briain (2018)	Najafi (2018)	Raj (2019)	Abujaber (2020)	Wan-Ting (2020)	Prosser (2020)	Fontoura Solla (2020)	Wu (2020)	Zeiler (2020)	Kim (2020	Amorim (2020)
White Blood Cell Count	2		Yes									Yes										
Hypoxemia	2				Yes																	yes
Hypotension	3				Yes															Yes		yes
Hypoxia	2							Yes												Yes		
Intracranial Pressure (ICP) monitor information (catchall)	3									Yes				Yes						Yes		
Cerebral perfusion pressure (CPP)	2									Yes				Yes								
Mean Arterial Pressure (MAP)	4									Yes		Yes	Yes	Yes								
Pressure Reactivity Index (PRx)	2									Yes										Yes		
Respiratory rate	2											Yes	Yes									

TABLE C.1: Included Study predictors

Lead Author (Year)		Dawes (2015)	Lu (2015)	Kelly (2015)	Thelin (2016)	Mueh. (2016)	Alsulaim (2017)	Han (2017)	Junior (2017)	Zeiler (2018)	Winans (2020)	O'Briain (2018)	Najafi (2018)	Raj (2019)	Abujaber (2020)	Wan-Ting (2020)	Prosser (2020)	Fontoura Solla (2020)	Wu (2020)	Zeiler (2020)	Kim (2020	Amorim (2020)
Alcohol-Blood Level	2														Yes				Yes			
Radiology:																						
Subdural hematoma	3	Yes	Yes					Yes														
Subarachnoid haemorrhage	4	Yes																	Yes	Yes		Yes
epidural haemorrhage	2																		Yes			Yes
subdural haemorrhage	2																		Yes			Yes
Epidural haematoma (Mass?)	2																		Yes	Yes		
Epidural hematoma	2	Yes	Yes																			

TABLE C.1: Included Study predictors

Lead Author (Year)		Dawes (2015)	Lu (2015)	Kelly (2015)	Thelin (2016)	Mueh. (2016)	Alsulaim (2017)	Han (2017)	Junior (2017)	Zeiler (2018)	Winans (2020)	O'Briain (2018)	Najafi (2018)	Raj (2019)	Abujaber (2020)	Wan-Ting (2020)	Prosser (2020)	Fontoura Solla (2020)	Wu (2020)	Zeiler (2020)	Kim (2020	Amorim (2020)
CT Diagnosis/Findings	2		Yes												Yes							
Cistern Status (traj, perf, penet)	2					Yes		Yes														
intraventricular haemorrhage	2					Yes															Yes	
Miscellaneous:																						
Hospital Fixed Effects	3	Yes										Yes			Yes							
Regional Fixed Effects	2			Yes															Yes			

Appendix D

Traumatic Brain Injury Systematic Review Model Performance Table

TABLE D.1: Systematic Review Model Performance Table

Author (Year)	Outcome	Prev. (%)	Study N.	MN	Var. Selection	Model Type	Training Split	VM	AUC	Sen (%)	Spec (%)	Acc (%)	NPV (%)	PPV (%)	Cal.	MDH
Abujaber (2020)	In hospital Mortality	12.5	1620	1	All Included	ANN	70/30	RS	0.94	62	96	92	96	66	-	CCA
Abujaber (2020)	In hospital Mortality	12.5	1620	2	All Included	SVM	70/30	RS	0.96	73	99	96	88	88	-	CCA
Alsulaim (2017)	In hospital Mortality	9.9	93,397	1	Hypothesis Based	LR	-	-	0.70	-	-	-	-	-	-	CCA
Alsulaim (2017)	In hospital Mortality	9.9	93,397	2	Hypothesis Based	LR	-	-	0.77	-	-	-	-	-	-	CCA
Alsulaim (2017)	In hospital Mortality	9.9	93,397	3	Hypothesis Based	LR	-	-	0.83	-	-	-	-	-	-	CCA
Alsulaim (2017)	In hospital Mortality	9.9	93,397	4	Hypothesis Based	LR	-	-	0.81	-	-	-	-	-	-	CCA
Alsulaim (2017)	In hospital Mortality	9.9	93,397	5	Hypothesis Based	LR	-	-	0.86	-	-	-	-	-	-	CCA
Alsulaim (2017)	In hospital Mortality	9.9	93,397	6	Hypothesis Based	LR	-	-	0.87	-	-	-	-	-	-	CCA
Amorim (2020)	14 Day Mortality	22.8	517	1	Previous Literature	NB	Unclear	5-F CV	0.91	-	-	-	-	-	-	MI
Amorim (2020)	14 Day Mortality	22.8	517	2	Previous Literature	BGLM	Unclear	5-F CV	0.88	-	-	-	-	-	-	MI
Amorim (2020)	14 Day Mortality	22.8	517	3	Previous Literature	PDA	Unclear	5-F CV	0.88	-	-	-	-	-	-	MI
Amorim (2020)	14 Day Mortality	22.8	517	4	Previous Literature	RF	Unclear	5-F CV	0.88	-	-	-	-	-	-	MI

TABLE D.1: Systematic Review Model Performance Table

Author (Year)	Outcome	Prev. (%)	Study N.	MN	Var. Selection	Model Type	Training Split	VM	AUC	Sen (%)	Spec (%)	Acc (%)	NPV (%)	PPV (%)	Cal.	MDH
Amorim (2020)	In hospital Mortality	30.9	517	5	Previous Literature	RF	Unclear	5-F CV	0.84	-	-	-	-	-	-	MI
Amorim (2020)	In hospital Mortality	30.9	517	6	Previous Literature	GPLM	Unclear	5-F CV	0.83	-	-	-	-	-	-	MI
Amorim (2020)	In hospital Mortality	30.9	517	7	Previous Literature	SGB	Unclear	5-F CV	0.82	-	-	-	-	-	-	MI
Amorim (2020)	In hospital Mortality	30.9	517	8	Previous Literature	PDA	Unclear	5-F CV	0.80	-	-	-	-	-	-	MI
Dawes (2015)	In hospital Mortality	38.8	822	1	Previous Literature	LR	-	-	0.94	-	-	-	-	-	-	MI
Fontoura-Solla (2020)	14 Day Mortality	22.8	517	1	Univariate Models	LR	-	-	0.85	-	-	-	-	-	HL and Brier	CCA
Fontoura-Solla (2020)	14 Day Mortality	22.8	517	2	Univariate Models	LR	-	-	0.81	-	-	-	-	-	HL and Brier	CCA
Han (2017)	14 Day Mortality	-	300	1	Univariate Models	LR	-	BS	0.81	69	82	-	75	77	HL and Brier	CCA
Han (2017)	14 Day Mortality	-	300	2	Univariate Models	LR	-	BS	0.84	76	83	-	79	80	HL and Brier	CCA
Han (2017)	6 month mortality	-	300	3	Univariate Models	LR	-	BS	0.81	80	71	-	75	77	HL and Brier	CCA

TABLE D.1: Systematic Review Model Performance Table

Author (Year)	Outcome	Prev. (%)	Study N.	MN	Var. Selection	Model Type	Training Split	VM	AUC	Sen (%)	Spec (%)	Acc (%)	NPV (%)	PPV (%)	Cal.	MDH
Han (2017)	6 month mortality	-	300	4	Univariate Models	LR	-	BS	0.84	82	72	-	77	77	HL and Brier	CCA
Junior (2017)	In hospital Mortality	19.1	1275	1	Unclear	LR	-	-	0.77	-	-	-	-	-	HL	-
Kelly (2015)	30 Day Mortality	21, 16	3496	1	Univariate Models	LR	-	-	0.84	-	-	-	-	-	-	CCA
Kelly (2015)	6 month mortality	21, 16	3496	2	Univariate Models	LR	-	-	0.82	-	-	-	-	-	-	CCA
Kim (2020)	In hospital Mortality	48.1	54	1	Univariate Models	DT	-	-	-	-	-	-	-	-	-	CCA
Lu (2015)	6 month mortality	25.2	115	1	Situation Based	ANN	-	10-F CV	0.81	62	90	-	-	-	CR	CCA
Lu (2015)	6 month mortality	25.2	115	2	Situation Based	NB	-	10-F CV	0.90	81	91	-	-	-	CR	CCA
Lu (2015)	6 month mortality	25.2	115	3	Situation Based	DT	-	10-F CV	0.78	70	92	-	-	-	CR	CCA
Lu (2015)	6 month mortality	25.2	115	4	Situation Based	LR	-	10-F CV	0.87	68	91	-	-	-	CR	CCA

