THE UNIVERSITY
of EDINBURGH

Population-based genotype-phenotype correlation to stratify incident cases of motor neurone disease in Scotland 2015-2017

Danielle Jane Leighton

THE UNIVERSITY *of* EDINBURGH

Doctor of Philosophy

University of Edinburgh

2019

DECLARATION

I submit this thesis for fulfilment of requirements for the degree of Doctor of Philosophy. The work in this thesis has not been submitted for any other degree or professional qualification. I declare that the thesis and research described within has been written by me, unless otherwise stated and highlighted in the text. Almost all experimental work and associated narrative is my own. Collaborative contributions and jointly-authored publications have been clearly indicated and acknowledged below and in the main body of the text.

The research presented in Chapter 2 is published in the journal *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* and has been adapted for this thesis. The publication is entitled "Clinical audit research and evaluation of motor neuron disease (CARE-MND): a national electronic platform for prospective, longitudinal monitoring of MND in Scotland". I helped to design the data platform, devised the concept of the manuscript, did the statistical analysis and wrote the manuscript. All co-authors contributed to platform design and data collection. The authors include: **Danielle Leighton**, Judith Newton, Shuna Colville, Andrew Bethell, Gillian Craig, Laura Cunningham, Moira Flett, Dianne Fraser, Janice Hatrick, Helen Lennox, Laura Marshall, Dympna McAleer, Alison McEleney, Kitty Millar, Ann Silver, Laura Stephenson, Susan Stewart, Dorothy Storey, Gill Stott, Carol Thornton, Carolyn Webber, Harry Gordon, Giulia Melchiorre, Laura Sherlock, Emily Beswick, David Buchanan, Sharon Abrahams, Anthony Bateman, Jenny Preston, Callum Duncan, Richard Davenport, George Gorrie, Ian Morrison, Robert Swingler, **Siddharthan Chandran** & **Suvankar Pal**.

The research presented in Chapter 3 is published in the *Journal of Neurology* and has been adapted for this thesis. The publication is entitled "Changing epidemiology of motor neurone disease in Scotland". I devised the concept of the manuscript and research questions, collected and assessed the data, did the statistical analysis and wrote the manuscript. All co-authors contributed to data collection and manuscript editing. The authors include: **Danielle J Leighton**, Judith Newton, Laura J Stephenson, Shuna Colville, Richard Davenport, George Gorrie, Ian Morrison, Robert Swingler, **Siddharthan Chandran**, **Suvankar Pal** on behalf of the CARE-MND Consortium.

The research presented in Chapter 5 is published in the journal *Neurobiology of Aging* and

has been adapted for this thesis. The publication is entitled "Genetic epidemiology of motor neuron disease-associated variants in the Scottish population". The publication has joint first authorship. I analysed the phenotype data, did the genotype-phenotype association analyses and wrote the introduction, methods and discussion relevant to these aspects. The genotyping (sequencing, filtering and variant calling) was led by Dr Holly Black at the Institute of Genetics and Molecular Medicine, Edinburgh University. Co-authors contributed to data collection, data analysis and manuscript editing. The authors include: Holly A Black*, **Danielle J Leighton* (*Joint First Authors)**, Elaine M Cleary, Elaine Rose, Laura Stephenson, Shuna Colville, David Ross, Jon Warner, Mary Porteous, George H Gorrie, Robert Swingler, David Goldstein, Matthew B Harms, Peter Connick, Suvankar Pal, Timothy J Aitman, Siddharthan Chandran.

_____

Danielle Jane Leighton

October 2019

ABSTRACT

Background: Motor neurone disease (MND) refers to a spectrum of rapidly progressive neurodegenerative diseases for which there remains no cure. A recognised and crucial barrier to more accurate diagnosis, prognosis and treatment relates to phenotypic heterogeneity. Recent discoveries in the genetic landscape of MND have resulted in an accelerated research investment exploring aetiology of disease and basis of phenotypic variation. Scotland benefits from a culture of longstanding MND data capture and an integrated healthcare system.

Methods: I helped to develop Clinical Audit Research and Evaluation of MND (CARE-MND), an evolution of the established Scottish MND Register. CARE-MND is a national electronic platform for prospective, longitudinal monitoring of MND in Scotland. All people with MND (pwMND) diagnosed in Scotland in 2015-17 were included in an epidemiological study of incidence and prevalence of the disease. Patients who consented to sharing their medical records via the Scottish MND Register were included for phenotypic characterisation and prognostic modelling. Patients also donated DNA samples for genetic research to the Scottish Regenerative Neurology Tissue Bank. Two cohorts were genotyped: i) a pilot cohort of patients diagnosed 1989-2014 who were studied using a limited six-gene panel and ii) an incident cohort of patients diagnosed 2015-17 who were genotyped using an extended 49-gene panel. Genotype-phenotype correlations were explored and the impact of genetics included in prognostic models.

Results: By the end of my study period, the CARE-MND electronic platform was fully integrated into routine clinical care across all 14 health boards in NHS Scotland. Using capture-recapture statistics, coverage of the CARE-MND platform was 99% making it a reliable resource for further study. Direct age-standardised incidence in 2015 was 3.42/100,000 (95% CI 2.99–3.91); in 2016, it was 2.89/100,000 (95% CI 2.50–3.34). This represents a rise in incidence in Scotland by 36.0% over a 25-year period. The standardised incidence was also 66.9% higher than Northern European estimates. Of 619 pwMND diagnosed 2015-17, 437 (70.6%) consented to shared their phenotypic data. The following variables significantly predicted mortality: rapid decline in the ALS Functional Rating Scale Preslope, older age of onset, family history of MND and exposure to heavy metals/pesticides.

Atypical MND phenotypes (PLS, PBP and PMA), a long time to diagnosis and having ever smoked predicted survival. Genetic epidemiology of a historical cohort (diagnosed 1989-2014) using a 6-gene panel revealed pathogenic or loss-of-function variants in 17%. Using an extended panel, up to 22% had pathogenic variants or variants of uncertain significance with pathogenic potential (VUS-P). The cohort was enriched for the Scottish p.I114T *SOD1* founder mutation. Gene carrier status was associated with a family history of MND and other neurological conditions (including Parkinson's disease and multiple sclerosis). Having a *C9orf72* repeat expansion was associated with an increased risk of cognitive impairment. Having a genetic mutation of any kind did not influence overall survival.

Conclusions: Through CARE-MND, stratification of the MND population has facilitated participation in observational studies and has established a platform for recruitment into drug trials. The epidemiological data show a changing landscape of MND in Scotland with a marked increase in incidence over 25 years. This is likely attributable to ascertainment in the context of improved neurological services in Scotland. Early disability, older age and a family history of MND are poor prognostic markers in the Scottish population, whereas a delay in time from onset to diagnosis, atypical subtypes of MND and a history of smoking are associated with longer survival. Clinical trial design in Scotland needs to reflect and control for these factors. Using an extended 49-gene panel, 22% of patients have a potentially pathogenic MND-associated variant. Diagnostic genotyping should be considered to inform patients' prognosis and guide management.

LAY SUMMARY

Motor neurone disease (MND) is a devastating neurological disorder, which results in rapidly progressive weakness, causing problems with speech, swallowing and breathing. On average, patients typically do not survive more than 2.5 years from developing first symptoms. People with MND can present with a wide spectrum of symptoms, depending on the part of the body affected. Some progress very quickly and some live much longer than expected. It is currently very difficult to predict the pathway each individual will follow. In addition, memory and behaviour change have become increasingly recognised as a problem affecting patients. About 10% of people with the condition have an inherited form, with symptoms caused by faulty genes ('variants') which can be passed to family members.

My PhD project aimed to gather detailed information about all people living with MND in Scotland between 2015 and 2017. With the consent of people with MND, I obtained information about the features of their disease. I also collected blood samples to test the DNA for abnormal genes that might have caused disease. This information was stored in an anonymised database, anchored at the University of Edinburgh, called CARE-MND (Clinical Audit Research and Evaluation for MND). Through this, I tried to answer four main questions:

1. How common is MND in Scotland?

I answered this question by completing an 'epidemiological' study to investigate the number of people newly diagnosed with MND each year (incidence), and the number of people living with MND in Scotland at any one time (prevalence). We now know that about 200 people are diagnosed in Scotland annually. The number of people living with MND at any one time ranges between 400-430. The incidence of MND in Scotland has increased (compared to 20 years ago), and is higher than other European countries. We think that the main reason for this is that our neurologists and nurses are better at diagnosing and recording MND. Another reason might be that there are now better treatments for other diseases in Scotland, such as heart disease, so people are more affected by conditions like MND.

2. What factors affect survival?

I looked at the clinical features that predict outcomes ('prognosis') in MND. Using CARE-MND, we were able to look at many different characteristics to build up a picture of MND in

Scotland ('phenotypes'). I found that having early disability, being older and having a family history of MND results in poorer outcomes. However, having an unusual type of MND, having a long time from start of symptoms to diagnosis (ie. slowly progressive disease) and being a current or ex-smoker results in better outcomes. Further studies are required to look at the influence of smoking in Scotland before coming to any conclusions about its impact.

3. What gene variants affect people with MND in Scotland?

I did two genetic studies in this PhD, collectively looking at DNA donated by 770 people with MND in Scotland. The most important genes in Scotland are called *C9orf72* and *SOD1*. They can affect people with, or without, a family history of MND. There is a specific fault that occurs in the *SOD1* gene that is particular to Scotland. This is important for us because awareness of the genetic fault can now be raised amongst healthcare teams and testing for the fault provided earlier. It may be a potential target for drug treatments for a small proportion of people in Scotland (4%). We can pick up a gene fault, including rare gene faults, in about two thirds of people with a family history of MND in Scotland. Up to 23% of people overall may have a gene abnormality. Knowing about gene faults is helpful as it can provide a more secure diagnosis of MND. It is also important to look carefully for these gene faults as they may affect other family members.

4. Are gene variants associated with particular phenotypes?

I studied the influence of carrying a gene variant ('genotype') on the clinical characteristics ('phenotype') of MND (genotype-phenotype correlation). I found that carrying any kind of gene variant is associated with a family history of MND, a family history of other neurological conditions, and a specific subtype of MND called Progressive Bulbar Palsy. People with the *C9orf72* variation are more likely to have problems with cognition. The presence of a gene variant likely puts an individual at risk of developing MND, but does not influence duration of disease.

In summary, information from this PhD thesis has allowed me to describe the clinical and genetic characteristics of people in Scotland. Knowing about the numbers of people with the disease and factors which predict outcomes helps us to organise clinical services in hospitals and in the community to make sure people get the care that they need. This information also allows us to plan for trials of more effective interventions for symptom control, as well

as potential drug treatments to slow down the rate of progression of MND and even improve survival.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude towards all the people with MND and their carers who contributed to this PhD thesis. Their generosity is boundless and their encouragement and faith gave me momentum throughout my studies. Each individual whom I visited at home, clinic, hospital ward or hospice expressed enthusiasm towards improving our knowledge of this terrible disease but, at the same time, knew that they were doing their bit for others and not themselves. Each story will last with me always. I would also like to thank Gordon Aikman, who, at the same age as I am now, sadly passed away from MND. His advice and guidance were invaluable and his presence gave personal meaning to my work.

I am also incredibly grateful to all the research teams who have supported this work and continue to do so:

- My supervisors Dr Suvankar Pal and Prof Siddharthan Chandran.
- The remainder of the CARE-MND Consortium: Judith Newton, Shuna Colville, Andrew Bethell, Gillian Craig, Laura Cunningham, Moira Flett, Dianne Fraser, Janice Hatrick, Helen Lennox, Laura Marshall, Dympna McAleer, Alison McEleney, Kitty Millar, Ann Silver, Laura Stephenson, Susan Stewart, Dorothy Storey, Gill Stott, Carol Thornton, Carolyn Webber, Harry Gordon, Giulia Melchiorre, Laura Sherlock, Emily Beswick, David Buchanan, Prof Sharon Abrahams, Dr Anthony Bateman, Dr Jenny Preston, Dr Callum Duncan, Dr Richard Davenport, Dr George Gorrie, Dr Ian Morrison, Dr Robert Swingler. Special thanks to Dr George Gorrie and Dr Rob Swingler for their support and encouragement.
- Dr Elaine Cleary, Dr Morad Ansari, Dr Austin Diamond, Dr Jon Warner, Professor Mary Porteous at the South East Scotland Genetics Service, Western General Hospital.
- Dr Holly Black, Dr David Parry, Professor Timothy Aitman at The Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh.
- Dr Matthew Harms, Dr David Goldstein, Dr Hemali Phatani, Dr Cristiane Moreno, Dr Helaina Lehrer at the Institute for Genomic Medicine, Columbia University, New York.

- My husband, Dr Samuel Leighton at the University of Glasgow for his statistical expertise.
- Professor Daniel Stahl at King's College London for his statistical tuition.
- Professor Malcolm Macleod at the Centre for Clinical Brain Sciences.
- Dr Anna Molesworth at the National CJD Research and Surveillance Unit.
- Funders and advisors at MND Scotland, in particular: Craig Stockton, Niamh Callan, Dawn Hamilton, Steven Karlsson-Brown
- My funders: The Chief Scientist Office for Scotland, MND Scotland, the MND Association

Finally, personal love and thanks to Sam, Jacqueline, Adam, Sophie and Anne with an E.

CONTENTS

LIST OF FIGURES AND TABLES

Figures

Tables

COMMON ABBREVIATIONS

ACMG = American College of Medical Genetics

ALS = Amyotrophic lateral sclerosis

ALSFRS-R = ALS Functional Rating Scale Revised

CARE-MND = Clinical Audit Research and Evaluation of MND

CI = Confidence Interval

CNS = Central nervous system

ECAS = Edinburgh Cognitive and Behavioural ALS Screen

EPR = Electronic Patient Record

FTD = Frontotemporal dementia

ISD = Information Services Division

LMN = Lower motor neurone

LoF = Loss of Function

MAF = Minor allele frequency

MND = Motor neurone disease

NHS = National Health Service

NICE = National Institute for Clinical Excellence

NIV = Non-invasive ventilation

PBP = Progressive Bulbar Palsy

pwMND = people with MND

PLS = Primary Lateral Sclerosis

PMA = Progressive Muscular Atrophy

SMART-MND = Scottish Motor Neurone Disease Audit, Research and Trials for Motor Neurone Disease

SMNDR = Scottish Motor Neurone Disease Register

SNV = Single Nucleotide Variant

UMN = Upper motor neurone

VUS = Variant of uncertain significance

This thesis is largely written in the passive voice. However, when referring to "we" it is in the context of collaborative or published work. "I" is used with reference to unpublished work carried out by the author.

1. INTRODUCTION

1.1 CURRENT UNDERSTANDING OF MOTOR NEURONE DISEASE (MND)

The Impact and Epidemiology of MND

Motor neurone disease (MND) refers to a spectrum of neurodegenerative diseases for which there remains no cure. All forms of MND involve progressive decay and death of motor neurones in the central nervous system (CNS). Upper motor neurones (UMNs) in the cerebral cortex and lower motor neurones (LMNs) in the brainstem and cervical, thoracic or lumbar spinal cord can be affected. This manifests clinically with a variety of symptoms including, as Jean-Martin Charcot first described, "paralysis and contracture"[1] (Table 1). The most likely common features, however, are rapidly progressive disability and significantly shortened life expectancy. People with a 'typical' form of MND usually have symptom onset in the seventh decade[2–4], a diagnostic delay from onset to diagnosis of 10-12 months[5,6] and a survival from diagnosis of 1-2 years[2,5,7]. During the short disease course, people with MND (pwMND) might expect to acquire difficulties with walking, speaking, eating and breathing, necessitating medical intervention to improve quality of life and help maintain relative independence in daily functioning. Cognitive/behavioural change is also a recognised feature in about 50% of pwMND, with approximately 15% developing frontotemporal dementia (FTD)[8].

In simple terms, there is currently a three-pronged approach to management of MND: i) pharmacological treatment with riluzole, ii) feeding management with insertion of a gastrostomy tube, and iii) respiratory assistance with non-invasive ventilation (NIV)[9,10]. Other interventions are available for symptom control and comfort – including medications for sialorrhoea, emotional lability and spasms – and are part of the palliative approach to MND care. Riluzole is currently the only licensed drug for MND in the UK. It has a modest benefit: based on four Class I trials, it is thought to prolong survival by 2-3 months[10–13]. Edavarone, an intravenous drug initially developed for the treatment of acute ischaemic stroke, was recently approved for use in MND in the USA and Japan[14]. However, the trials found a benefit only in a subgroup of patients and it is currently not licenced in the UK[15].

Gastrostomy tubes, including PEG (percutaneous endoscopic gastrostomy), RIG (radiologically inserted gastrostomy) and PIGG (per-oral image guided gastrostomy) tubes, are used to maintain weight and prolong survival[9,10,16,17]. NIV can be considered for symptomatic treatment of respiratory muscle weakness and has survival benefits[9,10,17,18]. These three management strategies all require careful consideration and adjustment. They all come with risk of side effects and complications. The urgency and volume of decisions and interventions is incredibly challenging for carers[19]. The consequent impact on individuals and families is tremendous.

	Upper Motor Neurone (UMN)	**Lower Motor Neurone (LMN)**
Limb Signs	Contractures Spasticity Weakness Brisk reflexes	Muscle atrophy Fasciculations Weakness Diminished reflexes
Limb Symptoms	Pressure sores secondary to contractures Pain secondary to spasticity	Cramps
Bulbar Signs	Spastic dysarthria Brisk jaw/gag reflexes	Tongue atrophy Tongue fasciculations Tongue weakness Facial weakness
Bulbar Symptoms	Emotional lability	Dysarthria Dysphagia

Table 1 Signs and symptoms associated with motor neurone disease

Historically, MND has been considered a "rare" disorder with a European incidence rate of approximately 2 per 100,000 of the population[20]. The median prevalence of disease is about 5 per 100,000, reflecting the typically terrible prognosis[20]. However, recent evidence suggest a UK lifetime risk of 1 in 400 and a rising global disease burden[21–23]. The economic impact is considerable, with annual health services costs exceeding that of patients following a stroke and equalling that of patients diagnosed with dementia[24].

In the context of this devastating disease, clinicians and clinician scientists have a duty to be attuned to the challenges of pwMND and target research towards improving i) symptoms

and care of those living with the disease and ii) prospects for future pwMND. Clinical research should ideally prioritise these goals equally, in parallel. Better knowledge of disease phenotypes and disease progression would enable clinicians to inform pwMND of individual prognosis, helping to plan care and address future risks. Better knowledge of disease aetiology would provide targets for early intervention, including pharmacological therapies, providing hope to newly diagnosed patients and at risk relatives.

The Heterogeneity of MND Classifiers and Clinical Phenotypes

MND is notoriously heterogeneous in clinical presentation[4,25]. Clinicians and researchers have defined the disease by certain phenotypic 'classifiers' and have attempted to prognosticate based on these markers. The most common manifestation of MND is classified as amyotrophic lateral sclerosis (ALS), which typically involves UMN and LMN degeneration resulting in weakness, and eventual paralysis, of muscles in the limbs, bulbar region (muscles of speaking, chewing and swallowing) and respiratory system[1,26]. A diagnosis of ALS is currently based on consensus clinical criteria, reflecting degree of diagnostic certainty and supportive electrophysiology information (the Revised El Escorial criteria)[27–29](Table 2). Interpretation of electrophysiological parameters may further guide diagnosis, such as that applied in the Awaji-Shima Consensus Recommendations[30]. However, in clinical practice, ALS is often simply dichotomised into "limb-onset" and "bulbar-onset" disease. The latter is thought preferentially to affect females and older individuals and is associated with poorer prognosis[31]. Classifying pwMND in this way aims to guide proactive management strategies such as early discussions regarding gastrostomy insertion to combat feeding difficulties. However, this system is likely over-simplified, as patients can present with a more global picture of ALS. Patients can present with respiratory symptoms, either acutely or insidiously[4,31,32]. Weight loss can also be a presenting feature, in the context of global atrophy and loss of appetite without focal muscle weakness[33].

ALS Definite	ALS Probable	ALS Possible
Evidence of UMN + LMN signs in bulbar region and at least 2 spinal regions Evidence of UMN + LMN signs in 3 spinal regions	UMN + LMN signs in 2 regions but there must be UMN signs above the LMN signs	UMN + LMN signs in 1 region UMN signs in 2 regions UMN signs above LMN signs, not attributable to other disease processes

Table 2 El Escorial Classification of ALS

Nomenclature of MND varies globally, with countries outside of the United Kingdom, Ireland and Australia using ALS interchangeably with MND. For the purposes of this research, MND is an umbrella term, which covers ALS and more 'atypical' phenotypes (MND-subtypes). ALS was first identified in 1887 by Jean-Martin Charcot[1] but since then four main MND-subtypes have been described:

i) Primary lateral sclerosis (PLS)

Although the term PLS was first documented by Charcot in a case report[34], diagnostic criteria were only formally reviewed and defined in 2006[35]. A diagnosis of PLS can be made if there are features of UMN degeneration (a manifestation of the discriminate involvement of corticospinal and corticopontine motor neurones), absence of LMN degeneration (on examination and on electromyographic (EMG) studies) and signs and symptoms which are persistent for >4 years[35]. In spite of this, emergence of clinical and EMG-supported LMN signs have been described 7.5-27 years after onset[36]. Mill's syndrome is thought to be a descriptive clinical identifier of a hemiplegic variant of PLS[37,38].

ii) Progressive muscular atrophy (PMA)

In contrast to PLS, PMA is an exclusively LMN disease affecting more than one body region[39]. However, PMA also has clinical overlap with ALS: a study reported that 22% of patients diagnosed with PMA developed UMN features (50% of these only one year after symptom onset)[39]. While the same study observed that patients who had persistently selective LMN disease did have phenotypic differences to people with ALS (longer survival), the pathological distinction between these two diseases remain unclear.

iii) Progressive bulbar palsy (PBP)

PBP refers to seemingly isolated bulbar-onset disease with no emergence of limb symptoms in the six month period after onset[40,41]. Patients are typically female, older and have rapidly progressive anarthria[40]. In reality, almost all of these individuals will develop limb symptoms eventually and 'convert' to a more typical ALS phenotype[42].

iv) Flail limb MND

First described in the context of upper limb disease, flail limb MND is a rare LMN predominant and localised presentation. In contrast to PBP it is seen much more frequently in males than females (ratio 9:1) and tends to have a slower progression and subsequent better prognosis[43].

Whilst being heterogeneous in presentation and survival, these MND classifiers clearly have significant phenotypic overlap. The prognostic implications of such labels are therefore not clear-cut.

Cognitive/behavioural change is now also a recognised feature of MND. This corresponds with frontal atrophy observed on imaging[44] and neuropathological findings in extra-motor brain regions[45]. Other studies have characterised the cognitive phenotype and have correlated with that seen in frontotemporal dementia (FTD)[46,47]. In particular, planning (executive function), speech (perseveration) and behaviour control (impulsivity) are features, leading some to argue that the cognitive features are in fact 'motor' in origin[1]. Over 50% of people with ALS are now thought to develop cognitive impairment in the course of disease, with approximately 15% developing FTD (usually behavioural variant FTD)[8,48]. Similarly, 10-15% of patients with FTD develop features of MND[48]. While the majority of studies in cognition include people with ALS only, there is evidence that people with PLS and PMA can develop cognitive impairment including frank FTD[49–51]. Co-existence of FTD has been shown to shorten life-expectancy and complicate decision-making[52]. The nature of behavioural change is such that families may describe a complete change in personality. These features, particularly apathy, impulsivity and disinhibition, are significant predictors of carer stress[53]. Identification and explanation of cognitive impairment and emotional lability are important i) to improve patient and carer understanding of disease, ii) to facilitate early and supported decision making and iii) to predict and plan for potentially shortened survival.

Phenotypic determinants have also included age of onset and survival. Juvenile MND (diagnosed before the age of 25) and young-onset MND (before the age of 45) tend to

harbour a genetic cause[54]. PwMND who are atypically "long surviving" are of interest as they may provide clues towards therapeutic targets or clinical interventions[55–57]; this population may constitute up to 20% of prevalent patients[58]. Within this group, age and site of onset are strong predictors of outcome[55].

It therefore becomes difficult to disentangle MND 'classifiers'. It is possible that these phenotypic characteristics interact and collectively contribute to a poorer prognostic outcome. For example, more widespread disease (ALS) may be associated with earlier-onset cognitive impairment; this, combined with bulbar-onset disease, may result in rapid progression and poor survival.

Neuropathological Features of MND

Typical neuropathological characteristics of MND include death of i) alpha LMNs from the anterior horn cells of the spinal cord and brainstem and ii) pyramidal UMNs (Betz cells) in the primary motor cortex[59]. This leads to amyotrophy of the muscles supplied and sclerosis of the corticospinal tracts, giving ALS its name[26]. On post-mortem examination, cytoplasmic ubiquitin deposits are a key signature of MND[60]. A particular breakthrough in understanding was the discovery of the 43kDa transactive response DNA binding protein (TDP-43) in both ALS and FTD cases, strengthening the argument for their being on a common spectrum of disease[61]. Mislocalised TDP-43 protein is now thought to be a signature of neuronal and glial pathophysiology in ~95% of MND cases (including MND-subtypes), leading us towards a classification of "TDP-43 proteinopathies"[61,62]. Further characterisation of pathological markers has occurred in parallel with genetic discoveries but there is a common theme of: misfolded proteins, abnormal protein aggregation and cytoplasmic/nuclear inclusion formation resulting in a toxic cellular atmosphere[63].

Study of disease progression suggests that pathology is initial focal and stochastic but propagates contiguously to involve other regions[59]. Pattern of UMN and LMN involvement supports a theory of localised pathological spread. For example, UMN signs and symptoms tend to progress from ipsilateral arm to leg, but LMN signs and symptoms progress from arm to arm, reflecting apposition of motor strips/nuclei in the frontal cortex and anterior spinal

cord cells respectively[59,63]. MND-subtypes such as PLS and PMA, where UMN or LMN are selectively affected, have less pathological burden, consistent with their expected longer survival. The theory of contiguous spread supports the hypothesis that MND may be a prion-like disease[64]. However, some presentations are more "multi-focal", suggesting that protein dysfunction is occurring randomly rather than through propagation[63].

Genetics of MND

Over the past decade, there has been a step change in our understanding of the inheritance and genetic aetiology of MND. The term "familial MND" was first described by Gowers to suggest an inherited pattern of MND, in spite of there being no supportive genetic evidence at the time[1]. In the early nineties, however, an association was recognised between MND inheritance and the Cu/Zn superoxide dismutase 1 gene (*SOD1)*[65,66]. About 9-16% of MND is considered inherited/familial which has naturally given genetic research necessary momentum[67,68]. Since the discovery of *SOD1,* gene discovery in MND has increased exponentially, largely due to improved access to next generation sequencing platforms, in terms of both cost and technological advancement[69,70]. Genomic technology has evolved from candidate gene Sanger sequencing to extended panel sequencing, whole exome sequencing (WES) and whole genome sequencing (WGS).

At the time of writing there are 126 genes associated (both strongly and tentatively) with MND[71,72]. The year 2011 marked the discovery of perhaps the most significant genetic mutation: expansion of a hexanucleotide GGGGCC repeat in intron one of the *C9orf72* (chromosome 9 open reading frame 72) region[73,74]. The threshold of repeat burden to reach pathogenicity has been difficult to quantify but it is currently accepted that more than 30 repeats is associated with symptoms, though the majority of patients have repeats in the order of thousands[75,76]. The *C9orf72* repeat expansion accounts for up to 10% of sporadic MND as well as being the most common mutation in familial MND[77].

Over the past two decades the following key genes have been implicated in the aetiology of MND: *ALS2* (alsin), *TARDBP* (encodes TDP-43 protein)[78], *FUS/TLS* (fused in sarcoma/translocated in liposarcoma)[79,80], *VAPB* (vesicle-associated membrane protein-

associated protein B)[81], *OPTN* (optineurin)[82], *UBQLN2* (ubquilin-2), *PFN1* (profilin 1)[83], *VCP* (valosin-containing protein)[84], *CHCHD10* (coiled-coil-helix-coiled-coil-helix domain containing 10). During the course of this PhD, two further genes were newly implicated in MND: *TBK1* (TANK-binding kinase 1)[85–87] and *NEK1*[88–90]. Each are thought to account for a small proportion of cases. The proposed molecular mechanisms of these genes are summarised in Figure 1 and Table 3.

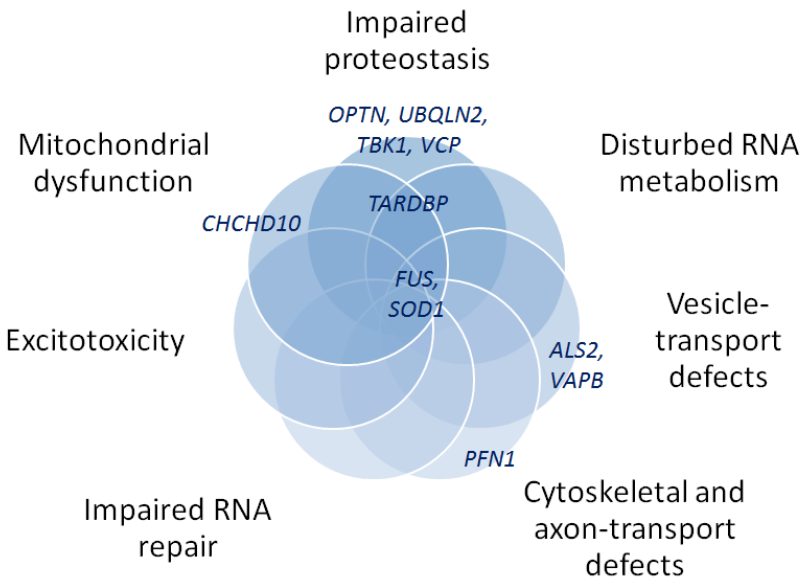


Figure 1 Molecular pathways implicated in motor neurone disease and associated genes. Adapted from Van Damme et al. 2016[91]

Oligogenic mutations are also reported, often in the context of the *C9orf72* expansion plus another gene variant[92–97]. Other genes are thought to be risk factors for MND development, for example the CAG repeat intermediate expansion in the ataxin-2 gene (*ATXN2*) which is more commonly associated with spinocerebellar ataxia[98].

Consequently, the clinical significance of *SOD1*, once thought to be a breakthrough discovery, has been attenuated. Further, carriers of *SOD1* mutations do not have the neuropathological

characteristic of TDP-43 proteinopathy, suggesting that they are in fact MND-spectrum outliers[99].

Our understanding of the genetics of MND continues to evolve, with an estimated doubling of the number of known associated genes every four years[100]. Applying this knowledge on a population scale may help to delineate geographical variations and targets for focussed genetic counselling and drug development.

Genetic Classifiers of MND: Familial and 'Apparently Sporadic'

Concomitant with this change in practice is a rich literature-base illuminating the genetic epidemiology of MND. Whilst the majority of cases of MND present without a family history, variants in the same genes are thought to contribute to the genetic aetiology of both familial and 'apparently sporadic' cases. Apparently sporadic genetic carriers (or simplex cases) are thought to be indistinguishable from familial cases; they were first described in Scotland in 1994[101]. In recent series, pathogenic variants in known genes have been found in 68% of cases with a family history and 11% of apparently sporadic cases[102]. The *C9orf72* repeat expansion remains the most commonly identified mutation, thought to be causative in 10% of apparently sporadic cases, 30% of familial cases and up to 88% of cases of MND with FTD[100,103]. A recent meta-analysis calculated that *SOD1* mutations were present in 18.9% (95% confidence interval (CI) 15.4-22.6%) of familial pwMND and 1.2% (95% CI 0.8-1.8%) of apparently sporadic cases[104]. Overall, approximately 10-15% of MND cases are now classified as having a monogenic aetiology in international case series/cohorts[77](see Table 3). These data lead us to believe that categorising people into "sporadic" and "familial" groups is partially artificial[100,105]. An important meta-analysis of "sporadic" MND-concordant monozygotic twins estimated a heritability of disease of 0.61 (95% CI 0.38-0.78); the unshared environment contributed 0.39 (95% CI 0.22-0.62)[106]. This study suggests that even apparently sporadic MND has a strong genetic component.

Genotype-Phenotype Correlation

It may be possible to associate particular genetic mutations, or sets of mutations, with a particular phenotypic picture: "genotype-phenotype correlation"[105]. In this way, early genetic testing of an individual might help to confirm diagnosis and predict future outcome of the disease. Some links have been postulated (genes listed in order of discovery; summarised in Table 3):

i. SOD1

In general, most people with a pathogenic/likely pathogenic *SOD1* mutation are thought to develop a "classical" form of ALS which progresses rapidly[107]. Onset is usually in the lower limbs and associated cognitive impairment is very rare[108]. Case reports of rare presentations are documented, including autonomic failure[109], ataxia, neuralgia and bladder disturbance[108]. Most mutations are dominant but the p.D91A variant has been identified in both heterozygous and homozygous states[110]. The homozygous form of the mutation is thought to originate from a Scandinavian population and, interestingly, has a less aggressive course than the heterozygous mutation[110,111]. The p.H48R variant found predominantly in Japanese populations has similarly long survival with rare bulbar involvement[112]. Conversely, the p.A5V mutation, the commonest *SOD1* variant in the USA and one of the first described, is characterised by rapid progression (survival <2 years from onset)[65,113,114]. Over 185 *SOD1* variants have been reported so far covering all five exons of the gene[108]. With increasing global interest in genome sequencing, genotype-phenotype associations will continue to become apparent. For example, the p.D12Y variant was relatively recently associated with distal muscle involvement and very slow progression[115,116].

ii. ALS2

Recessive or compound heterozygous mutations in this gene cause MND at a very young age resulting in the following phenotypes: infantile ascending hereditary spastic paraplegia (IAHSP), juvenile PLS (JPLS) and juvenile ALS (JALS)[117,118]. Many of the *ALS2* variant discoveries are a result of study of consanguineous families with unusual and severe

phenotypes, such as members of a family with a splice site mutation who had accompanying anarthria and dystonias[119].

iii. VAPB

VAPB mutations are rare and the phenotypic presentations heterogeneous and so no conclusive correlations have been postulated. A detailed study of a particular kindred found a phenotype which represented most closely adult-onset spinal muscular atrophy with autonomic features and movement disorder (constipation, sexual dysfunction and tremor)[120]. This phenotype is very distinct from typical ALS and perhaps suggests a separate disease process.

iv. TARDBP

Individuals with *TARDBP* mutations have heterogeneous features but all usually fall within the ALS spectrum[107]. The p.A328T mutation has been particularly studied and is thought to be a Sardinian founder mutation, accounting for 28.7% of all cases of MND (familial and apparently sporadic) in the region[121]. Individuals with this variant have been reported to have extrapyramidal features; however, the authors acknowledge that other mutations or neuroleptic medications might also account for these findings[121]. *TARDBP* mutations have also been found to cause MND with FTD[122].

v. FUS

FUS variants have been associated with younger onset disease. For example, the p.P525L variant is considered a particularly important cause of juvenile ALS (<25 years onset), with an aggressive disease course[123]. Although *FUS* mutations have been reported in atypical FTD (Basophilic Inclusion Body Disease)[124], cognitive impairment is not thought to be a feature of *FUS* MND[125].

vi. OPTN

Several *OPTN* variants have been reported in people with primary open angle glaucoma (POAG) and so it is important to separate these from those found in MND. Most of the *OPTN* studies originate from Japan with a slowly progressive phenotype[82,107].

vii. VCP

In 2010, an exome sequencing study identified individuals with ALS in families with multisystem disease including inclusion body myopathy with Paget disease of bone and frontotemporal dementia (IBMPFD)[84]. Manifestation of the *VCP* syndrome in family members is very variable, and ALS is now considered part of the wider phenotype. *VCP* mutations causing MND, however, are rare in the UK[126]. As shown in Figure 1, *VCP* is involved in proteostasis and delivery of proteins to the autophagosome. It operates with *OPTN*, *TBK1* and *UBQLN2* but yet has such a distinct clinical manifestation of disease. This might allude to the fragility and vulnerability of these pathological pathways and provide some explanation for the wide phenotypic spectrum of MND disease.

viii. C9orf72

Ethnic origin is a strong factor in *C9orf72* expansions: it is particularly enriched in the Finnish population, found in 28% of a population cohort of pwMND[73], but is relatively uncommon in Asian populations[127,128]. Indeed, the expansion is thought to originate from a single European founder mutation[129,130]. The phenotype of patients with the *C9orf72* expansion varies greatly within the literature. Certainly it is associated with an increased risk of cognitive impairment and FTD[103,131,132] and it may also cause psychotic symptoms without dementia[133] and/or a neuropsychiatric prodrome of MND/FTD[134]. Some studies also suggest that it is significantly associated with younger age of onset and shorter survival[132], faster decline in respiratory function[135] and that males with spinal onset disease have shorter survival[136]. Meanwhile, others suggest that it has no impact on age of onset, survival[129], or site of onset[137].

ix. UBQLN2

The only known X-linked MND-associated gene, *UBQLN2* mutations are also rare. They can affect both males and females but males present earlier, likely due to the hemizygosity[107,138]. Variants in this gene are also associated with MND-FTD[138].

x. PFN1

PwMND harbouring *PFN1* mutations tend to be younger at onset, have limb-onset disease and do not develop cognitive impairment[83,139].

xi. CHCHD10

CHCHD10 is the only MND-associated gene that causes mitochondrial dysfunction. Individuals with these mutations can have cognitive impairment, as well as cerebellar ataxia and myopathy, but, again, they are rarely seen in UK cohorts[140,141].

xii. TBK1

In 2015, a whole exome burden analysis newly identified an association between MND and the *TBK1* gene[86]. The *TBK1* protein interacts with products of *OPTN* and *TARDBP*, giving it a tantalising "narrative potential"[86]. *TBK1* MND has been replicated in multiple populations; variants are also associated with cognitive impairment (approximately 50% in one study) and frank FTD[85,87,142].

xiii. NEK1

The latest genetic association with MND was a result of pooled efforts of genomic consortium members[88]. A single variant was enriched in a small Dutch population but otherwise genotype-phenotype correlations are unclear.

Additionally, genetics have been used to refine the aforementioned definitions of "juvenile" MND by identifying those genes that cause MND before the age of 10 years (*ALS2, SETX, SIGMAR1* and *SPG11*) and between 10-24 years (*ANG, FUS, SETX, SOD1, SPG11, UBQLN2* and *VAPB*)[108].

To complicate matters, however, gene mutations have been documented to result in variable phenotypic traits, even within the same family (genetic pleiotropy)[143]. Examples of this have been described in *C9orf72*, *SOD1* and *VCP* carriers, with varying ages of onset, disease site of onset, rate of progression and burden of cognitive/behaviour change amongst first degree relatives with the same mutation[75,77,143,144]. As mentioned, phenotypes of genetic MND can also extend beyond the traditional spectrum of disease to other neurodegenerative disorders[1]. As well as FTD, Parkinsonism (tremor, rigidity and bradykinesia) is a well-reported feature which can coincide with or succeed MND signs and symptoms[145,146]. *C9orf72* expansions have been observed in patients presenting with phenotypic pictures that are more typical of Alzheimer's disease[147], Parkinsonism[148] or

Huntington's disease (HD)/HD phenocopy)[149]. The decision to consider these other neurodegenerative diseases as a "positive" family history is debatable[100] but certainly should be considered in view of the possible shared pathological mechanisms[146,150].

Gene	Chromo-some	Year of Association with MND	Protein function	Proposed impact on cellular function	Phenotype Association	Genetic Epidemiology Familial/Sporadic	
SOD1	21	1993	Homodimeric metalloenzyme which catalyses removal of superoxide into oxygen and peroxide[108,151]	Accumulation of misfolded protein/toxic gain of function; mitochondrial dysfunction[151,152]	"Classical" ALS; lower limb onset; cognitive impairment rare	19%[104]	1%[104]
ALS2	2	2001	Alsin comprises guanine-nucleotide exchange factors (GEFs) that activate GTPases[117]	Loss of function of alsin protein may impair vesicular transport/disrupt Golgi apparatus[117]	Juvenile ALS, PLS or IAHSP <10 years of age	Rare in adults[139]	Rare in adults[139]
VAPB	20	2004	A membrane protein that associates with microtubules, role in membrane transport[81]	Impaired intracellular membrane transport and secretion, with impaired neurotransmitter release[81,91]	Nil	Unclear	Unclear
TARDBP	1	2008	Active in RNA processing, transport, stress granule formation[151,153]	Accumulation of mislocalised cytoplasmic aggregates/toxic gain of function[151,153]	May have cognitive impairment	4%[99,107,154]	1%[99,107,154]
FUS	16	2009	Active in RNA processing[79,80,151,153]	Accumulation of mislocalised cytoplasmic aggregates/toxic gain of function[79,80,151,153]	Young age of onset; aggressive disease; no cognitive impairment	3%[79]	<1%[79]

OPTN	10	2010	Inhibits activation of nuclear factor kappa B (NF-kB)[82]	Negative feedback may upregulate NF-kB, inducing overexpression of *OPTN* and causing neuronal cell death[82]	Slow progression	0–4%[139]	Rare[139]
VCP	9	2010	Involved in maturation of ubiquitin-containing autophagosomes[84]	Mutant VCP expression leads to cytoplasmic accumulation of TDP-43[84]	Associated in families with IBMPFD	1-2%[84]	Unclear
C9orf72	9	2011	Noncoding intronic hexanucleotide repeat[73,74]	Repeat derived RNA toxicity and accumulation of dipeptide-repeat proteins[155]	Cognitive impairment	30%[100]	10%[100]
UBQLN2	X	2011	Role in proteosomal degradation[138]	Accumulation of ubiquitin-2 inclusions which bind with TDP-43[138]	Males present earlier; cognitive impairment	2%[156]	Rare[156]
PFN1	17	2012	Converts monomeric G-actin to filamentous F-actin[83]	Disturbs cytoskeletal pathway[83]	Younger onset; limb onset; cognitive impairment rare	0–2.6%[139]	Rare[139]
CHCHD10	22	2014	Mitochondrial protein located in intracellular space[140]	Thought to impair oxidative phosphorylation[140]	Can have cognitive impairment; otherwise heterogeneous	Rare[141]	Rare[141]
TBK1	12	2015	Autophagy, phosphorylates optineurin (OPTN), role in innate immunity signalling through NF-kB pathway [100]	Abrogates optineurin binding and may impair autophagy[85]	Cognitive impairment including MND with FTD[85]	2% overall[157]	

NEK1	4	2016	Formation and maintenance of cilia[100]	Interacts with ALS2 and VAPB proteins; impaired cytoskeletal organisation; DNA and cell cycle damage[158]	Nil	0.57%[89]	1.13%[159]

Table 3 Summary of the genotypic and phenotypic features of the top MND-associated genes to date

Phenotypes and Genotypes of MND: Conclusions from current understanding and unanswered questions

In spite of the above characterisations, a clear aetiology of MND has evaded the research community for many years. There is clear overlap between previously described classifications of disease, suggesting that this is insufficient. We can therefore conclude that there is a phenotypic spectrum of disease, with heterogeneous presentation but eventual unifying end-organ damage to muscles and brain/corticospinal tracts. There are commonalities in the neuropathological features of ALS and FTD, suggesting that FTD is also on the same spectrum. Recent discoveries in the genetic landscape of MND have provided important clues, have resulted in an acceleration in research investment and have raised the public profile of the disease. Causative genes have distinct pathological hallmarks suggesting that classification of disease by genotype, or proteinopathy, might provide the best clues for prognostication. However, most genetic mutations that are thought to be causative are autosomal dominant and adult-onset. This delay in onset suggests that the genetic burden imparts susceptibility, but that further triggers are required to overcome the disease threshold[63]. Recent studies support a multifactorial aetiopathogenesis with genetic and/or environmental risk factors: the "gene-time-environment (GTE)" hypothesis[143,160]. However, the interplay of multiple phenotypic and genotypic classifiers on prognosis is largely unknown.