TABLE D.1: Systematic Review Model Performance Table

Author (Year)	Outcome	Prev. (%)	Study N.	MN	Var. Selection	Model Type	Training Split	VM	AUC	Sen (%)	Spec (%)	Acc (%)	NPV (%)	PPV (%)	Cal.	MDH
Muehlschlegel (2016)	Inpatient Survival	42.4	413	1	Univariate Models	LR	-	-	0.93	-	-	-	-	-	-	CCA
Muehlschlegel (2016)	Inpatient Survival	42.4	413	2	Univariate Models	LR	-	-	0.96	-	-	-	-	-	-	CCA
Muehlschlegel (2016)	Inpatient Survival	42.4	413	3	Univariate Models	LR	-	-	0.96	-	-	-	-	-	-	CCA
Muehlschlegel (2016)	Inpatient Survival	42.4	413	4	Univariate Models	LR	-	-	0.95	-	-	-	-	-	-	CCA
Muehlschlegel (2016)	Inpatient Survival	42.4	413	5	Univariate Models	LR	-	-	0.97	-	-	-	-	-	-	CCA
Najafi (2018)	24hr Mortality	14	185	1	Univariate Models	LR	-	-	-	-	-	93	-	-	HL	CCA
O'Briain (2018)	In hospital Mortality	17.2	24148	1	Unclear	LR	-	-	0.88	-	-	-	-	-	-	CCA
O'Briain (2018)	In hospital Mortality	17.2	24148	2	Unclear	LR	-	-	0.88	-	-	-	-	-	-	CCA
Prosser (2020)	30 Day Survival	23, 8	355	1	Unclear	LR	-	-	0.88	-	-	-	-	-	-	MI

TABLE D.1: Systematic Review Model Performance Table

Author (Year)	Outcome	Prev. (%)	Study N.	MN	Var. Selection	Model Type	Training Split	VM	AUC	Sen (%)	Spec (%)	Acc (%)	NPV (%)	PPV (%)	Cal.	MDH
Raj (2019)	30 Day Mortality	19	472	1	Recursive Feature Elimination	LR	-	5-F CV	0.67	-	-	-	-	-	-	CCA
Raj (2019)	30 Day Mortality	19	472	2	Recursive Feature Elimination	LR	-	5-F CV	0.72	-	-	-	-	-	-	CCA
Thelin (2016)	Long Term GOS	20	417	1	Univariate Models	LR	-	-	-	-	-	-	-	-	-	MI
Thelin (2016)	Long Term GOS	20	417	2	Univariate Models	LR	-	-	-	-	-	-	-	-	-	MI
Thelin (2016)	Long Term GOS	20	417	3	Univariate Models	LR	-	-	-	-	-	-	-	-	-	MI
Thelin (2016)	Long Term GOS	20	417	4	Univariate Models	LR	-	-	-	-	-	-	-	-	-	MI
Thelin (2016)	Long Term GOS	20	417	5	Univariate Models	LR	-	-	-	-	-	-	-	-	-	MI
Thelin (2016)	Long Term GOS	20	417	6	Univariate Models	LR	-	-	-	-	-	-	-	-	-	MI
Wan-Ting (2020)	In hospital Mortality	24.7	438	1	Situation Based	LR	-	-	0.76	71	75	-	89	49	HL	CCA
Winans (2020)	In hospital Mortality	40	402	1	Unclear	LR	-	-	0.91	75	93	-	-	-	-	CCA

TABLE D.1: Systematic Review Model Performance Table

Author (Year)	Outcome	Prev. (%)	Study N.	MN	Var. Selection	Model Type	Training Split	VM	AUC	Sen (%)	Spec (%)	Acc (%)	NPV (%)	PPV (%)	Cal.	MDH
Wu (2020)	In hospital Mortality	9	212666	1	Previous Literature	SVM	80/20	10-F CV	0.79	75	83	79	-	31	-	CCA
Wu (2020)	In hospital Mortality	9	212666	2	Previous Literature	SVM	80/20	-	-	-	-	79	-	-	-	CCA
Wu (2020)	In hospital Mortality	9	212666	3	Previous Literature	SVM	80/20	-	-	-	-	80	-	-	-	CCA
Wu (2020)	In hospital Mortality	9	212666	4	Previous Literature	LR	80/20	-	-	-	-	80	-	-	-	CCA
Wu (2020)	In hospital Mortality	9	212666	5	Previous Literature	KNN	80/20	-	-	-	-	81	-	-	-	CCA
Wu (2020)	In hospital Mortality	9	212666	6	Previous Literature	DT	80/20	-	-	-	-	79	-	-	-	CCA
Wu (2020)	In hospital Mortality	9	212666	7	Previous Literature	GNBC	80/20	-	-	-	-	74	-	-	-	CCA
Wu (2020)	In hospital Mortality	9	212666	8	Previous Literature	LDA	80/20	-	-	-	-	81	-	-	-	CCA
Wu (2020)	In hospital Mortality	9	212666	9	Previous Literature	SVM	80/20	10F CV	0.76	71	-	81	-	27	-	CCA
Zeiler (2018)	6 month mortality	-	358	1	Hypothesis Based	LR	-	-	0.80	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	2	Hypothesis Based	LR	-	-	0.82	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	3	Hypothesis Based	LR	-	-	0.84	-	-	-	-	-	-	-

TABLE D.1: Systematic Review Model Performance Table

Author (Year)	Outcome	Prev. (%)	Study N.	MN	Var. Selection	Model Type	Training Split	VM	AUC	Sen (%)	Spec (%)	Acc (%)	NPV (%)	PPV (%)	Cal.	MDH
Zeiler (2018)	6 month mortality	-	358	4	Hypothesis Based	LR	-	-	0.86	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	5	Hypothesis Based	LR	-	-	0.86	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	6	Hypothesis Based	LR	-	-	0.85	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	7	Hypothesis Based	LR	-	-	0.76	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	8	Hypothesis Based	LR	-	-	0.80	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	9	Hypothesis Based	LR	-	-	0.80	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	10	Hypothesis Based	LR	-	-	0.81	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	11	Hypothesis Based	LR	-	-	0.82	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	12	Hypothesis Based	LR	-	-	0.81	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	13	Hypothesis Based	LR	-	-	0.76	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	14	Hypothesis Based	LR	-	-	0.80	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	15	Hypothesis Based	LR	-	-	0.80	-	-	-	-	-	-	-

TABLE D.1: Systematic Review Model Performance Table

Author (Year)	Outcome	Prev. (%)	Study N.	MN	Var. Selection	Model Type	Training Split	VM	AUC	Sen (%)	Spec (%)	Acc (%)	NPV (%)	PPV (%)	Cal.	MDH
Zeiler (2018)	6 month mortality	-	358	16	Hypothesis Based	LR	-	-	0.82	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	17	Hypothesis Based	LR	-	-	0.82	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	18	Hypothesis Based	LR	-	-	0.82	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	19	Hypothesis Based	LR	-	-	0.78	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	20	Hypothesis Based	LR	-	-	0.81	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	21	Hypothesis Based	LR	-	-	0.80	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	22	Hypothesis Based	LR	-	-	0.82	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	23	Hypothesis Based	LR	-	-	0.83	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	24	Hypothesis Based	LR	-	-	0.80	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	25	Hypothesis Based	LR	-	-	0.76	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	26	Hypothesis Based	LR	-	-	0.80	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	27	Hypothesis Based	LR	-	-	0.79	-	-	-	-	-	-	-

TABLE D.1: Systematic Review Model Performance Table

Author (Year)	Outcome	Prev. (%)	Study N.	MN	Var. Selection	Model Type	Training Split	VM	AUC	Sen (%)	Spec (%)	Acc (%)	NPV (%)	PPV (%)	Cal.	MDH
Zeiler (2018)	6 month mortality	-	358	28	Hypothesis Based	LR	-	-	0.82	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	29	Hypothesis Based	LR	-	-	0.82	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	30	Hypothesis Based	LR	-	-	0.80	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	31	Hypothesis Based	LR	-	-	0.74	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	32	Hypothesis Based	LR	-	-	0.74	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	33	Hypothesis Based	LR	-	-	0.74	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	34	Hypothesis Based	LR	-	-	0.74	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	35	Hypothesis Based	LR	-	-	0.75	-	-	-	-	-	-	-
Zeiler (2018)	6 month mortality	-	358	36	Hypothesis Based	LR	-	-	0.74	-	-	-	-	-	-	-
Zeiler (2020)	6 month mortality	20.7	193	1	Univariate Models	LR	-	-	0.71	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	2	Univariate Models	LR	-	-	0.78	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	3	Univariate Models	LR	-	-	0.81	-	-	-	-	-	-	Imp