Aetiological uncertainties have direct implications for provision of accurate diagnosis and prognosis, delivery of subsequent patient care, establishing appropriate services/clinical infrastructure, and crucially are a major obstacle to conducting clinically and genetically stratified clinical trials. Consequently, it would seem prudent to understand the genetic epidemiology and environmental influencers of a clinical MND population, to appreciate the potential burden of disease and structure care services accordingly.

1.2 CHALLENGES IN GENOTYPES AND PHENOTYPES OF MND

Challenges in identifying genotype-phenotype correlations

Successful genotype-phenotype correlation relies on having access to a full and rich database of patient characteristics with optimal acquisition of DNA samples. Historically this has been a major limiting factor in the discovery of potentially causative associations[108]. For example, in 2011 an Italian study proposed a link between UMN predominant MND and a particular single nuclear polymorphism within the kinesin-associated protein 3 (*KIFAP3*) gene[161]. These findings could not been replicated, however, due to poorly defined patient populations[105]. Variants in genes known to be associated with MND are not always 'pathogenic', or disease-causing. Supportive evidence requires information about the individual's clinical phenotype and a detailed family history[100]. Indeed some of the previously reported pathogenic mutations are now thought to have been polymorphisms, for example, in the *SETX* gene[162]. This could very well be the case in other MND-associated genes.

In spite of this, genotype-phenotype study remains challenging due to the rarity and breadth of the gene mutations concerned. Study of the *C9orf72* expansion has been possible due to its being the biggest monogenic contributor to MND. However, multi-centre or meta-analysed data are often required to reach conclusions[132,163,164]. Otherwise numbers are small and observations only hypothesis-generating, such as those speculated regarding particular *SOD1* variants[109,112,165].

Challenges in clinical genetic testing in MND

The growing acceptance of "genetic MND" has fuelled a need for routine clinical genetic testing. Global collaboration in the field continues to generate massive patient databases, allowing for new gene discovery, validation of previous findings and use of whole exome and whole genome technology to its full potential. In parallel, patient and clinician attitude to the disease has transformed. Historically, a diagnosis of MND required acceptance of

unknown aetiology[166]. Recent patient surveys suggest a move towards more proactive consultations, with 80% of patient responders having an understanding of MND genetics and 59% being aware of clinical testing options[167]. Further, enquiries about clinical, pre-symptomatic and pre-natal genetic testing are now being broached[168,169].

Clinicians at point-of-care, as well as genetic counsellors, have a duty to facilitate informed discussions but also to emphasise associated complexities (Figure 2). The European Federation of Neurological Sciences revised guidelines are clear that genetic testing is only recommended for individuals with a known family history of MND, or for individuals with symptoms in-keeping with the unusually slowly progressive phenotype seen in those with the *SOD1* p.D90A mutation[170]. However, reporting of family history of disease can be clouded by adoption, non-paternity, incomplete recollection, family estrangement, lack of recognition of MND phenotype or death before onset of symptoms[169]. National guidelines from the USA and UK do not address genetic testing[9,10]. Although it is universally agreed that indiscriminate genetic testing is unhelpful to patients and their families, it has been suggested that clinicians, along with multidisciplinary experts, introduce the concept of genetics and inheritance in all pwMND[166,168]. While 83% of pwMND agree with this approach, in reality only 35% of are offered genetic testing[167]. However, reservations exist due to perceived lack of benefit to patients and family members while genetically stratified treatments are not available. Lack of genetic counselling support is cited as a barrier, especially where variants of uncertain significance (VUS) are identified[171]. Clinicians agree that consensus guidelines for genetic testing would encourage implementation[171]. The unmet need is starting to be addressed, with the recent publication of recommendations for pre-symptomatic testing[172].

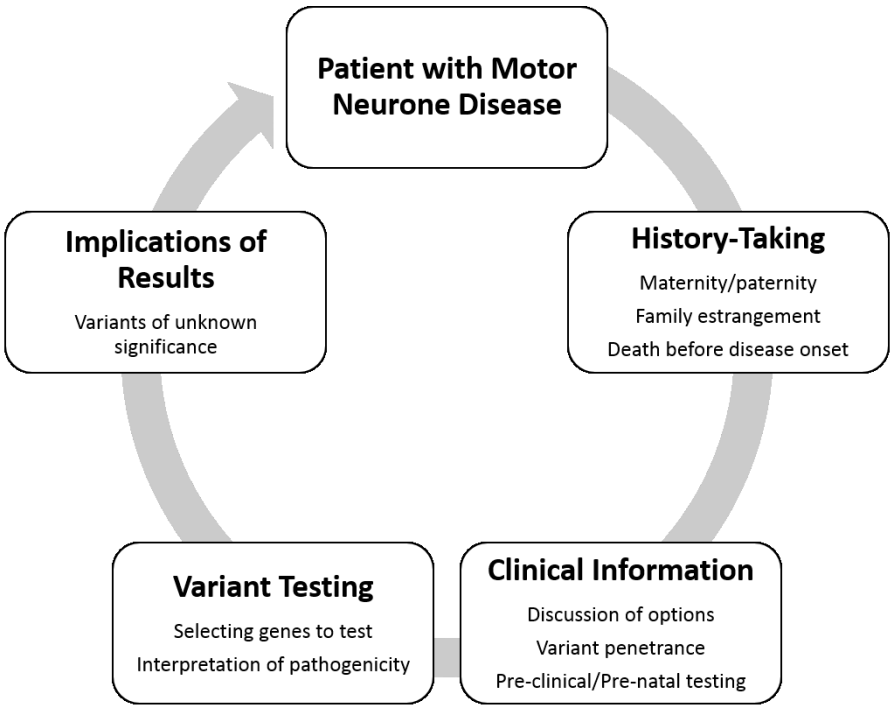


Figure 2 Barriers to clinical genetic testing of patients with motor neurone disease

Challenges in predicting outcomes in MND

As is clear from the heterogeneous nature of disease and the multitude of gene associations, predicting patient outcomes is complex. For pwMND, this is unsatisfactory. In blunt terms, individuals diagnosed with MND are unable to foresee likely intervention requirements, or estimate time until death. Further, identification of outcome predictors are essential for MND discovery. Investment in MND research is leading to a wave of new foundations for targeted drug trials[173–176]. In preparation, clinical trial-ready well-characterised patient populations will be key. Patient stratification for clinical trials depends on knowledge of key predictors of outcomes.

For prediction models to be useful, they must be:

i) Inclusive – all disease-relevant variables available for model development
ii) Generalisable - applicable to populations beyond the 'test' population
iii) Validated – externally validated in a plausibly related population

iv) Clinically relevant – intuitive for clinicians and informative for patients

Many previous prognostic modelling studies have focussed on clinical trial participants or patients recruited from tertiary referral centres[177]. However, these patients are known to be different to the general MND population, being typically younger, having longer survival and the cohorts are enriched for atypical (and longer surviving) phenotypes such as PLS and PMA[178]. A likely explanation for this is that these patients have milder or stable disease and they are therefore able to access site visits and engage in follow-up. Younger patients are perhaps more likely to explore research opportunities online and have support networks to facilitate enrolment. As such, models designed with trial/study data are likely not representative of the MND population and might fail to identify factors associated with rapidly progressive disease[179,180].

Identification of an appropriate outcome for prognostic modelling is the first hurdle. Ultimately, the most reliable endpoint is death. However, other markers have been explored such as time to NIV[181] and progression in the revised ALS Functional Rating Scale (ALSFRS-R)[182,183]. The ALSFRS-R is a measure of limb, bulbar and respiratory function in daily living. Although the course of decline in ALSFRS-R score is largely heterogeneous and unpredictable[183], its ease of use and acceptability to patients and care providers makes it an appropriate marker for clinical practice and research[184,185]. ALSFRS-R subscores may be more sensitive than patient self-report in predicting bulbar and limb symptoms; however, they do not appear to correlate well with respiratory function[186]. In view of its uncertain trajectory, its use as a prognostic endpoint is questionable.

Prediction modelling is becoming a key research priority and several models have been proposed[6,143,183,187,188]. The largest and most generalisable of these models was recently described by Westeneng et al.[6]. This large, multi-centre study examined the influence of 16 phenotypic variables on a composite outcome of death, NIV use >23 hours per day, or tracheostomy insertion. A backwards elimination process with bootstrapping was used to select predictors with imputation for missing data. Eight predictors were selected for the model, including: age at onset, forced vital capacity (FVC), diagnostic delay, ALSFRS-R slope of decline, bulbar onset disease, definite ALS, presence of FTD and presence of the *C9orf72* expansion. As this study was multicentre, an internal-external cross-validation

analysis could be conducted whereby results were trained on data with one site left out, and then tested on the remaining site. These data were used to develop a tool for classification of patients into a likely survival outcome, ranging from very short to very long survival. However, riluzole use, NIV and gastrostomy insertion were not considered as predictor variables.

There are many different statistical approaches for prediction modelling, from conventional statistics using binary and linear regression, to more advanced machine learning approaches. Regardless of the methods used, a priority from outset is to avoid a "sparse data bias" whereby there are insufficient individuals with the outcome of interest relative to the number of predictors (the Events Per Variable (EPV)[189]. An EPV of at least ten (the 'one-in-ten' rule) is generally considered satisfactory though there are exceptions[189]. Reduction of the number of predictors requires care; one approach is to remove non-significant variables from the model in a stepwise fashion (backwards elimination) but this risks excluding important confounding variables which might suppress the effects of remaining variables, thereby resulting in over-inflated, biased results[189,190]. Another recommendation is the use of penalisation, whereby the model coefficient estimates are 'shrunk' based on Bayesian theory, to reduce the variance within a model and ensure that the effects of a given variable are not inflated[189].

Further, missing clinical data in population databases are problematic, often because they are mishandled and therefore excluded, risking loss of valuable patient information[191]. For example, one population study of survival excluded 15% of pwMND because of missing data[7]. Missing data can be missing completely at random (MCAR), missing at random (MAR) or not missing at random (NMAR)[191,192]. Complete exclusion of MAR and NMAR data leads to systematic bias when predicting outcomes. Statistical methodology to overcome missing data is becoming more widely recognised, and includes both single and multiple imputation methods[191,193,194]. Imputed values are obtained by examining relationships between and within variables, again using a Bayesian approach.

Recent recommendations have been outlined by an expert committee for the reporting of prediction model methodology: the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines[195,196]. Published in 2015,

these will now provide structure and rigour to future studies and will allow for better replication between populations.

1.3 SCOTTISH MND INFRASTRUCTURE

The medical infrastructure of Scotland makes it an ideal location for the detailed study of rare diseases. We rely on three key strengths of Scotland: i) the National Health Service (NHS) in Scotland is a nationwide unified entity; ii) Scotland has a stable population structure with a relatively high incidence of certain diseases, including neurodegenerative diseases; iii) each patient has a unique community health index (CHI) identifier, enabling data linkage between community, hospital, prescribing and population health records[197]. The result is an integrated healthcare system, whereby patient data are uniformly collected across 14 health boards (Figure 3). Prospective, longitudinal monitoring and epidemiological analyses of rare diseases are therefore eminently possible. Scotland has one of the highest incidences of multiple sclerosis (MS) in the world and research into this disease has set precedent for other neurodegenerative diseases[198]. Scotland also has a history of robust epidemiological research into MND[199,200]. This set the tone for the development of the following population-wide initiatives.

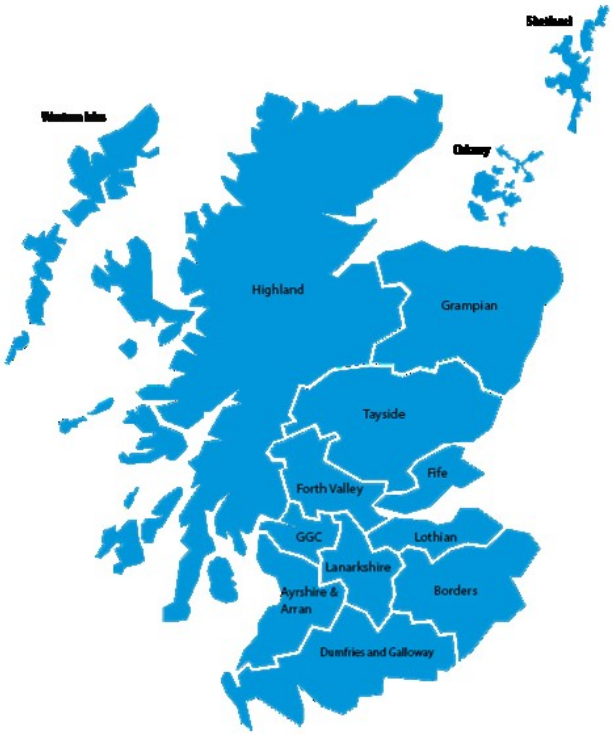


Figure 3 Fourteen health boards in Scotland

The Scottish MND Register (SMNDR)

The Scottish MND Register (SMNDR) is, to our knowledge, the first prospective national register of MND globally (the first retrospective national epidemiological study was in Israel 1959-1974[201]). In operation from January 1989, the register aimed to record and follow-up all incident cases of MND in Scotland; in view of the projected number (100-120 incident cases per year) this was considered achievable[202]. A full time nurse specialist and an IT team, as well as willing neurologists and neurophysiologists throughout Scotland, supported the register. Studies of the clinical characteristics and survival of the Scottish population were possible[203–205]. In combination with national hospital and death records from Information Services Division (ISD) Scotland, the register was able to achieve 98% case ascertainment[3].

In 2011, the register was modernised into a web-based interface – the Scottish Motor Neurone Disease Audit, Research and Trials (SMART-MND) register – hosted by the

Department of Clinical Neurosciences in Edinburgh making it more accessible to MND neurologists at point-of-care (usually outpatient setting) and permitting prospective data entry.

The Edinburgh Cognitive and Behavioural ALS Screen (ECAS)

The recognition of the substantial prevalence and impact of cognitive abnormalities in pwMND, including MND with FTD, is now well established[8]. This reflects in part the advent of standardised cognitive tools such as the recently validated Edinburgh Cognitive and Behavioural ALS Screen (ECAS) pioneered by Edinburgh Clinical Neuropsychology[206–208]. The ECAS has been incorporated into routine clinical care and MND research globally: there are now multiple versions and it has been translated into several different languages[209–211]. The ECAS can also be used to measure change in cognition over time[212,213].

The Scottish Regenerative Neurology Tissue Bank

The Scottish Regenerative Neurology Tissue Bank evolved as a result of growing interest in genetic research into neurodegenerative diseases, particularly MND. A study of the frequency of *SOD1* mutations in the Scottish population paved the way for subsequent creation of what is essentially an anonymised research 'DNA bank' to give patients the opportunity to contribute to this emerging field[214]. The Tissue Bank is hosted by the University of Edinburgh and has capabilities for storage of DNA from blood or saliva, and cerebrospinal fluid samples, for future research. DNA studies are overseen by the South East Scotland Clinical Genetics Service.

Genetic Testing in Scotland

In response to the urgent need for systematic clinical genetic testing of neurodegenerative diseases, a bespoke Scottish neurodegenerative gene panel was created and validated in

2014. This included many of the major genetic causes of MND: *C9orf72, SOD1, TARDBP, FUS, UBQLN2, VAPB, VCP, MAPT, GRN, CHMP2B* and *SQSTM1* (Pal & Porteous, University of Edinburgh (2014) [unpublished]). The selection of these genes was informed by sporadic diagnostic testing in Scottish MND and cognitive clinics.

Additionally, new and improved methods for detecting the *C9orf72* repeat expansion have been pioneered in Scotland[215]. The expansion is notoriously challenging to identify and quantify due to its GC-rich sequence. Indeed, blinded testing by 14 laboratories revealed significant discrepancies in detection results[216]. By incorporating a technique used for a similar repeat expansion syndrome (heat-pulse extension (HPE) used in Fragile X), the Edinburgh-based group developed a robust *C9orf72* detection assay, involving a flanking assay plus bidirectional repeat-primed PCR assays[215]. These methods are able to detect expansions in the order >100.

1.4 CURRENT KNOWLEDGE OF MND IN SCOTLAND

Phenotype

Through the SMNDR, inferences about the Scottish MND have been made. A study of incident pwMND diagnosed between 1989-1990 observed a median survival from onset of 2.5 years (95% CI 2.2-3.0)[203]. A diagnosis of PBP/bulbar symptoms was a significant poor prognostic marker. A long-term follow-up study between 1989-1998 allowed for accurate calculation of the incidence of disease, which was stable over this 10-year period (2.40 per 100,000 of the population (95% CI 2.22-2.58))[3]. In this population of 1226 patients, the male-to-female ratio was 1.19:1, 60% had spinal-onset disease and the frequency of familial MND was 4.8%. Median survival was 2.1 years (interquartile range (IQR) 1.3-3.2). The observation of poorer survival in the latter study prompted further analysis, which revealed that survival was actually declining over the years[217]. Contributors to this included bulbar-onset disease, as previously observed (Hazard Ratio (HR) 1.25 (95% CI 1.09-1.46)) and increasing age. Variables associated with better survival included longer diagnostic delay (HR 0.38 (95% CI 0.33-0.42)), neurology specialist involvement (HR 0.56 (95% CI 0.40-0.77)) and

riluzole use (HR 0.24 (95% CI 0.14-0.42))[217]. During this study period, the rate of intervention with gastrostomies for dietary supplementation doubled but did not confer any significant survival advantage[218].

A related study identified that 11% of this population were age 80 years or over, of which 50% had bulbar-onset disease[204]. Older people with MND were less likely to be prescribed riluzole (Odd's Ratio (OR) 0.12 (95% CI 0.02-0.89)) or be assessed by a neurologist (OR 0.76 (95% CI 0.67-0.86)).

The SMNDR historically has aimed to be a prospective and dynamic population register and this is evident from a study which evaluated the diagnostic certainty of register cases[205]. This identified that 8% of patients studied had alternative diagnoses on follow-up, making researchers aware of the diagnostic difficulties and the need for continuous re-assessment[205].

Genotype

Following the discovery of *SOD1* MND variants, an unselected population of 67 apparently sporadic and familial cases of MND in Scotland were directly sequenced for variants in the gene[214]. *SOD1* variants were detected in 50% (5) of familial cases and 7% (4) of the sporadic cases. Six of these unrelated patients had p.I114T variants. Subsequent analysis showed that these individuals had a shared haplotype, leading us to believe that the variant is a Scottish founder mutation[219].

As part of the development of the Scottish neurodegenerative disease gene panel in 2014, a pilot study examining for *C9orf72* expansions only, observed mutations in 8.3% (8/96) phenotypically diverse pwMND (Table 4).

Patient	Age of onset (years)	Duration of symptoms (years)	Family history (Y/N)	Motor symptoms	Cognitive symptoms (Y/N)	Psychiatric symptoms (Y/N)
1	45	22	Y	Limb	Y	Y
2	50	9	Y	Bulbar	Y	Y
3	61	4	Y	Limb &bulbar	N	N
4	60	2	Y	Limb	N	N
5	37	1	Y	Limb	Y	Y
6	58	4	Y	Limb	Y	N
7	67	3	Y	Limb	Y	N
8	61	4	Y	Limb	Y	Y

Table 4 Pilot data illustrating phenotypic heterogeneity amongst patients with pathogenic C9orf72 expansion amongst 96 incident Scottish patients with MND in 2014 (Pal & Porteous, 2014)

1.5 THESIS QUESTIONS, OBJECTIVES AND HYPOTHESES

In this introduction I have attempt to outline the current issues presented by phenotypic heterogeneity in MND, the recent explosion in knowledge of underlying genetic mutations and the lack of clarity in how these factors impact on survival. The current incidence and phenotypic components of the Scottish MND population is unknown. Similarly, in spite of accumulating DNA donations to the Scottish Regenerative Neurology Tissue Bank, gene panel testing of this population has not been exploited and so the incidence of gene carriers is also unknown.

During my PhD, I aimed to answer some of these questions with specific reference to the Scottish population. In order to achieve this, I helped to modernise the existing Scottish MND Register by creating a national electronic platform for integrative data capture of incident cases of MND. This platform (methods and utility outlined in Chapter 2) underpins subsequent study of the MND population and allows me to explore the following research questions:

Question 1: Has the incidence of MND in Scotland changed over time?

Objectives:

- To determine the incidence and prevalence of MND in Scotland 2015-2017.
- To analyse rate of change in incidence relative to the historical Scottish MND cohort (1989-1998).
- To compare incidence rates to global estimates.

Hypotheses:

- The global trajectory of MND incidence is rising due to aging populations[21]. I therefore expect the incidence of MND in Scotland to have risen over time.
- I expect that the incidence and prevalence of MND in Scotland in 2015-2017 will be comparable with other contemporary Northern European estimates.

Question 2: Can deep clinical phenotyping of the Scottish MND population reveal predictors of survival?

Objectives:

- To optimise on the breadth of phenotypic information that can be acquired regarding pwMND in Scotland (including disease characteristics, and endogenous and exogenous environmental exposures) and to describe phenotypic classifiers.
- To combine all phenotypic classifiers and environmental risk factors into a prognostic model to identify stratifiers of disease. Due to the expected median survival of pwMND in Scotland (2.1 years from onset[3]), survival studies of MND in Scotland will be achievable within the time-frame of this PhD project.

Hypotheses:

- The use of a standard phenotype proforma will allow a detailed characterisation of the Scottish MND population.
- Results will provide clarification about the impact of endogenous or exogenous environmental influences and which will be hypothesis generating for future studies.
- Results will identify predictors of survival, which will help inform patients, care-providers and future research studies, including clinical trials. Based on previous Scottish data, survival will be influenced by age, diagnostic delay and riluzole use[217].

Question 3: What is the genetic epidemiology of MND in Scotland? In particular, are there mutations unique to the Scottish population and are rare/newly described mutations present in this population?

Objectives:

- To test a historical research cohort of pwMND using a select panel of key MND genes to identify the proportion of genetic carriers.
- To test an incident research cohort of pwMND using an expanded and up-to-date neurodegenerative gene panel. To use these data to describe the genetic epidemiology of MND in Scotland.
- To develop a framework for the classification of pathogenicity of MND genomes and apply this to our genetic results.

Hypotheses:

- As with other European populations, I expect that that the most frequent genes in the Scottish MND population will be the *C9orf72* hexanucleotide expansion and *SOD1*. Based on previous study in Scotland, I expect the population to be enriched for the *SOD1* p.I114T variant.
- Based on the literature I expect to find rare variants and oligogenic gene carriers in the Scottish population.
- Results will highlight genes that are relevant, and irrelevant, to the Scottish MND population and will therefore inform future clinical genetic testing of pwMND and their family members.

Question 4: Can deep clinical phenotyping and extended genotyping of an incident MND cohort identify genotype-phenotype correlates (with particular regard to *C9orf72* and *SOD1*)?

Question 5: Does genotype influence survival in a multivariable prognostic model of MND in Scotland?

Objectives:

- To use the clinical and genotypic data of incident pwMND to undertake a national population based genotype-phenotype correlation study.
- To incorporate genetics into prognostic models to assess the relative influence of genetics on survival and thereby select the most appropriate statifiers for future studies including clinical trials.

Hypotheses:

- Analyses will identify phenotypic characteristics associated with key MND-associated genes such as *C9orf72* and *SOD1*.
- Patients with the *C9orf72* expansion are expected to have cognitive impairment. They may have younger age of onset and poorer survival.
- Patients with a *SOD1* variant are expected to have limb onset disease.
- Numbers of pwMND with rare gene mutations will be small and are unlikely to identify statistically significant associations but may identify characteristics which can be validated in future larger studies.
- The presence of a likely pathogenic gene mutation is expected to impact poorly on survival.

1.6 ETHICAL APPROVALS

This study is covered by ethical approvals for the SMNDR/Clinical Audit Research and Evaluation of MND (CARE-MND) and the Scottish MND DNA Bank/Regenerative Neurology Tissue Bank. SMNDR approvals are: MREC/98/0/56 1989-2010, 10/MRE00/78 2011-2015. In 2015 the ethical approvals were renewed by the Scotland A Research Ethics Committee 15/SS/0126 for this study. The Scottish MND DNA bank was approved by MREC/98/0/56 1989-2010, 10/MRE00/77 2011-2013, approved by the Scotland A Research Ethics Committee and the Scottish Regenerative Neurology Tissue Bank (including DNA collection) by MREC/98/0/56 1989-2010, 10/MRE00/77 2011 to 2013, 13/ES/0126 2013-2015, 15/ES/0094 2015-present, approved by the Chief Scientist Office Scotland. NHS Scotland Caldecott Guardians approvals were granted for all 14 participating health boards in Scotland. The author was granted relevant honorary contracts or local R&D approval for access to data from all participating health boards.

1.7 FUNDING

The author was awarded competitive funding for this PhD project through a Clinical Research Fellowship in Motor Neurone Disease. This is jointly funded by the Chief Scientist Office, Motor Neurone Disease Scotland and the Motor Neurone Disease Association (CAF/MND/15/01).

2. CARE-MND: A POPULATION DATA PLATFORM

The following content has been published in *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*[220]. It has been re-formatted for this thesis to include additional data relevant to thesis methodology.

Danielle Leighton, Judith Newton, Shuna Colville, Andrew Bethell, Gillian Craig, Laura Cunningham, Moira Flett, Dianne Fraser, Janice Hatrick, Helen Lennox, Laura Marshall, Dympna McAleer, Alison McEleney, Kitty Millar, Ann Silver, Laura Stephenson, Susan Stewart, Dorothy Storey, Gill Stott, Carol Thornton, Carolyn Webber, Harry Gordon, Giulia Melchiorre, Laura Sherlock, Emily Beswick, David Buchanan, Sharon Abrahams, Anthony Bateman, Jenny Preston, Callum Duncan, Richard Davenport, George Gorrie, Ian Morrison, Robert Swingler, Siddharthan Chandran & Suvankar Pal. Clinical audit research and evaluation of motor neuron disease (CARE-MND): a national electronic platform for prospective, longitudinal monitoring of MND in Scotland. *ALS FTD.* 2019 Mar; 20(3-4):242-250. https://doi.org/10.1080/21678421.2019.1582673.

2.1 BACKGROUND AND AIMS

Background

Scotland has a population of 5.3 million and benefits from an integrated healthcare system, in which patient data are uniformly collected across the 14 health boards of NHS Scotland[221]. Individuals are identified by the unique 10-digit CHI number, which can be referenced against local and national health records. Health boards have close academic links to the major University centres. Prospective, longitudinal monitoring and epidemiological analyses of rare diseases are therefore feasible. The Scottish Motor Neurone Disease Register (SMNDR) is a prime example. Launched in 1989, the SMNDR was the first prospective registry of people living with MND[222]. It gave robust estimates of regional incidence and prevalence of disease and guided care protocols[3,203–205].

In 2005, NHS Scotland commissioned a review of the care of people with neurological diseases, identifying that such diseases were given low priority for strategic service provision[223]. This led to Health Improvement Scotland (HIS) best practice standards[224]. Further to academic grant funding from MND Scotland, a web-based audit tool was designed to record SMNDR data, which could be used to compare against HIS guidelines (Scottish MND Audit, Research and Trials (SMART) study). This was the first national audit of MND care. It highlighted that over 91% of pwMND in 2011-13 were cared for within a specialist MND service. However, only 21% (3/14) of health boards had dedicated MND clinical pathways. HIS standards recommended contact with an MND specialist within two working days of diagnosis. These data were only available for 10% of incident cases in the audit period, with only 48% of these meeting this target. Indeed, a parallel Scottish study indicated that patients and carers felt that access to MND services was poor and that clinical specialists were time-pressured[225].

In 2015, a formal review of the register was conducted, with the intention of improving and modernising this system to facilitate a national, harmonised approach to MND care and to enable patient-involvement in clinical research. Patient-directed initiatives resulted in significant investment from the Scottish government to complement these efforts, including employment of a dedicated MND Nurse Consultant for Scotland and a doubling of the number of MND clinical specialist nurses/allied health professionals[226].

Recent commentary has outlined key features of an effective register and these informed our redevelopment: inclusivity to appreciate scope of disease; multi-source data capture; standardisation of data collection with respect to other national registries; long-term sustainability[227]. Through this, I helped to develop *Clinical Audit Research and Evaluation of MND (CARE-MND)*: a national electronic platform for prospective, longitudinal monitoring of MND in Scotland.

Aims

The CARE-MND platform is an overarching 'umbrella' study with the following aims:

1) To create a patient-centred approach to care based on recognised standards.
2) To standardise data sharing between healthcare professionals in 'real-time'.

3) To permit regular audit of care to facilitate timely improvements in service delivery.
4) To integrate existing Scottish MND biomarker research repositories:
    i. The Scottish MND Register
    ii. The Scottish Regenerative Neurology Tissue Bank
    iii. The Edinburgh Brain and Tissue Bank
5) To improve patient participation in research studies including observational and clinical trials.

I hypothesised that the CARE-MND would be a more efficient data platform than the historical SMART-MND platform. I anticipated that a more technologically-advanced and accessible system would better support audit and research, and therefore have direct application to patient care and service provision.

2.2 METHODS

Data Selection

Data fields for collection were reviewed in consultation with specialist healthcare professionals responsible for care of pwMND across Scotland via a national steering group meeting. Existing data collection proformas used in a clinical or research setting (including SMART data fields) were harmonised. Major national/international clinical registries and research repositories were scrutinised to ensure that our selected fields were comprehensive and internationally comparable. The following were selected due to their compatibility with the Scottish population and/or their direct collaborative links: UK MND DNA Bank, European Network for the Cure of ALS (ENCALS)/Project MinE, the Irish MND Register and Genomic Translation for ALS Care (ALS Association and Columbia University). Details regarding key disease progression markers were incorporated (onset, symptoms, classification of MND, MND-associated landmarks including insertion of gastrostomy and initiation of non-invasive ventilation (NIV)). In response to an emerging literature-base of potential environmental aetiological factors, we also included screening questions regarding past medical exposures

(medications, surgeries, blood transfusions[228–230]), toxin exposures (heavy metals and pesticides[231–233]) and other lifestyle variables (smoking and alcohol, exercise participation[234–237])). To complement our biomarker research (Scottish Regenerative Neurology Tissue Bank and Edinburgh Brain and Tissue Bank), we allowed for the inclusion of extended family history, focusing on family history of MND, dementia, other neurological diseases and psychiatric diseases. Validated measures of MND progress were also incorporated, including the Revised ALS-Functional Rating Scale (ALSFRS-R)[182] and the Epworth Sleepiness Scale (ESS) as a marker of respiratory muscle weakness[238]. With a growing evidence-base regarding pervasiveness of cognitive impairment in MND through all stages of disease, we recognised the need to integrate cognitive testing into routine assessment using the Edinburgh Cognitive and Behavioural ALS Screen (ECAS)[206,208,212,213]. Using these sources, elements were identified for inclusion in our national proforma. Finally, meetings with patient focus groups/other key stakeholders were arranged to ensure data fields were sensitive and appropriate to patient needs. With this knowledge, we developed a national care paper proforma for use in all 14 health boards of NHS Scotland (Appendix 1).

Pilot Phase: Paper Proforma

All adults living with MND spectrum disorders in Scotland are included in CARE-MND. The breadth of patient inclusion allows appreciation of the heterogeneous nature of the disease[25]. Inclusion criteria are as follows: i) individuals diagnosed by a neurologist with possible, probable or definite MND according to El-Escorial revised criteria [27,28] OR an MND subtypes (primary lateral sclerosis (PLS), progressive bulbar palsy (PBP), progressive muscular atrophy (PMA)), ii) adult individuals ≥16 years at diagnosis, iii) individuals resident in Scotland at the time of diagnosis or receiving long-term care in Scotland.

From September 2015, the national proforma was piloted in paper format across the 14 healthboards in Scotland. In parallel with the publication of MND National Institute for Clinical Excellent (NICE) Guidelines in February 2016, a final paper NICE-aligned proforma was disseminated among MND clinical specialists for routine use[9]. Standard Operating

Procedure documentation for proforma completion were developed and the MND nurse consultant provided one-to-one training.

Electronic Platform: CARE-MND

While electronic patient health records (EPRs) in Scotland are similar between health boards, regional disease-specific cross-talk is often not facilitated. The existing SMNDR web-based audit tool required data entry by dedicated research personnel and did not have a user-friendly interface for routine data entry. A modernised online platform was therefore designed. This electronic platform – CARE-MND – is hosted by the University of Edinburgh and uses a secure web-based password-protected interface to allow access by any authorised and trained MND clinical specialist in Scotland. The platform has three main goals: i) to enable and record patient recruitment to research, ii) to record and store prospective patient data for the purposes of audit and research, and iii) to make pwMND central to the process by allowing self-notification of interest and providing user-friendly information (Figure 4).

i) **Research recruitment:** The platform provides a centralised portal to enable and record research recruitment. All pwMND are automatically assigned a unique 'research-ready' ID code. Typically, MND nurse specialists/allied health professionals introduce research projects to pwMND, including the Scottish MND Register, Scottish Regenerative Neurology Tissue Bank and the Edinburgh Brain and Tissue Bank. Once the research team has received a consent form, participation is marked clearly on their CARE-MND record. Clinical specialists/allied specialists can view at-a-glance other studies that the patient is eligible for and those to which they have already contributed.

ii) **Prospective patient data:** In Scotland, there is a MND nurse specialist-to-prevalent-patient ratio of 1:26. Following an introductory meeting, the nurse specialist/allied health professional will document patient details using the structure of the national proforma. This is updated throughout the patient's journey until death.

iii) **Patient-centred approach:** The CARE-MND website (https://www.care-mnd.org.uk/) also hosts a patient-specific portal, providing key contact details for specialists, patient information leaflets, video information regarding research studies and an online form for patient self-registration of interest.

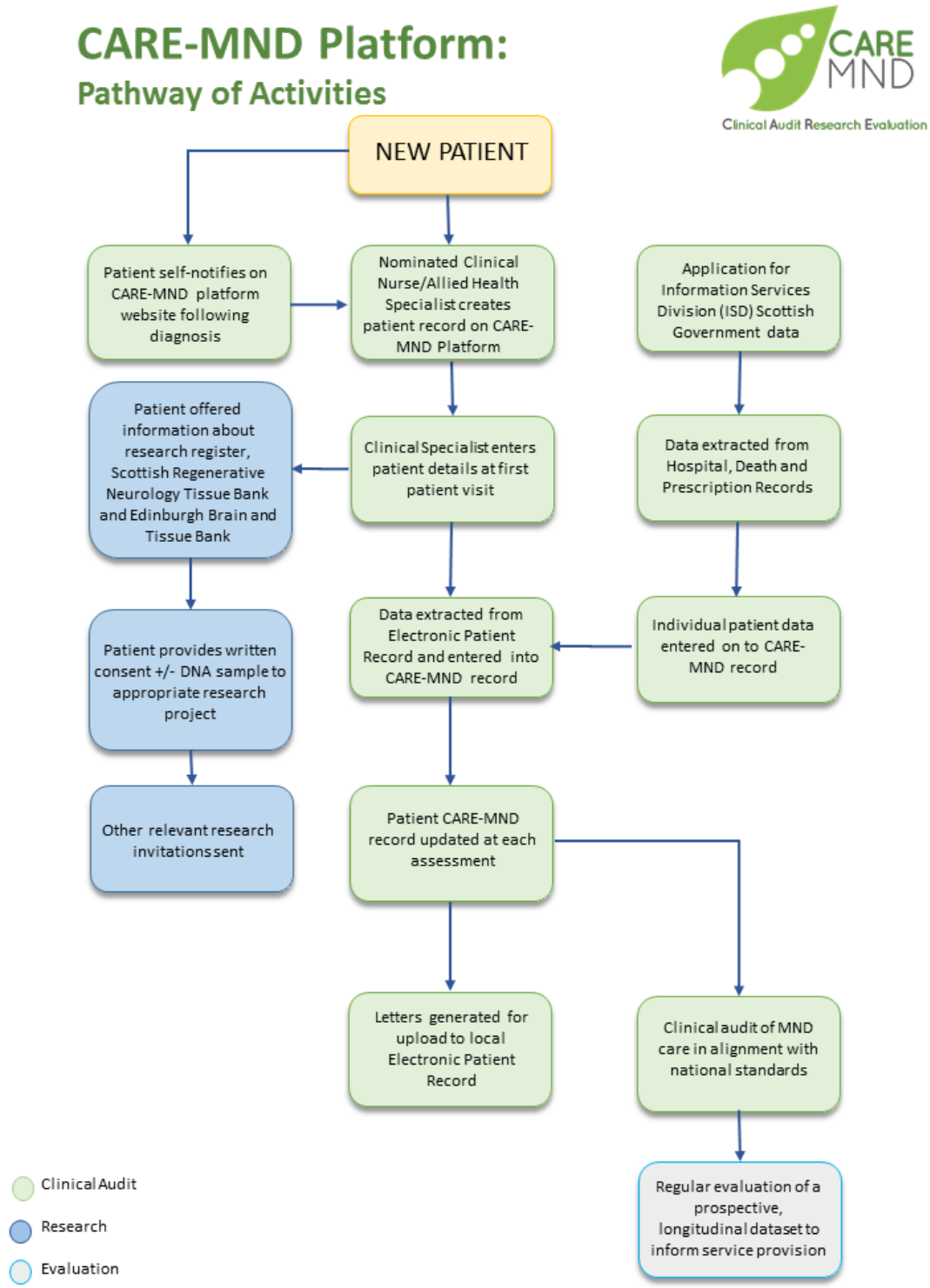


Figure 4 CARE-MND platform: pathway of activities

The CARE-MND platform was launched in February 2017. Clinical specialists received one-to-one training supported by an SOP (see Appendix 2) and piloted its use June to August 2017. During this interim period, data entry were supported by a research and audit team. The platform was supported by a named IT administrator at the University of Edinburgh who designed an in-built feedback system to allow users to generate alerts regarding glitches or suggest bespoke amendments. By Autumn 2017, all MND nurse specialists/allied health professionals were using a finalised model of the CARE-MND platform as their primary mode of recording patient details. The platform generates data and letters which can be uploaded to the health board-specific EPR.

National Health Records Linkage

The CARE-MND platform collates information from multiple sources: i) patient-reported data directly entered by MND nurse specialists/allied health professionals, ii) data extracted from EPR clinical reports (notably records from neurologists, neurophysiologists, radiologists, gastroenterologists, respiratory physicians and palliative care physicians) and iii) data directly obtained or entered by allied health professionals (neuropsychologists, speech and language therapists, dieticians, physiotherapists, ventilation nurse specialists). To ensure completeness, Scottish national government statistics from the Information Services Division (ISD) are sought. These routinely-collected prospective population records are a key strength of the SMNDR[3]. The following data items were sourced:

1. *Scottish Morbidity Records (SMR01)* – patients coded with International Classification of Diseases 10th revision (ICD-10) code G12.2 (Motor Neurone Disease) from inpatient and day case records from non-obstetric and non-psychiatric specialties. Data regarding "Ethnic Group" were also extracted from SMR01 records.
2. *National Records of Scotland (NRS) Deaths* – patients where the ICD-10 code G12.2 (Motor Neurone Disease) is listed as the primary or secondary cause of death. Data regarding "Date of Death" and "Place of Death" were also extracted from NRS records.
3. *Prescribing Information System (PIS)* – patients prescribed riluzole, a drug uniquely prescribed for this disease.

Relevant data are transcribed to the CARE-MND platform. Corroboration with medical records occurs if required. For the purposes of this study and PhD, ISD data were applied for on 3rd May 2017. A summary of the final CARE-MND platform is illustrated in Figure 4.

Ethical Approvals

Ethical approvals were obtained for the SMNDR as stated in Chapter 1, Section 1.6. In 2017, the 15/SS/0126 approval was amended to reflect the CARE-MND brand. The CARE-MND Project Manager was granted approval for use of ISD data by the Public Benefit and Privacy Panel (PBPP) for Health and Social Care, Scotland, on behalf of the University of Edinburgh and our funders at MND Scotland and in line with National Health Care Improvement Standards Scotland[239]. Researchers accessing ISD data completed an approved Medical Research Council Research Data and Confidentiality course.

Evaluation

To assess platform efficiency and sustainability, data capture pre- and post-CARE-MND implementation were audited with reference to MND NICE Guidelines. Seventeen data fields were identified related to MND nurse/allied health specialist contact, health demographics, MND outcomes and interventions (Table 5). The fields were selected because of their relevance to NICE recommendations for care and management, or their importance in guiding MND prognostication[9]. Although CARE-MND incorporates cognitive assessment, in line with NICE standards, these data were not available for the 2011-14 period.

The proportion of data captured using the historical SMART platform (2011-2014) was compared with that using CARE-MND (2015-2017). Pre- and post-intervention data capture was compared using Z-test of proportions. R v3.4.3 was used for all statistical analyses[240].

2.3 RESULTS

Information Services Division (ISD) National Records Linkage

ISD data obtained via PBPP for this study period were examined and processed to extract relevant information for inclusion on the CARE-MND database. The process of extraction and number of data entries examined at each stage are outlined in the flow chart in Figure 5. At the final stage of filtering/processing, raw ISD data were entered onto the electronic CARE-MND database (n = 1186 rows of data). Examples of common ICD-10 data uploaded to CARE-MND included:

- J69.0 Pneumonitis due to food and vomit
- I10 Essential (primary) hypertension
- K59.0 Constipation

Twenty one patients were identified who were not known to CARE-MND. The reasons for their being missed are discussed in Chapter 3.