TABLE D.1: Systematic Review Model Performance Table

Author (Year)	Outcome	Prev. (%)	Study N.	MN	Var. Selection	Model Type	Training Split	VM	AUC	Sen (%)	Spec (%)	Acc (%)	NPV (%)	PPV (%)	Cal.	MDH
Zeiler (2020)	6 month mortality	20.7	193	4	Univariate Models	LR	-	-	0.81	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	5	Univariate Models	LR	-	-	0.78	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	6	Univariate Models	LR	-	-	0.80	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	7	Univariate Models	LR	-	-	0.81	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	8	Univariate Models	LR	-	-	0.81	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	9	Univariate Models	LR	-	-	0.82	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	10	Univariate Models	LR	-	-	0.83	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	11	Univariate Models	LR	-	-	0.81	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	12	Univariate Models	LR	-	-	0.82	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	13	Univariate Models	LR	-	-	0.82	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	14	Univariate Models	LR	-	-	0.81	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	15	Univariate Models	LR	-	-	0.81	-	-	-	-	-	-	Imp

TABLE D.1: Systematic Review Model Performance Table

Author (Year)	Outcome	Prev. (%)	Study N.	MN	Var. Selection	Model Type	Training Split	VM	AUC	Sen (%)	Spec (%)	Acc (%)	NPV (%)	PPV (%)	Cal.	MDH
Zeiler (2020)	6 month mortality	20.7	193	16	Univariate Models	LR	-	-	0.82	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	17	Univariate Models	LR	-	-	0.67	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	18	Univariate Models	LR	-	-	0.77	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	19	Univariate Models	LR	-	-	0.82	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	20	Univariate Models	LR	-	-	0.82	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	21	Univariate Models	LR	-	-	0.75	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	22	Univariate Models	LR	-	-	0.78	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	23	Univariate Models	LR	-	-	0.79	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	24	Univariate Models	LR	-	-	0.79	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	25	Univariate Models	LR	-	-	0.81	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	26	Univariate Models	LR	-	-	0.82	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	27	Univariate Models	LR	-	-	0.81	-	-	-	-	-	-	Imp

TABLE D.1: Systematic Review Model Performance Table

Author (Year)	Outcome	Prev. (%)	Study N.	MN	Var. Selection	Model Type	Training Split	VM	AUC	Sen (%)	Spec (%)	Acc (%)	NPV (%)	PPV (%)	Cal.	MDH
Zeiler (2020)	6 month mortality	20.7	193	28	Univariate Models	LR	-	-	0.81	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	29	Univariate Models	LR	-	-	0.82	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	30	Univariate Models	LR	-	-	0.80	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	31	Univariate Models	LR	-	-	0.81	-	-	-	-	-	-	Imp
Zeiler (2020)	6 month mortality	20.7	193	32	Univariate Models	LR	-	-	0.81	-	-	-	-	-	-	Imp

Prev = Prevalence of outcome in sample, Study N. = Study sample size, MN = Model Number, Var. Selection Method = Variable Selection Method, VM = Validation Method, Sen = Sensitivity, Spec = Specificity, Acc = Accuracy, NPV = Negative Predictive Value, PPV = Positive Predictive Value, Cal. = Calibration, MDH = Missing Data Handling, CCA = Complete Case Analysis, MI = Multiple Imputation, Imp = Imputation, - = Not Reported, 10-F CV = 10-fold cross validation, 5-F CV = 5 fold cross validation, CR = Correction Rate, HL = Hosmer-Lemeshow P-Value Reported, Brier = Brier Score Reported, LR = Logistic Regression, SVM = Support Vector Machine, ANN = Artificial Neural Network, NB = Naïve Bayes Classifier, DT = Decision Tree, RF = Random Forrest, KNN = K-Nearest Neighbours, BGLM = Bayesian Generalised Linear Model , PDA = Penalised discriminant analysis, GPLM = Generalised Partial Least Squares, SGB = Stochastic Gradient Boosting, GNBC = Gaussian Naïve Bayes Classifier, LDA = Linear Discriminate Analysis, BS = Bootstrapping

Appendix E

Review of Systematic Reviews Search Strategy

TABLE E.1: Medline Search Terms: 68 Papers returned

Filters	2000-present, English Language, Reviews
Concept	**Terms**
Accidental Falls	exp Accidental Falls (27833)
Older Adults	exp "Aged, 80 and over"/ or exp Aged/ (3436262)
Systematic Reviews	exp "Systematic Review"/ (214346)
Multi-morbidity and chronic disease	exp Comorbidity/ or exp Multiple Chronic Conditions/ or exp Multimorbidity/ or exp Chronic Disease/ or exp neoplasms/ or exp musculoskeletal diseases/ or exp digestive system diseases/ or exp stomatognathic diseases/ or exp respiratory tract diseases/ or exp otorhinolaryngologic diseases/ or exp nervous system diseases/ or exp eye diseases/ or exp urogenital diseases/ or exp cardiovascular diseases/ or exp "hemic and lymphatic diseases"/ or exp "congenital, hereditary, and neonatal diseases and abnormalities"/ or exp "skin and connective tissue diseases"/ or exp "nutritional and metabolic diseases"/ or exp endocrine system diseases/ or exp immune system diseases/ (14293225)

TABLE E.2: Embase Search Terms: 103 Papers returned

Filters	2000-present, English Language, Reviews
Concept	**Terms**
Accidental Falls	exp falling/ (48625)
Older Adults	exp aged/ or exp home for the aged/ (3549498)
Systematic Reviews	exp "systematic review"/ (423082)
Multi-morbidity and chronic disease	exp chronic disease/ or exp comorbidity/ or exp multiple chronic conditions/ or exp neoplasm/ or exp musculoskeletal disease/ or exp digestive system disease/ or exp mouth disease/ or exp respiratory tract disease/ or exp ear nose throat disease/ or exp neurologic disease/ or exp eye disease/ or exp urogenital tract disease/ or exp cardiovascular disease/ or exp lymphatic system disease/ or exp skin disease/ or exp connective tissue disease/ or exp metabolic disorder/ or exp endocrine disease/ or exp immunopathology/ (19414193)

TABLE E.3: Web of Science Search Terms: 182 Papers returned

Filters	2000-present, English Language, Reviews
Concept	**Terms**
Accidental Falls	TI=(fall) OR AB =(fall) OR TI=(fallen) OR AB =(fallen) OR TI=(accidental fall) OR AB =(accidental fall) OR TI=(trip) OR AB =(trip) OR TI=(collapse) OR AB =(collapse) (695727)
Older Adults	TI=(Aged) OR AB =(Aged) OR TI=(Elderly) OR AB =(Elderly) OR TI=(over 65) OR AB =(over 65) OR TI=(older) OR AB =(older) (5179919)
Systematic Reviews	TI=(systematic review) OR AB =(systematic review) (379613)
Multi-morbidity and chronic disease	TI=(multimorbidity) OR AB =(multimorbidity) OR TI=(chronic) OR AB =(chronic) OR TI=(multimorbid) OR AB =(multimorbid) OR TI=(multiple health conditions) OR AB =(multiple health conditions) OR TI=(comorbidities) OR AB =(comorbidities) OR TI=(comorbid) OR AB =(comorbid) OR TI=(comorbidity) OR AB =(comorbidity) (1517019)