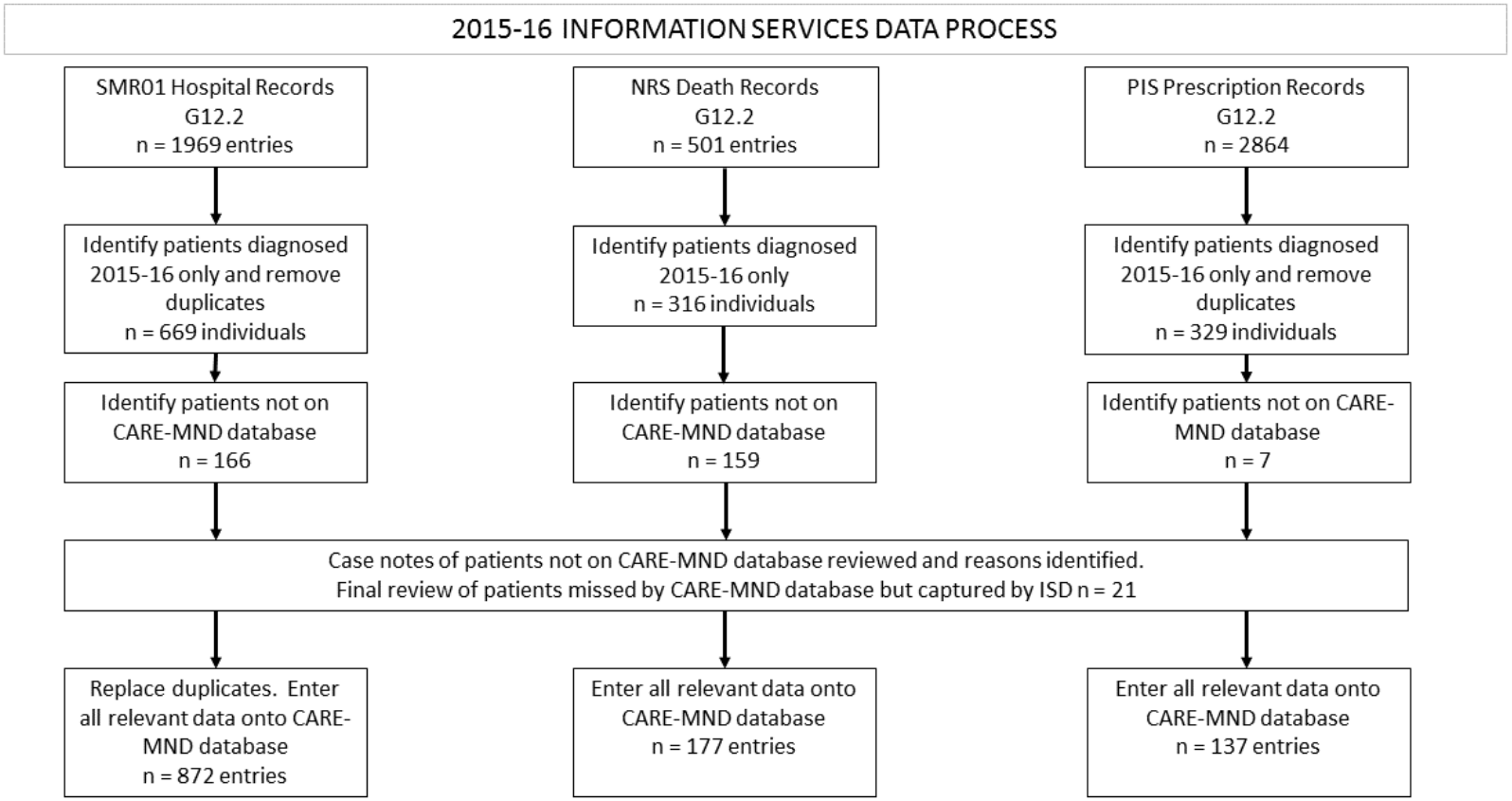


Figure 5 Process of extracting Information Services Division (ISD) data for the 2015-2016 patient cohort

CARE-MND Platform Evaluation

All pwMND in the 2015-17 CARE-MND (n=588) and 2011-14 SMART (n=703) platforms were studied (total n=1291; Table 5). Data field completion for 17 NICE-relevant data fields were extracted from the SMART (2011-14) and the CARE-MND databases (2015-17). Percentage completed fields using the SMART platform ranged from 4-95%; median completion 50%. CARE-MND capture ranged from 32-98%; median 87%. Fifteen of 17 fields were significantly more complete in CARE-MND ($p<0.001$). "Place of death" capture remained high (95%, 97%). "Forced Vital Capacity (FVC)" capture remained low (34%, 32%) (Table 5).

During 2011-14, only 21% (3/14) of health boards had dedicated MND clinical pathways[241]. However, through CARE-MND 100% of health boards now have a standardised proforma, which can guide care and management.

	CARE-MND 2015-2017 n = 588		SMART 2011-2014 n = 703		Statistical Difference (Z-Test of Proportions p-value)
Data type	Data available (number of patients)	% completed	Data available (number of patients)	% completed	
Clinical specialist name	575	98%	370	53%	<0.001
Date of referral to Clinical Specialist	359	61%	348	50%	<0.001
Date of first contact with Clinical Specialist	341	58%	190	27%	<0.001
Date of symptom onset	530	90%	524	75%	<0.001
Date diagnosis confirmed	571	97%	542	77%	<0.001
Electrophysiology	448	76%	212	30%	<0.001
Classification of MND	507	86%	514	73%	<0.001
Past medical history	509	87%	431	61%	<0.001
Family History of MND (Yes/No)	524	89%	415	59%	<0.001
Taking riluzole (Yes/No)	554	94%	434	62%	<0.001
Feeding tube inserted (Yes/No)	522	89%	329	47%	<0.001
Use of Non-invasive ventilation (Yes/No)	527	90%	279	40%	<0.001
Forced Vital Capacity	189	32%	240	34%	0.4845
Amyotrophic Lateral Sclerosis Functional Rating Scale	418	71%	321	46%	<0.001
Epworth Sleepiness Scale Score	225	38%	25	4%	<0.001
Body Mass Index	413	70%	26	4%	<0.001
Place of death	305/314	97%	593/621	95%	0.2987

Table 5 Comparison of data collection pre- and post-CARE-MND

2.4 DISCUSSION

Data access, utility and dissemination

As part of this PhD, I helped to develop, launch and implement CARE-MND, a national electronic platform for prospective, long-term monitoring of MND in Scotland. The CARE-MND platform has been delivered as a result of significant dedicated research funding, supporting investment in personnel and infrastructure. Its success, application, and sustainability is predicated on ongoing engagement from all MND nurse/allied health specialists in Scotland who have now incorporated CARE-MND into routine clinical practice. All recorded incident cases of pwMND in Scotland from 2015 have been phenotypically characterised longitudinally and offered participation in relevant research studies where appropriate.

To monitor sustainability, data entry has been audited monthly since launch. Median data capture as of November 2018 was 96%. Implementation of the CARE-MND platform has resulted in more complete ascertainment of clinical parameters emphasised in NICE Guidelines. This facilitates better coordinated care and continuous assessment of patients' needs together with prompt appropriate referral to relevant multidisciplinary team members[9]. Assessment of forced vital capacity (FVC), a measure of respiratory muscle weakness, is an exception. The low data-capture rate of 32% may reflect a general recent trend towards other methods of assessing respiratory function such as transcutaneous CO_2 monitoring. Anecdotal evidence from Scottish respiratory care-providers suggests that FVC assessment is dependent on accurate technique and results can be inaccurate. A need to transcribe respiratory measures from the local respiratory EPR to the CARE-MND platform by the clinical nurse/allied health specialist may also be a barrier, and partly explain the low capture rates of both FVC and the Epworth Sleepiness Scale Scores (38% capture).

All pwMND in Scotland who require respiratory input have access to appropriate referral pathways whereby they are reviewed at their local clinic or in the community. As a result of our findings we plan to allow respiratory care teams to enter clinical measurements on to

CARE-MND directly. Respiratory monitoring of MND in Scotland is also currently being audited to investigate regional variations (Table 6).

We also acknowledge that capture of other data fields, in particular dates (e.g. Date of referral to Clinical Specialist and Date of first contact with Clinical Specialist) is below average (61% and 58% respectively). On liaising with our clinical teams, our understanding is that these fields were given low priority in favour of seemingly more pertinent measures. However, the publication of NICE Guidelines and our parallel efforts to audit care have highlighted the importance of these fields and we anticipate that future completion will be much improved.

Due to the relative completeness of the CARE-MND dataset, national audits of MND care in Scotland have been possible (Table 6). Through CARE-MND, patient recruitment to the SMNDR, Scottish Neurology Tissue Bank and Edinburgh Brain and Tissue Bank can be recorded and appropriate patients approached for ongoing study. Local and collaborative teams can also apply for access to anonymised data from consenting patients for research. Scottish research projects have included the study of the epidemiology of MND in Scotland[242], genotype-phenotype correlations[243] (both of which are described in this thesis), cognitive profiles[212,213,244], neuropsychiatric phenotypes, and clinico-pathological studies[245]. CARE-MND has also provided data for collaborative work with international partners, including oropharyngeal secretion management[246] and genetics[247,248]. We have also been able to provide information for third sector and governmental organisation to inform service provision[249].

National Audits of MND Care in Scotland	Key Outcomes and Actions	Key Actions
Cognitive assessment	Neuropsychologists are completing assessments but sometimes these are not reflected on CARE-MND.	Neuropsychologists have been given access to CARE-MND for documentation of assessment.
Gastrostomy insertion and survival[250]	Incidence of gastrostomy insertion has increased over time. There is a lower 30-day mortality in 2015-16 than in 1989-1998. Use of riluzole alongside gastrostomy has survival benefits.	Measurement of quality of life was not included in this study and should be assessed subsequently.
Reasons for unexplained hospital admissions	Main reason for admission is falls.	A series of educational videos were created aiming to prevent falls in the community.
Riluzole prescription	Only 40% of people with MND in Scotland are taking riluzole.	Neurologists have been encouraged to document discussions regarding riluzole including reasons for not prescribing.
Augmentative and alternative communication (AAC) provision[251]	All health boards offer AAC in a timely manner. A third of patients use AAC.	A guide for allied health professionals detailing patient thresholds for AAC requirements is in development.
Respiratory monitoring	Patients would like to learn more about non-invasive ventilation through videos and leaflets.	Ongoing.

Table 6 List of national audits of MND care in Scotland since CARE-MND launch

Challenges and Future Development

We have developed an electronic platform that is fit for purpose as a contemporary audit and research tool. As the platform operates via a secure website, it is currently not possible for MND specialists/allied health specialists to access the website off-site (e.g. via tablet) unless they have in-built internet data on their device. As such, paper notes are required for some specialists visiting patients at home. These paper records, however, can be transcribed to CARE-MND on return to their work base. We have strived to ensure that the platform complies with local NHS health board governance requirements. This is coordinated and monitored by the MND nurse consultant for Scotland. Currently the platform does not integrate with existing NHS EPR portals but we are exploring this for the future. As a temporary solution, structured clinical letters can be automatically generated from CARE-MND content which can be printed and uploaded to local electronic records. The evolving nature of technology and the growing evidence-base in MND will inevitably necessitate further amendments.

2.5 CONCLUSIONS

In a condition for which systematic data collection is challenging and for which treatments are currently limited, we have evolved a sustainable platform in which patients throughout Scotland can be included prospectively and longitudinally in routine audit of care and research. All health boards in Scotland now have dedicated care pathways for MND and, through national monitoring, we have demonstrated an improvement in equity of care and service provision, as well as stratification of the population for research, including future clinical trials. In a technologically advancing climate, we aim to optimise available resources and pave the way for other neurological disorders.

3. EPIDEMIOLOGY OF MND IN SCOTLAND 2015-2017

The following content has been published in *Journal of Neurology*[242] and re-formatted for this thesis:

Danielle J Leighton, Judith Newton, Laura J Stephenson, Shuna Colville, Richard Davenport, George Gorrie, Ian Morrison, Robert Swingler, Siddharthan Chandran, Suvankar Pal on behalf of the CARE-MND Consortium. Changing epidemiology of motor neurone disease in Scotland. *J Neurol.* 2019 April; 266(4):817-825. https://doi.org/10.1007/s00415-019-09190-7.

3.1 BACKGROUND AND AIMS

Background

Recent discoveries in the genetics of MND have provided important clues for aetiopathogenesis and support a multifactorial pathophysiological process with genetic and/or environmental risk factors[143,160]. Regional variation in disease incidence could therefore be expected, corresponding with differing population ancestral origin and environmental exposures[252]. However, a recent meta-analysis of global registry data of MND (ALS *and* MND-subtypes) observed epidemiological homogeneity amongst populations of predominantly European origin[253].

Scotland benefits from a culture of longstanding MND data capture[199,200]. Annual incidence of MND in Scotland was 2.40 per 100,000 population (95% confidence interval (CI) 2.22-2.58) between 1989-1998, standardised to 1994 mid-year Scottish population estimates[3]. Multiple sources were used for case ascertainment including direct notification from MND clinical nurse/allied health specialists, neurologists and neurophysiologists and national hospital discharge records from Information Services Division (ISD) Scotland. A two source capture-recapture model demonstrated 97.8% national coverage[3].

SMNDR coverage declined between 1999 and 2014[227]. However, in 2015, the CARE-MND platform permitted more thorough data capture. This coincided with public awareness initiatives and government-led doubling of the number of MND clinical nurse specialists in Scotland.

Aims

I aimed to use CARE-MND to describe the epidemiological characterisation of MND in Scotland in 2015-2017, in comparison to the 1989-1998 study period. Phenotypic characterisation and social deprivation mapping of the complete 2015-16 cohort were also studied.

3.2 METHODS

Inclusion Criteria

All adults living with MND in Scotland are eligible for inclusion in CARE-MND. Inclusion criteria for this study were: i) diagnosed by a neurologist with suspected ALS, possible, probable or definite ALS according to El-Escorial revised criteria OR an MND subtypes (primary lateral sclerosis (PLS), progressive bulbar palsy (PBP), progressive muscular atrophy (PMA)) [27], ii) diagnosed between 1st January 2015 and 31st December 2017, iii) ≥16 years at diagnosis, iv) resident in Scotland at time of diagnosis or receiving care in Scotland during study period.

Case ascertainment

CARE-MND retrieves population data from five sources: two CARE-MND clinical sources, three Information Services Division (ISD) Scotland health records sources (Figure 6). This

system draws on the strengths of an integrated healthcare system in Scotland, whereby individuals are identified by a unique Community Health Index (CHI) number, which can be referenced against local and national health records. ISD data were sourced for incident patients in 2015 and 2016. Medical records for all pwMND unique to ISD were examined to clarify diagnosis and explore reasons for their not being known to CARE-MND.

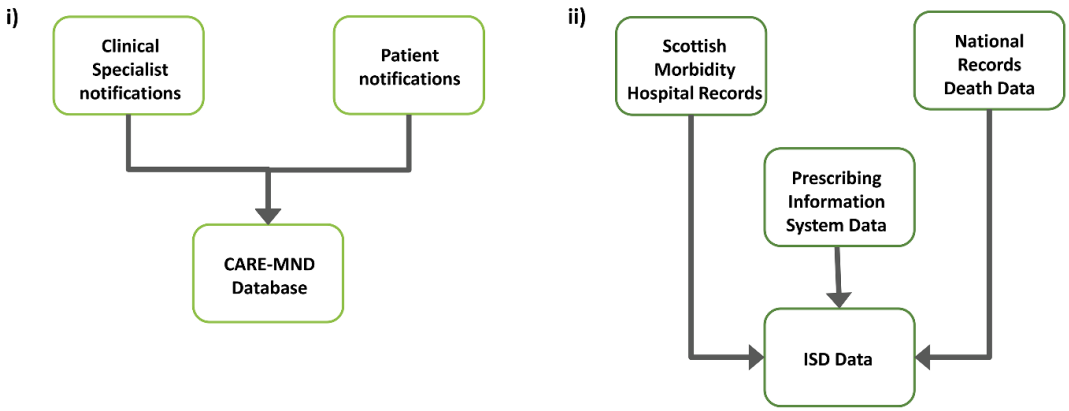


Figure 6 Multiple source algorithm for accurate ascertainment of MND cases: i) CARE-MND platform, ii) Information Services Division data

i) CARE-MND Database

MND nurse/allied health specialist notification: The CARE-MND platform is a prospective clinical tool and all cases are monitored by neurologists and nurse/allied health specialists. If a diagnosis is revised or revoked, the patient is removed from the database. CARE-MND notifications are therefore considered 'gold-standard'. Patients are referred to a local MND nurse/allied health specialist, who coordinates care. The CARE-MND electronic platform operates via a secure website hosted on an academic server. MND specialists enter information about incident patients following referral and until death. If a patient is diagnosed with MND shortly before, or after, death the relevant care provider can refer for retrospective data entry.

Patient self-notification: PwMND are invited to participate in the CARE-MND platform and can self-notify online. A member of the CARE-MND team contacts the patient's neurology consultant/clinical specialist to confirm diagnosis before entry onto the database.

ii) ISD Data

Information Services Division (ISD) Scotland data were extracted (see details Chapter 2.2 National Health Records Linkage). Patients were sourced if they were assigned International Statistical Classification of Diseases v10 (2016) (ICD-10) codes for Motor Neurone Disease (G12.2) or Frontotemporal Dementia (G31.0 and/or F02.0) on hospital records or as primary or secondary cause of death during 2015 and 2016. Patients prescribed riluzole, a drug uniquely used for MND, were also included.

Data Extraction and Statistical Analysis

CARE-MND clinical specialist and patient notifications are related sources and were collapsed into one measurement. ISD information sources were also combined. The result was two independent sources for capture-recapture analysis. Maximum likelihood estimates were used to determine total incidence rates and CARE-MND database coverage[3].

Incidence data were analysed by age and sex. Prevalence data were extracted directly from CARE-MND. Incidence and prevalence rates were calculated in reference to NRS mid-population estimates[221]. Incidence rates were standardised using the direct method to the US 2010 Census population to facilitate international comparison[254,255]. The US population has historically been used for Scottish cohort and other population standardisation[3,253]. The 2010 population was chosen to allow direct comparison with recent pooled incidence data[253]. Confidence intervals were calculated assuming a Poisson distribution[256].

The CARE-MND platform provides a standardised national proforma. The following fields were extracted: i) age of onset, ii) age of diagnosis, iii) sex, iv) site of onset, v) family history of MND (first, second or third degree relative), iv) family history of dementia (first, second or third degree relative), vii) clinical syndrome, viii) ethnic origin, ix) riluzole use, x) gastrostomy

insertion, xi) non-invasive ventilation (NIV) use. T-tests and Chi-square tests (parametric data) and Wilcoxon rank-sum test (nonparametric data) were used for univariate analyses of sex differences. Proportions were compared using Z-test of proportions. Individuals diagnosed in 2015-16 were followed up for at least one year (censorship 31st December 2017); 6-month and 12-month survival from onset and diagnosis were examined.

Patients were mapped to their corresponding Scottish Index of Multiple Deprivation (SIMD) 2016 data zone. The SIMD is a Scottish Government tool that ranks 6976 small data zones (760 people per zone) according to level of deprivation. Seven domains are used: employment, income, crime, housing, health, education and access (to medical care, public transport). SIMD rank was sourced for each patient and the corresponding SIMD quintiles (1 most deprived, 5 least deprived) analysed using Dunn's test of multiple comparisons using rank sums.

R version 3.4.3 was used for all statistical analyses (packages "epitools", "epiR", "DescTools" and "dunn.test")[240].

Ethics

As described previously, ethical approvals were obtained for the SMNDR/CARE-MND (MREC/98/0/56 1989-2010, 10/MRE00/78 2011-2015, Scotland A Research Ethics Committee (15/SS/0216) 2015-present). Access to ISD data was approved by the Public Benefit and Privacy Panel for Health and Social Care, Scotland.

3.3 RESULTS

Crude Results

Numbers of pwMND sourced using CARE-MND and ISD are presented in Table 7**.**

Year		Source	Number of pwMND
2015	**CARE-MND**	Clinical Nurse/Allied Health Specialist Notifications	221
		Total	*221*
	ISD	SMR01	179
		NRS Death Records	112
		PIS	92
		Total Unique	*211*
2016	**CARE-MND**	Clinical Nurse/Allied Health Specialist Notifications	185
		Total	*185*
	ISD	SMR01	107
		NRS Death Records	39
		PIS	69
		Total Unique	*152*
2017	**CARE-MND**	Clinical Nurse/Allied Health Specialist Notifications	189
		Patient Notifications	5
		Total	*192*
	ISD	SMR01	Na
		NRS Death Records	Na
		PIS	Na
		Total Unique	*Na*

Table 7 Sources of incident cases of MND 2015-2017

All CARE-MND patients were diagnosed by a consultant neurologist and were aware of their diagnosis. If a diagnosis is revised or revoked, the patient is removed from the database. CARE-MND notifications can therefore be classed as a 'gold-standard'. The CARE-MND database alone identified 406 true MND cases diagnosed in 2015-16 (Table 8). There were 596 prevalent patients in 2015-2016 coded with the G12.2 code in ISD datasets. Of these, 342 were 'true' incident cases (also present in CARE-MND 2015-16). A further 262 patients were coded with G12.2: 21 were incident MND cases unique to ISD and were included in our analyses (Table 7, Table 8); eight had MND but were diagnosed before 2015 and were not included. Case-notes of the remaining 233 patients were reviewed; none had MND. In summary, we identified 21/596 (3.5%) people with MND unique to ISD and 233/596 (39.1%) patients from ISD records who were coded with an ICD-10 MND code inaccurately. Forty two percent (99/233) of these patients had progressive supranuclear palsy (PSP). Other

diagnoses included pseudobulbar palsy secondary to cerebrovascular disease or dementia. Sensitivity of ISD code G12.2 for incident patients 2015-16 was therefore 0.89 (95% CI 0.86-0.92) with a positive predictive value (PPV) of 0.61 (95% CI 0.57-0.65). The specificity and negative predictive value were 100% and 1 respectively, as MND is rare in the general population. Sensitivity and PPV for hospital records were 0.74 (95% CI 0.69-0.78) and 0.66 (95% CI 0.62-0.71) respectively; for death records 0.42 (95% CI 0.37-0.47) and 0.55 (95% CI 0.49-0.61); for PIS records 0.40 (95% CI 0.35-0.45) and 0.96 (95% CI 0.92-0.98).

Average case ascertainment/coverage of the CARE-MND database was 98.9% (Table 8).

Year	2015	2016
CARE-MND notifications	221	185
Notifications unique to CARE-MND	19	45
Notifications unique to ISD	9	12
Notifications common to CARE-MND and ISD	202	140
Total notifications	230	197
Maximum Likelihood Estimate	0.8465	3.8671
CARE-MND coverage (%)	99.6	98.1
Mid-year population estimate	5373000	5404700
Crude incidence (per 100,000 population)	4.28	3.64

Table 8 Coverage of CARE-MND Platform and Crude Incidence Rates 2015-16

Case-notes of the 21 patients identified through ISD alone were examined. Sixteen (76.2%) patients were diagnosed shortly before death (median 3.5 days (interquartile range (IQR) 1.8-9.5)), in which case contact with a MND specialist might not have been made or pursued. Site of onset for these patients varied: bulbar (n=6, 37.5%), limb (n=6, 37.5%), respiratory (n=3, 18.8%), weight loss (n=1 (6.3%). Two patients were diagnosed posthumously: one from electromyography, one on post-mortem. The remaining three patients were: a nursing home resident, unknown to specialist teams; a patient with FTD-predominant disease receiving care from other specialists; a patient with PLS who declined specialist input.

Due to the excellent coverage of CARE-MND, incidence for 2017 was estimated using CARE-MND values alone: 192 new diagnoses, giving a crude incidence of 3.55/100,000 (3.07-4.09) (using 2016 mid-year population estimate as 2017 not yet available). Crude incidence over the three-year period was 3.83/100,000 person-years (3.53-4.14). On 31st December 2015, 409 people were living with MND in Scotland according to CARE-MND/ISD figures (crude prevalence 7.61 per 100,000 (6.89-8.39)). Similarly, prevalence rates in 2016 and 2017 were 413 (7.64/100,000 (6.92-8.42)) and 422 (7.81/100,000 (7.08-8.59)).

Direct Standardisation

Incidence rates for 2015 and 2016 were age and sex standardised to the US Census Population 2010 using the direct method (Table 9). This allows direct comparison with other populations.

		1989-98			2015			2016		
Sex	Age Range	Number of Cases in Cohort	Crude Incidence (per 100,000 population)	Standardised Incidence Value	Number of Cases in Cohort	Crude Incidence (per 100,000 population)	Standardised Incidence Value	Number of Cases in Cohort	Crude Incidence (per 100,000 population)	Standardised Incidence
Male	0-4	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
Female	0-4	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
Male	5-9	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
Female	5-9	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
Male	10-14	1	0.06	6.35	0	0.00	0.00	0	0.00	0.00
Female	10-14	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
Male	15-19	1	0.06	6.78	0	0.00	0.00	0	0.00	0.00
Female	15-19	1	0.07	7.52	0	0.00	0.00	0	0.00	0.00
Male	20-24	1	0.05	5.51	0	0.00	0.00	0	0.00	0.00
Female	20-24	1	0.05	5.29	0	0.00	0.00	0	0.00	0.00
Male	25-29	4	0.19	20.21	0	0.00	0.00	0	0.00	0.00
Female	25-29	3	0.15	15.70	0	0.00	0.00	0	0.00	0.00
Male	30-34	4	0.19	18.99	1	0.59	58.80	1	5.80	58.02
Female	30-34	1	0.05	4.98	0	0.00	0.00	1	5.57	55.49
Male	35-39	9	0.49	49.21	2	1.27	127.55	2	12.40	124.54
Female	35-39	5	0.27	27.37	0	0.00	0.00	1	6.01	60.89
Male	40-44	20	1.20	124.73	1	0.59	61.08	4	24.25	252.08
Female	40-44	5	0.30	31.49	1	0.56	58.43	2	11.58	121.56
Male	45-49	37	2.18	244.36	9	4.70	526.62	6	31.68	355.05
Female	45-49	15	0.87	100.05	2	0.98	112.60	1	4.93	56.70
Male	50-54	43	3.07	335.65	11	5.58	609.80	2	10.13	110.76
Female	50-54	30	2.05	232.98	5	2.40	273.10	5	23.89	271.55

Male	55-59	79	6.00	571.42	20	11.29	1075.69	8	44.26	421.55
Female	55-59	55	3.87	392.46	6	3.23	327.58	5	26.30	266.75
Male	60-64	86	7.05	569.46	20	12.97	1047.35	21	134.27	1084.53
Female	60-64	69	5.01	437.90	13	7.99	698.16	12	72.66	635.11
Male	65-69	123	11.31	661.92	31	20.46	1197.65	31	202.00	1182.20
Female	65-69	110	8.39	552.29	17	10.53	693.05	11	67.05	441.37
Male	70-74	120	13.16	558.51	18	16.68	707.84	16	143.02	606.99
Female	70-74	110	8.82	444.02	14	11.25	566.53	15	117.97	593.89
Male	75-79	80	15.47	492.32	18	21.76	692.63	12	145.43	462.81
Female	75-79	76	8.83	365.16	20	19.07	788.70	11	105.43	436.01
Male	80-84	48	13.90	318.92	6	10.91	250.20	10	177.93	408.24
Female	80-84	56	7.93	273.50	10	12.61	434.81	11	137.13	472.95
Male	85-89	8	5.79	73.76	1	3.69	46.96	4	141.41	180.13
Female	85-89	18	4.54	106.54	3	6.14	144.00	3	60.37	141.67
Male	90+	2	5.00	25.79	0	0.00	0.00	2	166.43	85.85
Female	90+	3	1.72	23.34	1	3.52	47.83	0	0.00	0.00

Table 9 Direct standardisation by age and sex for 2015, 2016 and 1989-98

Age-standardised incidence for 2015 in Scotland was 3.42/100,000 of the population (95% CI 2.99-3.91) (males and females). Age-standardised incidence for males was 2.00/100,000 (95% CI 1.68-2.39) and for females 2.64/100,000 (95% CI 2.12-3.27); age-adjusted male-to-female relative risk (RR) 0.76:1. Age-standardised incidence for 2016 was 2.89/100,000 (95% CI 2.50-3.34) overall; for males, 3.51/100,000 (95% CI 2.90-4.24), and for females, 2.26/100,000 (95% CI 1.78-2.86) with an age-adjusted male-to-female RR of 1.55:1. In both years, peak age-group incidence for males was 65-69 years. For females, peak incidence occurred later in the 75-79 age group. In 2016, however, it occurred earlier in the 60-64 age group (Figure 7).

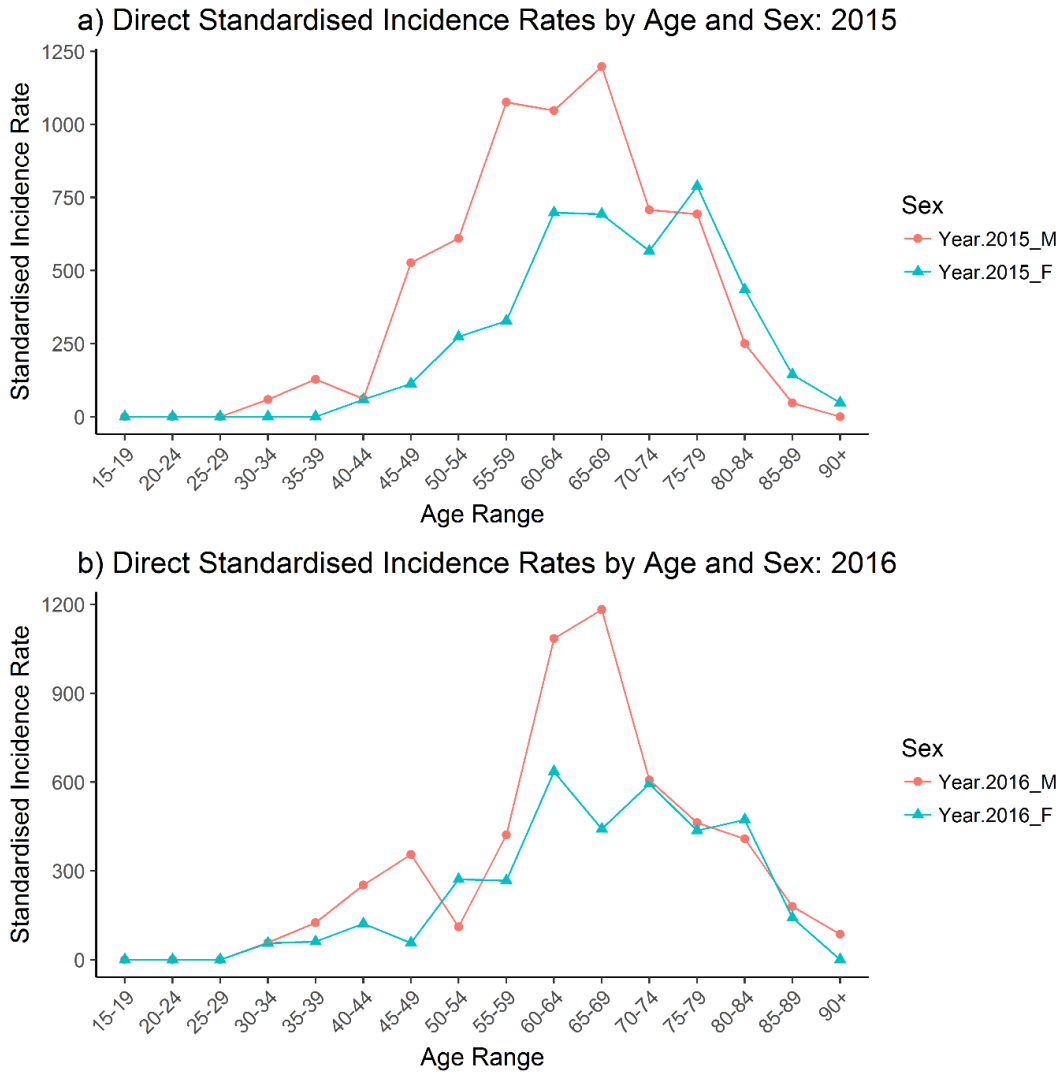


Figure 7 Age and sex rates direct standardised to the 2010 US Census population for a) 2015 and b) 2016 cohorts

Incidence rates for 1989-98 were also age-standardised to allow time period comparison (Figure 8). Overall, age-standardised incidence for the 1989-98 period was 2.32/100,000 (95% CI 2.26-2.37). In 2015-16, incidence was greater than 1989-98 across most age groups, but particularly in the 55-74 year cohort.

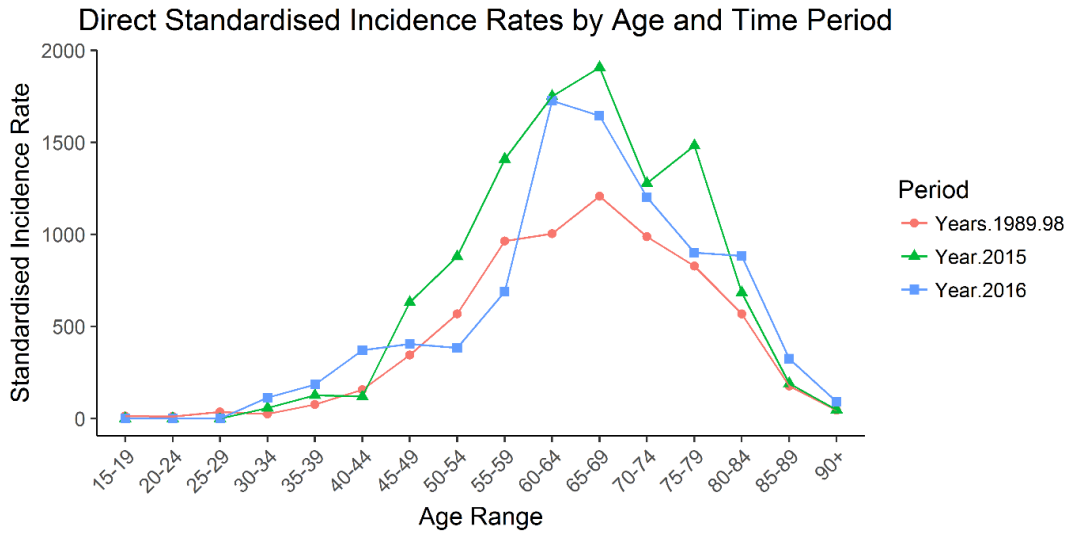


Figure 8 Time period comparison of incidence rates for i) 1989-98, ii) 2015 and iii) 2016 direct age standardised to the 2010 US Census population

Finally, age *and* sex time period comparison was plotted (Figure 9). This highlights that the clearest change is among males age 60-69. In particular, there was a 79-81% increase in incidence of males age 65-69 years having MND in 2015-16 versus 1989-98.

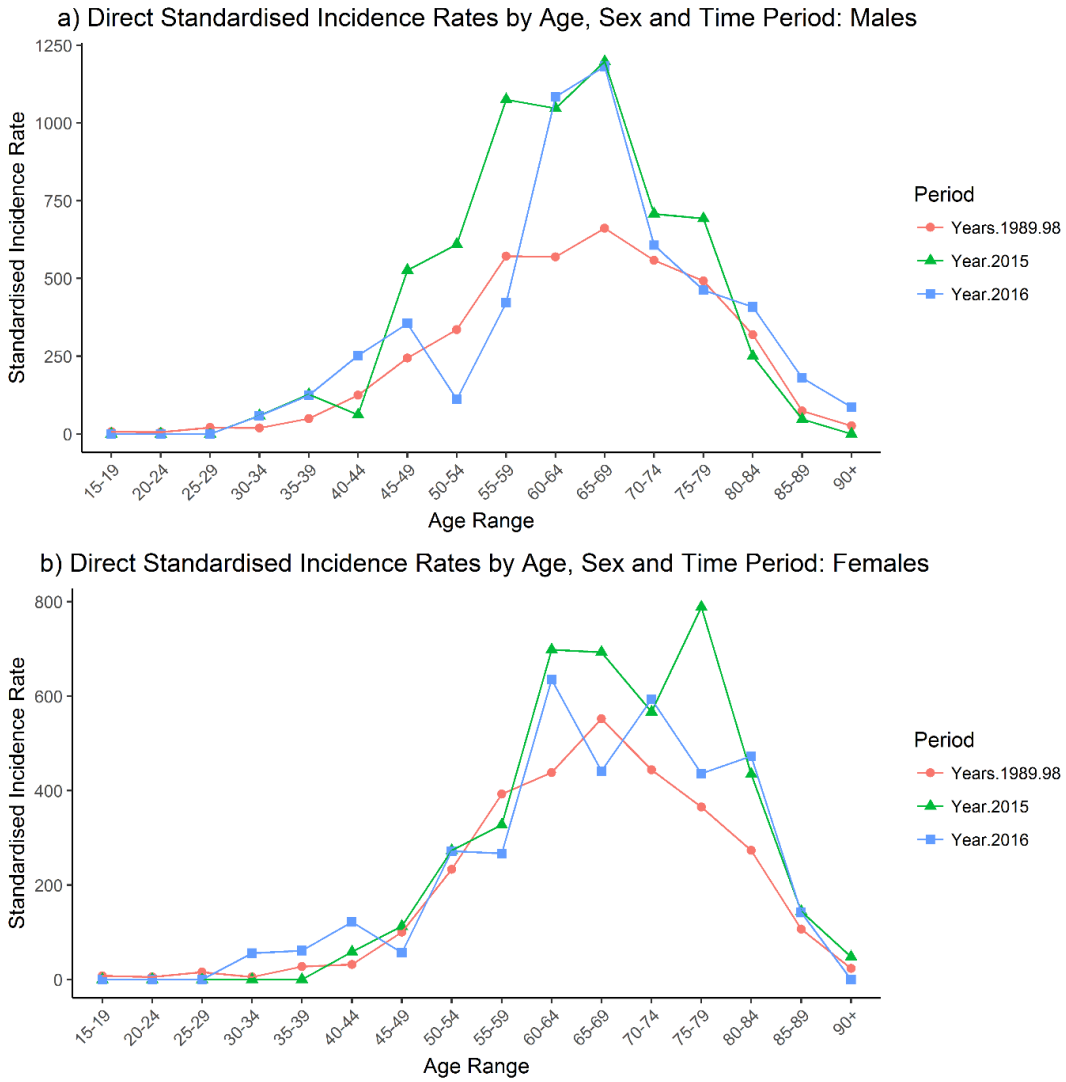


Figure 9 Time period comparison of incidence rates for i) 1989-98, ii) 2015 and iii) 2016 direct age and sex standardised to the US Census population 2010 (a) Males, b) Females)

Patient Characteristics

Phenotypic characteristics were evaluated from the 2015 and 2016 cohorts (Table 10). Statistical comparisons were made between males and females with Bonferroni correction for multiple testing. Females were significantly older at onset of symptoms and point of diagnosis (p=0.00086 and 0.00099 respectively). Proportionally, more female pwMND than males had bulbar-onset disease although this did not reach statistically significance

(p=0.0024). However, there were significantly fewer females than males with onset of disease in the upper limbs (p=0.00072). Although there was no sex difference for respiratory-onset disease, significantly fewer females received NIV ($p=3.27x10^{-5}$). Six-month survival from onset was significantly worse in females than males (p=0.00039); however, this did not remain the case at 12-months, or using survival from diagnosis.

Patient Characteristic (n=427)	% Complete Data	Total	Males (n=258)	Females (n=169)	Significance Test	p-value
Male-to-Female Relative Risk	100	1.53:1				
Mean Age of Onset (SD), years	97.7	65.3 (11.6)	63.7 (11.3)	67.6 (11.7)	t-test	0.00086*
Mean Age of Diagnosis (SD), years	100	66.8 (11.2)	65.4 (11.0)	69.0 (11.1)	t-test	0.00099*
Median Time to Diagnosis (IQR), months	97.7	11.0 (7.0-21.0)	11.0 (7.0-20.0)	11.0 (7.0-23.0)	Wilcoxon-rank sum test	0.92
Classification (%)	100				Chi square	0.039
- **ALS**		79.4	80.6	77.5	Z-test proportion	0.51
- **MND-FTD**		6.8	6.2	7.7	Z-test proportion	0.69
- **PBP**		4.9	2.7	8.2	Z-test proportion	0.018
- **PMA**		4.0	4.3	3.6	Z-test proportion	0.91
- **PLS**		3.3	3.5	3.0	Z-test proportion	0.98
- **Other (bibrachial/flail limb)**		1.6	2.7	0.0	Z-test proportion	0.077
Site of Onset	98.8				Chi square	0.0048
- **Bulbar**		28.2	22.7	36.7	Z-test proportion	0.0024
- **Lower limb**		28.4	28.1	28.9	Z-test proportion	0.16
- **Upper limb**		20.6	26.2	12.0	Z-test proportion	0.00072*
- **Mixed (upper limb, lower limb, bulbar)**		17.1	16.0	18.7	Z-test proportion	0.56
- **Cognitive change**		2.1	3.1	0.6	Z-test proportion	0.16
- **Respiratory**		1.7	1.6	1.8	Z-test proportion	1
- **Other (weight loss, camptocormia)**		1.9	2.3	1.2	Z-test proportion	0.64
Ethnicity (%)	90.6				Chi square	0.71
- **White Scottish**		75.2	75.1	75.3	Z-test proportion	1
- **White British/Irish/Not Specified**		23.3	23.2	23.3	Z-test proportion	1
- **Ethnic Minority**		1.3	0.9	1.3	Z-test proportion	1
Family History (%)						

- **MND**	96	8.5	5.6	12.8	Chi square	0.013
- **Dementia**	80	27.7	26.5	29.7	Chi square	0.22
Riluzole Medication Prescription (%)	98.1	37.9	37.6	38.4	Chi square	0.4
Gastrostomy Insertion (%)	99.5	31.5	32.0	30.8	Chi square	0.71
Non-invasive Ventilation (%)	95.3	33.9	41.6	21.7	Chi square	3.27x10-5*
6-Month Survival (%)						
- **From Onset**	97.6	94.2	95.3	92.1	Z-test proportion	0.00039*
- **From Diagnosis**	100	70.0	74.4	62.7	Z-test proportion	0.092
12-Month Survival (%)						
- **From Onset**	97.6	70.0	85.0	80.0	Z-test proportion	0.0033
- **From Diagnosis**	100	51.3	55.8	44.4	Z-test proportion	0.0032

*Table 10 Patient characteristics and sex comparison of the 2015-2016 cohort (n=427) *Bonferroni correct p-value <0.0016*

Scottish Index of Multiple Deprivation (SIMD) Mapping

PwMND for whom postcodes were recorded on the database (n=382) were assigned their corresponding SIMD social deprivation small area rank. Ranks ranged from 24 (most deprived) to 6953 (least deprived) (median 3512, IQR 1964-5210). There was an equal spread across deprivation quintiles (Figure 10). Dunn's test of multiple comparisons using rank sums confirmed that this was not significant (p=0.41).

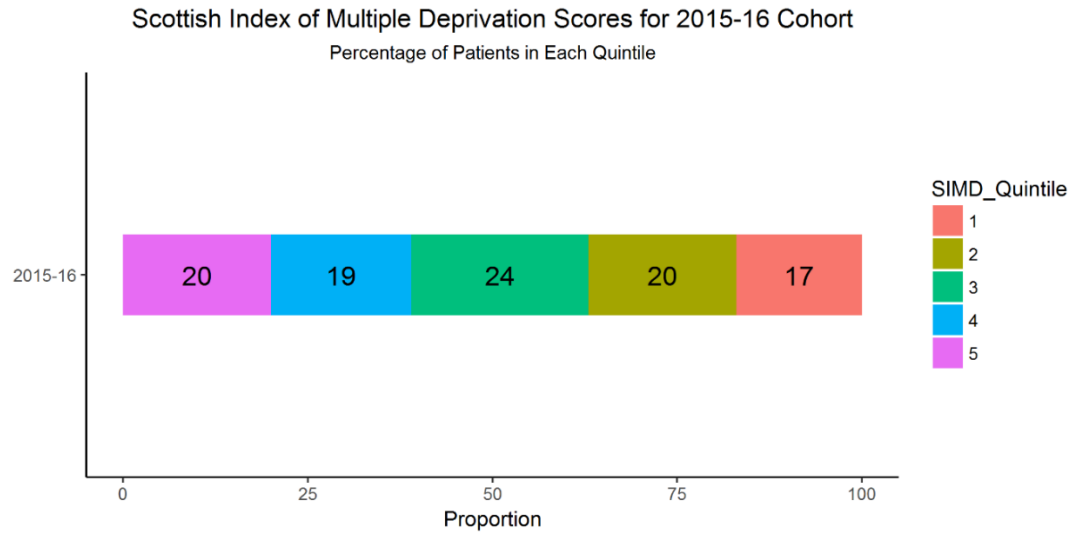


Figure 10 Scottish Index of Multiple Deprivation (SIMD) Scores for incident 2015-16 patients by quintiles from 1 (most deprived) to 5 (least deprived): 1) 1-1395, 2) 1396-2790, 3) 2791-4185, 4) 4186-5580, 5) 5581-6976)

3.4 DISCUSSION

Capture-Recapture

Through this work, I can report an up-to-date and standardised epidemiological analysis of the Scottish MND population. Case ascertainment methods were similar to those employed in the 1989-98 Scottish MND Register analysis and capture-recapture coverage is improved (98% in 1989-98, 99% in 2015-16)[204]. Some population registries report relatively low percentage coverage (81.1-89.5%) but our ascertainment is comparable with other small country/regional epidemiological MND studies[257–259]. We identified 21 pwMND not captured by CARE-MND: the reasons for their exclusion will inform future nursing care, ensuring that, for example, patients in nursing homes and those with FTD are offered access to adequate care support.

Validation of ISD Coding

With reference to the CARE-MND database, ISD records had relatively low PPVs for MND. Death records were likely under-representative due to the relatively short follow-up time for this incident cohort (median six months, maximum 18 months). The high predictive power of prescribing records is expected as riluzole is unique to the disease. One of the main problems with coding was inaccurate coding of patients with PSP. While individuals with this condition can develop a bulbar palsy similar to that seen in MND, their disease course and pathology is distinct. This error has been observed previously and has implications for both national MND and PSP statistics[260–262].

Incidence and Prevalence

Prevalence in 2015-17 ranged from 7.61-7.81/100,000 population. This is comparable with recently published cohorts in Italy, Cyprus and the Faroe Islands[5,263,264]. However, it is lower than prevalence reported in other Italian and Dutch studies (approximately 10 per

100,000 of the population)[257,265,266], perhaps suggesting poorer survival in Scotland. Future work comparing survival in the Scottish population with other European populations may guide management direction.

Direct standardised incidence in 2015-16 (average 3.16/100,000) has increased by 36.0% compared with the 1989-98 period. Although this contrasts with other historical 10 year studies[3,267,268], it mirrors longer and more recent analyses[21,22]. The age-standardised incidences of MND in Scotland for both 2015 and 2016 are the highest reported in the literature. Combined incidence for 2015-16 is 66.9% higher than pooled Northern European standardised rates[253].

Possible hypotheses for these observations can be categorised as follows: i) ascertainment, ii) environmental, iii) genetic.

i) Ascertainment

Ascertainment is the strongest argument for our relatively high incidence rates. The data suggest that no particular age groups are overlooked; older age groups are well-represented (11.9% of cohort ≥80 years at diagnosis). This challenges recent discourse suggesting that MND is underdiagnosed in older populations[269]. However, this does not explain the increase over time. Awareness of MND amongst health professionals and the public has heightened in recent years. In 2015, Scottish MND funding and care services were boosted through substantial government investment. This may, in part, explain the peak in incidence in 2015. 2016 marked a doubling of MND clinical nurse/allied health specialists in Scotland, with a specialist-to-prevalent-patient ratio of 1:26. Each nurse acquires approximately 12 new pwMND annually, compared with 31 new pwMND annually between 1989 and 1998. Similarly, in 1989 there was one neurologist for approximately every 231,500 people in Scotland but the ratio in 2016 was approximately 1:63,500. New patient referrals to neurology in Scotland rose by 50% from 2007 to 2016, suggesting better access to and more timely referral for tertiary review[270]. Indeed, median diagnostic delay in this study was better than pooled European delays (11.0 months vs 12.0 months)[252].

Concomitant with an increase in neurology service provision is better awareness of the extended phenotype of MND. In comparison to 1989-98, MND-FTD is a new addition (6.8% of 2015-16 cohort). Such patients were previously classified as having a "MND plus" disorder[205,227,259]. Better recognition of the cognitive phenotype, using tools such as the Edinburgh Cognitive and Behaviour ALS Screen, may explain some of the change in incidence over time[206].

ii) Environmental

The rise in incidence is dominated by males age 65-69, contradicting studies suggesting older women drive increased incidence in aging populations[227,257,271]. In a Scottish analysis of pwMND >80 years, standardised incidence was greater in men suggesting a possible localised geographical effect[204]. We found no association between MND incidence and social deprivation (SIMD). This agrees with recent findings in MND (including a historical Scottish case-control study) and a Scottish Parkinson's Disease cohort[272–275]. SIMD is a marker of residence, rather than individual environmental and lifestyle factors; the latter requires separate evaluation.

The change in incidence might be consequent on improved survival from competitive diseases (such as cardiovascular disease), allowing for increased manifestation of MND (a Gompertzian model)[271]. Age and sex adjusted mortality from heart disease has decreased by 35.5% in Scotland over the last 10 years, with noted decline across all health boards and SIMD quintiles[276]. Indeed previous studies observing rising incidences suggest that better care and treatment of people over the age of 60 may be contributory, especially in the context of MND for which there is no disease-modifying treatment[201].

Similarly high standardised incidence rates have been observed in Canterbury, New Zealand (where capture rates and population size are comparable[22]), the Faroe Islands and Finland which raises the possibility of latitudinal influence, similar to that found in multiple sclerosis (MS)[253].

iii) Genetics

Since the identification of 'genetic' MND, in particular, *SOD1* mutations and the *C9orf72* hexanucletoide repeat expansion, patients are increasingly undergoing genetic testing as part of the diagnostic work-up of MND. Clinicians are more aware of the phenotypic spectrum of disease of these mutations; for example, *C9orf72* carriers can present with a pure FTD phenotype but have subsequent sequential motor involvement[75]. The *C9orf72* expansion was first associated with MND in 2011 and so it is possible that these patients were missed from historical epidemiological cohorts[73].

A previously published *SOD1* genetic variant (p.I114T) is thought to be a founder mutation in the Scottish population[219]. While we do not anticipate that the frequency of this mutation has changed over time, it may suggest presence of other Scottish haplotypes which are driving disease. The relative ethnic homogeneity of Scotland may support this hypothesis. The ethnicity of this cohort is comparable with that of the general population in Scotland, 2015[277]. People who identify as white have been found to have a higher incidence of MND than ethnic minorities in more diverse populations, irrespective of socioeconomic status[273]. The relative ethnic homogeneity of Scotland does not allow us to come to such conclusions but does lend evidence to a possible genetic influence in Scotland. Genetic admixture seen in African American populations is thought to be protective against disease, whereas people of European origin are more likely to harbour homozygous and "probably damaging" genetic alleles[278]. Genetic data were not available at the time of publication of this part of the PhD but are explored in later chapters.

Phenotypic Characteristics

In this cohort, the male-to-female ratio and mean ages of onset/diagnosis are typical[3,5,199,243]. The majority of patients (79.4%) had an ALS form of MND, the remainder MND subtypes including MND-FTD (6.8%, a rate very similar to that observed in Ireland[268]). Family history of MND was higher in females than males although this did not

reach statistical significance. As in the previous Scottish cohort, limb-onset MND dominated (49.0% of cases). Pooled Northern European estimates for proportion of bulbar-onset patients is 45.4%[252]. We include a cohort of pwMND presenting with mixed signs and symptoms; historically these have been considered bulbar-onset for the purposes of dichotomisation. When combined with our proportions of bulbar patients the figures are indeed comparable[204].

Rates of gastrostomy insertion are similar to the LIGALS population, although NIV use is lower (33.9 vs 55.7%)[5], perhaps reflective of the relatively short follow-up period.

Examining between-sex differences, we observe significantly higher ages of onset and diagnosis in females, with significantly poorer survival of females at six months after onset. More females than males have bulbar-onset disease (although this did not reach statistical significance) which might explain the drop in survival shortly after onset. Interestingly, significantly fewer females than males made use of NIV; the reasons for this, and the impact on survival, will be explored in later chapters.

Limitations

While data capture methods are similar between the 1989-98 and 2015-17 cohorts, there were key differences in diagnostic inclusion criteria: the historical dataset was obtained before publication of revised El Escorial criteria, instead relying on modified World Federation of Neurology diagnostic criteria for half of the cohort, and original El Escorial criteria for the remainder[28,202]. Unfortunately, this is a limitation for many longitudinal MND population studies.

Data collection was less comprehensive between 1999 and 2014 and our study does not allow for age-period-cohort (APC) analysis. Nevertheless, it gives an accurate representation of the current climate of MND in Scotland. APC analysis highlights historical rather than existing potential environmental aetiology only. Recent APC analysis from Ireland, a population in geographical proximity to Scotland, showed no birth cohort effect[279].

3.5 CONCLUSIONS

This epidemiological study fulfils criteria for a population-based study in MND[227,271] and shows an increasing incidence of MND in Scotland, with the highest and most up-to-date standardised incidence rates reported in the literature. We have re-established the Scottish MND Register (now CARE-MND platform) and can achieve 99% capture of patients in Scotland. The high ascertainment rates obtained from a national prospective, multi-source register imply that findings can be generalised to other Northern European populations. This work and future CARE-MND analyses will help model demand for clinical MND services in Scotland and potential aetiological factors.

4. PHENOTYPING AND PROGNOSTIC MODELLING

4.1 BACKGROUND AND AIMS

Background

As discussed, MND populations are heterogeneous in presentation and prognosis. Characterisation of a population using a national register can therefore help to prioritise service needs and allocate care appropriately, as well as helping to increase the knowledge base for patients, funders and researchers. A national register is indiscriminate, capturing people at all stages of disease, in contrast to clinical trials, which may be biased towards patients who have access – either through service provision or because of physical ability – to specialised trial sites. The Joint Programming Initiative on Neurodegenerative Disease Research (JPND) – an international network geared toward the acceleration of discovery and treatment in neurodegenerative disease – recently highlighted the importance of longitudinal population data collection[280].

As introduced in Chapter 1.1, environmental risk factors are thought to contribute to the accumulation of burden of risk required for MND to manifest (the multistep and "Gene Time Environment" hypotheses)[143,160]. Additionally environmental influences may contribute to survival and speed of decline. Such factors can be categorised as endogenous or exogenous, though, of course, there can be interaction between the two.

i) Endogenous Factors

Endogenous variables include an individual's past medical history and family history of disease. Past history of autoimmune disease is thought to be linked with MND, with an increased frequency compared to controls[281]. Carriers of *C9orf72* expansions are also thought to have an increased prevalence of nonthyroid autoimmune diseases suggesting a shared inflammatory pathophysiology[282]. Epidemiological studies of ALS in cancer survivors have been undertaken and have not found an association between cancer and risk of ALS[283,284]. History of melanoma and tongue cancer results in poorer survival following

a diagnosis of ALS[283]. However, this may be related to exogenous factors such as specific cancer treatments, rather than the malignant process itself. MND has been found to be associated with previous cerebrovascular disease, suggesting a pathogenic continuum between tissue damage, triggering TDP-43 mislocalisation and aggregation, and subsequent MND[285].

ii) Exogenous Risk Factors

Brain vulnerability is also a factor in the consideration of exogenous triggers of disease. One example, which is debated, is the influence of head injuries on MND risk. Case-control studies and meta-analyses suggest a modest association, with increasing risk with accumulation of injuries[286]. The mechanistic link is plausible considering that TDP-43 accumulation is a common marker in patients with chronic traumatic encephalopathy (CTE) secondary to repeated head injury, and MND[287,288]. Recall bias in head injury is a concern; a study looking only at hospital records of serious head injury requiring admission suggested no association with MND[289]. Head injury may in fact be a prodromal sign of MND, for example in association with falls, and so it is important to clarify temporal relationship[289,290]. Indeed a Swedish register study showed increased risk of head injury in the one year preceding ALS diagnosis but not at three years prior to diagnosis[291]. Another related potential exogenous influence is physical exercise. Most recent studies suggest that physical activity itself does not pose a risk to development of MND[237,292,293]. One study found that strenuous activity >3 days per week shortened survival in postmenopausal women with ALS; however, the study did not adjust for site of onset or interventions[294].

Other exogenous elements that have been explored in MND include occupational exposure to chemicals and agricultural pesticides[295–297] and heavy metals such as lead and mercury[295,298,299]. So far, there is no firm evidence to suggest that these exposures contribute to MND risk or survival. A genotype-environmental link was proposed in the context of agricultural pesticides, specifically organophosphates: the paraoxonase gene (*PON1*) detoxicates paraoxon, the active ingredient in organophosphates. Mutations in the *PON1* gene were therefore thought to predispose to MND susceptibility; however, this has

not been proven[300,301]. In the case of lead, which is thought to be neurotoxic, increased exposure may even prolong survival in MND[298]. Other occupational considerations have been studied and jobs which involve direct contact with the general public have been implicated, perhaps due to exposure to infections[302]. A large retrospective study in Scotland revealed that military veterans were more likely to develop MND than non-veterans; however, the relative risk was very small (adjusted HR 1.49, 95% CI 1.01-2.21, p=0.046) and the study was not adjusted for other measures such as smoking[303].

Exposure to certain drugs, particularly statins, has also been explored[229,304–306]. Again, these studies have conflicting messages that do not allow us to come to conclusions regarding risk.

As discussed, there is thought that MND may be a prion-like disease due to its contiguous spread[63,64]. As such, the impact of blood transfusion from patients with neurodegenerative disease might influence disease onset. However, a large retrospective cohort study did not find association in neurodegenerative disease[228]. In terms of more conventional exogenous exposures, smoking is generally considered a risk factor for MND. However, studies are again conflicting with some suggesting an unequivocal relationship and a dose response[236,307,308] and others describing a trend towards association in women only[235].

The fact that all the above factors have been questioned in the development or progression of MND in recent years suggests that the environmental contribution is still very much open for debate. As such, an extensive longitudinal register such as CARE-MND has the potential to test and generate hypotheses about these factors in the Scottish population.

As discussion in Chapter 1.2, predicting outcomes in MND is fraught with difficulty as the population is heterogeneous in presentation and prognosis. Factors which are thought to be poor prognostic markers include older age at onset, bulbar or respiratory onset disease, ALS (over other forms of MND), cognitive impairment, early progression of ALSFRS-R score[58,185]. However, while there are supportive studies for these, there are also contradictory ones[58]. It is also unclear if cognition is an independent prognosticator or if this is related to uptake of interventions such as NIV and gastrostomy. Survival studies that

use patients recruited to trials are subject to selection bias and by default model only MND-specific variables[58]. Most studies involving patient data from electronic records have missing values[191,194]. Exclusion of these variables/individuals can further introduce bias. The best approach is considered to be population-wide recruitment using multiple sources of data with multiple imputation for missing data[58,194].

Aims

Clinical Audit Research and Evaluation of MND (CARE-MND) is a newly relaunched national platform for prospective data collection of demographics and health-related variables, including past medical and drug histories, of people with pwMND in Scotland. I aimed to characterise phenotypically the population in the first few years of inception (2015-2017), with the following intentions:

i) To describe a national cohort in terms of demographics, clinical phenotypes and environmental exposures.
ii) To undertake a survival analysis using selected features described in the literature to identify predictors of disease prognosis (prognostic modelling).

4.2 METHODS

Inclusion Criteria

All available data stored on the CARE-MND database were sourced for this component of the study. Only patients who had provided written informed consent to share their data were included. Patients who lacked capacity at diagnosis were therefore not included in this part of the analysis. Patients consented to examination of medical notes, processing of this information and storing of the data securely, via the CARE-MND database (see MND Register Invitation Letter and Consent Form, Appendix 3). The database is managed by data

processors who are based at the University of Edinburgh. Application for data requires submission of a form detailing requirements and project specifications including ethical approvals and funding. This is subsequently reviewed by the CARE-MND Research Project Manager who has the power to grant release of data. From 25th May 2018 all data handling adhered to General Data Protection Regulation guidelines.

Variables and Pre-processing

Data are anonymised using a unique patient code. Released data are 'raw' ie. As entered prospectively onto the CARE-MND platform. Pre-processing is therefore required to 'clean' the data and make it more interpretable[309]. Pre-processing is a key component of data mining in medical datasets and involves various steps including i) data cleaning, ii) missing data imputation and iii) feature selection[309].

During platform development, we aimed to include mainly binary/drop-down options to facilitate easy data interpretation. However, by its nature, CARE-MND is a prospective source for MND clinical/allied health specialists and 'free text' boxes are required for elaboration and explanation of factors relevant to patient care. Data cleaning is therefore essential for raw medical data stored on CARE-MND. The variables included and specifics regarding pre-processing are outlined below. All environmental and past medical history data refers to pre-morbid state prior to recorded date of MND onset.

i) Demographic Variables

Demographic variables related to patient sex and ethnicity. Ethnicity was coded as per ISD guidelines[310]; indeed some of this information was extracted from ISD records (see Chapter 2.3). Due to the low variance in ethnicity in the population studied (infrequency of ethnic minorities), this variable was reclassified into the following: i) Group A – White, 1A White Scottish, ii) Group A – White, 1B Other British, iii) Group A – White Other (including 1C Irish, 1K Gypsy/Traveller, 1L Polish, 1Z Other white ethnic group) iv) Group C – Asian, Asian

Scottish or Asian British, v) Group G – Refused/Not provided by patient and Group H - Unknown. Group G was classed as missing data.

ii) Environmental Variables

Environmental variables included those related to social circumstances (co-habitation, occupation, military service, exercise participation) and exogenous exposures (smoking, alcohol, recreational drugs, heavy metal and pesticide exposure). Co-habitation was dichotomised into a Yes (Married/Living with partner) or No (Single/Divorced/Separated) variable with reference to a recent retrospective analysis[311]. Occupational histories were recorded as free text. The patient's primary occupation was analysed; where this was unclear the highest skilled occupation was selected. Rather than make artificial assumptions about occupation type, each individual's occupation was coded as per the UK Office for National Statistics using the Standard Occupational Classification coding index 2010[312]. This system ranks occupations into nine major groups (level from 9, low skill, to level 1 high skill). Military service was dichotomised into Yes/No variables and required patients to have partaken in active service. Exercise participation was crudely screened using None/Light/Moderate/Heavy levels. Examples of "Light" exercise included dog-walking and "Heavy" exercise was considered as strenuous physical activity more than five times per week or professional/semi-professional sportsmanship.

Smoking was recorded as Yes/Ex-smoker/Never smoked and then dichotomised into Ever Smoked/Never Smoked variables. Alcohol use was examined from the perspective of hazardous drinking (>14 units per week for men and women) and harmful drinking (>35 units per week for women, >50 units per week for men) as per NICE Guidelines for Alcohol-use disorder[313]. Recreational drug misuse, heavy metal exposure and pesticide exposure were all dichotomised into Ever Yes/Ever No variables.

iii) Past Medical History

Based on a review of the literature, past medical history data were interrogated for the following: surgical intervention within one year preceding symptom onset (Yes/No); past medical history of malignancy (Yes/No), cardiovascular disease (Yes/No), autoimmune disease (Yes/No), central nervous system surgery (Yes/No), psychiatric illness (Yes/No), nervous system infection (Yes/No), significant pre-morbid head injury (Yes/No) and pre-morbid receipt of blood products by transfusion (Yes/No); number of pre-morbid co-morbidities; number of pre-morbid surgical interventions. Co-morbidities were defined as chronic conditions requiring medical input (excluding acute infections or injuries). Accurate past medical and surgical histories including dates were obtained from patient self-report, GP records and ISD data (see Chapter 2.3).

Medication history was sourced but for the purposes of analysis, only pre-morbid statin use (Yes/No) was examined. Accurate medication history was obtained from patient self-report, GP records and Emergency Care Summary (ECS) records on EPRs.

iv) Family history

Family history was obtained from patients, with corroboration from a family member where possible. History was explored across four generations (patient's grandparents to children), though documentation of this was sometimes limited by patient recall or family estrangement. Patients were questioned specifically about family history of MND (considered positive (Yes/No) if there was any family member with MND), other neurological diseases (Yes/No) with a separate analysis for neurodegenerative diseases (multiple sclerosis (Yes/No) and Parkinson's disease (Yes/No)), dementia (Yes/No) with a separate analysis for early-onset dementia (<65 years) or FTD (Yes/No) and psychiatric diseases (Yes/No).

v) Characterisation of Disease

Variables related to disease onset and trajectory included those from diagnosis (mean age of onset, mean age of diagnosis) to initial assessment (site of onset, El Escorial classification of disease[27], baseline ALSFRS-R score[182], presence of cognitive impairment and ECAS ALS-specific and ALS-non-specific scores[206,208]. Site of symptom onset was classified as: bulbar; upper limb; lower limb; mixed; cognition; respiratory; other. 'Mixed' onset refers to pwMND for whom it was not possible to identify an initial one symptom. 'Other' refers to a small proportion of patients for whom symptom onset could not be classified into other categories.

Additionally, rate of ALSFRS-R decline was calculated using the concept of the 'preslope' or ALSFRS-R-based linear estimate of rare of disease progression ie. The rate of decline in function between the day before symptom onset (date of onset minus one) and the date of the first ALSFRS-R after diagnosis (providing this is within six months of diagnosis) (see Figure 11). The 'preslope' is a recognised phenomenon and has been used for prediction modelling of MND in observational studies and clinical trials[58,184,185,314].

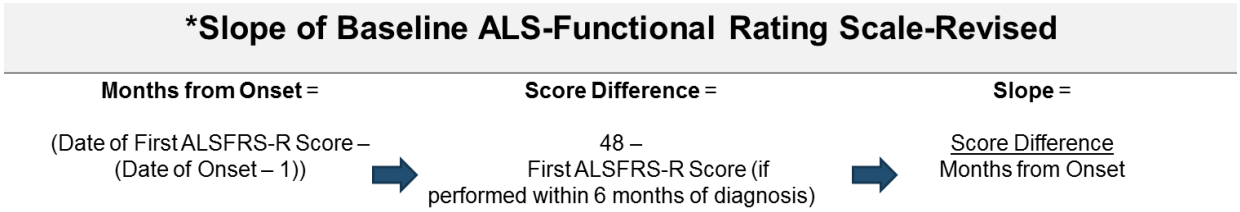


Figure 11 Process of calculation of ALSFRS-R slope of decline at diagnosis (preslope)

vi) Markers of Care

Through CARE-MND, we can also assess quality of care delivery and align these to NICE standards. Data regarding the following were collected: time to neurology clinic review; time to MND specialist clinic review; referral to a neuropsychologist (Yes/No); referral to a dietician (Yes/No); discussions undertaken regarding feeding tubes (Yes/No); referral to a speech and language therapist (Yes/No). Place of death was also recorded as a surrogate marker of anticipatory care planning (home, hospital, hospice or care home).

vii) Management

Again, in line with NICE recommendations, proportion of patients undergoing interventions relevant to quality of life or survival was examined. This included: proportion of patients undergoing ECAS assessment (Yes/No); patient decision regarding riluzole treatment (Yes/No/Undecided at time of censorship); proportion of patients taking riluzole at censorship (excluding patients who had taken riluzole at any point and had subsequently discontinued treatment); proportion of patients who had taken taken riluzole but subsequently discontinued treatment; patient decision regarding feeding tube placement (Yes/No/Undecided); proportion of patients undergoing feeding tube insertion by time of censorship; patient decision regarding non-invasive ventilation (NIV) (Yes/No/Undecided); proportion of patients who had commenced NIV at time of censorship.

viii) Survival

Finally, summary survival data were sourced to assess the data in advance of further survival analysis. The following were summarised: proportion of deaths and proportion of patients alive at censorship; median survival in days and months; proportion of patients who had died within two years of symptom onset.

Missing Data

From study of the CARE-MND database (Chapter 2), missing data were anticipated. While the analysis of CARE-MND compared with MND-SMART revealed a significant improvement in ascertainment across almost all domains (Chapter 2.3), some of the variables on the CARE-MND proforma were new to both MND clinical nurse specialists/allied health professionals and patients. It is well-recognised that new population registers take time to achieve adequate data capture[227] and we suspected that capture of new variables might be low in the 2015-2017 time period. Real-world, routinely-collected data is rarely complete but ignoring variables with missingness can be misleading and can introduce bias[191,309]. In a recent review of missing data in prediction studies for diabetes, 62.5% of studies did not acknowledge missing data; of the studies that did acknowledge missing data, 43.8% excluded all individuals with missing data[192].

Potential reasons for missingness in our dataset include:

- Incomplete CARE-MND proformas due to patients not being asked relevant questions by clinical specialists due to:
    - Time constraints
    - Prioritisation of other variables in a patient who is struggling to communicate
- Patient declining to answer certain questions
- Data not available at time of censorship due to proforma being in progress

Following pre-processing, descriptive statistics and percentage missingness were calculated for each variable.

Statistical Modelling of Data

Subsequent to descriptive analysis, variables were incorporated into models to identify predictors of survival. Based on review of the literature, only selected clinically-relevant variables which taken forward for modelling. Further computational pre-processing was

undertaken to select appropriate features. Variables with zero variance were excluded. A correlation matrix was generated to identify highly correlated variables using Pearson's correlation coefficients; coefficients >0.5 were considered significant[315]. A maximum variable limit was set to ensure that the models did not exceed the 'one in ten' rule[195,316]. Values were standardised (centred and scaled) for model analysis. Categorical variables were dummy coded into dichotomous variables.

In order to analyse data in the context of missingness, imputation was required to replace data values with statistically plausible estimates (required when >15% missing data)[191,192,194]. Our missing values were assumed to be missing at random (MAR); while missingness was not biased towards any particular patient groups, there was a trend toward missingness for variables new to CARE-MND (eg. Exposure to heavy metals and pesticides, cognitive screening).

Multiple imputation is considered best practice and a recognised method is Multiple Imputation by Chained Equations (MICE)[191,193,194]. Standard MICE software incorporates predictive mean matching (PMM) which is a parametric and non-parametric approach; it has been shown to be a good option for Cox regression models where missingness ranges between 10-50%[317]. M=10 imputations were used and results pooled using Rubin's rules[318]. Sensitivity analyses were performed whereby models were imputed and tested using all variables with i) ≤25% missing data, ii) variables with ≤50% missing data and iii) all variables. Cox regression modelling was used to assess simultaneously the effect of multiple variables, both quantitative and categorical, on survival. Due to the conflicting evidence in the literature regarding risk factors for MND, a data-led hypothesis-free approach was taken whereby as many variables as possible were included in the model. This approach also provides more information for multiple imputation[194].

Model development was outlined with reference to TRIPOD guidelines[195,196]. R v3.4.3 was used for all statistical analyses (in particular, packages "caret", "mice", "DescTools", "mitools" and "survival")[240].

4.3 RESULTS

Descriptive statistics

Of the 619 pwMND diagnosed in Scotland in 2015-2017, 437 (70.6%) consented to share their medical record data. The dataset was split into two cohorts for censorship. 2015-16 data were censored on 16th April 2018. To maximise follow-up, 2017 patients were censored on 20th July 2018. The mean follow-up time was 23.0 months (SD 9.2, range 6.0 – 39.0 months). Descriptive statistics by phenotypic category are presented in Table 11. Values and percentages relate to completed fields, excluding missing data.

Ethnic minorities included people of Pakistani origin (n=3), Chinese origin (n=1), Indian origin (n=1) and Other Asian (non-Chinese) origin (n=1). For the individuals in whom Site of Onset was defined as "Other" (n=4), sites included: weight loss (n=2), truncal weakness/camptocormia and generalised fasciculations. Of the 41 patients who had "Mixed" onset disease, 21 (51.2%) had bulbar symptoms at onset. Therefore 144 (33.3%) of pwMND had bulbar symptoms at onset.

Premorbid surgeries were defined as major or minor surgical procedures performed prior to disease onset. While the median number of surgeries is zero, the maximum number was 13.

Twenty one pwMND were referred for genetic counselling; if this field was blank it was assumed that it was not carried out.

Phenotypic Character	Missing data/Not disclosed (%)	Value (n=437)
Demographics		
Sex	0	
- Males (%)		275 (62.9)
- Females (%)		162 (37.1)
Ethnicity	3.7	
- White All (Group 1A, 1B and 1 Other) (%)		415 (98.6)
- White Scottish Group 1A (%)		296 (70.3)
- White British Group 1B (%)		71 (16.9)
- Ethnic minority Group C (%)		6 (1.4)
Patient Environment		
Cohabiting, Yes (%)	6.9	307 (75.4)
Highest Occupational Classification	13.5	
- 1 (%)		44 (11.6)
- 2 (%)		63 (16.7)
- 3 (%)		50 (13.2)
- 4 (%)		23 (6.1)
- 5 (%)		86 (22.8)
- 6 (%)		24 (6.3)
- 7 (%)		12 (3.2)
- 8 (%)		37 (9.8)
- 9 (%)		35 (9.3)
- 10 Unemployed (%)		4 (1.1)
History of Active Military Service, Yes (%)	42.1	32 (12.6)
Ever Smoked, Yes (%)	10.3	209 (53.3)
History of Alcohol Intake Exceeding 14 units/week, Yes (%)	31.8	79 (26.5)
History of Harmful Alcohol Intake Exceeding 35 units/week for women, 50 for men, Yes (%)	31.8	36 (12.1)
Recreational Drug Use, Yes (%)	47.4	24 (10.4)
Premorbid Exercise Participation	19.7	
- Grade 4: Heavy (%)		69 (19.7)
- Grade 3: Moderate (%)		148 (42.2)
- Grade 2: Light (%)		115 (32.8)
- Grade 1: None (%)		19 (5.4)
History of Heavy Metal Exposure, Yes (%)	32.0	46 (15.5)
History of Pesticide Exposure, Yes (%)	33.2	35 (12.0)
Past Medical History (PMH)		
Median number of co-morbidities (IQR)	0	3.0 (1.0-5.0)
PMH Cardiovascular Disease, Yes (%)	0	206 (47.1)
PMH Autoimmune Disease, Yes (%)	0	66 (15.1)
PMH Malignancy, Yes (%)	0	37 (8.5)
PMH Psychiatric Condition, Yes (%)	0	88 (20.1)
PMH Nervous System Infection, Yes (%)	0	15 (3.4)
Premorbid Statin Use, Yes (%)	0	54 (12.4)
Median number of premorbid surgeries (IQR)	0	0.0 (0.0-1.0)

Premorbid Surgery, Yes (%)	0	202 (46.2)
Surgery <1 year before Disease Onset, Yes (%)	0	13 (3.0)
PMH of Central Nervous System Surgery, Yes (%)	0	8 (1.8)
PMH of Significant Premorbid Head Injury, Yes (%)	30.9	78 (25.8)
PMH of Blood Transfusion, Yes (%)	34.6	28 (9.8)

Family History

Family History of MND, Yes (%)	2.7	39 (9.2)
Referred for Genetic Counselling, Yes (%)	0	21 (4.8)
Family History of Dementia, Yes (%)	7.1	111 (27.3)
- Family History of Early Onset Dementia/FTD, Yes (%)		16 (3.9)
Family History of Neurological Conditions, Yes (%)	9.6	107 (27.1)
- Family History of Parkinson's Disease, Yes (%)		37 (9.4)
- Family History of Multiple Sclerosis, Yes (%)		21 (5.3)
Family History of Psychiatric Conditions, Yes (%)	16.9	54 (14.9)

Characterisation of Disease

Mean age of Onset, years (SD)	1.6	63.9 (10.9)
Mean age of Diagnosis, years (SD)	0	65.5 (10.7)
Median Time to Diagnosis, months (IQR)	1.6	12.0 (8.0-23.0)
Site of Onset	1.1	
- Bulbar (%)		123 (28.5)
- Upper limb (%)		103 (23.8)
- Lower limb (%)		144 (33.3)
- Mixed (%)		41 (9.5)
- Cognition (%)		12 (2.8)
- Respiratory (%)		5 (1.2)
- Other (%)		4 (0.9)
Mean Baseline ALSFRS-R (SD)	24.0	37.4 (7.2)
Mean Time from Diagnosis to Baseline ALSFRS-R, months (SD)	24.0	0.9 (1.5)
Median Rate of Decline in ALSFRS-R from Onset to First Assessment (IQR)	24.0	0.6 (0.3-1.1)
MND Classification	0	
- ALS El Escorial Definite (%)		75 (17.2)
- ALS El Escorial Probable (%)		134 (30.7)
- ALS El Escorial Possible (%)		51 (11.7)
- ALS Clinician Diagnosis (%)		78 (17.8)
- MND-FTD (%)		25 (5.7)
- PBP (%)		25 (5.7)
- PLS (%)		19 (4.3)
- PMA (%)		19 (4.3)
- Other (%)		11 (2.5)
Cognitive Impairment, Yes (%)	42.3	99 (39.3)
Median ECAS Total Score (IQR)	50.6	109.0 (96.0-116.0)
- Median ALS-Specific Score (IQR)	51.0	81.0 (70.3-87.0)
- Median ALS Non-Specific Score (IQR)	51.0	27.0 (24.0-30.8)

Markers of Care

Median Time from Disease Onset to Neurology Clinic Review, months (IQR)	36.4	8.0 (5.0-16.0)
Median Time from Disease Onset to MND Specialist Clinic Review, months (IQR)	36.2	13.0 (8.0-26.0)
Referral to Neuropsychologist, Yes (%)	5.9	198 (48.2)

Referral to Dietician, Yes (%)	9.2	298 (75.1)
Feeding Tube Discussed, Yes (%)	7.6	282 (69.8)
Referral to Speech & Language Therapist, Yes (%)	6.9	315 (77.4)
Place of Death (n=242)	11.1	
- Home (%)		74 (34.4)
- Hospital (%)		83 (38.6)
- Hospice (%)		43 (20.0)
- Care Home (%)		15 (7.0)
Management		
ECAS Assessment, Yes (%)	1.6	244 (56.7)
Patient Decision Regarding Riluzole	15.3	
- Yes (%)		225 (60.8)
- No (%)		118 (31.9)
- Undecided (%)		27 (7.3)
Taking Riluzole at Censorship, Yes (%)	0.9	173 (40.0)
Riluzole Discontinued Ever, Yes (%)	0	42 (9.6)
Patient Decision Regarding Feeding Tube	8.9	
- Yes (%)		165 (41.5)
- No (%)		167 (42.0)
- Undecided (%)		66 (16.6)
Feeding Tube Inserted by Censorship, Yes (%)	0.7	136 (31.3)
Patient Decision Regarding NIV	18.3	
- Yes (%)		132 (37.0)
- No (%)		115 (32.2)
- Undecided (%)		110 (30.8)
NIV Commenced by Censorship, Yes (%)	0.7	118 (27.2)
Survival		
Death During Follow-up	0	
- Death (%)		242 (55.4)
- Censorship (%)		195 (44.6)
Median Survival from Onset All	1.6	
- Days (IQR)		914.5 (647.5-1354.0)
- Months (IQR)		30.0 (21.0-44.0)
Median Survival from Onset Death Only	1.2	
- Days (IQR)		753.0 (559.0-1003.0)
- Months (IQR)		24.0 (18.0-32.5)
<2 year Survival from Onset Death Only, Yes (%)	1.2	117 (49.0)

Table 11 Phenotypic characteristics of consenting patients with MND diagnosed 2015-17

Prognostic Modelling

Variables relevant to survival were selected for modelling and are summarised in Table 12. As described in the table, variables were simplified for ease of modelling and interpretation. In particular, levels of classification of “Site of Onset” and “Classification” were simplified to test hypotheses relevant to survival (as described in review of the literature in Chapter 1.1). For example, bulbar-onset disease is considered to have poorer survival than limb-onset disease, but the impact of ‘other’ sites of onset (respiratory, weight loss) is unclear. Further, more atypical forms of MND (PLS, PMA, PBP) and thought to confer a longer survival benefit than typical ALS presentations but people with MND-FTD are considered to have poor prognosis.

Demographics	**Environment**	**Past Medical History**	**Family History**	**Characterisation of Disease**	**Care/Management**
Sex Ethnicity *Group A or Group C*	Smoking Exercise Toxin Exposures *Including Heavy Metals and Pesticides*	Number of co-morbities Malignancy Autoimmune disease Cardiovascular disease Psychiatric conditions Head injury Blood transfusion	MND Dementia Early-onset dementia Other neurological conditions Psychiatric conditions	Age of onset Site of Onset *Bulbar, Limb or Other* Diagnostic delay Classification *ALS, MND-FTD, Other* ALSFRS-R Preslope ECAS ALS Specific Score ECAS ALS Non-Specific Score	Time to Neurology Clinic Time to MND Clinic Rilzuole Use Feeding Tube Insertion Non-invasive Ventilation Use

Table 12 Variables considered for prognostic modelling (n=30)

After removing variables with near-zero variance, there remained 27 variables with a total of 28 predictors taking into account variables with multiple levels. Correlated variables were then reviewed using a correlation matrix (Figure 12). Of each correlated pair, the most clinically applicable variable was preserved. As such, the following variables were excluded based on a correlation coefficient cut-off of 0.5: Time from Onset to Neurology Clinic, Time from Onset to MND Clinic, and ALS Non-Specific Score. Number of comorbidities was also excluded as this was closely correlated with past medical history of cardiovascular disease (0.48). The final number of predictors was 24.

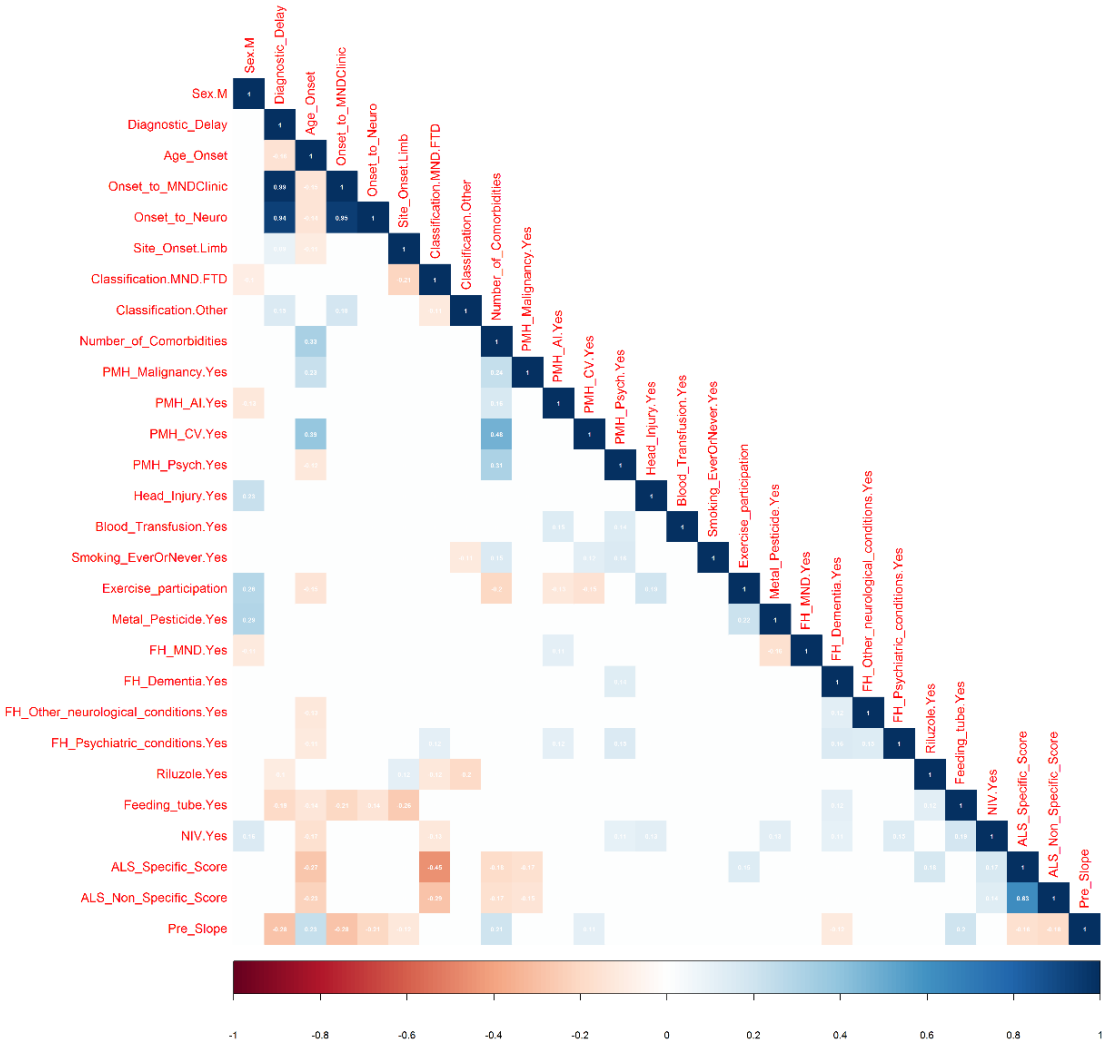
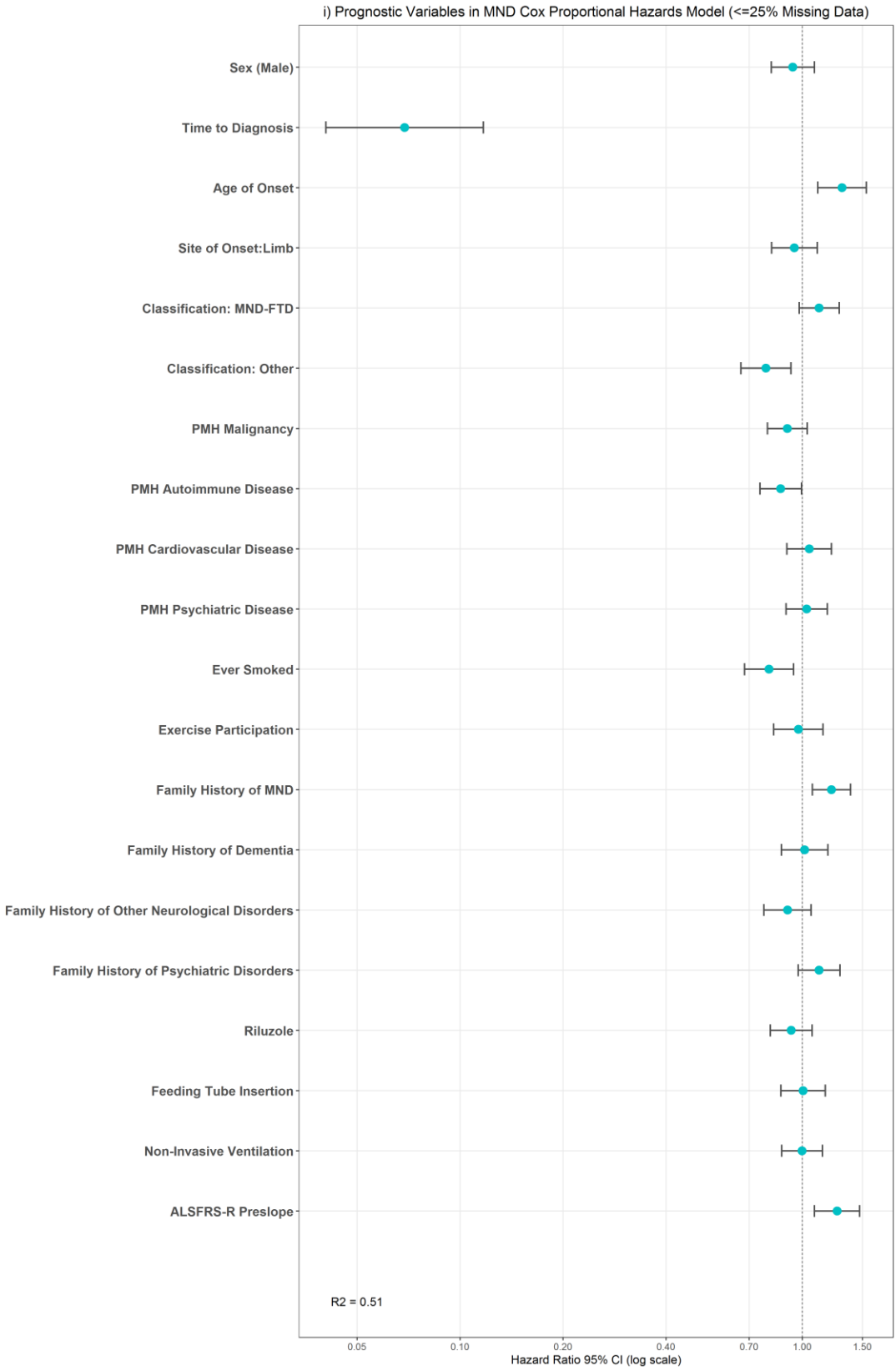


Figure 12 Correlation matrix of phenotypic variables with Pearson's correlation coefficient

The Cox regression model was first tested using ≤25% missing data. Twenty predictors were featured in this model (p = 20) (Figure 13 (i)). The following were significant predictors of death: increasing Age of Onset (p=0.0013), Family History of MND (p=0.0026) and steeper ALSFRS-R Preslope (p=0.0025). In contrast, a longer Time to Diagnosis (p<0.00001), Classification: Other (which included PLS, PMA and PBP) (p=0.0044), Past Medical History of Autoimmune Disease (p=0.043) and a history of Ever Smoked (p=0.0081) were associated with better prognosis. The pseudo-R^2 for the Cox model was 0.511 (95% CI 0.506-0.517), suggesting that 51% of the variance in survival could be explained by these predictors.