TABLE E.4: PROSPERO Search Terms: 136 Papers returned

Filters	None
Concept	Terms
Accidental Falls	MeSH DESCRIPTOR Accidental Falls EXPLODE ALL TREES OR accidental falls OR falls
Older Adults	MeSH DESCRIPTOR Aged EXPLODE ALL TREES OR MeSH DESCRIPTOR Aged, 80 and over EXPLODE ALL TREES OR Elderly OR 65+ OR aged OR 65 OR 65 and over OR aged 65 and over
Multi-morbidity and chronic disease	MeSH DESCRIPTOR Chronic Disease EXPLODE ALL TREES OR MeSH DESCRIPTOR Multiple Chronic Conditions EXPLODE ALL TREES OR MeSH DESCRIPTOR Comorbidity EXPLODE ALL TREES OR MeSH DESCRIPTOR Multimorbidity EXPLODE ALL TREES OR comorbidity OR chronic disease

TABLE E.5: CINAHL Search Terms: 17 Papers returned

Filters	2000-present, English Language, Reviews
Concept	Terms
Accidental Falls	(MH "Accidental Falls") (26,182)
Older Adults	(MH "Aged+") OR (MH "Aged, 80 and Over+") (939,334)
Systematic Reviews	(MH "Systematic Review") (118,910)
Multi-morbidity and chronic disease	(MH "Mental Disorders, Chronic") OR (MH "Chronic Disease+") OR (MH "Comorbidity") OR (TI multimorbidity) OR (AB multimorbidity) (143,544)

Appendix F

Fall Definition Codes

TABLE F.1: Fall Definition Code Descriptions

Category	ICD-10 Code Description
Diagnosis Codes	Contusion/abrasion*
	Dislocation/fracture/joint injury/amputation*
	Head injury*
	Laceration
	Muscle/tendon injury
	Soft tissue inflammation
	Soft Tissue Inflammation (Hip) (Bilateral)
	Sprain/ligament injury
	Sprain/ligament injury (Cervical spine) (Not applicable)
Treatment Codes	Wound Closure/Dressing
	Wound closure (excluding sutures) - other (e.g. clips)
	Wound closure (excluding sutures) - wound glue
	Wound closure (excluding sutures) - steristrips
	Sutures - primary sutures
	Sutures - secondary/complex suture
	Dressing - dressing minor wound/burn/eye
	Breaks/ Dislocations following trauma
	Splint
	Plaster of Paris - application Plaster of Paris
	Manipulation - manipulation of upper limb fracture
	Manipulation - manipulation of lower limb fracture
	Manipulation - manipulation of dislocation
	Sling/collar cuff/broad arm sling
	Loan of walking aid (crutches)

Appendix G

Frailty Index Codes

TABLE G.1: Charlson Comorbidity Index Codes

Condition	**ICD-10 Codes**	Points
AIDS	B20, B21, B22, B24	6
Any malignancy, including lymphoma and leukemia, except malignant neoplasm of skin	C00, C01, C02, C03, C04, C05, C06, C07, C08, C09, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, C25, C26, C30, C31, C32, C33, C34, C37, C38, C39, C40, C41, C43, C45, C46, C47, C48, C49, C50, C51, C52, C53, C54, C55, C56, C57, C58, C60, C61, C62, C63, C64, C65, C66, C67, C68, C69, C70, C71, C72, C73, C74, C75, C76, C81, C82, C83, C84, C85, C88, C90, C91, C92, C93, C94, C95, C96, C97	2
Cerebrovascular disease	G45, G46, H34.0, I60, I61, I62, I63, I64, I65, I66, I67, I68, I69	1
Chronic pulmonary disease	I27.8, I27.9, J40, J41, J42, J43, J44, J45, J46, J47, J60, J61, J62, J63, J64, J65, J66, J67, J68.4, J70.1, J70.3	1
Congestive heart failure	I09.9, I11.0, I13.0, I13.2, I25.5, I42.0, I42.5, I42.6, I42.7, I42.8, I42.9, I43, I50, P29.0	1
Dementia	F00, F01, F02, F03, F05.1, G30, G31.1	1
Diabetes with chronic complication	E10.2, E10.3, E10.4, E10.5, E10.7, E11.2, E11.3, E11.4, E11.5, E11.7, E12.2, E12.3, E12.4, E12.5, E12.7, E13.2, E13.3, E13.4, E13.5, E13.7, E14.2, E14.3, E14.4, E14.5, E14.7	2
Diabetes without chronic complication	E10.0, E10.1, E10.6, E10.8, E10.9, E11.0, E11.1, E11.6, E11.8, E11.9, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.1, E13.6, E13.8, E13.9, E14.0, E14.1, E14.6, E14.8, E14.9	1
Hemiplegia or paraplegia	G04.1, G11.4, G80.1, G80.2, G81, G82, G83.0, G83.1, G83.2, G83.3, G83.4, G83.9	2
Metastatic solid tumour	C77, C78, C79, C80	6
Mild liver disease	B18, K70.0, K70.1, K70.2, K70.3, K70.9, K71.3, K71.4, K71.5, K71.7, K73, K74, K76.0, K76.2, K76.3, K76.4, K76.8, K76.9, Z94.4	1
Moderate or severe liver disease	I85.0, I85.9, I86.4, I98.2, K70.4, K71.1, K72.1, K72.9, K76.5, K76.6, K76.7	3
Myocardial infarction	I21, I22, I25.2	1
Peptic ulcer disease	K25, K26, K27, K28	1
Peripheral vascular disease	I70, I71, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9	1
Renal disease	I12.0, I13.1, N03.2, N03.3, N03.4, N03.5, N03.6, N03.7, N05.2, N05.3, N05.4, N05.5, N05.6, N05.7, N18, N19, N25.0, Z49.0, Z49.1, Z49.2, Z94.0, Z99.2	2
Rheumatic disease	M05, M06, M31.5, M32, M33, M34, M35.1, M35.3, M36.0	1

TABLE G.2: Electronic Frailty Index Codes

EFI Deficits	**ICD-10 Codes**
Activity limitation	Assumed Present
Anaemia and haematinic deficiency	D50, D51, D52, D53, D56, D58, D59, D61, D62, D63, D64
Arthritis	M00, M05, M06, M13, M15, M16, M17, M18, M19
Atrial fibrillation	I48
Cerebrovascular disease	G45, G46, I60, I61, I62, I63, I64, I67, I68, I69,
Chronic kidney disease	N04, N08, N18, Q61
Diabetes	E10, E11, E14
Dizziness	R42
Dyspnoea	R06, W84
Falls	W00, W01, W03, W04, W05, W06, W07, W08, W10, W17, W18, W19, Y30
Foot problems	S90, S91, S92, S93, S99
Fragility fracture	M80
Hearing impairment	H61, H83, H90, H91, H93
Heart failure	I27, I42, I43, I50, I51
Heart valve disease	I05, I07, I08, I34, I35, I36, I37, I38, Q22
Housebound	Assumed present
Hypertension	I10
Hypotension/syncope	I95, R55
Ischaemic heart disease	I24, I25
Memory and cognitive problems	F00, F05, F06, F70, F71, F79, G30, R41
Mobility and transfer problems	Assumed present
Osteoporosis	M80, M81
Parkinsonism and tremor	G20, G21, G23
Peptic ulcer	K27
Peripheral vascular disease	I73, Q27
Polypharmacy	Assumed present
Respiratory disease	J9, C78, J06, J22, J39, J42, J43, J44, J47, J60, J61, J67, J80, J84, J94, J95, J96, J98, J99, Q33
Skin ulcer	L89, L97, L98
Sleep disturbance	G47
Social vulnerability	Assumed present
Thyroid disease	E03, E05, E06, E07
Urinary incontinence	R32
Urinary system disease	C68, D41, N20, N21, N30, N32, N35, N39, N48, N73, N76, N81, N89, N90, N95, Q54, Q63, R39, T83
Visual impairment	H01, H02, H04, H05, H10, H17, H18, H20, H21, H25, H26, H28, H31, H35, H36, H47, H49, H51, H54, H55, H57, H58, S05, W44,
Weight loss and anorexia	F50

TABLE G.3: Health Frailty Risk Score Codes

ICD-10 code	**HFRS Points**	**Code Description**
F00	7.1	Dementia in Alzheimer's disease
G81	4.4	Hemiplegia
G30	4	Alzheimer's disease
I69	3.7	Sequelae of cerebrovascular disease
R29	3.6	Other symptoms and signs involving the nervous and musculoskeletal Systems
N39	3.2	Other disorders of urinary system (includes urinary tract infection and urinary incontinence)
F05	3.2	Delirium, not induced by alcohol and other psychoactive substances
W19	3.2	Unspecified fall
S00	3.2	Superficial injury of head
R31	3	Unspecified haematuria
B96	2.9	Other bacterial agents as the cause of diseases classified to other chapters (secondary code)
R41	2.7	Other symptoms and signs involving cognitive functions and awareness
R26	2.6	Abnormalities of gait and mobility
I67	2.6	Other cerebrovascular diseases
R56	2.6	Convulsions, not elsewhere classified
R40	2.5	Somnolence, stupor and coma
T83	2.4	Complications of genitourinary prosthetic devices, implants and grafts
S06	2.4	Intracranial injury
S42	2.3	Fracture of shoulder and upper arm
E87	2.3	Other disorders of fluid, electrolyte and acidbase balance
M25	2.3	Other joint disorders, not elsewhere classified
E86	2.3	Volume depletion
R54	2.2	Senility
Z50	2.1	Care involving use of rehabilitation procedures
F03	2.1	Unspecified dementia
W18	2.1	Other fall on same level
Z75	2	Problems related to medical facilities and other health care
F01	2	Vascular dementia
S80	2	Superficial injury of lower leg
L03	2	Cellulitis
H54	1.9	Blindness and low vision
E53	1.9	Deficiency of other B group vitamins
Z60	1.8	Problems related to social environment
G20	1.8	Parkinson's disease
R55	1.8	Syncope and collapse
S22	1.8	Fracture of rib(s), sternum and thoracic spine
K59	1.8	Other functional intestinal disorders
N17	1.8	Acute renal failure
L89	1.7	Decubitus ulcer

Z22	1.7	Carrier of infectious disease
B95	1.7	Streptococcus and staphylococcus as the cause of diseases classified to other chapters
L97	1.6	Ulcer of lower limb, not elsewhere classified
R44	1.6	Other symptoms and signs involving general sensations and perceptions
K26	1.6	Duodenal ulcer
I95	1.6	Hypotension
N19	1.6	Unspecified renal failure
A41	1.6	Other septicaemia
Z87	1.5	Personal history of other diseases and conditions
J96	1.5	Respiratory failure, not elsewhere classified
X59	1.5	Exposure to unspecified factor
M19	1.5	Other arthrosis
G40	1.5	Epilepsy
M81	1.4	Osteoporosis without pathological fracture
S72	1.4	Fracture of femur
S32	1.4	Fracture of lumbar spine and pelvis
E16	1.4	Other disorders of pancreatic internal secretion
R94	1.4	Abnormal results of function studies
N18	1.4	Chronic renal failure
R33	1.3	Retention of urine
R69	1.3	known and unspecified causes of morbidity
N28	1.3	Other disorders of kidney and ureter, not elsewhere classified
R32	1.2	Unspecified urinary incontinence
G31	1.2	Other degenerative diseases of nervous system, not elsewhere classified
Y95	1.2	Nosocomial condition
S09	1.2	Other and unspecified injuries of head
R45	1.2	Symptoms and signs involving emotional state
G45	1.2	Transient cerebral ischaemic attacks and related syndromes
Z74	1.1	Problems related to care-provider dependency
M79	1.1	Other soft tissue disorders, not elsewhere classified
W06	1.1	Fall involving bed
S01	1.1	Open wound of head
A04	1.1	Other bacterial intestinal infections
A09	1.1	Diarrhoea and gastroenteritis of presumed infectious origin
J18	1.1	Pneumonia, organism unspecified
J69	1	Pneumonitis due to solids and liquids
R47	1	Speech disturbances, not elsewhere classified
E55	1	Vitamin D deficiency
Z93	1	Artificial opening status
R02	1	Gangrene, not elsewhere classified

R63	0.9	Symptoms and signs concerning food and fluid intake
H91	0.9	Other hearing loss
W10	0.9	Fall on and from stairs and steps
W01	0.9	Fall on same level from slipping, tripping and stumbling
E05	0.9	Thyrotoxicosis [hyperthyroidism]
M41	0.9	Scoliosis
R13	0.8	Dysphagia
Z99	0.8	Dependence on enabling machines and devices
U80	0.8	Agent resistant to penicillin and related antibiotics
M80	0.8	Osteoporosis with pathological fracture
K92	0.8	Other diseases of digestive system
I63	0.8	Cerebral Infarction
N20	0.7	Calculus of kidney and ureter
F10	0.7	Mental and behavioural disorders due to use of alcohol
Y84	0.7	Other medical procedures as the cause of abnormal reaction of the patient
R00	0.7	Abnormalities of heart beat
J22	0.7	Unspecified acute lower respiratory infection
Z73	0.6	Problems related to life-management difficulty
R79	0.6	Other abnormal findings of blood chemistry
Z91	0.5	Personal history of risk-factors, not elsewhere classified
S51	0.5	Open wound of forearm
F32	0.5	Depressive episode
M48	0.5	Spinal stenosis (secondary code only)
E83	0.4	Disorders of mineral metabolism
M15	0.4	Polyarthrosis
D64	0.4	Other anaemias
L08	0.4	Other local infections of skin and subcutaneous tissue
R11	0.3	Nausea and vomiting
K52	0.3	Other noninfective gastroenteritis and colitis
R50	0.1	Fever of unknown origin

Appendix H

Calderon-Laranaga 2017 Groupings

TABLE H.1: Calderon-Laranaga ICD-10 Code Groups

Grouping	**ICD-10 Codes**
ALLERGY	J301,J302,J303,J304,J450,K522,L20,L23,L500,Z516
ANEMIA	D50,D51,D52,D53,D55,D56,D57,D58,D59,D60,D61,D63,D64
ASTHMA	J45
ATRIAL FIBRILLATION	I48
AUTOIMMUNE DISEASES	I731, L10, L12, L40, L41, L93, L94, L95, M30, M31, M32, M33, M34, M35, M36
BLINDNESS, VISUAL IMPAIRMENT	H54,Z442,Z970
BLOOD AND BLOOD FORMING ORGAN DISEASES	D66, D67, D68, D69, D71, D720, D730, D731, D732, D74, D750, D761, D763, D77, D80, D81, D82, D83, D84, D86, D89
BRADYCARDIAS AND CONDUCTION DISEASES	I441,I442,I443,I453,I455,Z950
CARDIAC VALVE DISEASES	I05, I06, I07, I08, I091, I098, I34, I35, I36, I37, I38, I390, I391, I392, I393, I394, Q22, Q23, Z952, Z953, Z954
CATARACT AND OTHER LENS DISEASES	H25,H26,H27,H28,Q12,Z961
CEREBROVASCULAR DISEASE	G45,G46,I60,I61,I62,I63,I64,I67,I69
CHROMOSOMAL ABNORMALITIES	Q90,Q91,Q92,Q93,Q95,Q96,Q97,Q98,Q99
CHRONIC INFECTIOUS DISEASES	A15, A16, A17, A18, A19, A30, A31, A50, A52, A53, A65, A66, A67, A692, A81, B20, B21, B22, B23, B24, B381, B391, B401, B572, B573, B574, B575, B65, B92, B94, J65, M863, M864, M865, M866
CHRONIC KIDNEY DISEASES	I120, I130, I131, I132, I139, N01, N03, N04, N05, N07, N08, N11, N183, N184, N185, N189, Q60, Q611, Q612, Q613, Q614, Q615, Q618, Q619, Z905, Z940
CHRONIC LIVER DISEASES	B18, K70, K713, K714, K715, K717, K721, K73, K74, K753, K754, K758, K761, K766, K767, K778, Q446, Z944
CHRONIC PANCREAS, BILIARY TRACT AND GALLBLADDER DISEASES	K800, K801, K802, K808, K811, K86, Q440, Q441, Q442, Q443, Q444, Q445, Q450
CHRONIC ULCER OF THE SKIN	I830, I832, L89, L97, L984
COLITIS AND RELATED DISEASES	K520, K528, K551, K552, K572, K573, K574, K575, K578, K579, K58, K590, K592, K62, K634, K64,
COPD, EMPHYSEMA, CHRONIC BRONCHITIS	J41, J42, J43, J44, J47
DEAFNESS, HEARING IMPAIRMENT	H80, H90, H911, H913, H919, Q16, Z453, Z461, Z962, Z974
DEMENTIA	F00, F01, F02, F03, F051, G30, G31