Using variables with ≤50% missing data (p =23), again increasing Age of Onset (p=0.0013), Family History of MND (p=0.00077) and ALSFRS-R Preslope (p=0.0028) predicted death and increasing Time to Diagnosis (p<0.00001), Classification: Other (p=0.0031), Past Medical History of Autoimmune Disease (p=0.029) and a history of Ever Smoked (p=0.0077) predicted survival. The pseudo-R^2 for this model was 0.521 (95% CI 0.514-0.527) (Figure 13 (ii)).

Finally, by including and imputing all predictors (p=24), predictors of outcome remained relatively stable with increasing Age of Onset (p=0.012), Family History of MND (p=0.0023) and ALSFRS-R Preslope (p=0.0010) predicting death and increasing Time to Diagnosis (p<0.00001), Classification: Other (p=0.0020) and a history of Ever Smoked (p=0.0013) predicting survival (Figure 13 (iii)). In this final model, a Past Medical History of Autoimmune Disease no longer predicted survival (p=0.058) but exposure to Heavy Metals or Pesticides was associated with poorer survival (p=0.042). The pseudo-R^2 for this model was 0.538 (95% CI 0.531-0.545). This model was considered the best in view of the higher pseudo-R^2 value. Hazards ratios and confidence intervals for this model are summarised in Table 13, ranked by hazard ratio from most significant predictor of death to most significant predictor of survival. R scripts can be found in Appendix 4.


i) Prognostic Variables in MND Cox Proportional Hazards Model (<=25% Missing Data)
Sex (Male)
Time to Diagnosis
Age of Onset
Site of Onset:Limb
Classification: MND-FTD
Classification: Other
PMH Malignancy
PMH Autoimmune Disease
PMH Cardiovascular Disease
PMH Psychiatric Disease
Ever Smoked
Exercise Participation
Family History of MND
Family History of Dementia
Family History of Other Neurological Disorders
Family History of Psychiatric Disorders
Riluzole
Feeding Tube Insertion
Non-Invasive Ventilation
ALSFRS-R Preslope
R2 = 0.51
0.05
0.10
0.20
0.40
0.70
1.00
1.50
Hazard Ratio 95% CI (log scale)

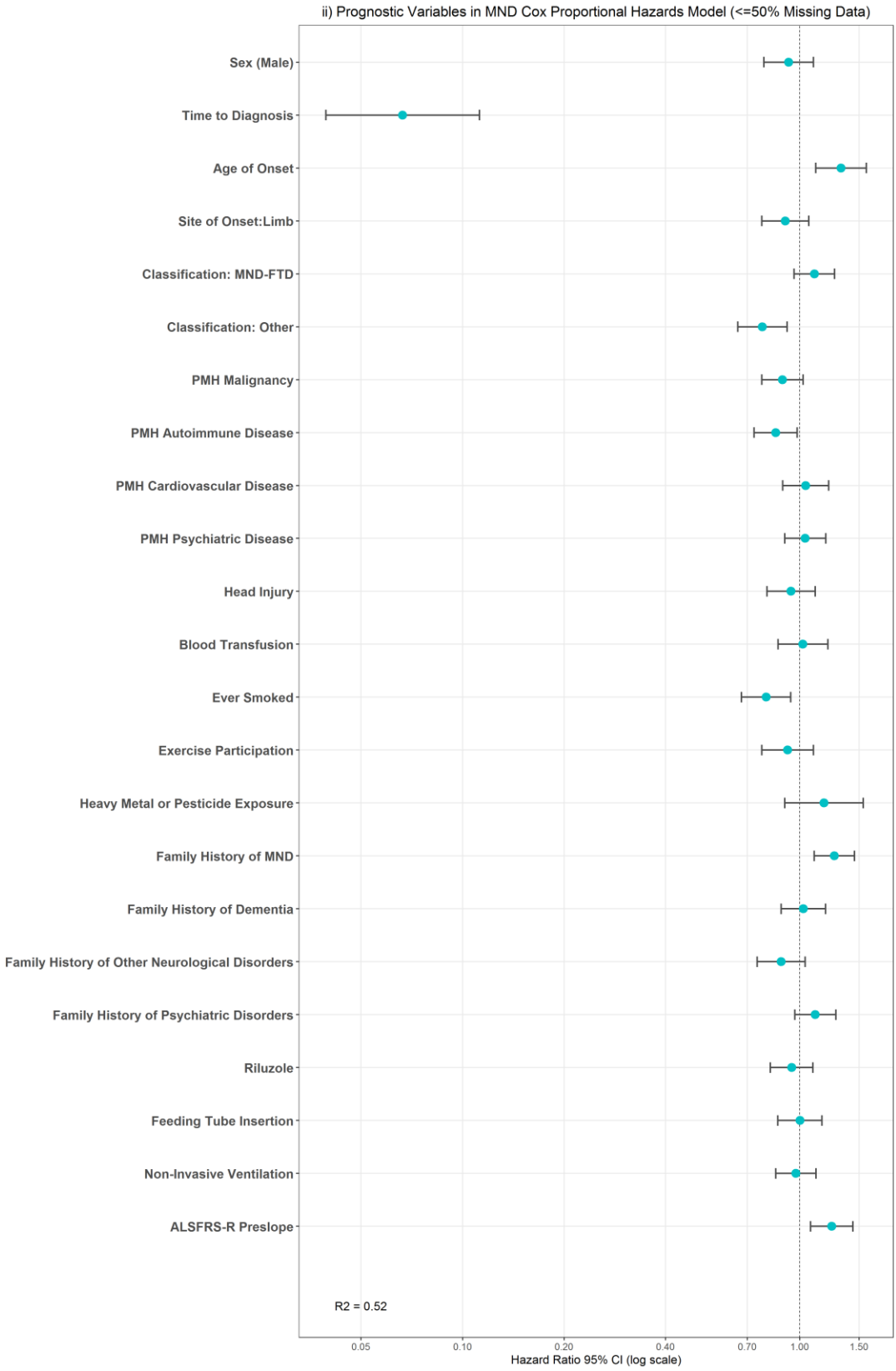

ii) Prognostic Variables in MND Cox Proportional Hazards Model (<=50% Missing Data)
Sex (Male)
Time to Diagnosis
Age of Onset
Site of Onset:Limb
Classification: MND-FTD
Classification: Other
PMH Malignancy
PMH Autoimmune Disease
PMH Cardiovascular Disease
PMH Psychiatric Disease
Head Injury
Blood Transfusion
Ever Smoked
Exercise Participation
Heavy Metal or Pesticide Exposure
Family History of MND
Family History of Dementia
Family History of Other Neurological Disorders
Family History of Psychiatric Disorders
Riluzole
Feeding Tube Insertion
Non-Invasive Ventilation
ALSFRS-R Preslope
R2 = 0.52
0.05
0.10
0.20
0.40
0.70
1.00
1.50
Hazard Ratio 95% CI (log scale)

Figure 13 Cox Regression Model predicting survival i) With <= 25% missing data, ii) With <=50% missing data, iii) With all missing data imputed

Predictor	Hazard Ratio	95% CI
ALSFRS-R Preslope	**1.29**	**1.11-1.50**
Age of Onset	**1.28**	**1.06-1.56**
Family History of MND	**1.24**	**1.08-1.43**
Heavy Metal or Pesticide Exposure	**1.20**	**1.01-1.43**
Family History of Psychiatric Disorders	1.12	0.96-1.30
PMH Cardiovascular Disease	1.06	0.91-1.24
Family History of Dementia	1.06	0.91-1.24
Classification: MND-FTD	1.03	0.86-1.23
PMH Psychiatric Disease	1.03	0.89-1.19
Blood Transfusion	1	0.83-1.20
Feeding Tube Insertion	0.99	0.84-1.16
Non-Invasive Ventilation	0.97	0.84-1.12
Head Injury	0.96	0.81-1.14
Exercise Participation	0.96	0.80-1.14
Riluzole	0.95	0.82-1.10
Sex (Male)	0.92	0.78-1.09
Site of Onset: Limb	0.92	0.79-1.09
Family History of Other Neurological Disorders	0.90	0.77-1.04
PMH Malignancy	0.87	0.75-1.01
PMH Autoimmune Disease	0.86	0.74-1.00
ECAS: ALS Specific Score	0.82	0.59-1.13
Ever Smoked	**0.79**	**0.68-0.91**
Classification: Other	**0.76**	**0.65-0.91**
Time to Diagnosis	**0.062**	**0.034-0.11**

Table 13 Hazard ratios of prognostic modelling predictors using model (iii). ***bold*** *= significant predictor*

4.4 DISCUSSION

Phenotypic Characteristics

The male-to-female ratio in this research cohort of pwMND was 1.70:1; this is higher than that seen in the unselected incident cohort described in Chapter 3 (1.53:1) suggesting that more males than females were willing to, or able to, consent to research. To our knowledge, this is not a recognised observation and may be unique to this particular research cohort. Similarly, research participants were slighter younger than the overall incident cohort (age of onset and diagnosis 63.9 and 65.5 years versus 65.3 and 66.8 years respectively), as has been observed in other studies[178]. However, the figures remains fairly typical compared with other population cohorts[3,5,199,243]. There were comparable proportions of patients with bulbar onset MND and MND-FTD between the two cohorts (28.5% and 5.7% in the research cohort versus 28.2% and 6.8% in the incident cohort) showing that these factors were not barriers to research recruitment. The ethnic homogeneity of the research cohort was comparable with the incident cohort and with Scottish government data[277].

This study does not benefit from a matched control population to which phenotypic characteristics can be directly compared; as such the influence of environmental factors on MND onset cannot be assessed. However, Scottish government statistics are available for some of the features to allow us to contextualise the cohort. Government statistics are broken down by age and sex; considering our population, males age 65-74 are likely the most directly comparable.

Seventy five percent of the Scottish MND cohort were cohabiting; this is higher than Scottish census statistics for 2011 (50.4%)[319]. This may reflect the level of disability and dependence of some pwMND. In accordance with our analysis of social deprivation in the Scottish MND population (Chapter 3), there also seemed to be an even distribution across occupational classifications in the cohort. A small proportion of patients (12.6%) had participated in previous military service but there was a high proportion of missing data (42.1%) and insufficient details regarding type and duration of service making further inclusion of this variable in statistical modelling difficult.

In the 2017 Scottish Health Survey, 32% of men aged 65-74 reported having had any form of cardiovascular disease[320]. In comparison, 47.1% of our MND cohort had had some form of cardiovascular disease. The higher proportion in our study may reflect sources of information: ISD and GP records were sourced for the MND cohort whereas the Scottish Health Survey is self-reported only. However, as discussed, vascular disease may be a risk factor for MND[285]. In the same survey, 17% of all adults had self-reported psychiatric symptoms as studied using the Global Health Questionnaire (GHQ); however, this proportion was lower in adults >65 years (12-13%)[320]. In our MND population, 20.1% had a documented or reported psychiatric illness. Recent data from CARE-MND and other studies suggest that neuropsychiatric disease may be a prodrome for both neurodegeneration associated with MND and with specific cognitive and/or behavioural changes and this may explain the relatively high proportion of premorbid psychiatric disease in the MND cohort[321,322]. In 2017, 44% of all adults were current or ex-smokers compared with 53% of the MND cohort[320]. However, 57% of males age 65-74 were current or ex-smokers which is more comparable. Our cohort figures, however, are higher than in other European populations (Cyprus (33%) and Italy (47%)[263,323]).

Prognostic Modelling

Through this study, I aimed to identify a model which would optimise on the breadth of CARE-MND data. Using a hypothesis-free data-driven approach, a Cox regression survival model (measuring survival in days) was tested. Sensitivity analysis revealed that inclusion of all variables in the imputation process provided more information than excluding variables with different missingness thresholds. This highlights the statistical power of multiple imputation. By excluding variables with missing data the dataset is subject to increased variance due to chance without correcting for bias[194].

The most informative model is model (iii) (Figure 13). By defaulting to this model, it is possible to conclude that ALSFRS-R score can predict survival, with a steeper decline from baseline to first assessment (within six months of diagnosis) predicting an earlier death (HR 1.29, 95% CI 1.11-1.50). Similarly, longer time to diagnosis (with each increasing month) strongly predicts survival (HR 0.06, 95% CI 0.03-0.11), and so the reverse predicts more rapid

decline. El Escorial Classification: Other might also be considered to be related to these predictors: a more slowly progressive, atypical form of MND (PLS, PMA or PBP) is associated with longer survival (HR 0.76, 95% CI 0.65-0.91). Collectively, these observations suggest that early widespread manifestation of combined upper and lower motor neurone degeneration will result in a faster time to death. Whilst this is not surprising - and indeed has been observed in other survival studies[7,184,187,324] - it is important to know that these factors predict survival independent of all other medical and environmental influences. Such patients can be identified at point of recruitment into clinical trials, and their progress studied separately to measure impact of the intervention on their survival. Patients with a family history of MND have significantly poorer prognosis (HR 1.24, 95% CI 1.08-1.43); inclusion of genetic data in the model may provide further clarification on this observation (discussed in Chapter 7). Being older at onset of disease is also associated with poorer outcomes (HR 1.28, 95% CI 1.06-1.56, with each increasing year of age). This has been observed in the Scottish population previously[3]. However, on this occasion, we observe its influence independent of other comorbidities such as cardiovascular disease and history of malignancy, perhaps supporting prior theories that older people with MND are more vulnerable due to aging or depleted motor neurones[178,204].

Interestingly, two environmental features had prognostic influence. A history of ever smoking was associated with better prognosis (HR 0.77, 95% CI 0.64-0.92). This feature was stable across all three Cox regression models. Our findings mirror observations in another neurodegenerative disease – Parkinson's disease – where smoking is found to be protective in a dose-dependent fashion[325]. A possible biological explanation for this relates to changes in the gut microbiome: cigarette smoking is thought to reduce gut inflammation (as seen in ulcerative colitis) and by extension decreases protein misfolding and propagation of misfolded alpha-synuclein through the CNS[325,326]. As discussed in Chapter 1, the role of misfolded TDP-43 and the apparent contiguous spread of ALS could represent a parallel model of disease. While there has been a recent surge in interest in the microbiome in neurodegenerative diseases, including MND, the biological interactions are as yet unclear[327–329]. The fact that a Past Medical History of Autoimmune Disease was a potential protective marker (significant in models (i) and (ii) though not significant in the final model; HR 0.86, 95% CI 0.74-1.00) may also hint at a potential link between MND and inflammation. The role of inflammation in MND is indeterminate but the utility of measuring

CSF inflammatory biomarkers is a current topic of debate[330]. Nevertheless, the findings in relation to smoking contradict previously published populations studies[307,308,323], although these studies did not control for other endogenous and exogenous influences and potential modifiers of MND progression and survival. Most recently, a large case-control study examined pack-year history and time from stopping smoking from three pooled European population cohorts, controlling for sex, age, site, alcohol intake and educational level[331]. Results from this work suggest that there is an increased risk of developing ALS with smoking and that this risk diminishes with time after stopping smoking. However, the effect of smoking on survival was not studied. This study used detailed questionnaires to capture smoking history; the authors acknowledge that this may limit patient inclusion (by excluding, for example, fast progressors and those unable to complete questionnaires). In contrast, data collection in Scotland was via a simple screening question (Smoking: Yes/Ex/Never; Provide Details). Future work in Scotland to explore the smoking hypothesis could learn from the more structured approach to this variable.

Exposure to Heavy Metals or Pesticides reached statistical significance in the final imputed model (HR 1.20, 95% CI 1.01-1.43). As this variable was an amalgamation of two different observations of toxin exposure, and as it was unstable (being absent from other models), it is not possible to draw firm conclusions about its impact on survival. However, it warrants further study using case-control methods using the CARE-MND population. Interestingly, a historical case-control Scottish survey of environmental exposures found a significant increase in exposure to heavy metals and pesticides in MND cases versus controls (OR 3.2 (95% CI 1.5-7.3)[275]. However, again, these data were subject to recall bias. It may be that patients with MND are more hyper-vigilant to potential causative agents and might scrutinise their past history more carefully. A case-control study using occupational records and with an attempt to quantify exposures would be more beneficial.

Other exogenous factors including exercise participation and history of blood transfusion did not influence survival. This is unsurprising considering the conflicting evidence regarding their impact in MND. However, these factors were self-reported, relying on patient recollection and awareness of exposures, and so may not be wholly reliable as anything more than a screening tool.

Site of onset was not a significant predictor of outcome. Our results therefore agree with previous dialogue, which proposes that conventional stratification of pwMND for clinical trials by region of onset (eg. Bulbar or spinal onset) is artificial and inadequate[58]. This may relate to poor recall of first symptoms – indeed in the Scottish cohort many patients report a 'mixed' presentation at onset. However, from our other observations, it appears that *burden* of motor neurone degeneration, rather than the origin of symptoms is of greater importance. It is possible that individuals with fast and global expression of disease have a stronger underlying predisposition (such as a genetic cause) or other biological or environmental factors, which make their CNS more vulnerable to a cascade of decay.

Further, initiation of riluzole, gastrostomy feeding or NIV did not confer any survival benefit in this analysis. Although the follow-up times may be too short to comment conclusively, it appears that predetermined markers of disease, which are present at diagnosis, overpower any survival benefit of current available interventions. Previous studies have excluded NIV and gastrostomy from survival models because of assumptions that they are markers of advanced disease and would confound results[178]. However, the proportions of gastrostomy (31.3%) and NIV (27.2%) by censorship time (mean follow-up 23.0 months) in our study provides an argument for their inclusion.

4.5 CONCLUSIONS

Using a hypothesis-free, data-driven approach and the wealth of clinical phenotypic information collected via CARE-MND, I have been able to stratify patients in Scotland to identify outcomes that independently predict survival. These variables should be used to stratify patients for Scottish clinical trial study groups, rather than more conventional but arbitrary group categorisation. Future models should build on this approach but also include weight/BMI at diagnosis, electrophysiological markers and, with emerging research, imaging biomarkers such as iron deposition quantification[332,333]. Further study into gut microbiota in pwMND may clarify our findings that smoking is protective. Finally, the contribution of genetic information to these survival models should be explored and will be discussed in Chapter 7.

5. GENOTYPE-PHENOTYPE PILOT STUDY OF HISTORICAL COHORT (1989-2014)

The following content has been published in *Neurobiology of Aging* and re-formatted for this thesis:

Holly A Black*, **Danielle J Leighton* (*Joint First Authors)**, Elaine M Cleary, Elaine Rose, Laura Stephenson, Shuna Colville, David Ross, Jon Warner, Mary Porteous, George H Gorrie, Robert Swingler, David Goldstein, Matthew B Harms, Peter Connick, Suvankar Pal, Timothy J Aitman, Siddharthan Chandran. Genetic epidemiology of motor neuron disease-associated variants in the Scottish population. *Neurobiol Aging.* 2017 Mar;51:178.e11-178.e20.

Genotyping (sequencing, filtering and variant calling) was led by Dr Holly Black at the Institute of Genetics and Molecular Medicine, Edinburgh University. Phenotyping was undertaken by the author (Dr Danielle Leighton).

5.1 BACKGROUND AND AIMS

Background

The major genetic contributors of monogenic MND in the UK population and European populations are expansions of an intronic hexanucleotide repeat in *C9orf72*, and missense variants in *SOD1* and *TARDBP*[102,334,335]. In Scotland, the *SOD1* p.I114T variant has been described as a founder mutation[144,214,219,336] and, more recently, *C9orf72* expansions were found in 11% of Scottish cases[215]. The incidence of MND-associated variants specifically in the Scottish population is unknown.

Aims

To investigate this knowledge gap further, we investigated a historical cohort of individuals with MND in Scotland diagnosed between 1989 and 2014, in collaboration with researchers at the Institute of Genomic and Molecular Medicine, University of Edinburgh. This work acted as a pilot study of a limited panel of MND-associated genes before embarking on a more detailed study using an incident cohort of patients (Chapter 6). We aimed to determine the contribution of variants in a selected group of genes commonly associated with MND to the burden of cases, and their association with disease phenotype. This included studying two genes that were found to be associated with MND after the inception of this PhD project - *TBK1* and *NEK1*[86,88,89].

5.2 METHODS

MND cases were recruited through the SMNDR. Recruited cases included individuals aged ≥16 with possible, probable or definite amyotrophic lateral sclerosis and individuals with MND subtypes (PBP, PMA and PLS). Individuals provided written consent for DNA extraction and genetic studies.

Four hundred forty-one samples, obtained from cases diagnosed with MND in Scotland in the years 1989–2014, were included in this study, which included three pairs of related individuals (two brothers, two first cousins, and two first cousins once removed). Case records were examined for seven phenotypic characteristics: sex, age at onset, age at diagnosis, time to diagnosis, duration of disease (until death or final data review [20th April 2016]), site of onset (bulbar or spinal), and family history of MND. Individuals were classified as having a family history of MND if a first, second, or third degree relative had been known to have MND. The study did not include the presence of frontotemporal dementia (FTD) alone as a criterion for positive family history. Five individuals were lost to follow-up due to relocation from Scotland, and survival dates were censored to date of last contact. The cohort was screened for expansions of the *C9orf72* intronic hexanucleotide repeat, as described by Cleary et al.[215]. A subset of the cohort had been screened for variants in *SOD1* in several previous studies[214,219].

Five MND cases, each with a variant in one of the five genes sequenced, were included as positive controls; these were taken from the cohort described by Cirulli et al.[86]. Four hundred ethnicity and sex-matched healthy controls were selected from the Generation Scotland Donor DNA databank[337]. The selected controls were aged ≥56 at the time of collection (20% aged 50–56, 66% aged 60–65, 14% aged 66+); an older cohort was chosen to minimize inclusion of young subjects who could go onto develop MND later in life.

Genotyping

Five genes were sequenced: *SOD1*, *TARDBP*, *OPTN*, *TBK1*, and *NEK1*; these will be referred to hereafter as the MND gene panel. The panel includes the recently associated *TBK1* and *NEK1*, alongside genes that are among the largest contributors to cases in UK and European populations (*SOD1*, *TARDBP*, and *OPTN*), after *C9orf72*[102,334,335]. All samples were also screened for *C9orf72* hexanucleotide expansions using repeat-prime PCR methods published by Cleary et al[215]. Expansions >30 repeats were considered pathogenic.

Genotyping, variant calling and variant classification were undertaken by a team at the Institute of Genomic and Molecular Medicine, University of Edinburgh, led by Dr Holly Black. The following methods were outlined by Dr Holly Black:

The primers for 120 amplicons were designed according to the Fluidigm Access Array protocol (Fluidigm). The primers amplified the coding regions of these five genes, excluding a total of 312bp of coding sequence across the five genes, due to primer design constraints. The Fluidigm Access Array was used for amplification. The amplicon library was sequenced on an Illumina MiSeq with 2×150-bp reads. Each batch of 48 samples included at least one water blank as a negative control. The positive control samples were each run in two independent batches. Quality control of the amplicon sequencing data was performed using FastQC (version 0.11.2)[338]*. Primer sequences were removed from the 5' ends of reads using the cutadapt tool (version 1.7.1) with the 'anchor' option*[339]*. The maximum error rate for primer sequences was set to 10%. Reads were mapped to the human genome reference sequence hs37d5 using BWA MEM (version 0.7.10)*[340]*. Picard (version 1.85)*[341] *and DepthOfCoverage GATK tool (version 3.3–0)*[342] *were used to collect alignment and amplicon coverage statistics. The median amplicon coverage of the negative control sample*

was used to determine the threshold for including samples from the same batch in further analysis. Only samples with a median coverage that was >10× that of the negative control were included. This removed seven cases and 11 controls, leaving 434 cases (432 independent) and 389 controls for further analysis. The UnifiedGenotyper tool, as implemented in GATK version 2.6 [342]*, was used for variant calling. ANNOVAR (version 2014 Nov 12) was used to provide functional annotation of the variants*[343].

Intronic and synonymous variants, variants with a population frequency greater than 1% in either the 1000 Genomes (October 2014 release)[344] or ExAC v0.2[345] data sets and variants with a frequency >5% in our cohort were excluded from further analysis. Per-sample variant calls were filtered to exclude calls with a read depth <50 or an allele balance <0.3. Filtered variants were validated by Sanger sequencing. One sample failed to amplify using PCR at variant validation stage and was excluded from further analysis, leaving 433 (431 independent) cases. Nine false positives were identified in *SOD1* exon 1, and this exon was excluded from the analysis. Following exclusion of this exon, five false positives remained, which were excluded from downstream analysis. Validated variants were submitted to ClinVar. The known variants in the positive controls were identified in both assays in which they were tested.

Assessing variant pathogenicity

Stop-gain, frameshift, and splice site variants were categorized according to their predicted effect on protein function. Missense variants reported as disease-causing in association with MND in the Human Gene Mutation Database[346] were categorized as pathogenic. The remaining missense variants were categorized according to *in silico* scores of pathogenicity and conservation. The scores used are listed below, with the thresholds for supporting pathogenicity/conservation in brackets: SIFT (=D), PolyPhen HDIV (=P/D), LRT (=D), Mutation Taster (=D), Mutation Assessor (=M/H), FATHMM (=D), CADD phred (>15), GERP (>2), phyloP (>2), and SiPhy (>10). Variants for which 7–10 *in silico* measures supported pathogenicity/conservation were categorized as likely pathogenic, variants with 4–6 measures supporting pathogenicity/conservation were categorized as uncertain significance,

and variants with 0–3 measures supporting pathogenicity/conservation were categorized as likely benign.

The ExAC v0.3.1 reference exome data were used to provide an additional, larger control population. It contains exome-wide variants identified in 60,706 individuals. For both *TBK1* and *NEK1*, the total number of loss-of-function or loss-of-function + missense alleles in ExAC were recorded (including filtered [non-pass] variants), excluding variants found in >1% individuals. The number of ExAC individuals was reduced to reflect the average percentage of individuals covered at 30× across each gene. For *TBK1*, an average of ∼69% of individuals was covered at 30×, giving 41,848 individuals assumed to have sufficient depth for variant calling. For *NEK1*, it was ∼50%, giving 30,298 individuals.

Statistical analysis

Fisher's exact tests were used in comparisons between the number of variants in cases and controls. One-tailed tests were used because of the prior assumption that cases contain more pathogenic variants than controls. Genotype-phenotype association testing was used to compare clinical phenotypes with different MND genotypes. Only unrelated individuals (n = 431) were included in the analysis, with the first individual recruited to the SMNDR from each pair retained for analysis. Variables were examined for collinearity using Pearson's correlation coefficient. Age of onset and age at diagnosis were highly correlated (Pearson's correlation r=0.98, $p<5\times10^{-7}$), as were duration of disease from onset and duration of disease from diagnosis (r=0.87, $p<5\times10^{-7}$). Disease onset is more important to disease biology than date of diagnosis, which relies on clinical services; therefore, variables related to diagnosis were excluded from genotype-phenotype association testing. Time to diagnosis was also excluded from genotype-phenotype analysis, as this is a derived variable. Univariate analysis was carried out for the following variables: sex, age of onset, duration of disease from onset, site of onset, and family history. Two-tailed Fisher's exact test was used for categorical data and *t*-test or Mann Whitney *U* test was used for parametric and non-parametric continuous data, respectively. Variables with significant univariate association at $p \leq 0.1$ were inputted into binomial logistic regression models dependent on gene. Logistic regression analysis was used to test the hypothesis that significant variables were

independently associated with having a pathogenic or loss-of-function variant in one of the genes tested. For the purposes of the model, age of onset was grouped by decade. Results from logistic regression modelling were considered significant if $p<0.05$. SPSS Statistics version 21 was used for statistical analysis.

5.3 RESULTS

Phenotypes

After sample filtering, 433 MND cases (431 independent) remained, of which complete phenotypic data were obtained for 428 (99%) (Table 14). Site of onset was unrecorded for one case. Family history was unrecorded for three cases and was unknown for one case, who was adopted. The cohort had a male-to-female ratio of 1.4:1 (Table 14). The mean age of onset was 59.5 years. Of the 443 cases in this cohort, 367 (85%) had died before censorship date; median duration of disease from onset for all the cases was 42 months. Of the 430 independent cases for which site of onset was recorded, 304 (71%) had spinal onset MND. Of the 429 cases with recorded family history, 44 (10%) had a family history of MND.

Phenotypic Characteristic	Summary Statistic	Values
Sex	*Female (%)*	177/433 (41)
Age of onset (years)	*Mean (SD)* *Range*	59.5 (12.9) 14-94
Age of diagnosis (years)	*Mean (SD)* *Range*	61.1 (12.6) 19-94
Time to diagnosis (months)	*Median (IQR)* *Range*	12 (6, 22) 0-386
Duration of disease from onset (months)	*Median (IQR)* *Range*	42 (25, 73.5) 4-583
Duration of disease from diagnosis (months)	*Median (IQR)* *Range*	25 (12, 55.5) 0-309
Site of onset	*Bulbar (%)*	126/432 (29)
Family history of MND	*Yes (%)*	44/429 (10)

Table 14 Phenotypic Characteristics of MND cases (n=433)

Identification of variants in MND genes

Targeted sequencing with the MND gene panel achieved a mean coverage of 6280× per amplicon per sample. Following variant filtering and validation, at least one rare stop-gain, splice site, frameshift, or missense variant was identified in 57/433 cases and 19/389 controls. Following sample filtering, the case population contained two pairs of related individuals. The first pair (brothers) both carried the *SOD1* G94R variant, whereas the second pair (first cousins) did not carry any variants in the MND gene panel that passed variant filtering. Therefore, the results for the related individuals were concordant.

Thirty-seven unique variants were identified in either cases or controls across the MND gene panel; of these, three were stop-gain, three were splice site, three were frameshift, and 28 were missense variants. All stop-gain, splice site, and frameshift variants were categorized as loss-of-function, except for three variants. A stop-gain variant

in *TARDBP* was categorised as uncertain significance, as it is in the final exon of the gene and therefore not expected to result in nonsense-mediated decay. There were also two splice site variants in *NEK1* predicted to result in the in-frame loss of a single exon[347], which were categorized as uncertain significance. Therefore, of the 37 unique variants identified, six were categorized as loss-of-function, seven pathogenic, 11 likely pathogenic, eight uncertain significance, and five likely benign (Table 15).

Gene	Variant	Variant type	Pathogenicity classification	Number of cases	Number of controls
NEK1	c.1948delC;p.Q650fs	frameshift	Loss-of-function	1	0
NEK1	c.T386G;p.I129S	missense	Likely Pathogenic	1	0
NEK1	c.G695A;p.R232H	missense	Likely Pathogenic	0	1
NEK1	c.G782A;p.R261H	missense	Likely Pathogenic	5	2
NEK1	c.G827T;p.C276F	missense	Uncertain significance	0	1
NEK1	c.G1021A;p.A341T	missense	Uncertain significance	4	3
NEK1	c.T1137A;p.D379E	missense	Likely Benign	0	1
NEK1	c.T1789A;p.F597I	missense	Likely Pathogenic	0	1
NEK1	c.G2137A;p.V713M	missense	Likely Benign	1	1
NEK1	c.T2235G;p.N745K	missense	Likely Pathogenic	3	4
NEK1	c.A2306G;p.H769R	missense	Likely Benign	0	1
NEK1	c.G2368A;p.A790T	missense	Likely Benign	0	1
NEK1	c.C2392G;p.L798V	missense	Likely Benign	1	0
NEK1	c.C3140T;p.S1047L	missense	Likely Pathogenic	1	0
NEK1	c.214+1G>A	splice site	Loss-of-function	1	0
NEK1	c.1750-5T>C	splice site	Uncertain significance	1	0
NEK1	c.1911+1->TATA	splice site	Uncertain significance	1	0
NEK1	c.C481T;p.R161X	stop-gain	Loss-of-function	1	0
OPTN	c.A280C;p.K94Q	missense	Likely Pathogenic	0	1
OPTN	c.A941T;p.Q314L	missense	Pathogenic	1	1
OPTN	c.A1337G;p.E446G	missense	Uncertain significance	1	0
OPTN	c.T1403G;p.M468R	missense	Likely Pathogenic	2	0
SOD1	c.G112A;p.G38R	missense	Pathogenic	1	0
SOD1	c.G280C;p.G94R	missense	Pathogenic	2*	0
SOD1	c.A302G;p.E101G	missense	Pathogenic	1	0
SOD1	c.T341C;p.I114T	missense	Pathogenic	18	1
SOD1	c.C437A;p.A146D	missense	Likely Pathogenic	1	0

TARDBP	c.G859A;p.G287S	missense	Pathogenic	3	0
TARDBP	c.G1043T;p.G348V	missense	Pathogenic	1	0
TARDBP	c.T1122G;p.Y374X	stop-gain	Uncertain significance	1	0
TBK1	c.1427delA;p.E476fs	frameshift	Loss-of-function	1	0
TBK1	c.2114_2126del;p.A705fs	frameshift	Loss-of-function	1	0
TBK1	c.C452T;p.S151F	missense	Likely Pathogenic	1	0
TBK1	c.C829G;p.L277V	missense	Likely Pathogenic	1	0
TBK1	c.A1135G;p.I379V	missense	Uncertain significance	0	1
TBK1	c.C1508T;p.T503I	missense	Uncertain significance	1	0
TBK1	c.C1330T;p.R444X	stop-gain	Loss-of-function	1	0

* pair of brothers

Table 15 Variants identified across the MND gene panel, with their pathogenicity classification and frequency in cases and controls

In total, 31 MND cases and two controls carried at least one loss-of-function or pathogenic variant across the MND gene panel (Fisher's p=1.761×10^{-7}) (Table 16). Taken together with prior findings that 44 of the 431 independent cases contain a pathogenic intronic hexanucleotide repeat expansion in *C9orf72*, 74 independent cases (17%) carried at least one pathogenic or loss-of-function variant. This represented 26/42 (62%) of cases with a family history and 47/385 (12%) of cases with no family history (Figure 14).

	Loss-of-function		Pathogenic		Likely Pathogenic		Uncertain significance		Likely benign		Total^	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	**Cases**	**Controls**
SOD1	0	0	22*	1	1	0	0	0	0	0	**23***	**1**
TARDBP	0	0	4	0	0	0	1	0	0	0	**5**	**0**
OPTN	0	0	1	1	2	1	1	0	0	0	**4**	**2**
TBK1	3	0	0	0	2	0	1	1	0	0	**6**	**1**
NEK1	3	0	0	0	10	8	6	4	2	4	**21**	**16**

*Table 16 Variants identified across the MND gene panel, with their pathogenicity classification and frequency in cases and controls; *including pair of brothers*

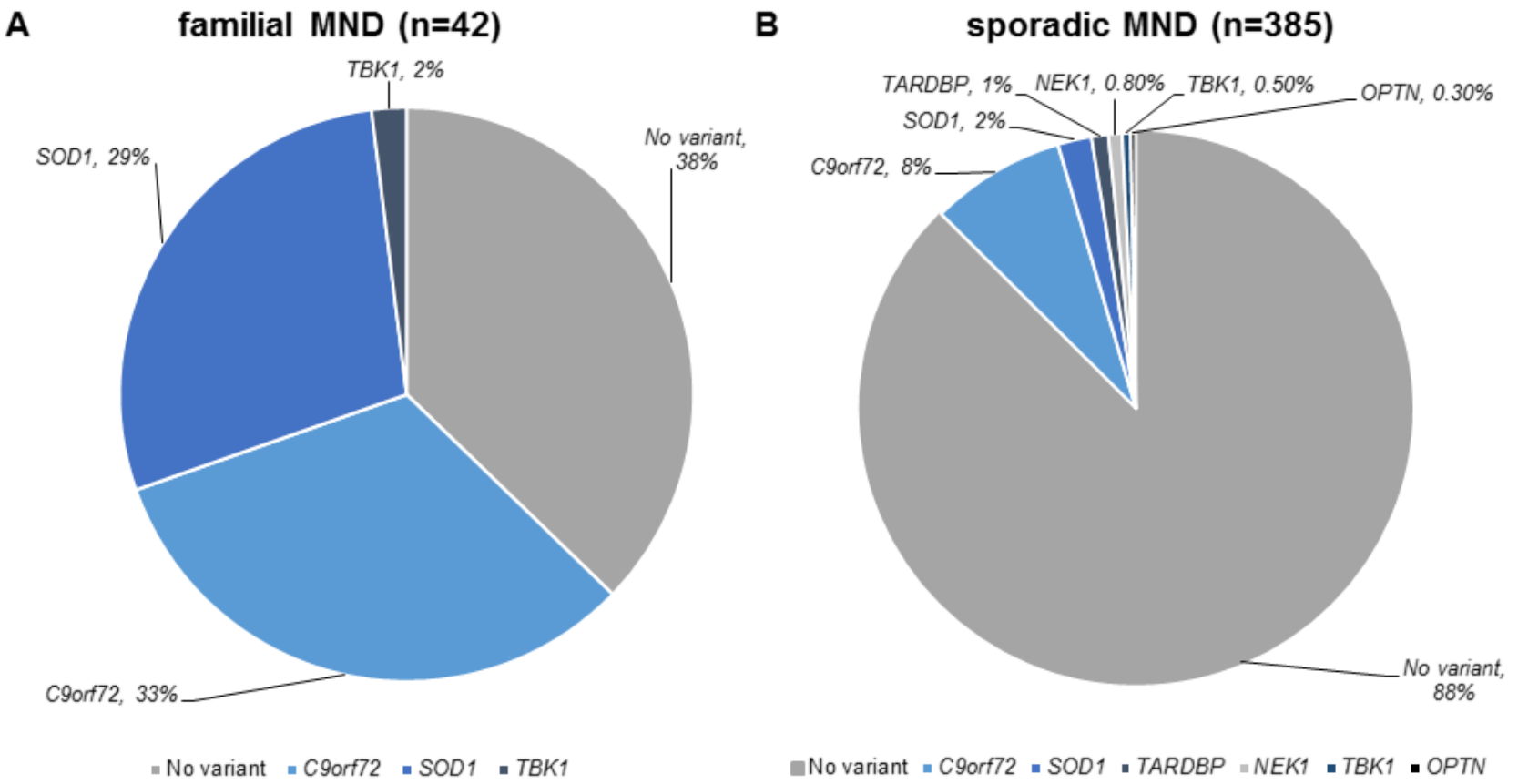


Figure 14 Proportion of cases with a pathogenic or loss-of-function variant in genes on the MND gene panel or in C9orf72. (A): cases with a family history; (B): sporadic cases with no known family history

Variants in *SOD1*, *OPTN*, and *TARDBP*

After *C9orf72*, the largest genetic contributor to cases in our study was *SOD1*, with pathogenic or likely pathogenic variants observed in 22/431 independent MND cases (5%) and 1/389 controls (0.3%) (Table 15, Table 16). The p.I114T Scottish founder mutation was observed in 18 cases (4%) and one control (0.3%). Of all cases, 29% of familial and 2% of sporadic cases carried a pathogenic variant in *SOD1*.

Pathogenic variants in *TARDBP* were observed in 4/431 (1%) cases and no controls. This is in addition to a stop-gain variant of uncertain significance, observed in one case. For *OPTN*, at least one pathogenic or likely pathogenic variant was observed in 3/431 (1%) cases and 1/389 controls (0.3%). The single pathogenic variant in *OPTN*, p.Q314L, was observed in one case and one control. There is no functional evidence to support the association of this variant with MND and, although in previous studies, *in silico* evidence has suggested a pathogenic role for the variant in MND[348], the occurrence of the variant in one case and one control in our data suggests that the variant either has variable penetrance or is not pathogenic for MND. *OPTN* therefore appears to contribute to very few cases of MND in the Scottish population.