DEPRESSION AND MOOD DISEASES	F30, F31, F32, F33, F34, F38, F39, F412
DIABETES	E10, E11, E13, E14, E891
DORSOPATHIES	M40, M41, M42, M43, M47, M48, M49, M50, M51, M53, Q675, Q761, Q764
DYSLIPIDEMIA	E78
EAR, NOSE, THROAT DISEASES	H604, H661, H662, H663, H701, H71, H731, H741, H810, H831, H832, H95, J300, J31, J32, J33, J341, J342, J343, J35, J37, J380, J386, K051, K053, K07, K110, K117, Q30, Q31, Q32, Q35, Q36, Q37, Q38
EPILEPSY	G40
ESOPHAGUS, STOMACH AND DUODENUM DISEASES	I85, I864, I982, I983, K21, K220, K222, K224, K225, K227, K230, K231, K254, K255, K256, K257, K264, K265, K266, K267, K274, K275, K276, K277, K284, K285, K286, K287, K293, K294, K295, K296, K297, K298, K299, K311, K312, K313, K314, K315, Q39, Q40, Z903
GLAUCOMA	H401, H402, H403, H404, H405, H406, H408, H409
HEART FAILURE	I110, I130, I132, I27, I280, I42, I43, I50, I515, I517, I528, Z941, Z943
HEMATOLOGICAL NEOPLASMS	C81, C82, C83, C84, C85, C86, C88, C90, C91, C92, C93, C94, C95, C96
HYPERTENSION	I10, I11, I12, I13, I15
INFLAMMATORY ARTHROPATHIES	M023, M05, M06, M07, M08, M09, M10, M11, M12, M13, M14, M45, M460, M461, M468, M469,
INFLAMMATORY BOWEL DISEASES	K50, K51
ISCHEMIC HEART DISEASE	I20, I21, I22, I24, I25, Z951, Z955
MIGRAINE AND FACIAL PAIN SYNDROMES	G43, G440, G441, G442, G443, G448, G50
MULTIPLE SCLEROSIS	G35
NEUROTIC, STRESS-RELATED AND SOMATOFORM DISEASES	F40, F41, F42, F43, F44, F45, F48
OBESITY	E66
OSTEOARTHRITIS AND OTHER DEGENERATIVE JOINT DISEASES	M15, M16, M17, M18, M19, M362, M363
OSTEOPOROSIS	M80, M81, M82
OTHER CARDIOVASCULAR DISEASES	I09, I281, I310, I311, I456, I495, I498, I70, I71, I72, I790, I791, I950, I951, I958, Q20, Q21, Q24, Q25, Q26, Q27, Q28, Z958, Z959
OTHER DIGESTIVE DISEASES	K660, K900, K901, K902, K911, K93, Q41, Q42, Q43, R15, Z904, Z980

OTHER EYE DISEASES	H022, H023, H024, H025, H04, H05, H104, H17, H184, H185, H186, H187, H188, H189, H193, H198, H201, H21, H310, H311, H312, H318, H319, H33, H352, H353, H354, H355, H357, H358, H359, H36, H47, H48, H49, H51, Q10, Q11, Q13, Q14, Q15, Z947
OTHER GENITOURINARY DISEASES	B901, N200, N202, N209, N210, N218, N219, N22, N301, N302, N303, N304, N31, N320, N323, N328, N329, N33, N35, N393, N394, N480, N484, N489, N701, N711, N731, N734, N736, N761, N763, N81, N88, N895, N905, N952, Q54, Q620, Q621, Q622, Q623, Q624, Q627, Q628, Q638, Q639, Q640, Q641, Q643, Q644, Q645, Q646, Q647, Q648, Q649, Z906, Z907, Z960
OTHER METABOLIC DISEASES	E20, E21, E22, E23, E24, E25, E26, E27, E28, E29, E31, E34, E35, E40, E41, E42, E43, E44, E45, E46, E64, E70, E71, E72, E74, E75, E76, E77, E79, E80, E83, E84, E85, E88, E85, E85, E85, E89, K903, K904, K908, K909, K912, M83, M88, N25
OTHER MUSCULOSKELETAL AND JOINT DISEASES	B902, M212, M213, M214, M215, M216, M217, M218, M219, M22, M23, M24, M252, M253, M357, M61, M652, M653, M654, M700, M720, M722, M724, M750, M751, M753, M754, M797, M841, M89, M91, M93, M94, M96, M94, M94, M94, M99, Q65,Q66, Q68, Q71, Q72, Q73, Q74, Q77, Q78, Q796, Q798, Q87, S382, S48, S58, S68, S78, S88, S98, T05, T096, T116, T136, T147, T90, T91, T92, T93, T94, T95, T96, T97, T98, Z440, Z441, Z891, Z892, Z893, Z894, Z895, Z896, Z897, Z898, Z899, Z946, Z966, Z971
OTHER NEUROLOGICAL DISEASES	B900, D482, G041, G09, G10, G11, G12, G13, G24, G25, G26, G32, G37, G51, G52, G53, G70, G71, G723, G724, G728, G729, G73, G80, G81, G82, G83, G90, G91, G938, G939, G95, G99, M471, G99, G99, G99, Q00, Q01, Q02, Q03, Q04, Q05, Q06, Q07, Q760
OTHER PSYCHIATRIC AND BEHAVIORAL DISEASES	F04, F06, F07, F09, F102, F106, F107, F112, F116, F117, F122, F126, F127, F132, F136, F137, F142, F146, F147, F152, F156, F157, F162, F166, F167, F172, F176, F177, F182, F186, F187, F192, F196, F197, F196, F196, F196, F50, F52, F60, F61, F62, F63, F68, F70, F71, F72, F73, F78, F79, F80, F81, F82, F83, F84, F88, F89, F95, F99
OTHER RESPIRATORY DISEASES	B909, E662, J60, J61, J62, J63, J64, J65, J66, J67, J684, J701, J703, J704, J84, J92, J941, J953, J955, J961, J98, Q33, Q34, Z902, Z942, Z943, Z963
OTHER SKIN DISEASES	L13,L28,L301,L43,L508,L581,L85,Q80,Q81,Q821,Q822,Q829
PARKINSON AND PARKINSONISM	G20,G21,G22,G23
PERIPHERAL NEUROPATHY	B91, G14, G54, G55, G56, G57, G58, G59, G60, G628, G629, G63, M472, M531, M541
PERIPHERAL VASCULAR DISEASE	I702,I73,I792,I798
PROSTATE DISEASES	N40,N411,N418

SCHIZOPHRENIA AND DELUSIONAL DISEASES	F20,F22,F24,F25,F28
SLEEP DISORDERS	F510,F511,F512,F513,G47
SOLID NEOPLASMS	C, D00, D01, D02, D03, D04, D05, D06, D07, D09, D320, D321, D329, D330, D331, D332, D333, D334, Q85
THYROID DISEASES	E00,E01,E02,E03,E05,E062,E063,E065,E07,E350,E890
VENOUS AND LYMPHATIC DISEASES	I780,I83,I87,I89,I972,Q820

Appendix I

Project Ethical Approval

The University Of Sheffield.

Downloaded: 04/10/2023
Approved: 19/11/2021

Samuel Watchorn
Registration number: 200273200
Division of Population Health
Programme: NIHR ARC PhD

Dear Samuel

PROJECT TITLE: Modelling Fall Patterns in Elderly Care Home Residents for Use in Risk Stratification and Prediction Models.
APPLICATION: Reference Number 043450

This letter confirms that you have signed a University Research Ethics Committee-approved self-declaration to confirm that your research will involve only existing research, clinical or other data that has been robustly anonymised. You have judged it to be unlikely that this project would cause offence to those who originally provided the data, should they become aware of it.

As such, on behalf of the University Research Ethics Committee, I can confirm that your project can go ahead on the basis of this self-declaration.

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since full ethical review may be required.

Yours sincerely

Molly Girvan
Departmental Ethics Administrator

Appendix J

Trusted Research Environment Authorisation Agreement

Trusted Research Environment

Hosted by Durham University

Approved Data User Authorisation Agreement

Version 1 FINAL October 2020

Trusted Research Environment (TRE)

TRE Proposed Data User Information

The purpose of this document is to gather all relevant information to ensure that users of the TRE have appropriate permissions and training in place.

1. Name: Samuel Watchorn

2. Student

3. Institution and Department: The University of Sheffield, School of Health and Related Research

4. Project Title for which TRE access is required: Modelling fall patterns in elderly care home residents for use in risk stratification and prediction algorithms

5. The following training needs to be completed prior to accessing the TRE (with the exception of the AIMES TRE training, which begins on first accessing the TRE). If you have already completed relevant training, please indicate the date of completion and provide your course certificate with this form. (Training should have been completed / refreshed within the last 12 months).

Training required	**Completion Date**
MRC e-learning module Research Data and Confidentiality	29/04/21
DPST Toolkit requirements met (Must include 'Data Security Awareness and Protection')	01/07/21
	Date due to start
AIMES TRE training	07/07/21

6. Please provide a brief description of measures you will take to ensure that privacy of information is maintained while you are accessing data held within the TRE. If any additional equipment or other arrangements need to be put in place in order for you to access the information securely (e.g. privacy screen), please indicate in the space below.

I will be accessing the data under the University of Sheffield governance regulations and have completed the required training (in information governance, protecting information, protecting research data, GDPR research and confidentiality). I will be working alone and on a university computer. I do not require any other equipment.

7. Software Tools. Where an Approved Data User works is employed by an academic institution, they may request use of existing Durham University statistical / analytical

software packages available within the TRE for their academic-led studies. Additional terms may apply. If such software is required, please list below.

Access to R, R-Studio and Python

TRE Data User Declaration

Proposed Data User: ("the Applicant") Samuel Watchorn

The above-named Applicant hereby applies to be an approved data user of the Durham University hosted Trusted Research Environment ("TRE"), with the right of access to datasets held in the TRE in order to undertake analysis for approved purposes in support of the Project. Access is subject to the terms of this Declaration as set out below.

"Approved Data User" shall mean a person who has signed the declaration below. The Approved Data User shall either be a Principal Investigator ("PI") of a Project or a person who is authorised by the Project PI to have access to Project research data held in the TRE and who the TRE Operations Group is satisfied can be given access to the TRE.

Note - Where the Applicant is an employee of Durham University, they are required to complete sections a), b) and c) (as appropriate) of the declaration only. Applicants not employed by Durham University must obtain a signature from an authorised signatory for and on behalf of their institution (the Institution named in section c) of the declaration).

Durham University hosts data on a server in the TRE and provides Approved Data Users with secure remote access to data held by the TRE.

The Applicant hereby acknowledges and accepts the responsibilities as set out below:

Approved Data User Responsibilities

A) The Applicant is aware of the sensitive nature of the data being accessed and shall maintain the security and confidentiality of any datasets held by the TRE in accordance with the terms of this Declaration, the requirements of data protection legislation and the Data Protection Principles (as outlined in the appropriate data sharing / collaboration agreement associated with the Project).

B) The Applicant shall report any events that are in breach of the terms of the Declaration. Reports by the Applicant must be made to a TRE designated Duly Authorised Person in the first instance as soon as becoming aware of the incident.

C) The Applicant agrees:
1) To complete approved information governance training:

- the MRC e-learning module Research Data and Confidentiality;
- IG toolkit
- AIMES TRE training

2) To confirm completion of information governance training by sending a completed course certificate to the TRE Operations Group prior to being given access to the TRE;

3) Not to reuse the Project Dataset for any purpose which is outside of the project's original scope for which the data was obtained, without the formal agreement of the relevant Data Controller if applicable;

4) Not to share data or any derived data set with colleagues who are not Approved Data Users for that particular Project;

5) Not to attempt to link the data to other datasets;

6) Not to attempt to de-anonymise / de-pseudonymise the data;

7) To ensure that individual-level data is not transferred into or outside the TRE via any means including, for example, SFTP , photographs, voice recording, screen grabbing or note taking; unless explicit agreement is given by the Data Controller;

8) Not to share or disclose their TRE login details;

9) Not to allow people who are not Approved Data Users access to individual level data within the TRE;

10) To make every effort to stop people who do not have a right of access from viewing data on the TRE screen;

11) To ensure that the TRE and the relevant health or social care body responsible for initially providing patient service user data are acknowledged as data sources in all resulting reports and publications. E.g. "We acknowledge the support of the TRE, Durham University, for managing and supplying the anonymised data, and NHS data supplier (please specify) for the original data source."

D) The Applicant, once accepted as an Approved Data User, may be given access to more than one research project dataset. A new User Declaration must be completed for each Project.

Definitions

Project: means a unique research, service improvement or evaluation study with a PI, specified cohort, aims and methods that is logged onto the TRE Project Management System and has obtained all required governance approvals.

Project Dataset: means the research data that has been pseudonymised, anonymised uniquely and is specifically for use within a Project. The Project Dataset shall relate to the cohort and purpose defined for the Project and the terms of the relevant collaboration and information sharing agreements.

TRE Operations Group: means a group convened by the Executive Dean of the Faculty of Social Sciences and Health to independently assess applications for access to the TRE. Representation from Durham University's Research Policy, Information Governance, Legal Services, and Computing and Information Services will be called upon as appropriate.

Signatures

a) Declaration by Applicant for Approved and Authorised Data User status

By signing, the Applicant accepts the terms set out above.

Delete as applicable:

1. Applicant hereby confirms that they are a staff member at Durham University

Sign section a) below and section b) where the Applicant is not the PI of the Project.

2. Applicant hereby confirms that they are a student at Durham University

Sign section a); supervisor to sign section c).

3. Applicant confirms that they are not a staff member or student at Durham University and have received authorisation from their own institution.

Sign section a); section b) where the Applicant is not the PI of the Project; section c) supervisor to sign where applicable, and section d) authorised signatory to sign to indication Institutional approval.

Any breach of the terms of this Declaration by the Applicant will result in access being withdrawn, and a review will be undertaken by the TRE Governance Group who will decide on any other action deemed necessary. The TRE Governance Group has a duty to report serious legal or regulatory breaches to the appropriate authorities (such as the Data Protection Commissioner, Employers and professional regulatory bodies).

Name: Samuel Watchorn ("the Applicant")

Position: PhD Student

Institution: University of Sheffield

Signature [redacted]

Date signed: 01/07/2021

b) Declaration by Project Principal Investigator

Note: Where the Approved Data User is not the PI of the Project, this Declaration must be signed by the PI below.

By signing and dating below, you acknowledge that the Applicant named above has read and understood the terms of this Declaration.

Name: Professor Suzanne Mason

Position: Professor of Emergency Medicine

Institution: University of Sheffield

Signature: [redacted]

Date signed: 07/07/2021

c) Declaration by Student Supervisor

Note: Where the Approved Data User is a student, this Declaration must be signed by the student's supervisor.

By signing and dating below, you acknowledge that the Applicant named above has read and understood the terms of this Declaration.

Name: Professor Suzanne Mason

Position: Professor of Emergency Medicine

Institution: University of Sheffield

Signature: [redacted]

Date signed: 07/07/2021

d) Declaration by External Institution

Applicants for Approved Data User status who are not employees of Durham University must have this section signed for and on behalf of their institution by an authorised signatory.

The Institution named below hereby agrees that the Applicant named in section a) above is a bona fide employee or student of this Institution engaged in a reputable data analysis project for which all relevant required permissions have been granted, and that the Project Dataset requested can be entrusted to this person in the knowledge that they will conscientiously discharge their obligations in regard to the confidentiality of the Project Dataset.

This Institution agrees to abide by the terms of this Declaration and shall take responsibility for ensuring that the proposed Approved Data User complies with the terms of this Declaration, relevant data sharing agreements, and all applicable statutory and regulatory permissions and Data Protection requirements, and the Institution agrees to provide a secure working environment and suitable technical resources to meet this obligation.

The Institution agrees that a breach of this Declaration may lead to the withdrawal of access to the TRE for the Institution, its staff and students, and that Durham University, as host of the TRE, has a duty to report serious legal or regulatory breaches to the appropriate authorities (such as the Data Protection Commissioner and professional regulatory bodies).

The Institution has completed, and had approved by the Durham TRE Operations Group, a corresponding TRE Institutional Access Form/ is a signatory to a corresponding Collaboration Agreement for the project for which User Access is required.