Variants in *TBK1*

Several previous studies have reported dominant variants in *TBK1* in cases of MND, FTD, and MND-FTD, accounting for approximately 1% of all cases[85,87,159,248,349–355]. In our study, we observed six cases and one control with a rare stop-gain, splice site frameshift, or missense variant in *TBK1* (Fisher's p=0.080). When compared with the ExAC reference data set, this result is not significant (Fisher's p=0.238). When looking only at loss-of-function variants, 3 cases and no controls contained a loss-of-function variant in *TBK1* (Fisher's p=0.145). When compared with the ExAC reference data set, where the frequency of loss-of-function variants was 0.03%, this represents a statistically significant excess (Fisher's $p=4.370\times10^{-4}$).

Variants in *NEK1*

Loss-of-function variants in *NEK1* have been associated with MND in two recent studies of both familial and sporadic MND cases[88,89], following an earlier study of mostly sporadic cases that highlighted *NEK1* as a candidate gene[159]. These three studies estimate that loss-of-function variants in *NEK1* contribute to approximately 1% of cases. In this study, we observe three cases and no controls with a variant predicted to result in loss-of-function of *NEK1* (Fisher's p=0.145) (Table 15, Table 16). When compared with the ExAC reference data set, where the frequency of loss-of-function variants was 0.3%, our result remains nonsignificant (Fisher's p=0.143). Regarding missense variants in *NEK1*, we observe 16 cases and 16 controls with an *NEK1* missense variant, which is not statistically significant (Fisher's p=0.453).

Digenic cases

There are several reports of MND cases that carry MND-associated variants in more than one gene[92,94–97,356]. We observed five cases and one control carrying two different rare stop-gain, frameshift, splice site, or missense variants either across the MND gene panel or in *C9orf72* (Table 17). Two cases, MND-0040 and MND-0434, and no controls carry two variants that are predicted to be either pathogenic or loss-of-function; these will be referred to as digenic cases. Both digenic cases have an age of onset within two standard deviations of the mean and a typical disease duration. However, as discussed, the pathogenicity of the variant p.Q314L in *OPTN* in MND-0040 is unclear.

Of the remaining cases with two rare variants in MND-associated genes, case MND-0119 carries one pathogenic variant and one variant of uncertain significance (VUS) and has a typical age of onset and disease duration. Case MND-0158, who carries one pathogenic and one likely pathogenic variant, had a young age of onset (26 years, >2.5 standard deviations from the mean) and long disease duration (142 months, above 90th percentile). MND-0211 also carries one pathogenic and one likely pathogenic variant and had a typical age of onset,

but disease duration at the lower end of the spectrum (16 months, equal to the 10th percentile). Interestingly, three of the five cases (60%) carrying two rare stop-gain, splice site, frameshift, or missense variants in MND genes (including *C9orf72* expansions) had a bulbar site of onset, which is higher than the 29% observed across all MND cases in the cohort.

CONTROL-0325, who was aged 56–60 at the time of collection, carries two missense variants in OPTN; one considered pathogenic; and one likely pathogenic. The pathogenicity of variant p.Q314L, as discussed, is unclear. The data available do not allow confirmation of whether the two *OPTN* variants are found in cis or in trans and, therefore, if both allelic copies of *OPTN* carry a variant. It is also unclear whether both variants are expected to contribute to MND pathogenesis; if so, they would be expected to be associated with variable penetrance. As no follow-up information is available for controls, there is a small chance the control developed MND after sample donation, although this is unlikely given the lifetime risk of MND.

Sample Information		Phenotype Information				Variant 1				Variant 2			
Sample ID	**Sample Type**	**Age of Onset (years)**	**Duration of disease from onset (months)**	**Site of Onset**	**Family History**	**Gene**	**Variant**	**Variant Type**	**Pathogenicity**	**Gene**	**Variant**	**Variant Type**	**Pathogenicity**
MND-0040	Case	46	23	Bulbar	No	*OPTN*	c.A941T; p.Q314L	missense	Pathogenic	*TBK1*	c.2114_2126del; p.A705fs	frameshift	Loss-of-function
MND-0434	Case	44	34	Limb	Yes	*C9orf72*	(GGGGCC)n	intronic HRE	Pathogenic	*TBK1*	c.1427delA; p.E476fs	frameshift	Loss-of-function
MND-0158	Case	26	142	Limb	No	*C9orf72*	(GGGGCC)n	intronic HRE	Pathogenic	*TBK1*	c.C829G; p.L277V	missense	Likely Pathogenic
MND-0211	Case	65	16	Bulbar	No	*C9orf72*	(GGGGCC)n	intronic HRE	Pathogenic	*NEK1*	c.T2235G; p.N745K	missense	Likely Pathogenic
MND-0119	Case	64	44	Bulbar	No	*TARDBP*	c.G859A; p.G287S	missense	Pathogenic	*NEK1*	c. G1021A; p.A341T	missense	Uncertain significance
CONTROL-0325	Control	N/A	N/A	N/A	N/A	*OPTN*	c.A941T; p.Q314L	missense	Pathogenic	*OPTN*	c.A280C; p.K94Q	missense	Likely pathogenic

Table 17 Samples containing two variants that potentially contribute to MND pathogenicity

Genotype-phenotype associations

Genotype-phenotype association testing was used to determine the relationship between carrying a variant in a specific gene and five phenotypic markers (Table 18). MND cases carrying the *SOD1* p.I114T founder variant were also analysed independently of other *SOD1* variants in view of the high incidence of this specific variant in the Scottish population. Only unrelated cases were included in the analysis (n=431).

Univariate analyses comparing cases with a pathogenic or loss-of-function variant (including *C9orf72*expansions) to cases without such a variant found associations between carrying a pathogenic or loss-of-function variant and a family history of MND (Fisher's $p<5\times10^{-7}$), a younger age of onset (*t*-test $p=0.018$), and female sex (Fisher's $p=0.028$) (Table 18). A logistic regression model evaluating the possible independent associations of these variables with variant status explained 21% of the variance between cases with and without a pathogenic or loss-of-function variant (Nagelkerke R^2). Independent associations were found for positive family history ($p<5\times10^{-7}$, OR 10.88, 95% CI 5.38–22.01), decreasing age of onset by decade ($p=0.018$, OR 0.77, 95% CI 0.62–0.96), and female sex ($p=0.028$, OR 1.88, 95% CI 1.07–3.30) (Table 19).

A high proportion of cases carrying a pathogenic *SOD1* variant had spinal-onset disease (95%) compared with the frequency of spinal-onset in all other case (Fisher's $p=0.012$). Carrying a pathogenic *SOD1* variant was also associated with having a family history of MND, compared to all other cases (Fisher's $p<5\times10^{-7}$) (Table 18). When examining *SOD1* p.I114T carriers in isolation, these two factors remained significant (site of onset Fisher's $p=0.031$, family history Fisher's $p=9\times10^{-6}$) (Table 18). Twenty-seven percent of the variance between *SOD1* pathogenic variant carriers and all other cases was explained by the model (Nagelkerke R^2). Spinal onset disease ($p=0.044$, OR 8.25, 95% CI 1.06–64.49) and family history of MND ($p<5\times10^{-7}$, OR 16.21, 95% CI 6.21–42.33) were independently associated with carrying a *SOD1* pathogenic variant. After initial modelling for the *SOD1* p.I114T variant (using sex, site of onset, and family history), sex and site of onset did not achieve significance at the $p=0.1$ threshold and were removed. On revised modelling, family history remained significant ($p=1\times10^{-6}$, OR 11.39, 95% CI 4.23–30.67) (Table 19), explaining 16% of the variance (Nagelkerke R^2).

Univariate analysis showed carrying a *C9orf72* expansion was associated with having a family history of MND (Fisher's $p=1\times10^{-5}$) (Table 18). Site of onset did not reach statistical significance ($p<0.05$), but met the criteria required for inclusion in the logistic regression model (Fisher's $p=0.082$). The logistic regression model explained 12% of the variance between carriers of *C9orf72* expansions and those without (Nagelkerke R^2). Bulbar onset disease ($p=0.021$, OR 2.23, 95% CI 1.13–4.42) and family history ($p=1\times10^{-6}$, OR 6.86, 95% CI 3.19–14.77) were significantly and independently associated with carrying a *C9orf72* expansion (Table 19).

Phenotypic Characteristic	Statistic	No Path (n=357)	Path/ LoF (n=74)	*C9orf72* (n=44)	*SOD1* inc. I114T (n=21)	*SOD1* I114T (n=18)	*TARDBP* (n=4)	*TBK1* (n=3)	*NEK1* (n=3)	Digenic (n=2)
Sex	*Female (%)* Fisher's	*138 (39)* -	*39 (53)* **p=0.028**	*21 (48)* p=0.419	*12 (57)* p=0.171	*11 (61)* **p=0.090**	*3 (75)* p=0.310	*2 (67)* p=0.571	*1 (33)* p=1.000	*1 (50)* p=1.000
Age of onset (years)	*Mean (SD)* Mean diff CI T-test	60.2 (12.9) -	*56.3 (12.1)* -0.39 -0.72 – -0.07 **p=0.018**	*55.6 (11.4)* -0.44 -0.84 - -0.04 p=0.266	*55.9 (14.4)* -0.38 -0.94 – 0.18 p=0.186	*59.4 (12.0)* -0.01 -0.62 – 0.59 p=0.967	*61.3 (7.8)* 0.18 -1.09 – 1.44 p=0.786	*48.7 (6.4)* -1.09 -2.55 – 0.36 p=0.141	*66.7 (10.6)* 0.72 -0.74 – 2.18 p=0.332	*45.0 (1.4)* -1.46 -3.24 – 0.32 p=0.108
Duration of disease from onset (months)	*Median (IQR)* Mann-Whitney	42.0 (48) -	*37.0 (45)* U=11784.0 p=0.144	*38.5 (53)* U=7644.0 p=0.266	*37.0 (67)* U=3847.5 p=0.411	*38.0 (89)* U=3483.0 p=0.652	*39.5 (17)* U=834.0 p=0.936	*34.0 (-)* U=558.5 p=0.698	*31.0 (-)* U=610.5 p=0.884	*28.5 (-)* U=271.5 p=0.370
Site of onset	*Bulbar %* Fisher's	*103 (29)* -	*23 (31)* p=0.779	*18 (41)* **p=0.082**	*1 (5)* **p=0.012**	*1 (6)* **p=0.031**	*2 (50)* p=0.584	*2 (67)* p=0.207	*0 (0)* p=0.559	*1 (50)* p=0.501
Family history of MND	*Yes (%)* Fisher's	*16 (5)* -	*26 (36)* **$p<5x10^{-7}$**	*14 (33)* **$p=1x10^{-5}$**	*12 (57)* **$p<5x10^{-7}$**	*9 (50)* **$p=9x10^{-6}$**	*0 (0)* p=1.000	*1 (33)* p=0.268	*0 (0)* p=1.000	*1 (50)* p=0.187

*Table 18 Genotype-phenotype tests of association comparing phenotype with pathogenic or loss-of-function variants in different MND genes; values in **bold** are p<0.1*

Variant type	Predictor	p-value	OR (CI)
Path/LoF (n=74)	*Female Sex*	0.028	1.88 (1.07-3.30)
	Age of Onset (decade)	0.018	0.77 (0.62-0.96)
	Family History of MND	$<5x10^{-7}$	10.88 (5.38-22.01)
***C9orf72* (n=44)**	*Bulbar Onset*	0.021	2.23 (1.13-4.42)
	Family History of MND	$1x10^{-6}$	6.86 (3.19-14.77)
***SOD1* (n=21)**	*Spinal Onset*	0.044	8.25 (1.06-64.49)
	Family History of MND	$<5x10^{-7}$	16.21 (6.21-42.33)
***SOD1* I114T (n=18)**	*Family History of MND*	$1x10^{-6}$	11.39 (4.23-30.67)

Table 19 Logistic regression modelling of genotype-phenotype tests of association

The influence of carrying a pathogenic or loss-of-function variant in one of the MND genes on survival was visualized using a Kaplan-Meier plot (Figure 15). Cases were grouped by genotype. Survival was plotted for all cases, including the 15% of cases for whom duration of disease was calculated from onset to date of last contact/study truncation. No significant difference was found between groups (log rank p=0.276).

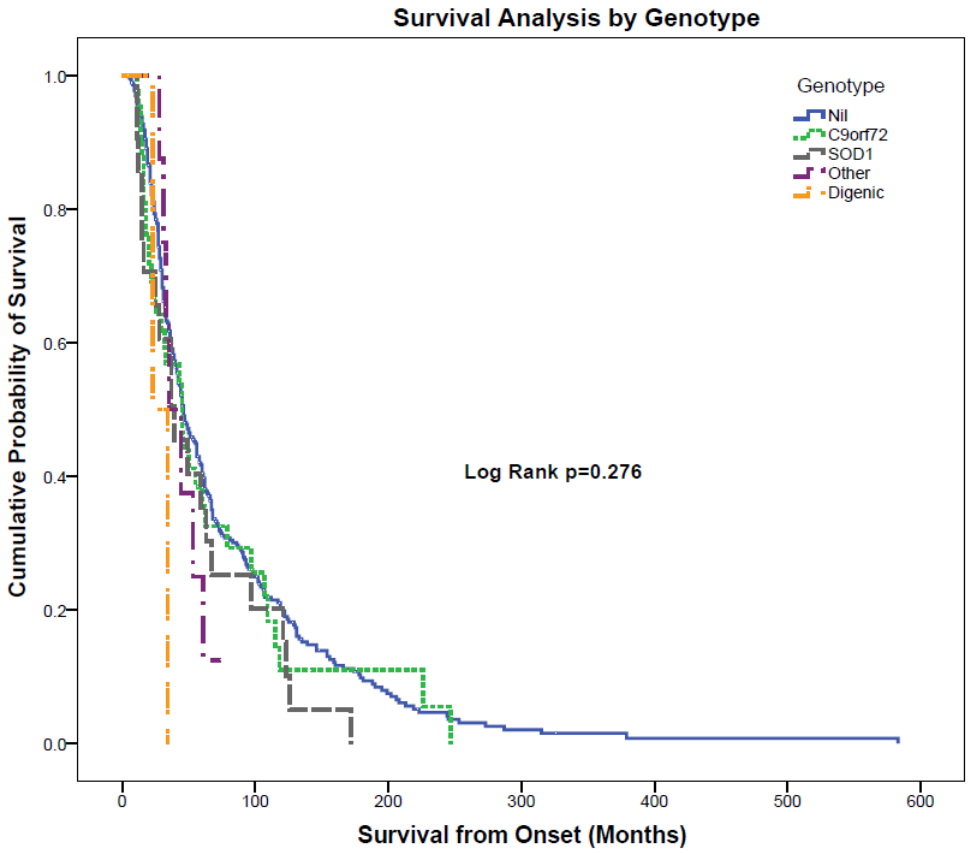


Figure 15 Kaplan-Meier survival plot of MND cases grouped by genotype

5.4 DISCUSSION

Genetic epidemiology of MND in Scotland

This study provides a description of the genetic pathology of MND in Scotland using a six-gene panel. We have identified 31 independent cases carrying a pathogenic or loss-of-function variant across our MND gene panel. The identification of pathogenic variants in two controls (0.5%) is comparable to previous studies and is explained by the variable penetrance observed with variants in genes associated with MND[94,144].

The historical Scottish MND population were comparable with other European MND populations on demographic and disease characteristics[357–359]. A positive family history in 10% of Scottish MND cases was comparable recent population estimates (16% in the Irish population and 9% in a US population of European ancestry[67,68]. Our data show that, after *C9orf72*, variants in *SOD1* contribute to the largest proportion of MND cases (5%) in the Scottish population, which is in the upper range compared with other populations of European ancestry[102,360]. This reflects the high frequency of the p.I114T variant, which had previously been identified as a founder mutation with variable penetrance in the Scottish population[144,214,219,336]. Although this variant is less frequent than was observed previously in a smaller cohort[219], it is still found in 4% of cases when assessed across our larger cohort and is found in 24% of those cases in which a pathogenic variant was identified. In comparison, pathogenic variants in *TARDBP* contribute to <1% of cases, comparable to previous estimates in populations of European ancestry[102], and pathogenic variants in *OPTN* are extremely rare, which is similar to a previous study of the UK population[361].
Our data support a growing literature associating loss-of-function variants in *TBK1* with MND. Several previous studies have associated dominant coding variants in *TBK1* with MND. The first study reported a significant excess of variants in *TBK1*, excluding those predicted to be benign, in mostly sporadic cases compared to controls[159]. However, the biggest excess was observed for loss-of-function variants. This was supported by a second study, which found a significant excess of loss-of-function variants in *TBK1*, but only in familial cases[85]. Several other studies, looking at MND, frontotemporal dementia (FTD) and MND-FTD cases from both European and Asian populations, have also reported variants in *TBK1,* mostly loss-

of-function[248,349–351,353–355,362]. In our study, although the number of loss-of-function variants identified in *TBK1* was not statistically significant when compared with our 389 controls, a significant difference was observed when compared with the larger ExAC reference data set and the percentage of cases carrying a loss-of-function variant in *TBK1* was comparable to previous studies (0.7% vs. <1% in previous studies[85,159]). Therefore, our data support the association of loss-of-function variants in *TBK1* with MND.

Similarly, the frequency of *NEK1* loss-of-function variants in our cases (0.7%) is comparable to previous studies (1%)[88,89,159]; however, the observation of three loss-of-function variants in *NEK1* in our cases was not significantly different to that of our 389 controls or the ExAC reference data set. Although the result does not provide formal statistical significance for the association of loss-of-function variants in *NEK1* with MND, as the frequency is comparable to other studies, our result is supportive of these previous findings. We observed the same number of missense variants in *NEK1* in both cases and controls, with little difference in the assigned pathogenicity classifications. This is comparable to a recent study which showed that, although one specific missense variant, p.R261H (classified as likely pathogenic in our study), was associated with MND, collectively, missense variants showed very little difference in frequency between sporadic cases and controls[88]. In addition, the other variants in *NEK1* causally related to the development of MND in previous studies are loss-of-function variants. There may therefore be some specific missense variants that result in impaired *NEK1* function, but collectively, this appears not to be the case.

We identified two cases each carrying two variants classified as pathogenic or loss-of-function in different MND genes (digenic cases). As controls were not screened for hexanucleotide repeat expansions in *C9orf72*, we were biased against identifying digenic controls, so it is not clear whether our two digenic cases represent a statistically significant excess. As each digenic case has a different variant combination, it is not possible to obtain a detailed assessment of how digenicity influences MND phenotype. Further investigation of a larger number of digenic cases is required to determine whether each variant contributes to disease pathogenesis, whether variants have an additive effect and to determine how different variant combinations influence disease presentation.

Genotype-phenotype associations

Our genotype-phenotype analysis confirms the expected association that MND cases with a family history of MND are more likely to carry a pathogenic or loss-of-function variant than cases with no family history. The results also indicate that a young age of onset of MND is a significant independent predictor of carrying a pathogenic variant. Although this finding requires replication, it suggests that, in addition to the presence of a family history, clinicians should use a lower threshold for genetic testing in cases developing symptoms of MND before the mean age of onset. The analysis also suggests that female sex is significantly associated with carrying a pathogenic or loss-of-function variant. This most likely reflects the equal number of males and females carrying variants pathogenic for MND, within an MND case population which overall has a male bias. This result is also consistent with the liability threshold genetic model, which would suggest that, as females have an overall lower risk than males of developing MND, they must accumulate a larger burden of risk factors to pass the threshold for disease onset[363]. The logistic regression model concerned only explains 21% of the difference between cases with a pathogenic or loss-of-function variant and those without, suggesting that additional phenotypic markers are acting as confounders. Our study confirms that *C9orf72* expansion carriers are more likely to present with bulbar disease than other MND cases[132,163,364]. However, our model explained only 12% of the variance, and substantial phenotypic heterogeneity among *C9orf72* carriers, particularly in terms of cognitive symptoms, is widely recognized[75]. Cognitive profile may be an important parameter to explore further in characterizing this cohort of cases. *SOD1* variant carriers in our cohort were relatively homogeneous in terms of site of onset, with a bias toward spinal onset MND. This mirrors clinical accounts of lower limb onset disease as a common presentation of *SOD1* MND[365]. A high proportion of p.I114T carriers also had spinal onset disease, despite failing to achieve statistical significance on logistic regression modelling. However, heterogeneity in p.I114T carriers for age of onset (range 42–84 years) and duration of disease from onset (11–172 months) was notable and suggests any biological link to these features is absent or extremely weak.

Limitations

Patients were recruited for this study between 1989 and 2014. Although demographics are comparable with other population cohorts, it is likely that this cohort is enriched for patients who are more interested in genetic research (for example, those with a family history of disease). Some patients were previously recruited for *SOD1* studies in Scotland[214,336]. Analysis of an incident cohort of new, unselected patients might give a more accurate estimate of the true genetic epidemiology in Scotland. This will be explored in Chapter 6.

Our methods for interpretation of variant pathogenicity are outlined – these are based on arbitrary rules guided by available evidence from research and genetics communities. However, there is clearly a need for standardisation of methods. The recent American College of Medical Genetics guidelines[366] sought to address this; however, they may require additional consideration for complex disorders such as MND, where there is a preponderance for sporadic cases and considerable variability in the penetrance of pathogenic variants. The lack of segregation data and large number of singleton variants meant that using *in silico* predictors of variant pathogenicity and conservation were most appropriate for assessing variant pathogenicity in this particular study, although it is clear that standardisation of the methods used to assign pathogenicity for potential MND-associated variants is required. This will also be explored in Chapter 6.

One limitation of the combined use of the Fluidigm Access Array and Illumina MiSeq is the number of false positives we observed. The large number in *SOD1* exon 1 is most likely due to the high GC content of this exon (Timothy J Aitman, unpublished data). Exclusion of this exon could have reduced the overall variant detection rate for *SOD1*, as many pathogenic missense variants have been reported in this exon[346]. After excluding this exon, 6% of the remaining variants also failed to validate. This could be due to the high number of cycles in the two rounds of PCR (35 and 15 cycles respectively) required to generate the sequencing library. Therefore, it is essential to use Sanger sequencing to validate variants identified through this protocol, as is typically applied following high-throughput sequencing.

Survival analysis between MND cases carrying pathogenic or loss-of-function variants in different genes did not achieve statistical significance. Possible limitations to this analysis

include sample size and lack of data relating to interventions that influence survival (e.g., non-invasive ventilation, gastrostomy insertion, and use of riluzole).

5.5 CONCLUSIONS

In summary, we identified a pathogenic or loss-of-function variant in an MND gene in 17% of our cohort of MND cases from the Scottish population. Our data give supporting evidence for the association of loss-of-function variants in *TBK1* and *NEK1* with MND. Genotype-phenotype association testing has highlighted that MND cases with a family history or with a young age of onset are significantly more likely to carry a genetic variant pathogenic for MND and suggests that cases presenting with a young age of onset should be referred for genetic testing, in addition to cases with a family history.

Study of an existing historical cohort of pwMND (Chapter 5) has provided us with pilot data from which I can generate hypotheses to test using an incident cohort of newly diagnosed patients. New hypotheses include the following and will be addressed in Chapters 6 and 7:

i) The incident cohort will have a similar or lower proportion of patients with familial MND (as the historical cohort may have been enriched for patients with a family history).

ii) The proportion of patients genotyped with an extended gene panel with a mutation will be higher than those genotyped with a limited six-gene panel (greater than 17%). *TBK1* and *NEK1* mutations will be found in the incident population.

iii) A systematic variant classification system will provide greater certainty regarding the pathogenicity of MND-associated variants.

iv) Having a gene mutation will be associated with a family history of MND and younger age of onset. We expect that the association with female sex may not be stable in future models of genotype-phenotype correlation.

v) Having a *C9orf72* expansion will be associated with bulbar-onset disease; inclusion of cognitive testing may reveal a significant association between

C9orf72 expansions and cognitive impairment. Having a *SOD1* mutation may be associated with limb-onset disease.

6. GENOTYPING 2015-2017

6.1 BACKGROUND AND AIMS

Background

Work leading up to this part of the project suggests that genetics appears to be an important factor in MND, with the potential for genetic testing to elucidate genotype-phenotype associations and inform prognosis. Seventeen percent of Scottish pwMND (both familial and apparently sporadic cases) had a potential genetic cause of their disease using a limited gene panel only (Chapter 5)[243]. In 2014, a neurodegenerative gene panel comprising 11 genes was incorporated into clinical practice in Scotland; this was available for clinical testing of pwMND and at risk relatives (details described in Chapter 1.3). However, with the emergence of new discoveries regarding genetic associations in MND as a consequence of advances in next generation sequencing techniques and large-scale population analyses (Chapter 1.1), this has quickly become outdated and may not capture the breadth of potential causative genes in an MND population.

As outlined in Chapter 1.2, pwMND are becoming increasingly aware of, and interested in, pursuing genetic testing. For individuals who proceed with genetic testing, interpretation of variant implications brings significant challenges. The majority of MND genetics variants are autosomal dominantly inherited but there is evidence for recessive disease in carriers of *OPTN*, *SPG11*, *FUS* and *SOD1* mutations[169]. Mutations in the *UBQLN2* gene are X-linked but can affect males and females[138]. Variant penetrance is variable and often age-dependent; study of an extended pedigree of individuals with the *SOD1* p.I114T variant calculated penetrance of 50% by age 60 and 88% by age 80[144]. Indeed, this can result in inherited MND skipping generations[100]. Even having a *C9orf72* repeat expansion does not guarantee manifestation of MND[100] and quoted inheritance is about 50%[103]. Quality of evidence may be limited or restricted to certain geographical locations[123,350]. The possibility of oligogenic MND also has significant implications for genetic testing, requiring the clinical team to be mindful not to stop investigation if a single gene test is positive[100].

Classification of variant pathogenicity is a problematic domain for many genetic diseases but becomes particularly difficult within the scope of a rare disease with multiple genetic links such as MND[367]. Typical classification systems include variants that are Pathogenic, Likely Pathogenic, Benign, Likely Benign and Variants of Uncertain Significance (VUS)[366]. VUS are inevitable and bring their own diagnostic difficulties. For diseases in which invasive interventions are currently available, such as breast cancer, a VUS result presents a management dilemma[368]. For example, 11% of individuals undertaking *BRCA1* and *BRCA2* gene testing for breast or ovarian cancer were carriers of a VUS[369]. Ten percent of the women in this cohort subsequently opted to undergo prophylactic mastectomies, 21% bilateral salpingo-oopherectomies. The extrapolated implications for MND are clear when we consider the imminent advent of genetic stratified therapies, which have the potential to involve prolonged inpatient and intravenous treatments[173,176,370]. On a more immediate level, however, there is the concern of burdening patients and their relatives with the anxiety of an uncertain future risk[169]. For many pwMND who have impaired cognition, FTD, or other neuropsychiatric comorbidities the possibility of discovering a VUS requires additional consideration.

Barriers to firm classification have been acknowledged, including the relative paucity of functional studies and large pedigrees for assessment of co-segregation[371]. In spite of this, as mentioned in Chapter 5.4, there remains no consensus classification system for assessment of MND variant causality[372]. Many population-based genotype studies have reported expected pathogenicity of novel variants by considering the following criteria: i) presence of variant in control populations, ii) presence of variant in public databases such as the ExAC database of genetic variants[352], 1000 Genomes[344], ESP6500[373] or ALS Online Database[335], and iii) assessment of variant using a varying number of *in silico* functional predictions[374–377]. More arbitrary methods have included publication of the variant in more than one study[378] and publication consensus combined with minor allele frequency (MAF)[372]. Latterly, the American College of Medical Genetics (ACMG) guidelines have been adopted[243,366,379]. The ACMG workgroup devised standards and guidelines for the interpretation of variants in disease. The utility of these recommendations for MND is yet to be determined.

The cumulative evidence indicates a growing need for improved resources for genetic testing and risk assessment, and consequent better communication with patients. It is the duty of the multidisciplinary team of MND specialists, geneticists and counsellors to address these requirements, in parallel to, and perhaps as a priority over, ongoing discovery. The importance of these efforts to clinicians and patients will only become clearer with emerging genetic stratification of pwMND and treatments.

Chapter 5 outlined the genetic epidemiology in Scotland using six key genes only. However, the burden of MND-associated rare genes in this population is unknown. This information is required to inform clinical and diagnostic testing and to outline priorities for future disease modelling studies. Crucially for pwMND, we need to be aware of the proportion of gene carriers in Scotland for which there might be genetically-targeted treatments in the near future. Accessibility to research is a key barrier to population-based genetic studies. However, with the revival of the Scottish MND Register through CARE-MND, prospective, unselected recruitment for population-based genetic study is possible.

Aims

i) To develop a new and updated neurodegenerative disease gene panel for testing pwMND in a research and clinical capacity.

ii) To review existing methods for variant classification in MND genomes and to outline a framework for disease and gene-specific classification.

iii) To describe the genetic epidemiology of the Scottish MND population in 2015-2017 using an inclusive, contemporary gene panel.

6.2 METHODS

Gene Panel Selection

A review was undertaken to update the existing 11-gene neurodegenerative disease gene panel. In addition to a review of the literature (Chapter 1.1), existing UK-based MND-related gene panels and resources were examined. Resources included: FTD and/or ALS and dementia disorders gene panels from the Institute of Neurology in London the King's College London[380], the ALS with or without FTD gene panel from the Sheffield Diagnostic Genetics Service[381], the ALS panel from the National Centre for Biotechnology Information (NCBI) Gene Testing Registry[382], the list of major genes from the King's College London ALS Online Database[72,335] and genes related to ALS and FTD according to the OMIM database[383]. The final panel consisted of 49-MND associated genes for research study. Figure 16 illustrates the components of both the original clinical gene panel and the extended research panel. Results of the research panel would aim to inform an updated clinical panel.

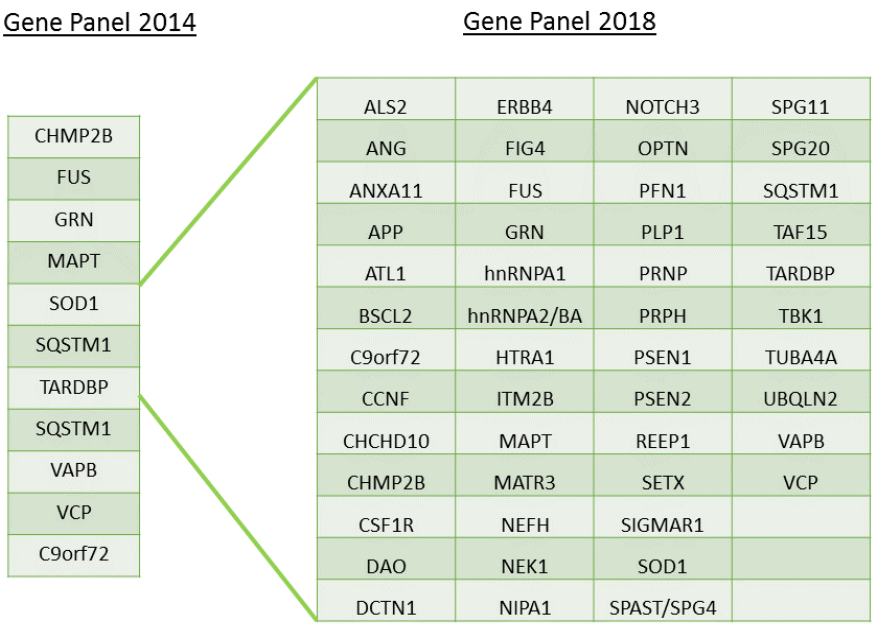


Figure 16 Neurodegenerative disease gene panel in Scotland in 2014 and in 2018

DNA Collection

DNA was acquired for this national genotypic study through the Scottish Regenerative Neurology Tissue Bank. For the majority of samples, DNA was acquired by the author by travelling to patients in all 14 health boards in Scotland, meeting them at home, in clinic or during inpatient hospital/hospice stays. DNA collection was supported by an MND Research Nurse and, latterly, by the MND Nurse Consultant and local clinical nurse/allied health specialists. We provided training for local specialists in DNA acquisition and consent and by the end of the study period, DNA acquisition was maintained sustainably at a local level. Third sector parties, in particular MND Scotland, publicised this research through their website and magazines to generate interest. I also delivered talks and interactive question-and-answer sessions at charity-led support group meetings for patients and carers throughout Scotland.

The consent process involved a period of discussion regarding the benefits and implications of research genetic testing, before the patient was invited to provide written consent and donate a DNA sample (See Appendix 5 for DNA Bank information sheet and consent form). During these discussions, it was emphasised that patients would not receive results from genetic testing from these research samples. This is a stipulation of the ethical approvals for the Scottish Regenerative Neurology Tissue Bank for the following reasons: i) discoveries in MND genetics are emerging and dynamic and the implications of genetic mutations as yet unclear; ii) genetically-targeted treatments are not yet available for MND and the benefits of relaying this information to patients is as yet unclear; iii) genetic services for MND in Scotland are not get equipped to provide counselling for routine testing of pwMND – this may require provisions for extended family counselling, preimplantation genetic diagnosis counselling and review appointments to reassess variants of uncertain significance (VUS).

DNA was acquired from all consenting patients for: i) testing against the neurodegenerative disease panel and ii) storage for potential future studies. All data recorded for the DNA database was anonymised using a non-identifiable patient code. DNA was stored in the Scottish Regenerative Neurology Tissue Bank, which also has approval from the Scotland A Research Ethics Committee. This database is an amalgamation of “The Scottish Motor

Neurone Disease DNA Bank" and the "Cognitive Disorders Clinic – Diagnosis, Audit, Research and Treatment (CDC-DART) DNA Bank.

Genotyping

Samples were genotyping using QiaSeq Amplicon Sequencing. The following methods were employed and sequencing was carried out by Dr Morad Ansari in the South East Scotland Genetics Service, Western General Hospital, Edinburgh:

Sequence analysis of a panel of 48 genes causally associated with neurodegeneration was carried out using a custom-designed QIAseq assay for library construction as per manufacturer's instructions (QIAGEN). In brief, 80ng of DNA was fragmented followed by adaptor ligation. Target enrichment was carried out by single primer extension, followed by sample indexing and amplification. Equal volumes of libraries were combined, and quantified using a Quantus™ Fluorometer as per manufacturer's instructions. Paired-end sequencing of the resulting DNA library (at a concentration of 10pM) was performed using an Illumina MiSeq instrument. Alignment and variant calling was performed using the QIAGEN CLC Genomics Workbench as per in-house standard operating procedure. Sequence read coverage was assessed against a browser extensible data (BED) file containing the genomic regions of interest.

Coding regions of the genes were covered with -15 bp and +10 bp at exon-intron boundaries with the exception of hnRNPA1 and hnRNPA2 (where sequencing was limited to specific known hotspots: *hnRNPA1* gene (NM_031157_exon_8_10_chr12_54677596 and NM_031157_exon_9_10_chr12_54678042), *hnRNPA2* gene (NM_031243_exon_2_10_chr7_26232871)). Further, *MAPT* was given wider boundaries to account for known intronic variants (NM_001123066_exon_1_10_chr17_44039687, exon_2_44049225, exon_3_44051751, exon_4_44055741, exon_5_44060544, exon_6_44064406, exon_7_44067244, exon_8_44068826, exon_9_44071290, exon_10_44073765, exon_11_10_chr17_44087676, exon_12_44091609, exon_13_44095984, exon_14_44101322).

All samples were also screened for *C9orf72* hexanucleotide expansions using repeat-prime PCR methods published by Cleary et al[215]. Expansions >30 repeats were considered pathogenic.

Variant Classification – Framework Development

There is no consensus classification system for assessment of causality of MND-associated genetic variants. American College of Medical Genetics (ACMG) Guidelines have been employed but their use has not been validated in MND[366]. I therefore aimed to validate the ACMG criteria for MND-associated variants and to assess the potential benefits and limitations of adopting this classification system. The following work outlines methods and results of the development of a framework of recommendations for clinicians assessing novel and established variants in MND-associated genes. This summary aims to aid clinicians in difficult discussions with patients about genetic testing in MND.

This part of my PhD was undertaken in collaboration with clinicians and academics at Columbia University, New York. During my PhD, I undertook a training period at the Institute for Genomic Medicine, Columbia University as the Inaugural International Rowling Fellow for ALS/MND Genomics. The fellowship was under the supervision of Dr David Goldstein, Director of the Institute for Genomic Medicine and Professor of Genetics and Development, and Dr Matthew Harms, assistant professor of neurology in the Division of Neuromuscular Medicine.

Variant Sourcing

Based on clinical consensus, 10 genes were considered securely causative for MND: *ALS2, CHCHD10, FUS, OPTN, PFN1, SOD1, TARDBP, UBQLN2, VAPB* and *VCP*. Newly associated genes *TBK1, CCNF, NEK1* and *ANXA11* were also considered for exploratory post-hoc analysis. The *C9orf72* mutation was excluded as it is an intronic repeat expansion. For each of the 10 MND-associated genes, all reported MND-associated variants were downloaded from the

following three public databases in March 2016: Human Gene Mutation Database (HGMD)[346], National Center for Biotechnology Information (NCBI) ClinVar database[384] and the King's College London ALS Online Database (ALSoD)[72,335]. To ensure assembly compatibility, ClinVar variants were converted from GRCh38 to GRCh19 using the University of California, Santa Cruz (UCSC) Genome Bioinformatics Liftover Tool[385]. Datasets were merged using Galaxy Project tools[386]. Prediction data were assigned to each single nucleotide variant (SNV) using the SeattleSeq Annotation online tool[387] for PolyPhen, Grantham, GERP and CADD scores, Mutation Assessor[388] for Functional Impact (FI) and FI score, and PROVEAN (v1.1.3)[389] for SIFT and PROVEAN scores. Splice site impact was assessed using MaxEntScan and dbscSNV obtained from ANNOVAR[343,390] (Figure 17).

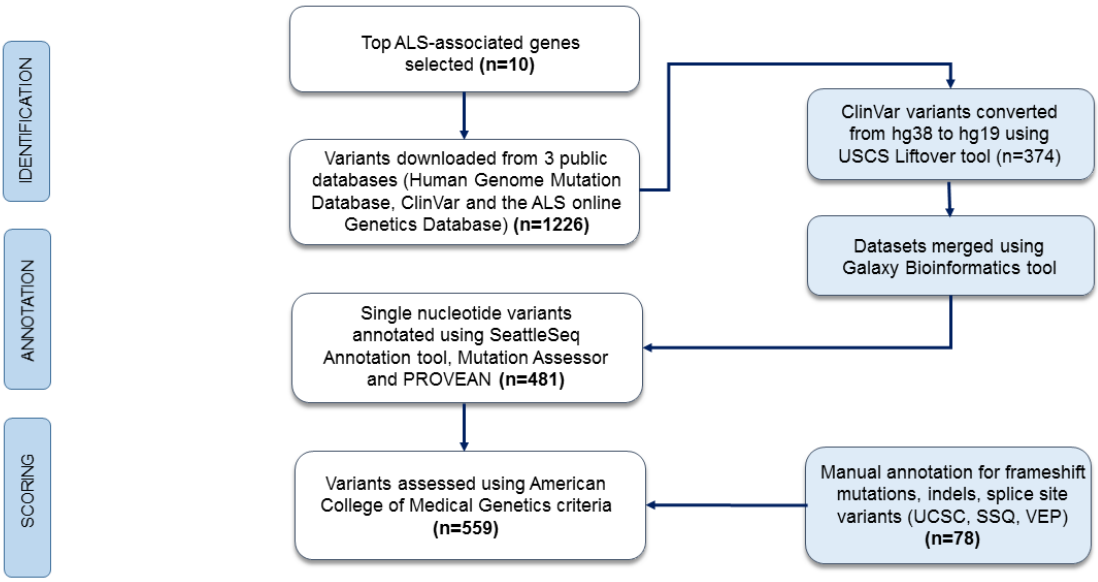


Figure 17 Variant sourcing methods

Variant Classification Methods

Each variant was systematically reviewed using the ACMG 28-point guidelines; criteria were combined to give a level of evidence (Pathogenic, Likely Pathogenic, Benign, Likely Benign, VUS). For each variant, publications associated with the variant were sourced and reviewed

(using HGMD PubMed ID and a subsequent literature search to ensure all publications captured). Details of methods for ACMG classification can be found in Appendix 6. A team of three clinician-scientist variant classifiers (the author being the primary rater) adopted a modified Delphi method[391,392], to develop a consensus approach to variant classification (Figure 18).

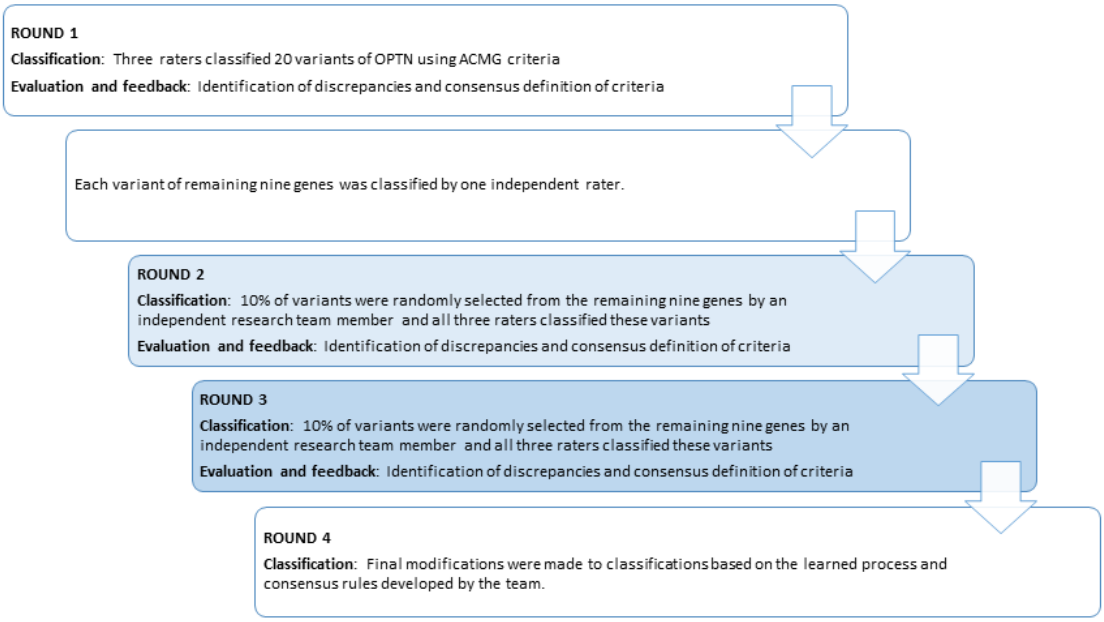


Figure 18 Modified Delphi method for training of rater classification

Post-Hoc Analyses

Genes associated with MND after initial study design (*TBK1, NEK1, CCNF, ANXA11*) were also evaluated. Variants were sourced as described above. No variants for *NEK1* and *ANXA11* were reported in the ALSoD or HGMD databases, leaving only case-reports with minimal evidence in ClinVar. They were therefore excluded from further analyses. Variant lists for the original 10 genes were also updated by sourcing newly reported variants in ALSoD and HGMD, making our analysis up-to-date as of August 2017.

Our methods generated many VUSs and so further analysis looked at whether the variant was trending towards pathogenic or benign despite not achieving requirements for ACMG classification. VUS with pathogenic potential were referred to as VUS-P variants; variants that trended towards being benign VUS-B. For a variant to be classified as a VUS-P, >1 criteria suggesting pathogenicity needed to be met, with the absence of any criteria suggesting that the variant was benign. The reverse was true for VUS-B variants. Variants meeting both pathogenic and benign criteria remained VUSs.

The ACMG methods use *in silico* prediction algorithms in classification of pathogenicity. We assessed each SNV variant using three methods: i) multiple prediction algorithms and a 'leave-one-out' approach, ii) meta-LR and iii) meta-SVM. Meta-LR and Meta-SVM are ensemble-based prediction approaches which integrate multiple scoring methods using logistic regression (LR) and support vector machine (SVM) models[393]. We aimed to test if we could simply our testing approach to use these predictions only, as a surrogate for reviewing multiple prediction algorithms individually.

Statistical Analyses

Rater classification data were collected using an online Google Form (https://goo.gl/forms/5RvB8B0HNVMH18lt1). Data were formatted and analysed using R version 3.4.3. Krippendorff's alpha (k-alpha) statistic (R package "irr") was used to assess formally inter-rater reliability; k-alpha score ranges from 0 (no concordance) to 1 (complete concordance) with good agreement considered ≥0.80[394]. In view of the small n-numbers of some variables, Fisher's exact tests were used for association testing.

Results

Table 20 summarises the number of variants sourced for each gene. Consensus on variant classification was initially assessed using 20 *OPTN* gene variants. Concordance between all three raters was achieved for 50% of the variants, with an error rate of 7.0% of all criteria assessed (28 criteria per variant). Mean k-alpha statistic for the 20 *OPTN* variants was 0.67

(95% CI 0.63, 0.71). Lowest k-alpha statistics resulted from the assessment of loss of function variants (mean k-alpha 0.56) and informed the next round of the Delphi procedures. Key disagreements concerned: interpretation of functional studies and degrees of cosegregation; having no pre-agreed cut-offs for allele frequencies of Pathogenic and Benign variants; and interpretation of multiple prediction and splice site algorithms. Consequently, we adopted rules for assessment of PS3, PP1, PP3 and BP4 (see Appendix 6); these rules reflect recognised classification difficulties[393,395,396].

On second round of the Delphi process, concordance rose to 63%, with an error rate of 6.7% and a mean k-alpha of 0.77 (95% CI 0.72, 0.82). Areas of concern included definition of a variant "hotspot" (PM1), and how best to determine if a gene which has a low rate of benign missense variants (PP2). This led to the refinement of these classification rules (Appendix 6). On third round, the concordance was 84%, with an error rate of 3.0% and a mean k-alpha of 0.91 (95% CI 0.87, 0.95). In a recent review of use of the ACMG criteria among nine laboratories, average intra-laboratory k-alpha was 0.91[397]. As our third-round k-alpha was ≥0.80 and compatible with clinical sequencing laboratory agreements, the Delphi procedure was considered complete.

Gene	*ALS2*	*CCNF*	*CHCHD10*	*FUS*	*OPTN*	*PFN1*	*SOD1*	*TARDBP*	*TBK1*	*UBQLN2*	*VAPB*	*VCP*
Number of variants	**60**	**10**	**18**	**76**	**45**	**8**	**190**	**57**	**96**	**26**	**6**	**25**
% of all MND-associated variants identified	**10**	**2**	**3**	**12**	**7**	**1**	**31**	**9**	**16**	**4**	**1**	**4**

Table 20 Number of MND/MND-spectrum disorder variants sourced from three databases (HGMD, ClinVar and ALSoD) (excluding synonymous variants)

All remaining variants were reviewed by an individual rater and assigned an ACMG classification. Results are presented in Figure 19. Only 19% of all variants met criteria for Pathogenic/Likely Pathogenic. Seventy eight percent of all variants were VUS. These VUS were further examined for direction of pathogenicity, to identify VUS-P and VUS-B (Figure 20). Twenty three percent of variants remained unclassifiable.

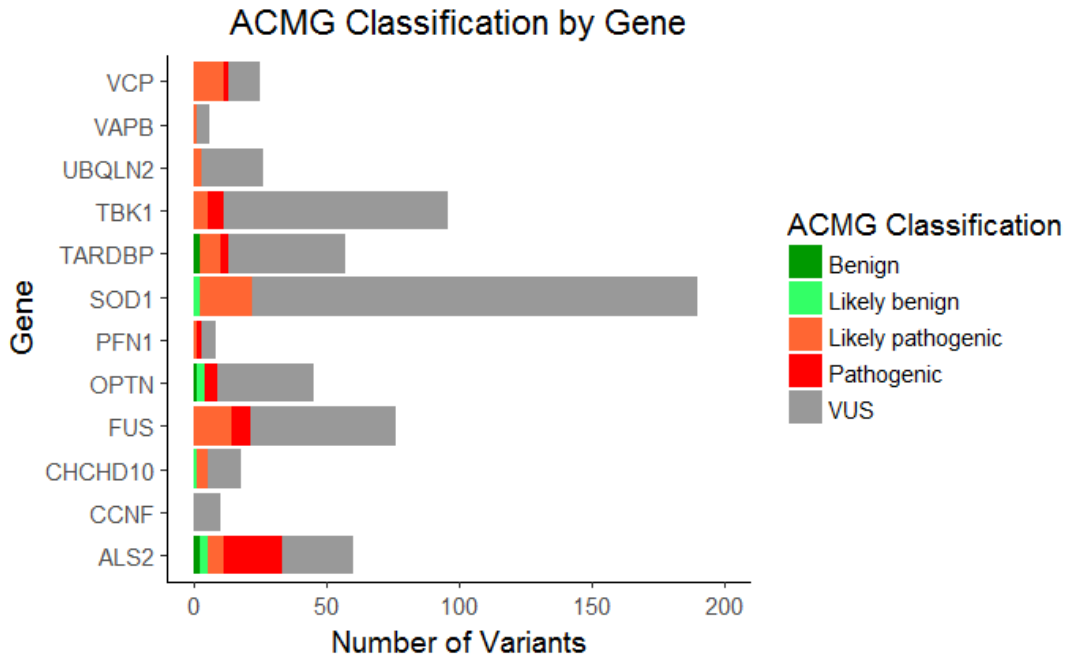


Figure 19 ACMG classifications by gene

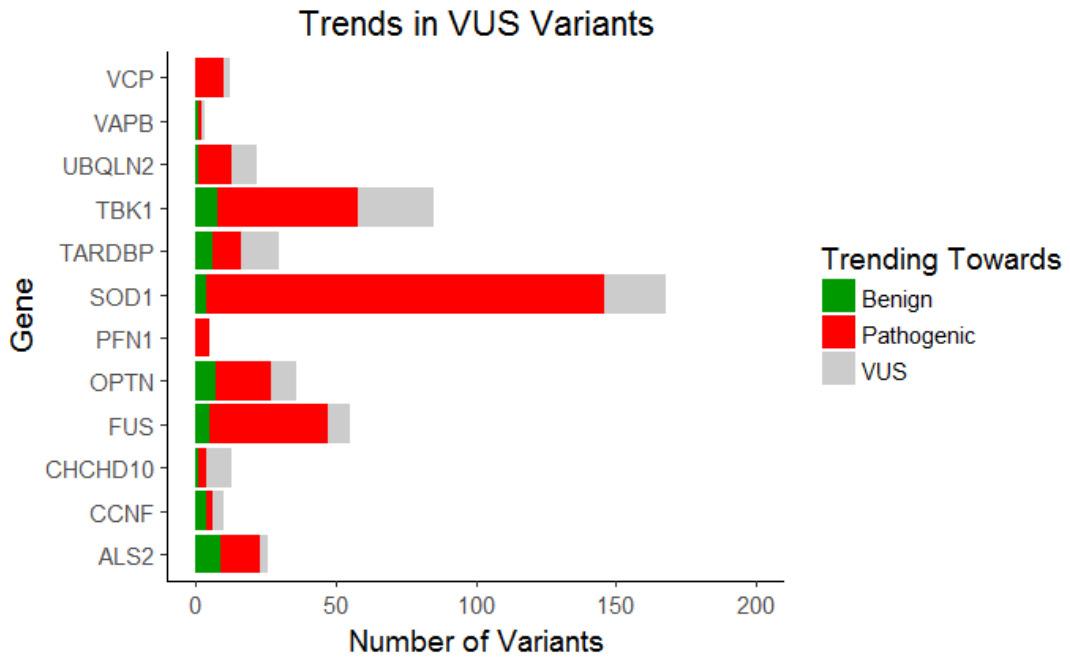
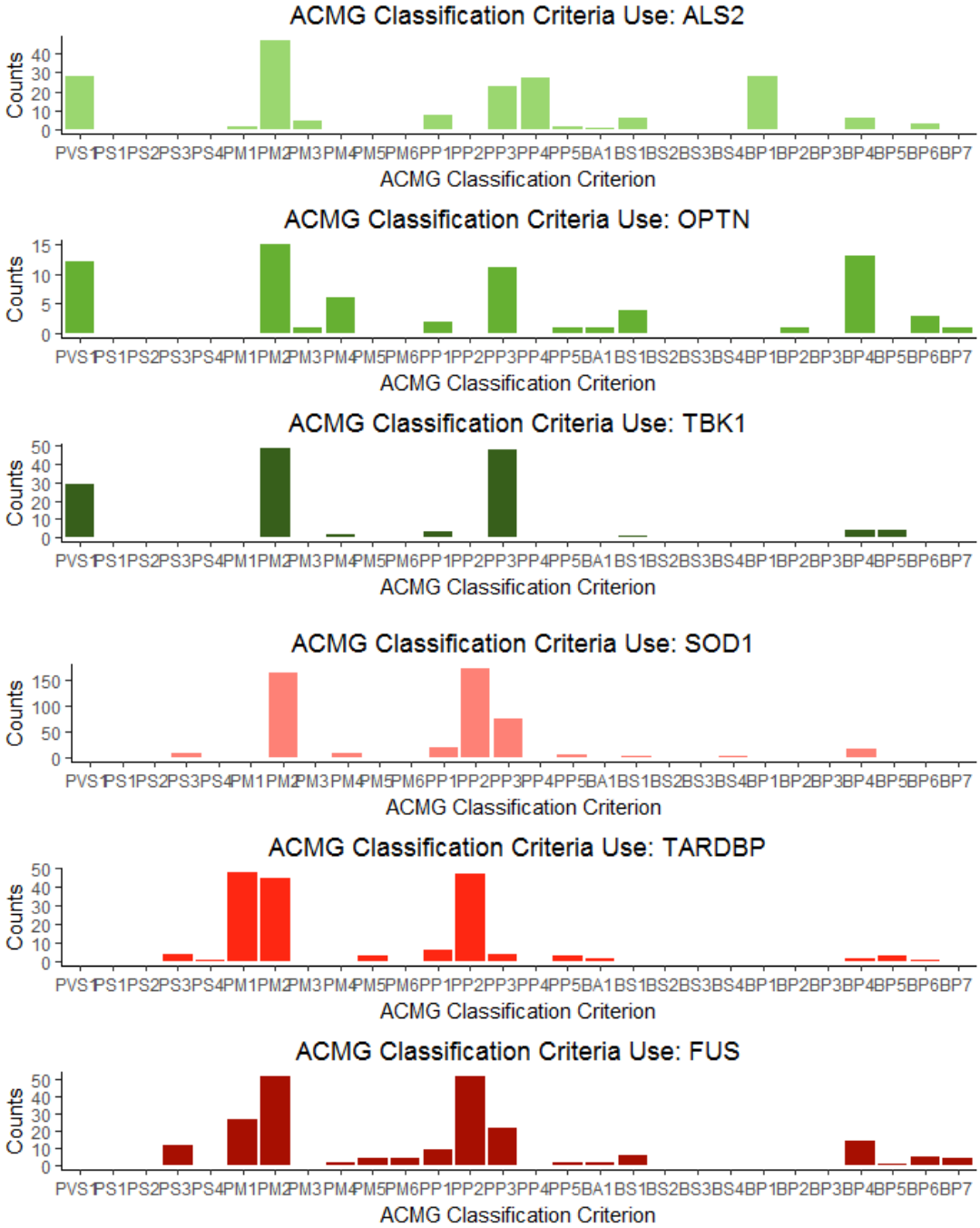


Figure 20 Direction of 'trend' of variant

We plotted frequency of use of individual ACMG criteria by gene (Figure 21). The top four classification criteria (used >100 times) were i) PM2 (absence from ExAC reference database), ii) PP2 (novel missense variant), iii) PP3 (computational evidence of pathogenicity) and iv) PM1 (variant located in a mutational hotspot). In contrast PS1 (amino acid change seen in a previously reported pathogenic variant with a different nucleotide change), BS2 (variant observed in a healthy individual in a fully penetrant disease), BS3 (functional studies showing no effect of protein function) and BP3 (in-frame indel in a repetitive region with known function) were not used for any of our MND-associated variants. The plots (Figure 21) were used to identify gene-specific features.


ACMG Classification Criteria Use: ALS2
Counts
ACMG Classification Criterion
ACMG Classification Criteria Use: OPTN
Counts
ACMG Classification Criterion
ACMG Classification Criteria Use: TBK1
Counts
ACMG Classification Criterion
ACMG Classification Criteria Use: SOD1
Counts
ACMG Classification Criterion
ACMG Classification Criteria Use: TARDBP
Counts
ACMG Classification Criterion
ACMG Classification Criteria Use: FUS
Counts
ACMG Classification Criterion

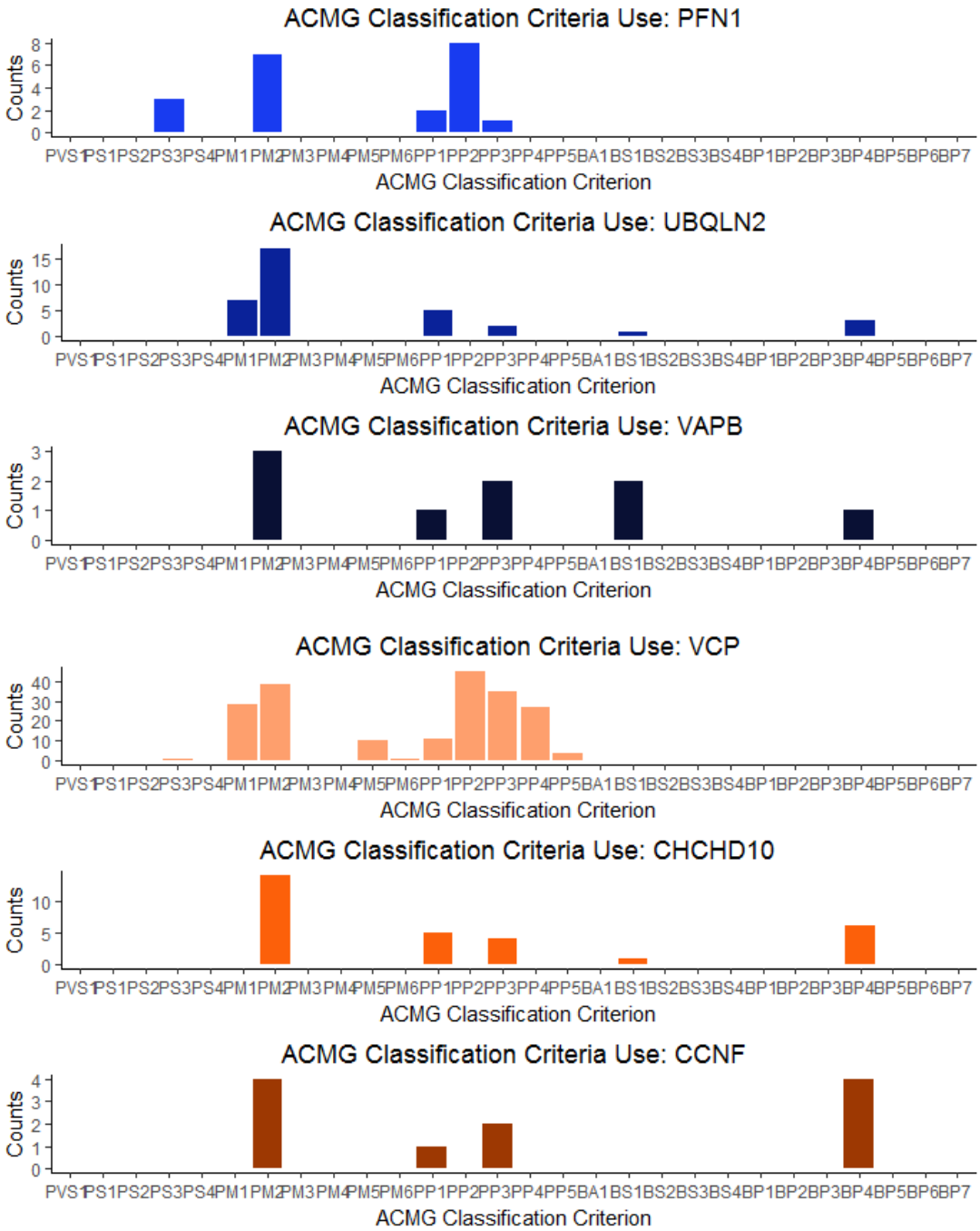


Figure 21 ACMG classification criteria use by gene; green genes = loss of function genes; red genes = missense genes; blue genes = rare genes; orange genes = unusual/new MND associated genes

Finally, our prediction algorithm assessment methods were analysed. PP3 and BP4 classifications were compared against meta-LR and meta-SVM values for each SNV. Fisher's exact tests showed strong association between meta-methods and our methods for assessing PP3 and BP4: meta-LR compared more favourably than meta-SVM (meta-LR p=0.0004 (PP3) and p=<0.00001 (BP4); meta-SVM p=0.017 (PP3), and p<0.00001 (BP4).

Review of Methods

Our results present a comprehensive and systematic review of the ACMG classification system for MND-associated variants. Using multiple raters and stages of 'learning' using a modified Delphi approach, we identified key challenges in assessment. One particular area was assessment of cosegregation (PP1). Interpretation of inheritability of variants is fraught with subjectivity. Recommendations in original ACMG recommendations are unclear[366]. However, a follow-up report suggested methods for calculating probability of inheritance of a variant occurring due to chance, which acts as a surrogate for likelihood of cosegregation[396]. The authors suggested cut-offs for level of evidence (supporting, moderate and strong) using a simplified Bayes Factor method. We found these calculations of probability easy to apply to our process and removed subjectivity. Difficult pedigrees were discussed between raters to ensure accurate calculation.

Another problematic area was interpretation of *in silico* prediction algorithms (PP3, BP4). We included nine algorithms (including splice site algorithms). There are no consensus cut-offs for these predictions. We use a combination of suggested cut-offs[393] and clinical experience to generate relatively strict thresholds aiming to capture true pathogenic and benign variants for the purposes of clinical practice. Assessment of seven algorithms for SNVs is time-intensive and might be a barrier to clinician interpretation. The meta-LR score, however, is easy to interpret (output: damaging or tolerant) and was strongly associated with our assessments.

We encountered potential difficulty in interpretation of *PM5: Novel missense change at the same site as a different pathogenic amino acid missense change.* This criterion can only be met if a corresponding variant is Pathogenic. The process relies on harmonized classification

of all variants at this site. Our methods made this possible but it is an additional consideration classification of single patient results.

Disease-Specific Review

We propose that the ACMG system is appropriate for use in MND. Specific recommendations are summarised in Table 21.

Variant Feature	Corresponding ACMG criteria	Considerations
Animals models of disease	PS3	Reliable models include rodent models that recapitulate disease (motor or cognitive symptoms, impact on survival, or proven loss of motor neurones pathologically)
Co-segregation of variant with disease	PP1	Extended and well-documented pedigrees are helpful Methods proposed by Jarvik and Browning[396] correctly identify familial clustering of MND variants Consider DNA banking in proband and relevant family members to allow future co-segregation studies
Prediction algorithms	PP3, BP4	Meta-LR can be used a simple and easily interpretable surrogate for variant pathogenicity instead of using multiple algorithms with subjective cut-offs
Specificity of disease phenotype	PP4	Relevant for mutations in: VCP - inclusion body myositis, frontotemporal dementia and/or Paget's disease ALS2 - infantile ascending hereditary spastic paraplegia (IAHSP).
Reputable source recently reports variant as pathogenic/benign	PP5, BP6	ClinVar reports acceptable but only if sample if from clinical testing independent to literature evidence and which reflects recent evidence-base

Table 21 Disease specific framework

A potential criticism of this work is that ACMG guidelines are suitable only for Mendelian disorders, and are ineffective for diseases with incomplete variant penetrance, such as MND. We observed that extensive pedigrees provided strong evidence for cosegregation of variants, thereby identifying "familial clustering" typically seen in MND[26]. However, well-documented pedigrees are relatively rare and are additionally confounded by late onset disease, absence of living relatives and, in view of the terminal nature of disease, variable uptake of genotyping by extended family members. Interestingly, only one *SOD1* variant (p.H80R) met criteria for *PS2: De novo variant (both maternity and paternity confirmed) in patient with the disease and no family history,* suggesting that there is a notable absence of trio studies in apparently sporadic MND cases. Lack of segregation (BS4) could not be applied in this disease in view of incomplete penetrance. In view of these challenges, we propose that DNA banking should be explored in all pwMND and relevant family members to facilitate future testing should disease emerge in later life[169].

We sourced reliable reviews of animal models of MND (PS3)[395,398,399]. However, many models in MND fail to recapitulate key disease markers such as motor symptoms[26]. To detect *true* clinically relevant disease-causing variants, we limited our use of this criterion to rodent models which resulted in an MND-spectrum disease phenotype (total use of criterion = 29).

Specificity of disease phenotype (PP4) is challenging for a heterogeneous spectrum-disorder such as MND. We only determined this relevant for patients with *VCP* mutations who had typical disease syndrome features (inclusion body myositis, FTD and/or Paget's disease) and for juvenile patients with *ALS2* mutations who had features of infantile ascending hereditary spastic paraplegia (IAHSP).

Finally, we agreed that classification of a variant by a clinically reputable laboratory that was reported in ClinVar was sufficient evidence for PP5/BP6. Often classification methods are not described for ClinVar reports and so there is likely to be great variability in interpretation of pathogenicity. However, this criterion provides supporting evidence only and so ACMG methods, appropriately, do not rely heavily on external or prior classification. Timing of report is crucial as some predate recent evidence, particularly for long-recognised genes such as *SOD1*. Further, we identified that it was important to examine the source of the report,

as some were simply duplications of literature that had been used as evidence for other ACMG criteria.

Gene-Specific Review

The following is summarised in Table 22.

ALS2

Many MND-associated mutations in this gene result in a premature stop codon in the transcript, creating a truncated protein (48% of all reported variants and 100% of Pathogenic/Likely Pathogenic variants). Missense variants can therefore be considered as probably benign (BP1). *ALS2* mutations are also highly specific for development of autosomal recessive forms of three early-onset neurodegenerative diseases: juvenile ALS (JALS), juvenile primary lateral sclerosis (JPLS), and infantile ascending hereditary spastic paraplegia (IAHSP). In view of the young-onset of disease, this is the only gene for which BS2 (disease is fully penetrant at an early age) can be considered. Homozygous and compound heterozygous variants are more common in this gene than other MND-associated genes. The ACMG criteria do not sufficiently take into consideration homozygous variants and so variants in this gene failed to meet pathogenicity in spite of apparently convincing evidence.

CCNF

CCNF is a gene newly associated with MND: all reported variants are from the same study[400]. ACMG evidence was limited and consequently all variants were VUS. Multiple prediction algorithm agreement was met. Accounting for this, however, only two variants trended towards pathogenicity.

CHCHD10

Many variants in *CHCHD10* were VUS (72%) and remained VUS on further assessment of trend. Explanation of this includes the gene's recent association with MND and absence of functional studies.

FUS

Thirty six percent of MND-associated variants in the *FUS* gene were found at the nuclear localization sequence in exon 15 which is considered to be a hotspot[401]. There is strong evidence for *in vitro* modelling of cytoplasmic localization/nuclear exclusion of *FUS* variants and these were considered exceptional evidence for criterion PS3[79,401,402].

OPTN

OPTN is associated with both MND and primary open angle glaucoma (POAG). Many reported variants in the public databases are in fact associated with POAG, leaving only 12% of Pathogenic/Likely Pathogenic variants in MND-associated variants. The majority of reported MND-associated variants are SNVs (65%) but all of the Pathogenic/Likely Pathogenic variants are loss-of-function variants, suggesting that this is the true mechanism of disease.

PFN1

In the context of a relatively high Z-score (2.97), all reported variants in *PFN1* were SNVs. Two variants in particular (p.C71G and p.M114T) met pathogenicity due to recent robust mouse models and strong cosegregation data[403,404].

SOD1

Since its association with ALS in 1993[65,66], 190 variants have been reported. Variants are found throughout the gene and so there are no hotspots, although there are multiple variants within the same domain eg. Cys6. Many of the earlier reports of rare variants are limited by lack of sufficient information regarding pedigrees, paternity or maternity. Further, some novel mutations have not been replicated. Several *SOD1* animals models exist but are difficult to interpret – for example the p.A4V variant causes early and rapid onset disease in humans but late onset disease in mice[399]. In spite of these indeterminate factors, prediction algorithms tend to favour pathogenicity: 86% of SNVs, with meta-LR and meta-SVM values “Damaging” for all. In light of this conflicting evidence, no *SOD1* variants met criteria for pathogenicity using ACMG guidelines (11% are Likely Pathogenic). Eighty eight percent of variants are VUS; of these 86% of these trend towards pathogenicity.

TARDBP

The majority (84%) of MND-associated *TARDBP* mutations are in exon six, which is considered a mutational hotspot[405,406]. *TARDBP* has a high Z-score (4.33): 88% of reported variants are SNVs (including all Pathogenic/Likely Pathogenic variants). Prediction algorithms are problematic in *TARDBP* – most algorithms trend towards benign, but the gene is highly conserved. Meta predictions are equally difficult to interpret with meta-LR and meta-SVM in disagreement in 12% of SNVs. Our recommendation would therefore be to rely on other components of the guideline over *in silico* modelling.

TBK1

Many *TBK1* variants have been associated with MND and ALS-FTD since 2016 [159]. Thirty percent of variants are loss-of-function. A significant proportion of missense variants (75%) trended towards pathogenicity, in spite of not fulfilling full ACMG criteria. The potential for disease-causation of this gene is therefore apparent.

UBQLN2

Mutations in *UBQLN2* are found in patients with dominant X-linked MND, and MND with FTD. The PXX domain is considered a hotspot[407]. No *UBQLN2* variants met criteria for pathogenicity, primarily due to conflicting evidence regarding allele frequency and prediction algorithms. Three variants with strong cosegregation evidence met criteria for likely pathogenicity (p.P497H, p.P497S and p.T487I). Stronger literature evidence is needed to classify variants in this gene.

VAPB

Only six MND-associated variants were identified in *VAPB*. It is a small and tolerant gene and evidence is limited for classification – individual ACMG criteria were met on only nine occasions. Based on strong segregation and prediction data, the p.P56S variant was Likely Pathogenic. Further sequencing of this gene in MND populations is required to determine disease-association.

VCP

The *VCP* gene had the highest frequency of Pathogenic/Likely Pathogenic variants (56% of all variants identified). This is consequent on mutational hotspots (exons 3-5 and the N-terminal

domain[408], a defined syndromic phenotype and well-documented extended pedigrees. It is an intolerant gene, with multiple concordance of prediction algorithm data. One caution in interpretation is that several papers report novel variants, when in fact the variant is a new missense change at the same site as a previously reported amino acid change. Systematic use of ACMG guidelines (specifically, PM5) helps to make sense of these variants.

Gene	Key ACMG criteria	Considerations
ALS2	PVS1, BS2, BP1	Pathogenic variants results in a truncated protein. While variant penetrance can be variable, young-onset disease is best characterised. Missense variants are probably benign.
CCNF		All reported variants are VUS. All variants are from the same study – replication required.
CHCHD10		Paucity of evidence and functional studies.
FUS	PS3, PM1	Evidence for *in vitro* modelling of cytoplasmic localization/nuclear exclusion of FUS variants. 1/3 of variants are found at the nuclear localization sequence in exon 15.
OPTN	PVS1	Pathogenic variants tend to be loss-of-function but caution as many missense variants are reported. Caution regarding evidence as variants associated with alternative phenotypes (POAG).
PFN1	PS3, PP1, PP2	Few variants reported but all are missense. Good emerging evidence for cosegregation and gene function.
SOD1	PS3, PP3, PP5	Caution regarding animal models which do not recapitulate disease. Caution regarding historically reported variants with little robust evidence which have not been replicated. Prediction algorithms favour pathogenicity.
TARDBP	PM1, PP2, PP3	Hotspot in exon 6. Majority of variants are missense.

		Prediction algorithms are conflicting.
TBK1	PVS1, PM5	Loss-of-function and missense variants are potentially disease-causing.
UBQLN2	PM1, PP3	PXX domain is a hotspot. Prediction algorithms are conflicting.
VABP		Few MND-associated variants in this gene – insufficient to make recommendations.
VCP	PM1, PM5, PP1, PP4	Mutational hotspots (exons 3-5 and the N-terminal domain). Novel missense variants at the same site as a previously reported amino acid change. Strong cosegregation evidence. Phenotype highly specific.

Table 22 Gene-specific framework

Variant Classification – Application

This framework was applied to classify incident samples collected in Scotland 2015-2017. To streamline methods, VarSeq Golden Helix software was used for variant filtering and annotation of variant call filter (VCF) files[409]. Filters were selected based on the learned framework and are summarised in Table 23. Variants were annotated using multiple *in silico* prediction algorithms from the Database for Nonsynonymous SNPs and their Functional Predictions (dbNSFP)[410] and included: SIFT, PolyPhen2 HDIV and HVAR, Mutation Taster, Mutation Assessor, FATHMM, PROVEAN, GERP and PhastCons. MetaLR and MetSVM were also sourced as a simplified approach to classification. However, individual algorithms were reviewed in addition to ensure that they did not differ significantly from Meta classification. For mutations where prediction information was not available (all except SNVs, but particularly intronic variants with missing splice site predictions), variants were reviewed using the Variant Effect Predictor (VEP) tool[411]. Measures of impact on splice site were also extracted from VarSeq and included scores derived from adaptive boost (Ada) and random forest (RF) models[412].

Variant interpretation procedures and cut-offs for each ACMG classification point are described in Appendix 6. For this analysis, the more up-to-date population reference database gnomAD (Genome Aggregation Database)[413] was used instead of the Exome Aggregation Consortium (ExAC)[352]. Probability of loss of function (pLI) scores and Z-scores for missense variants for each gene were extracted from gnomAD.

The control population for this study was the Lothian Birth Cohort of 1921 and 1936, comprising of 1383 individuals. As per ACMG guidelines, this is considered an appropriate population in view of its being i) a large cohort (>1000 individuals) and ii) race-matched[366].

As outlined in the ACMG-guided framework, the proportion of MND-associated variants being classified as Pathogenic or Likely Pathogenic are low. However, some variants are "hot" VUS, or VUS with pathogenic potential (VUS-P). Instead of making assumptions about their pathogenicity (ie. assuming they are Likely Pathogenic or rejecting them as unqualified VUS), these will be discussed separately in the results.

Filter	VarSeq Source	Inclusions	Exclusions
Zygosity	VarSeq genotype zygosity algorithm	Heterozygous Homozygous	Reference
Sequence ontology	RefSeq Genes 105 Intermin v1, NCBI	3’ UTR variant 5’ UTR premature start codon gain variant 5’ UTR variant Frameshift variant Inframe deletion Inframe insertion Intergenic variant Intron variant Missense variant Splice acceptor variant Splice donor variant Splice region variant Stop gained variant	Synonymous variants
Effect	RefSeq Genes 105 Intermin v1, NCBI	Loss of function Missense Other	
Alternate Allele Frequency	gnomAD Genomes Variant Frequencies 2.0.1v3, BROAD	<0.01 =0.01 Missing	>0.01
Clinical Significance	ClinVar 2017-09-05, NCBI	Pathogenic Likely Pathogenic Conflicting Interpretations of Pathogenicity	Benign Benign/Likely Benign Likely Benign

		Uncertain Significance Not Provided Risk Factor Other Missing	
Functional Predictions	dsNSFP Functional Prediction Voting	0/6 Predicted as Tolerated 1/6 Predicted as Tolerated 2/6 Predicted as Tolerated 2/6 Predicted as Damaging 3/6 Predicted as Damaging 4/6 Predicted as Damaging 5/6 Predicted as Damaging 6/6 Predicted as Damaging Missing	3/6 Predicted as Tolerated 4/6 Predicted as Tolerated 5/6 Predicted as Tolerated 6/6 Predicted as Tolerated 0/6 Predicted as Damaging 1/6 Predicted as Damaging
Meta LR Meta SVM	dbNSFP Functional Predictions and Scores, 3.0, GHI	Damaging Missing	Tolerated
Ada Score	dbscSNV Splice Altering Predictions 1.1, GHI	>0.6 = 0.6 Missing	<0.6
RF Score	dbscSNV Splice Altering Predictions 1.1, GHI	>0.6 =0.6 Missing	<0.6

Table 23 Variant Filtering Inclusion and Exclusion Criteria

6.3 RESULTS

The number of samples donated by incident pwMND 2015-2017 was 339; this is representative of 54.8% of the incident MND cohort 2015-17. All samples underwent gene panel sequencing and were screened for the *C9orf72* hexanucleotide expansion.

C9orf72

Sequencing for the *C9orf72* expansion identified 29/339 (8.6%) individuals with >30 expansions. Of these, one patient had an unusual intermediate-length expansion (70 repeats). One further patient had 28 expansions which was initially considered abnormal but of indeterminate significance. Examples of unexpanded repeats (two and eight repeats, a typical combination in the Scottish population[215]) and a typical repeat prime expansion trace are illustrated in Figure 22.

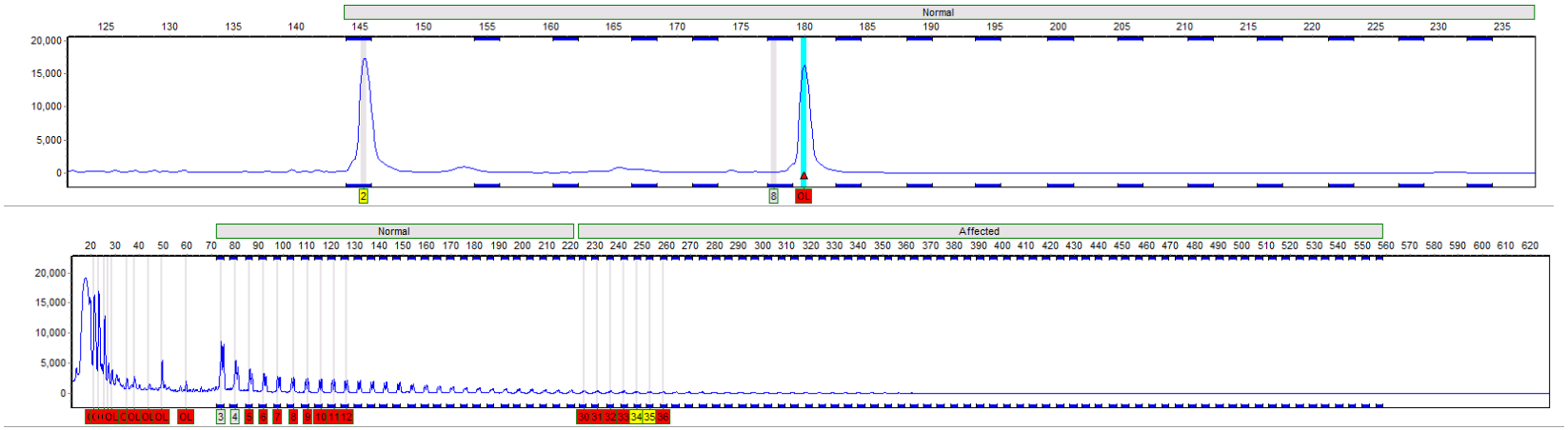


Figure 22 Top figure: a normal, unexpanded C9orf72 trace with 2 GGGGCC repeats on 1 allele, 8 on the other; Bottom Figure: an abnormal C9orf72 trace with 2 GGGGCC repeats on 1 allele, and expansion >30 repeats on the other (note difference in scales)

Panel Sequencing: MND-Associated Genes

On gene panel sequencing, depth of coverage (≥ 20X) was, on average, 98% across the regions of interest. After VarSeq variant filtering and ACMG-based classification, 503 variants were identified in 339 samples; the median number of variants per sample was one (range 0-5). Post-filtering variants were identified in 277/339 (81.7%) of samples; 103 individuals (103/339, 30.4%) had a Pathogenic or Likely Pathogenic variant or VUS. Fifty three (53/339, 15.6%) individuals had a Pathogenic or Likely Pathogenic variant or a potentially pathogenic VUS (VUS-P); of these 5/339 (1.5%) had two variants of interest. The variants identified in this cohort are presented in Table 24. However, some of these variants were in genes which typically have a recessive inheritance in disease-causing states (*ALS2*, *BSCL2*, *HTRA1* and *SPG11*); the significance of these variants is therefore significantly diminished when they are present on one allele only. Excluding these variants, 47/339 (13.9%) of individuals had a potential genetic explanation for their disease. *The phenotypes of variant carriers will be discussed in Chapter 7.*

SOD1

After the *C9orf72* expansion, the most common variant in this MND cohort was the *SOD1* p.I114T variant (n=9), previously described as a founder mutation in the Scottish population (Chapter 5). Three further *SOD1* variants were identified. One of these had previously been seen in our pilot study (p.A146D; Chapter 5) – using the newer classification system this is a VUS-P but we previously described it as Likely Pathogenic. In view of its recurrence in the Scottish MND population and the absence of the variant in control populations (as well as its being absent in gnomAD) the pathogenic potential of this variant is strong. Two further *SOD1* missense variants were found in cases: the p.Q23H variant was first described in a Japanese patients with ALS[414], the p.G73C described in a patient with an usual ALS phenotype mimicking a myopathy[415]. These variants have not previously been identified in the Scottish population.

FUS, OPTN, TBK1 and VCP

Variants were also identified in *FUS*, *OPTN* and *TBK1*, genes that were previously included in our pilot gene panel (Chapter 5). *FUS* variants included a missense variant and a loss of function frameshift variant, both on exon 14. The missense variant (p.R485W) has been reported in a sporadic ALS patient, has a low MAF in gnomAD and most prediction algorithms suggest pathogenicity (including meta-SVM and meta-LR)[96]. The loss of function variant (p.Y479Mfs*50) is a novel variant in a genomic location near to previously described MND-associated frameshift mutations. Although the impact on protein function is unclear due to the absence of functional studies, this variant is potentially causative of disease.

Only one *OPTN* variant was seen in this cohort (p.V572M). This variant was absent from controls, had a very low MAF in gnomAD and 8/9 *in silico* predictions suggested pathogenicity (meta-SVM and meta-LR damaging). However, this is a missense variant in a gene in which truncating variants primarily cause disease (Chapter 6.2) and so its pathogenic potential is attenuated. Similarly, there was one variant found in the *TBK1* gene (p.Q565P) in exon 15. This novel variant has a low MAF in gnomAD and is in a conserved region but remains a missense variant in a gene for which the mechanism of disease is unclear, but thought to be loss of function[416].

Another confirmed MND-associated gene is *VCP* and, indeed, a pathogenic variant in this gene was identified in one case (p.R155C). This missense mutation, located at a mutation hotspot on exon 5, has been previously described and is reported in clinical cases on Clinvar[384,417,418]. All *in silico* predictions suggest pathogenicity.

NEK1, CCNF and ANXA11

Variants were identified in genes newly associated with MND: *NEK1*, *CCNF* and *ANXA11*[89,400,419]. Two individuals were found to have loss of function variants in exon 21 of *NEK1*, which were classified as Pathogenic (p.E634Kfs*11). As far as we can determine from patient histories, they were unrelated. Four further *NEK1* variants were identified, three of which were missense variants (p.F135V, p.R714C, p.L106S). All these missense

variants were absent in controls, had low MAF or were absent from gnomAD and were located in conserved regions with multiple prediction algorithms supportive of pathogenicity. The final *NEK1* variant was a loss of function splice donor variant associated with exon 10, which was present in two cases and zero controls. The variant had a very low MAF in gnomAD and splice impact scores indicated a significant impact on protein function (Ada score 0.999, RF score 0.94).

Both variants identified in the *CCNF* gene were novel missense variants (p.G161R and p.D51N), absent in controls and gnomAD and in conserved regions. Information regarding this gene's role in MND is emerging, but a recent study suggests that a single missense mutation can adversely affect the autophagy degradation pathway which is implicated in MND pathogenesis[420].

One mutation in the *ANXA11* gene (p.L337H) was considered to have pathogenic potential due to a low MAF, some supportive *in silico* prediction evidence, and being located in a conserved region. Patients with *ANXA11* mutations have been found to have A11-positive protein aggregation in spinal cord motor neurones[419]. Similar missense variants have been more recently described in Chinese populations[421,422]. Seven other variants were identified in this gene but these were considered VUS or VUS-B. These comprised an intronic variant found in four samples (c.1029+13C>G), an intronic variant in one sample and one control (c.561+14C>T), a splice site variant with low impact found in two samples (c.745-7C>G), a missense variant with a MAF of 0.002 found in 14 controls (p.T244M), another missense variant in four samples and 31 controls (p.P8L; European non-Finnish MAF 0.009), a missense variant in two samples and 12 controls (p.E369L; European non-Finnish MAF 0.006) and a missense variant with a MAF of 0.01 in the South Asian population (p.R452W). Overall, there were 16/339 (4.7%) *ANXA11* variants in cases and 58 /1383 (4.2%) controls.

Rare MND-syndrome Genes: DAO, DCTN1, ERBB4, MATR3, NEFH, PRPH, SETX, SQSTM1

A variant in the *DAO* gene (p.R115W) was absent from controls and all prediction algorithms suggested pathogenicity. This variant has not previously been described in relation to MND. Novel MND-associated variants are emerging in the literature but it is the p.R199W variant

which has been more extensively studied in disease models[423]. A variant at the same site but with a different amino acid change - p.R199Q - was seen in our MND cohort and was classified as a likely benign VUS because of its relative high population MAF (0.003 in South Asian and 0.002 in Ashkenazi Jewish populations). Prediction indicators were conflicting and the variant was present in two control samples. This variant has been described in a previous cohort of patients of European origin[378].

Three mutations were found in the *DCTN1* gene. This gene is associated with Perry syndrome, characterised by parkinsonism, depression and central respiratory hypoventilation, but is also rarely associated with neurodegenerative diseases such as MND, FTD and Parkinson's disease[424]. The previously reported variants do not segregate with disease; due to this and its rarity in MND populations, *DCTN1* remains a tentative gene with pathogenic potential only.