Name:

Position:

Signature: [redacted]

Date signed: 18th August 2021

For and on behalf of:
(The "Institution") The University of Sheffield

Appendix K

Overdispersion Tests Results

As described in Section 4.4.2 before the results of the count data regression models were interpreted, the level of over-dispersion present was identified. The results of this over-dispersion testing for the Poisson regression models is shown in Table K.1. The final column in this table (RD/DF) indicates there is clear over-dispersion present in the fall count data. These results indicate the variance of the fall count data exceeds the mean, meaning Poisson regression is not appropriate for use in this instance. Therefore, as discussed in Section 4.4.2, Negative Binomial regression was used instead. As a result of the additional term incorporated in the equidispersion relationship, there is no over-dispersion present in the negative binomial models. This is shown in the final column in Table K.2 where all values are ≤ 1. Therefore, the results presented in the five main effects models presented in Chapter 5 result from negative binomial regression models.

TABLE K.1: Poisson Regression Over-dispersion Table

Poisson Model	**RD**	**DF**	**RD/DF**
CCI	5479	3998	1.37
EFI	5342	3996	1.34
HFRS	5455	3997	1.36
K-Means Clusters	5396	3992	1.35
Chronic Condition Shortlist	5285	3980	1.33

RD: Residual Deviance, DF: Degrees of Freedom, CCI: Charlson Comorbidity Index, EFI: Electronic Frailty Index, HFRS: Hospital Frailty Risk Score

TABLE K.2: Negative Binomial Regression Over-dispersion Table

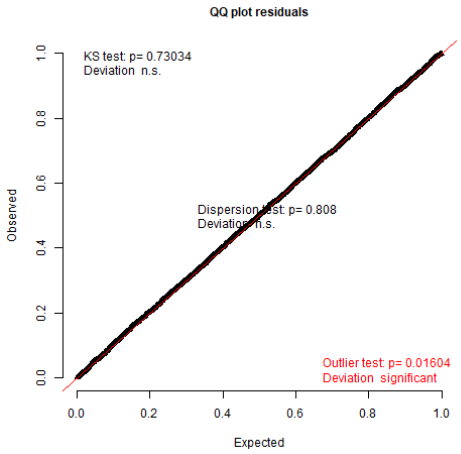
Negative Binomial Model	**RD**	**DF**	**RD/DF**
CCI	3709	3998	0.93
EFI	3712	3996	0.93
HFRS	3713	3997	0.93
K-Means Clusters	3715	3992	0.93
Chronic Condition Shortlist	3724	3974	0.94

RD: Residual Deviance, DF: Degrees of Freedom, CCI: Charlson Comorbidity Index, EFI: Electronic Frailty Index, HFRS: Hospital Frailty Risk Score

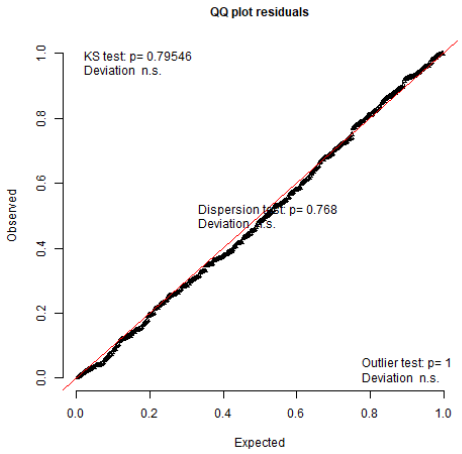
Appendix L

DHARMa Residual Plots for Interaction Models

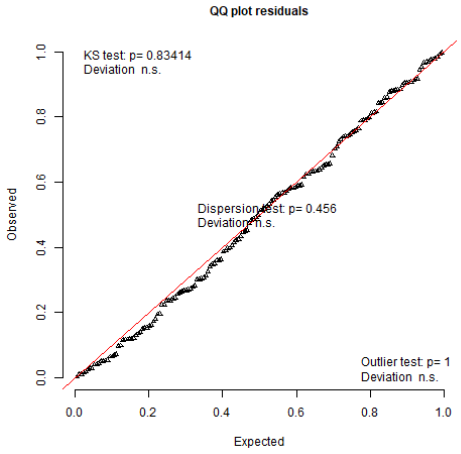
FIGURE L.1: DHARMa Q-Q Plots Interaction Models



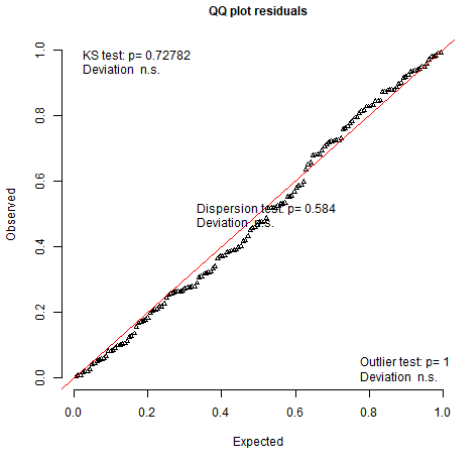
(A) DHARMa QQ Plot for Top 5 by Prevalence Interaction model



(B) DHARMa QQ Plot for Cardiovascular cluster Interaction model

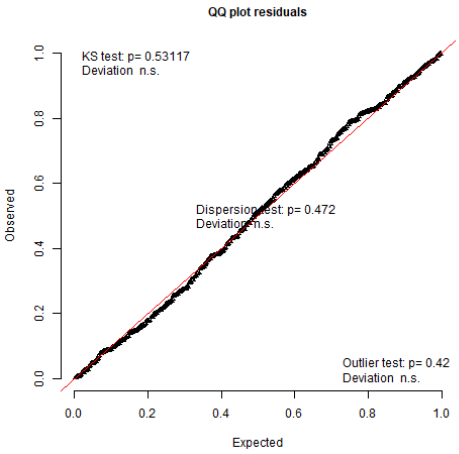


(C) DHARMa QQ Plot for Cardiovascular-Metabolic cluster Interaction model

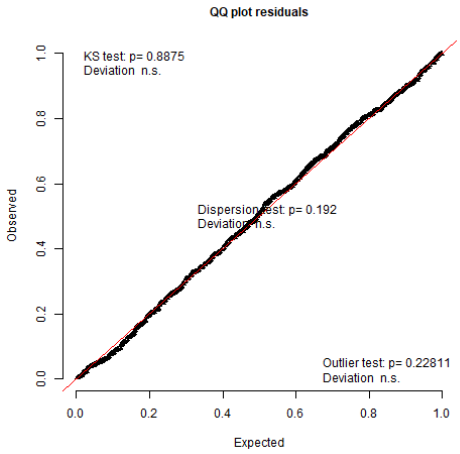


(D) DHARMa QQ Plots for N-S-High-Burden cluster Interaction model

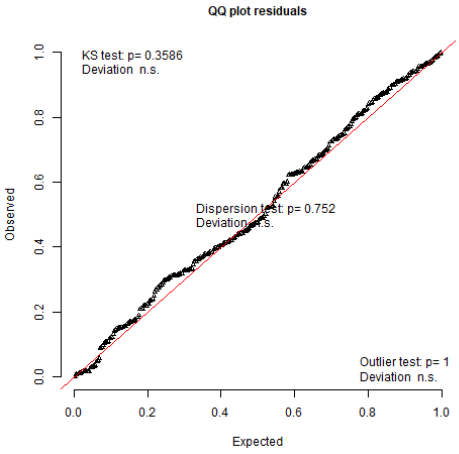
FIGURE L.2: DHARMa Q-Q Plots for Interaction Models Continued



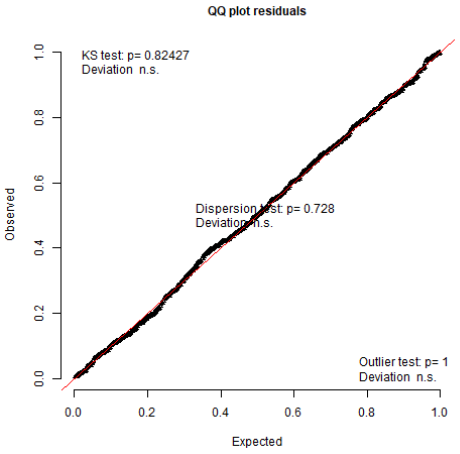
(A) DHARMa QQ Plot for Central cluster Interaction model



(B) DHARMa QQ Plot for Low-Neuro-Psychiatric cluster Interaction model



(C) DHARMa QQ Plot for High-Neuro-Psychiatric cluster Interaction model



(D) DHARMa QQ Plot for Low-Cardio-Neuro cluster Interaction model