One variant in the *ERBB4* gene was identified in cases (p.I892V), with a low MAF and supportive *in silico* predictions. Z score for this gene is 3.25. Few MND-associated *ERBB4* mutations have been described but recent immunohistochemistry studies of spinal cord pathology samples from MND patients suggest that activity of ErbB4 correlates with TDP-43 mislocalisation[425,426].

The *MATR3* gene interacts with TDP-43; heterozygous missense variants are thought to be a rare cause of both familial and apparently sporadic pwMND of European origin[427–429]. Z score of this gene is 2.73. One individual carried a missense variant (p.P776S) of pathogenic potential – although this has not been reported in MND previously, it was absent from controls and had a low MAF.

A novel frameshift loss of function variant was identified in the *NEFH* gene (p.F15Vfs*83); this variant was absent from gnomAD and controls. However, the *NEFH* gene is tolerant of mutations and several VUS are reported[378]. *NEFH* mutations associated with increased MND risk are typically deletions affecting the tail end of the protein[430].

In contrast, the *PRPH* gene encodes an intermediate filament protein involved in peripheral nerve integrity. A missense variant (p.R435W) was identified in this gene but in the second

last exon. While in silico predictions were supportive, the impact of this variant on protein function remains uncertain.

Dominant missense variants which segregate in families with MND have been described in the *SETX* gene and cause juvenile onset ALS[431]; however, this gene is normally associated with spinal cerebellar ataxia with axonal neuropathy. A novel missense variant in this cohort (p.L1111W) is potentially pathogenic. However, an observed intronic variant (c.6843-5delT) is most certainly a single nucleotide polymorphisms (SNPs) (Table 26). Recent literature suggests that many previously described *SETX* are indeed SNPs/benign[162].

SQSTM1 missense variants are described in familial and apparently sporadic MND[432], as well as in FTD without MND. The first variant observed (p.P146L) is rare and all predictive algorithms suggested pathogenicity. The second (p.R267H) was seen in two of our MND cases and zero controls and has been described in an MND case previously as a VUS[433].

FTD Genes: GRN, MAPT

Mutations in two genes which are normally associated with FTD were seen in our panel screen. In spite of the phenotypic overlap, *GRN* mutations are almost exclusively seen in people with FTD, rather than MND[434]. Missense variants, as seen in this cohort, can be pathogenic. Our two observed variants are close in location, near to other previously described variants in gnomAD (p.C475R; p.R478C) and have low MAFs and supportive prediction algorithms. Similarly, missense mutations in *MAPT* are found in people with FTD. The variant in this cohort (p.G107V) has previously been described in a patient with FTD[435].

Other Dementia Genes: APP, CSF1R, NOTCH3, PSEN1, PSEN2

Several genes normally associated with other cognitive phenotypic profiles were identified in this cohort. One missense variant in the amyloid precursor protein (APP) gene was seen (p.D219V) – this variant was found on exon 5 in a conserved region and all bar one prediction predictions predicted pathogenicity, although meta-LR and meta-SVM scores did not. The

APP gene is causative in early-onset Alzheimer's disease (<65 years onset). A missense mutation was also observed in the *CSF1R* gene – missense variants in this gene are normally associated with hereditary diffuse leukoencephalopathy with spheroids. Presentations of gene carriers can include phenotypes similar to FTD[436]. Several variants were seen in *NOTCH3*, the gene associated with cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). These were all missense variants at various locations throughout the gene that were absent from controls and had low population MAFs. The Z-score for *NOTCH3* is 3.53 implying a low rate of benign missense mutations.

PSEN1 and *PSEN2* mutations are seen in patients with early-onset Alzheimer's disease. The two variants observed in this cohort (*PSEN1* p.S132A; *PSEN2* p.T128I) have low MAFs and supportive *in silico* markers. The latter was present in controls (n=1) but it is not possible to know the cognitive status of control individuals.

Digenic Cases

Two samples had *C9orf72* expansions plus another variant of interest. A further five samples harboured two VUS-Ps. These are described in Table 25. Overall, digenic cases comprised 7/339 (2.1%) of the MND cohort.

Gene	Genomic Position	Variant DNA Change	Variant Protein Change	Variant Type	Pathogenic Classification	Number cases	Number controls
Pathogenic and Likely Pathogenic Variants							
SOD1	21:33039672T>C	c.341T>C	p.Ile114Thr	Missense	Pathogenic	9	0
NEK1	4:170428210TC>-	c.1900_1901delGA	p.Glu634Lysfs*11	LoF frameshift	Pathogenic	2	0
VCP	9:35065361G>A	c.463C>T	p.Arg155Cys	Missense	Pathogenic	1	0
Variants of Uncertain Significance with Pathogenic Potential (VUS-P)							
ANXA11	10:81923309A>T	c.1010T>A	p.Leu337His	Missense	VUS-P	1	0
APP	21:27423322T>A	c.656A>T	p.Asp219Val	Missense	VUS-P	1	0
CCNF	16:2487264G>A	c.481G>A	p.Gly161Arg	Missense	VUS-P	1	0
CCNF	16:2481265G>A	c.151G>A	p.Asp51Asn	Missense	VUS-P	1	0
CSF1R	5:149441322C>T	c.1717G>A	p.Glu573Lys	Missense	VUS-P	1	0
DAO	12:109283278C>T	c.343C>T	p.Arg115Trp	Missense	VUS-P	1	0
DCTN1	2:74605284C>T	c.122G>A	p.Arg41Gln	Missense	VUS-P	1	0
DCTN1	2:74595092A>C	c.2015+5G>A		Intronic splice site	VUS-P	1	0
DCTN1	2:74593111C>T	c.2795G>A	p.Arg932His	Missense	VUS-P	1	0
ERBB4	2:212293178T>C	c.2674A>G	p.Ile892Val	Missense	VUS-P	1	0
FUS	16:31202343C>T	c.1453C>T	p.Arg485Trp	Missense	VUS-P	1	0
FUS	16:31202325T>-	c.1435delT	p.Tyr479Metfs*50	LoF frameshift	VUS-P	1	0
GRN	17:42429718T>C	c.1423T>C	p.Cys475Arg	Missense	VUS-P	1	0
GRN	17:42429727C>T	c.1432C>T	p.Arg478Cys	Missense	VUS-P	1	0
MAPT	17:44055753G>T	c.320G>T	p.Gly107Val	Missense	VUS-P	1	0
MATR3	5:138661306C>T	c.2326C>T	p.Pro776Ser	Missense	VUS-P	1	0
NEFH	22:29876289->C	c.41dupC	p.Phe15Valfs*83	LoF frameshift	VUS-P	1	0
NEK1	4:170510659A>C	c.403T>G	p.Phe135Val	Missense	VUS-P	1	0
NEK1	4:170398485G>A	c.2140C>T	p.Arg714Cys	Missense	VUS-P	1	0
NEK1	4:170501992C>G	c.868+1G>C		LoF splice donor	VUS-P	2	0
NEK1	4:170511956A>G	c.317T>C	p.Leu106Ser	Missense	VUS-P	1	0
NOTCH3	19:15272414G>A	c.6025C>T	p.Arg2009Trp	Missense	VUS-P	1	0
NOTCH3	19:15298776C>T	c.1522G>A	p.Val508Met	Missense	VUS-P	1	0
NOTCH3	19:15288789G>C	c.3950C>G	p.Pro1317Arg	Missense	VUS-P	1	0

NOTCH3	19:15302615C>G	c.743G>C	p.Gly248Ala	Missense	VUS-P	1	3
OPTN	10:13178846G>A	c.1714G>A	p.Val572Met	Missense	VUS-P	1	0
PRPH	12:49691776C>T	c.1303C>T	p.Arg435Trp	Missense	VUS-P	1	0
PSEN1	14:73640329T>G	c.394T>G	p.Ser132Ala	Missense	VUS-P	1	0
PSEN1	1:227073265C>T	c.383C>T	p.Thr128Ile	Missense	VUS-P	1	1
SETX	9:135203653A>C	c.3332T>G	p.Leu1111Trp	Missense	VUS-P	1	0
SOD1	21:33040863C>A	c.437C>A	p.Ala146Asp	Missense	VUS-P	1	0
SOD1	21:33032151G>C	c.69G>C	p.Gln23His	Missense	VUS-P	1	0
SOD1	21:33038809G>T	c.217G>T	p.Gly73Cys	Missense	VUS-P	1	0
SQSTM1	5:179250993C>T	c.437C>T	p.Pro146Leu	Missense	VUS-P	1	0
SQSTM1	5:179260077G>A	c.800G>A	p.Arg267His	Missense	VUS-P	2	0
TBK1	12:64889529A>C	c.1694A>C	p.Gln565Pro	Missense	VUS-P	1	0
Heterozygous Variants in Recessive Genes							
ALS2	*2:202614512C>T*	*c.1738G>A*	*p.Val580Ile*	*Missense*	*VUS-P*	*1*	*0*
ALS2	*2:202574703T>C*	*c.4181A>G*	*p.Tyr1394Cys*	*Missense*	*VUS-P*	*1*	*0*
ALS2	*2:202626404C>T*	*c.313G>A*	*p.Ala105Thr*	*Missense*	*VUS-P*	*1*	*0*
BSCL2	*11:62458810A>C*	*c.755T>G*	*p.Phe252Cys*	*Missense*	*VUS-P*	*1*	*0*
HTRA1	*10:124266357C>T*	*c.928C>T*	*p.Arg310Cys*	*Missense*	*VUS-P*	*1*	*0*
SPG11	*15:44949428AT/-*	*c.733_734delAT*	*p.Met245Valfs*2*	*LoF frameshift*	*Pathogenic*	*1*	*0*
SPG11	*15:44855329A>G*	*c.7322T>C*	*p.Leu2441Pro*	*Missense*	*VUS-P*	*1*	*0*
SPG11	*15:44925771A>G*	*c.1667T>C*	*p.Phe556Ser*	*Missense*	*VUS-P*	*1*	*0*
SPG11	*15:44876486C>T*	*c.5392G>A*	*p.Glu1798Lys*	*Missense*	*VUS-P*	*1*	*0*

Table 24 Variants identified in incident MND cohort 2015-2017; LoF = Loss of function

Gene 1	Variant Protein Change 1	Gene 2	Variant Protein Change 2
C9orf72	Expansion	*SOD1*	p.Ile114Thr
C9orf72	Expansion	*ALS2*	p.Val580Ile
SOD1	p.Ala146Asp	*SETX*	p.Leu1111Trp
MATR3	p.Pro776Ser	*SQSTM1*	p.Arg267His
NEK1	p.Arg714Cys	*DCTN1*	p.Arg932His
FUS	p.Tyr479Metfs*50	*SPG11*	p.Glu1798Lys
ALS2	p.Ala105Thr	*NOTCH3*	p.Pro1317Arg

Table 25 Digenic MND cases

Single Nucleotide Polymorphisms

Four variants were found in multiple MND case samples and were thought to be single nucleotide polymorphisms (SNPs); these are described in Table 26.

Gene	Genomic Position	Variant DNA Change	Variant Protein Change	Variant Type	Pathogenic Classification	Number cases	Number controls
Single Nucleotide Polymorphisms in MND population							
SETX	9:135152544A>-	c.6843-5delT		Intronic splice site	Benign	143	317
FIG4	6:110053825T>-	c.447-3delT		Intronic splice site	Benign	46	0
NEK1	4:170506525C>T	c.782G>A	p.Arg261His	Missense	VUS-B	9	7
DCTN1	2:74590116C>T	c.3529+5G>A		Intronic splice site	Likely benign	9	31

Table 26 Single nucleotide polymorphisms in the Scottish MND population

In summary, 22.4% of individuals carried a potentially pathogenic variant. This percentage includes people with *C9orf72* expansions, Pathogenic or VUS-P variants as assessed using the extended gene panel and those individuals carrying more than one pathogenic or potentially pathogenic expansion or variant (digenic carriers). This is illustrated in Figure 23. The proportion of Pathogenic and VUS-P by gene name is illustrated in Figure 24. Although different variant classification methods were adopted in this analysis compared with our published analysis in Chapter 5, the proportion of potentially pathogenic mutations in the incident cohort is significantly more (1989-2014 74/431 cases; 2015-2017 82/339 cases; Z-test of proportions p=0.021 (95% CI 0.13-0.0097). As described in Chapter 5, the genetic epidemiology of MND is often described by sub-categorising "familial" (n=28) and "apparently sporadic" (n=305) cohorts and this is illustrated in Figure 25. In those considered to have familial MND, 67.9% had a potentially pathogenic mutation, whereas 20.3% of apparently sporadic patients had a potentially pathogenic mutation.

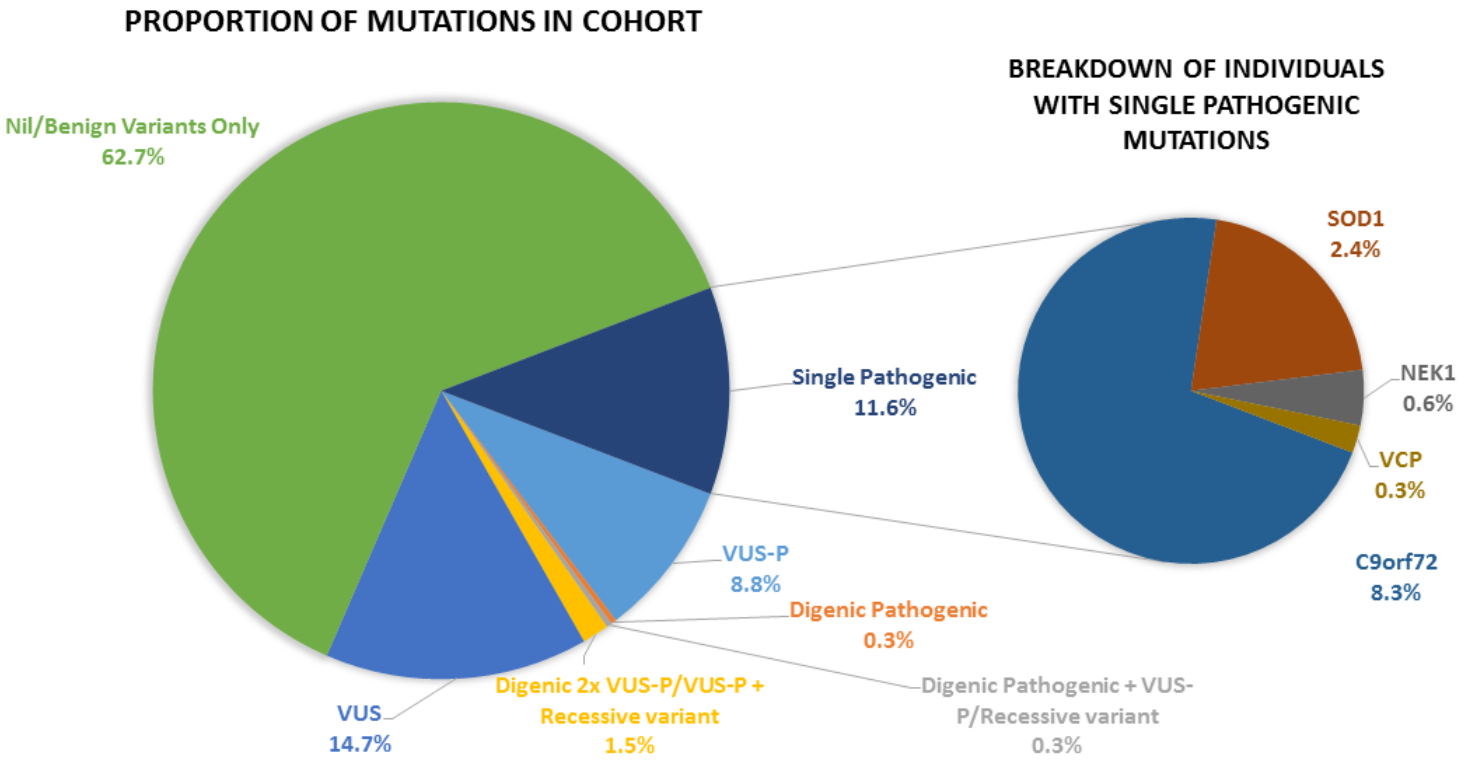


Figure 23 Genetic epidemiology of MND cohort 2015-17 (n=339)

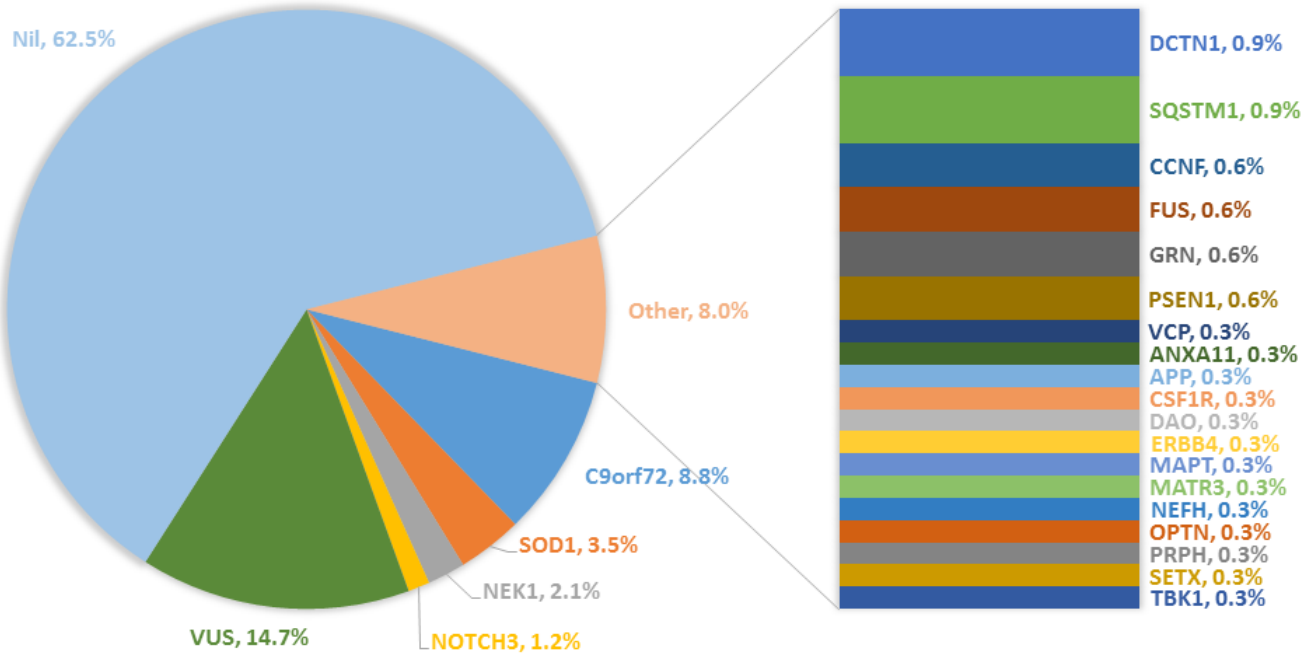
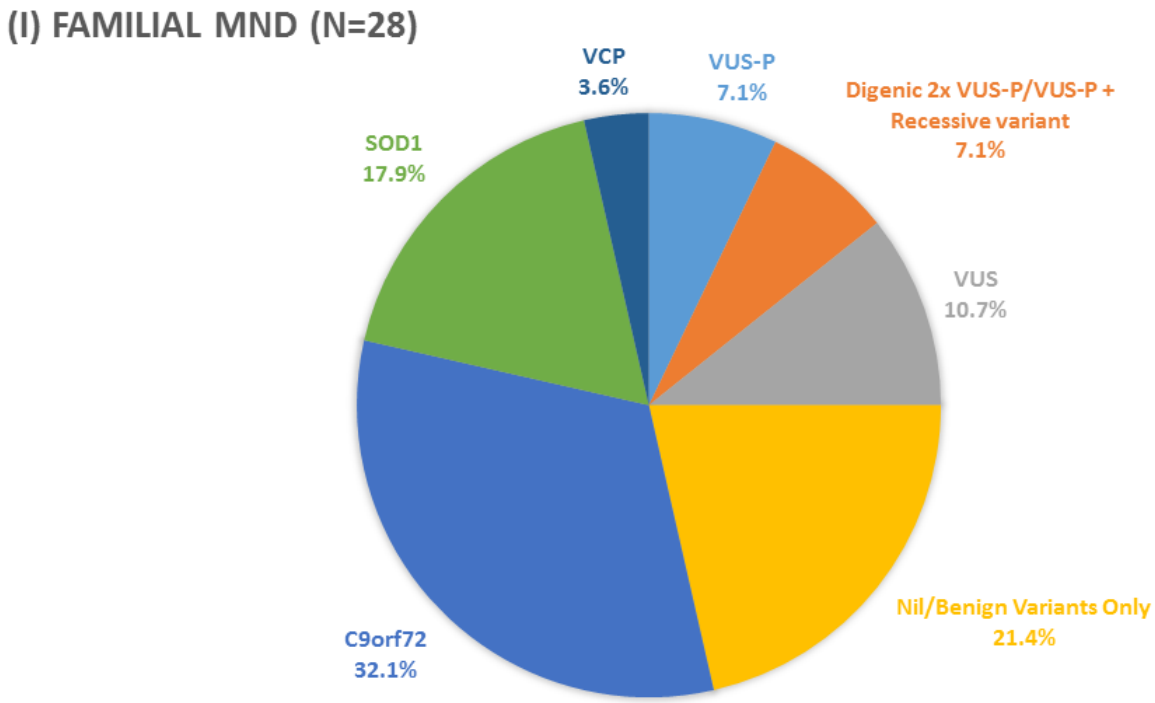


Figure 24 Genetic epidemiology of MND cohort 2015-17 (n=339) by gene name


(I) FAMILIAL MND (N=28)
VCP
3.6%
VUS-P
7.1%
Digenic 2x VUS-P/VUS-P +
Recessive variant
7.1%
SOD1
17.9%
VUS
10.7%
Nil/Benign Variants Only
21.4%
C9orf72
32.1%

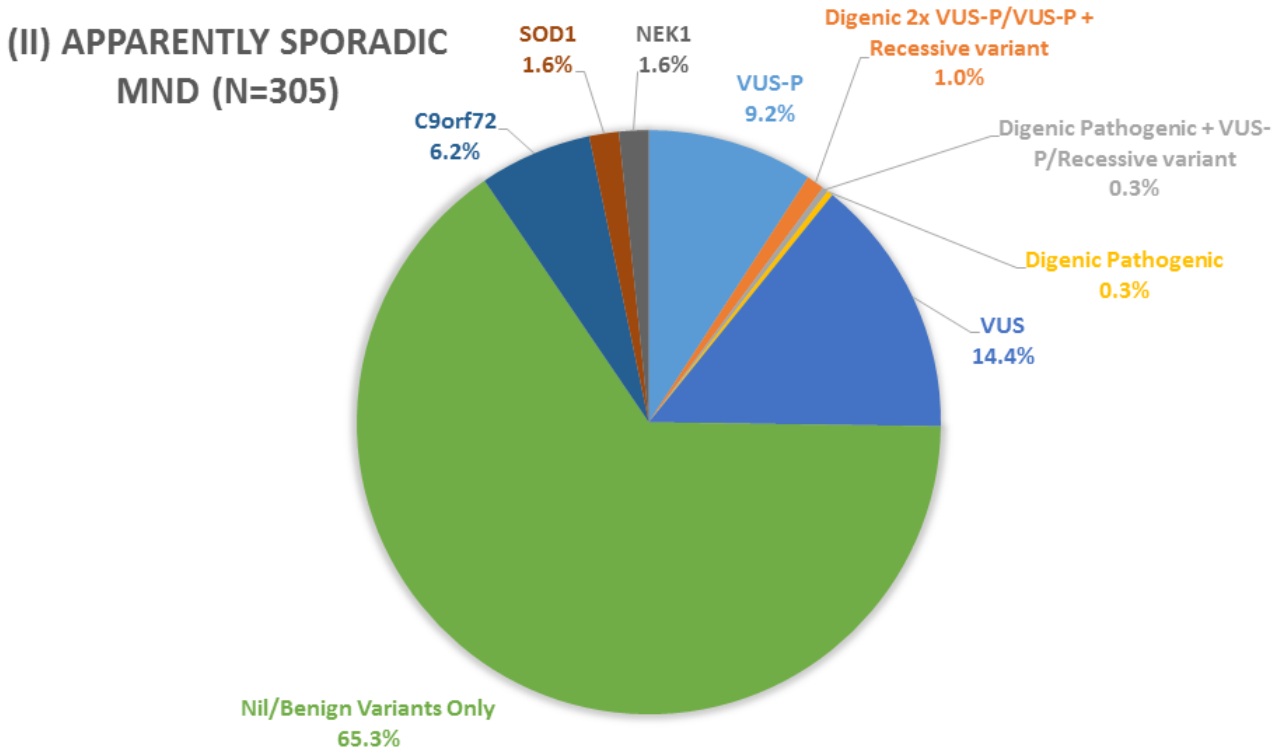


Figure 25 Genetic epidemiology of MND cohort 2015-17 subdivided by (i) Familial (n=28) and (ii) Apparently Sporadic Cases (n=305)

6.4 DISCUSSION

We have shown the utility of the ACMG guidelines for 12 MND-associated genes and have outlined a framework for focused interpretation of variants. To our knowledge, this is the first systematic analysis of classification of MND genomic data. We applied the framework to an incident cohort of MND patients diagnosed in 2015-2017. Just over half of the incident MND population (54.8%) contributed to this genetic study. Reasons for not achieving higher ascertainment might include patient choice, limited discussions regarding research options prior to the CARE-MND initiative (prior to August 2015) and rapidly deteriorating disease, meaning that patients were less willing to devote time and efforts to research. Currently, genetic research in Scotland does not offer feedback of results and does not lead to treatment modification and so benefits to patients at an individual level are limited. Allowing for these factors, we consider our recruitment figures to be appropriate and reflective of the generosity of the Scottish MND patient community.

The proportion of patients found to have a potentially pathogenic mutation/expansion was 22%; this is significantly higher than the proportion obtained using a limited gene panel (17%), as hypothesised. However, the figure obtained in the incident cohort includes VUS-P mutations. By using a systematic ACMG-guided framework, we aimed to outline a framework that might lead to improved increased accuracy of variant classification. However, ACMG-guided methods are in fact more stringent and result in a more conservative estimation of pathogenicity. One benefit of this is that, on reviewing previously described MND-associated variants, we can identify with certainty those that are Pathogenic and those that are Benign. This can be used as a reference for future MND genetic work. We tried to provide some clarity by qualifying VUS by describing VUS-P and VUS-B. Ideally, future MND genetic studies should make reference to VUS-P to inform the research community better. This would help in the identification of ACMG criteria that might help to change this classification to Likely Pathogenic (eg. Focus research towards trio or segregation studies or animal/cell models of disease). Clinician understanding of VUS-P is also essential, as this may better facilitate informed discussions regarding potentially inheritability of disease (even in the context of variable penetrance).

The disease-specific and gene-specific use of the ACMG guidelines parallels that adopted by the colorectal cancer community. The International Society for Gastrointestinal Hereditary Tumours (InSiGHT) adopted similar methods to ours (modified Delphi) but on a wider and more collaborative scale[392]. Research and clinical groups were invited to submit unpublished variants, which were recognised by microattribution. This resulted in an open-access annotated database for active and ongoing variant interpretation. Although 32% of variants remained unclassifiable at the end of the study, public availability of this information has facilitated ongoing reassessment. This group also found that missense VUSs were problematic. However, by putting an increased weight on highly accurate *in silico* predictors, they could potentially reclassify ~50% of VUSs into Likely Pathogenic variants[392]. Elements of this platform are echoed in the ALS Online Database; however, in the current ALS website, nuances of variant pathogenicity are not outlined[71,72]. While generic variant databases (such as ClinVar) do exist, the InSiGHT database highlights the benefits of a disease specific platform to which experts can continuously contribute emerging knowledge.

The most important regions of interest in this incident Scottish cohort are: the *C9orf72* expansion, the *SOD1* gene and the *NEK1* gene. The proportion of *C9orf72* expansion carriers (8.6%) was lower than in our pilot study (Chapter 5; 10.2%). However, the 2015-17 cohort is more unselected and population-based than our historical cohort, and so this figure is likely more representative. We identified two individuals with intermediate-length repeat expansions (70 repeats and 28 repeats). Previous studies regarding intermediate repeat-length expansions are conflicting. A case of an 89 year old man without MND or FTD who had a 70-repeat expansion has been described[437]. However, more recently, a meta-analysis suggested that intermediate expansions 24-30 repeats in length are associated with MND (p=0.02)[438]. In view of this new evidence, our intermediate length samples were both considered pathogenic giving a final population frequency of 30/339 (8.9%). The *C9orf72* expansion is the commonest cause of familial MND affecting 32.1% of cases, but also 6.2% of apparently sporadic cases. In a large study of people of European, USA and Australian origin, higher figures were identified (expansions in 39.3% familial and 7.0% of apparently sporadic pwMND)[439] but in a more recent study of Italian patients, only 6.9% of patients were *C9orf72* expansion carriers overall[440].

SOD1 mutations were identified in 3.5% of cases overall, but 17.9% of familial cases. This is lower than expected relative to findings in our pilot study 1989-2014 cohort (5% of all cases, 29% of familial cases) though figures are still higher than global population estimates (12% familial and 1-2% sporadic cases[102,440]). As before, the p.I114T variant is the biggest contributor to this observation though clearly other *SOD1* variants are found in the Scottish population. The relative ethnic homogeneity of the Scottish MND population (Chapters 3, 4, 7) is likely a factor in the persistence of this variant. This has significant implications as antisense oligonucleotide (ASO) gene-modifying treatments for *SOD1* carriers appear promising[441]. Functional studies of the p.I114T variant may be particularly beneficial to the Scottish population – if this variant results in toxic-gain-of-function pathophysiologically, then ASOs may be a potential treatment for 3-4% of the Scottish MND population. The p.A4V mutation – the most common *SOD1* mutation in the USA – does not appear to be present in the Scottish population[65,113,114].

As hypothesised in Chapter 5, our population also appeared to be enriched for potentially pathogenic *NEK1* mutations, present in seven cases (2.1%) versus zero controls. None of these individuals had a family history of MND. Four of these were loss-of-function mutations (two frameshift, two splice donor); loss of function is considered mechanistic for this gene[90]. Loss of function variants were present in 1.2% of cases, which parallels previous population studies (rates 0.7-1.2%)[88,90]. Of all the potentially pathogenic *NEK1* variants identified, none were seen in the pilot study cohort (Chapter 5). Two mutations were identified in both cohorts (p.R261H and p.A341T). The former is described in other studies: first identified in an isolated community in the Netherlands, it is now thought to be a risk variant for ALS using meta-analysed data ($p=4.8 \times 10^{-5}$, OR 2.4 in cases versus controls)[88]. In the 1989-2014 cohort, five cases and two controls had this variant and it was considered Likely Pathogenic. In the 2015-17 cohort, it was present in nine cases and seven controls. Other clinical samples reported to ClinVar suggest that the variant may be a VUS, Likely Benign or Benign. In view of this information, we would now classify this variant as a VUS-B, present in a total of 1.8% of cases and 0.7% of controls (1989-2017). The p.A341T variant was previously identified in the Scottish population as a VUS, present in four cases and three controls (Chapter 5). In the latest cohort, we identified the variant in four cases and 14 controls. *In silico* predictions suggest that it is benign and it has a MAF in the European non-Finnish population of 0.003. We now predict that this variant is therefore a VUS-B.

Similarly, the previously described *DAO* p.R199Q variant is a VUS-B based on our analysis but previously thought to be potentially pathogenic[378].

Potentially pathogenic rare variants were identified in several MND-associated and dementia-associated genes. Phenotypic characterisation of these cases will be important to identify any correlations with specific MND phenotypes (eg. Those seen in those with *VCP* and *DCTN1* mutations) and with cognitive profiles. This will be discussed in Chapter 7. No loss of function variants were identified in the *TBK1* gene, contrary to results in Chapter 5. Recent studies suggest that *TBK1* variant population frequencies range from 0.5%-9%[416] No pathogenic variants were identified in genes related to hereditary spastic paraparesis (HSP). By extension, we can be confident that the CARE-MND platform reliability identifies patients with a diagnosis of MND, as opposed to MND mimics.

Finally, 2.1% of the population were oligogenic gene carriers, although some of the variants were in recessive genes. This is similar to other population rates[94,442]. None of these had *TBK1* variants, in spite of recent research suggesting that *TBK1* variants are found in oligogenic cases[416]. Again, phenotypic correlation of these cases is key. Interestingly, we identified one patient with a *C9orf72* expansion and the *SOD1* p.I114T variant. We might expect this individual's disease to manifest at an earlier age with a more aggressive course due to the burden of pathogenic variants.

As with other published MND genotype studies, our findings highlight that there is an increasing number of MND-associated genes but that variants in these genes are rare. The novelty of these genes means that insufficient information is available to allow conclusive classification of pathogenicity. As we have seen, further study necessitates reclassification of variants. As clinical diagnostic testing of MND genomes because more established in Scotland, provisions will need to be made to allow reassessment and revision of an individual's gene status.

Some variants have changed classification between the 1989-2014 and 2015-2017 analysis. We therefore recommend that all pwMND undergoing genetic testing should be counselled about the continuing emergence of new data about MND genetics and the current high

proportion of VUS. Trends towards pathogenic or benign status, as outlined in our methods, may be useful to aid explanations. Our analyses are up-to-date as of August 2017 but the dynamic nature of the field necessitates continuous review of each variant. The requirement to follow-up individuals with VUS and reassess the variant of interest at regular intervals should be explained[443]. In view of cognitive impairment in MND, careful assessment of capacity and involvement of family members is crucial. When faced with results of MND genotyping, we suggest referring to our disease and gene-specific considerations. We have outlined key areas that are clinically relevant and interpretable, and hope that it can be used to inform better our patients and their families.

Limitations

The ACMG-guided framework adopted for this part of the thesis is more strict than the approach taken in the pilot study. One major limitation of this is that it generates more uncertainty, by classifying most variants as VUS. This makes the translation of our results into clinical practice difficult. The argument for a more stringent and transparent approach is: i) that is provides a more realistic estimate of the number of true pathogenic, causative gene mutations, ii) it highlights the uncertainty and lack of information within the field and identifies areas for future collaborative research and knowledge building; iii) it avoids the blanket dismissal of potentially clinically relevant variants.

We also acknowledge that that we have not confirmed variants identified in the incident cohort using Sanger sequencing. While the concordance between next generation sequencing (NGS) and Sanger techniques is now excellent, the risk of false positives with NGS may be 1.3%[444]. One reason that this was not pursued was that sequencing was performed as part of a research study only, with results not being fed back to patients. In Scotland, all patients with MND are encouraged to consider storing a DNA sample in an NHS clinical-approved laboratory. Research participants were reminded of this option at point of recruitment to the Scottish Regenerative Neurology Tissue Bank. Theoretically, if a drug were to become available for an MND-associated gene, the patient or their family members could activate confirmatory testing through the clinical laboratory. Sanger sequencing is labour

intensive compared with NGS techniques and, although this was not possible during the course of this PhD, it will be explored in future work.

6.5 CONCLUSIONS

In summary, we have tested an updated and comprehensive gene panel on an unselected cohort of incident pwMND. We have shown that the ACMG classification system can be applied to MND genomes. However, certain classification criteria are more relevant than others, and so a disease or even gene-specific approach should be adopted to avoid ignoring potentially pathogenic VUS. We have proposed changes to classifications of previously described variants, highlighting the importance of continuous reassessment in genes for which evidence is emerging. Our results suggest that, while the *C9orf72* expansion is important in the Scottish MND population, it should certainly not be the only target for future genetically-stratified treatments. *SOD1* p.I114T carriers (3%) and *NEK1* loss of function variant carriers (1%) make up a small but important subset of patients. As *NEK1* carriers were all apparently sporadic cases, future work into functional and pathological correlates is certainly merited. Potentially pathogenic rare variants in MND and dementia-associated genes are found in the Scottish population. Further exploration of their phenotypes may provide justification for their inclusion in a diagnostic panel. Digenic cases are found in the Scottish population; the *SOD1* p.I114T variant and the *C9orf72* expansion can co-exist and so anyone considered pursuing genotyping to inform patients or families of risk of heritability should ensure that both screening for *C9of72* expansions and MND panel genes occurs simultaneously.

7. GENOTYPE-PHENOTYPE CORRELATION AND PROGNOSTIC MODELLING

7.1 BACKGROUND AND AIMS

Background

As described, it has been possible to achieve deep clinical phenotyping of an unselected population cohort of pwMND at and around diagnosis, and subsequently genotype a subset of these individuals using a comprehensive gene panel. Gene carriers of sufficient number can be studied for genotype-phenotype associations. The benefit of the CARE-MND platform is the broad selection of variables which can be studied. However, in this case, the number of variables are high relative to the number of individuals with the outcome of interest. As described in Chapter 5.2, one way to determine association between variables and a binary outcome of interest it to perform multiple univariate analyses of all variables against the outcome, and then to input near-significantly associated variables into a multivariable model such as a logistic regression model. These methods can be applied but they are generally now considered to be undesirable because they do not take into account between-variable association[445]. A preferred approach is to undertake multivariable modelling from the start. However, this limits the number of predictors that one can include in the analyses due to the 'one in ten' rule[195,316]. Machine learning methods, however, can be used to select the most appropriate predictors (feature selection), where the number of predictors (p) is high relative to the number of events (n).

Aims

1. To identify genotype-phenotype associations in mutation carriers.
2. To test the use of machine learning methods for feature selection of phenotypic variables in CARE-MND.
3. To examine case studies of rare gene mutation carriers.

4. To appreciate the impact of having a genetic mutation in a multivariable survival model.

7.2 METHODS

Phenotype Variables

All patients studied in Chapter 6 (n=339) were included for genotype-phenotype analysis. The variables considered for prognostic modelling in Chapter 4.4 (Table 12) were revisited and some minor modifications were undertaken. The variable examining El Escorial Classification of disease was expanded to include Primary Lateral Sclerosis (PLS), Progressive Muscular Atrophy (PMA) and Progressive Bulbar Palsy (PBP), rather than grouping these into one variable ("Other") in case specific genetic mutations were related to these phenotypes. Variables previously found to be correlated with others (Time to Neurology Clinic, Time from Onset to MND clinic, Number of Comorbidities) were excluded from the outset. Although the ECAS ALS Non-Specific Score was previously removed from the prognostic models due to its correlation with the ALS Specific Score, this was reconsidered for genotype-phenotype studies in view of their being several genes related to cognitive phenotypes. In summary, 26 phenotypic variables were considered for genotype-phenotype analyses.

Genotype-Phenotype Correlations Analyses

In view of the relatively small sample sizes in this study, gene carriers were summarised based on frequency of mutation. Thereby, the following groups were studied: i) pwMND carrying Pathogenic or potentially pathogenic mutations (including *C9orf72* expansions and VUS-P); ii) pwMND carrying Pathogenic mutations (including *C9orf72* expansions), iii) pwMND carrying *C9orf72* expansions only, iv) pwMND carrying *SOD1* mutations only, v) pwMND carrying the *SOD1* p.I114T variant only, vi) pwMND carrying *NEK1* mutations, and vii) pwMND who carried more than one mutation (digenic carriers). Descriptive statistics for phenotypic variables by group were summarised. For groups (i), (ii) and (iii), where a significant

proportion of patients had the outcome of interest, statistical modelling was explored to determine if significant phenotypic predictors were associated with the outcome using multivariable analyses.

Machine learning methods were incorporated using elastic net regularisation with variable selection[446]. The aim was to undertake penalised regression to reduce the variability of the regression estimates. The machine learning algorithm selects the best penalisation hyperparameter (lambda (λ)) over a range of alpha (α) hyperparameters (ranging between ridge (α=0) and lasso (α=1)) which minimises both variance and bias, and therefore overfitting.

The process for this analysis involves pre-processing to remove variables with near-zero variance and dummy coding of categorical variables, as described in Chapter 4.2. A nested leave one site out cross-validation (LOSOCV) approach was taken, whereby the model was trained on part of the dataset and then tested on unseen data. The data were split by "Health Board"; 15 health board sites were included accounting for the 14 health boards in Scotland plus a proportion of patients who moved out of Scotland after receiving initial care in Scotland. By training the data on each combination of 14 sites and testing on the unseen site, it is internally-externally validated[447]. This process is repeat 10 times (10-fold cross-validation) over a 10x10 grid of λ and α hyperparameters to identify the optimal model. Single imputation was used to account for missing data using the k-nearest neighbour method[191]. The stability and significance of the model is checked and, if satisfactory, the features common to all models (the most generalisable and stable features) are extracted. These features can then be taken forward for regression modelling.

The selected variables were then tested for independent association with the outcome of interest using a binomial logistic regression model. Comparison was made against the more traditional method of inputting near-significant ($p \leq 0.1$) variables from univariate analyses into a logistic regression model. For groups (iii)-(vi), where the number of individuals with the outcome of interest was small, univariate analyses only were performed using two-tailed Fisher's exact tests for categorical data and *t*-test or Wilcoxon Rank-Sum/Mann Whitney *U* test for parametric and non-parametric continuous data, respectively. Correction for multiple testing was undertaken using the Bonferroni method.

Finally, genotype (group (i)) was included in the Cox regression models described in Chapter 4.2 to test the hypothesis that having a genetic mutation is an independent predictor of survival.

All statistical analyses were undertaken using R statistical programming[240], in particular packages “caret”, “glmnet” and “survival”.

7.3 RESULTS

Pathogenic or VUS-P carriers

Individuals carrying a Pathogenic variant, *C9orf72* expansions or VUS-P were studied (n=82) and compared against those who did not have one of these types of mutations (n=257). Descriptive statistics are summarised in Table 28. Across 29 variables, the median proportion of missing data was 1.8% (range 0-44.5). An elastic net regularised regression with nested cross-validation split by health board site (n=15) was undertaken. These models were significant (p=0.0013). The simplest model within one standard error of the model with the ‘best’ tuning parameters was selected to avoid over-fitting (Breiman’s one standard error rule). This selected out seven features which were common to all models (listed in order of importance): Family History of MND, Family History of Other Neurological Conditions, ALSFRS-R Preslope, Classification: Progressive Bulbar Palsy, Feeding Tube Inserted, Sex and Time to Diagnosis. A logistic regression model was undertaken using these seven variables (p=8 predictors taking into account all levels of the Classification variable included after removing for near-zero variance). After modelling, variables significantly associated with having a genetic mutation were: Family History of MND (p<0.00001), Classification: PBP (p=0.018) and Family History of Other Neurological Conditions (p=0.027). Significantly fewer mutation carriers underwent Feeding Tube Insertion compared with non-mutation carriers (p=0.017). The pseudo-R^2 for the model was 0.141 (95% CI 0.135-0.147) implying that these variables explained 14% of the variance between mutation carriers and non-mutation carriers. The Akaike information criterion (AIC) (an estimate of relative quality of models) was 340.4 (95% CI 337.9-342.6). The model is illustrated in Figure 26.

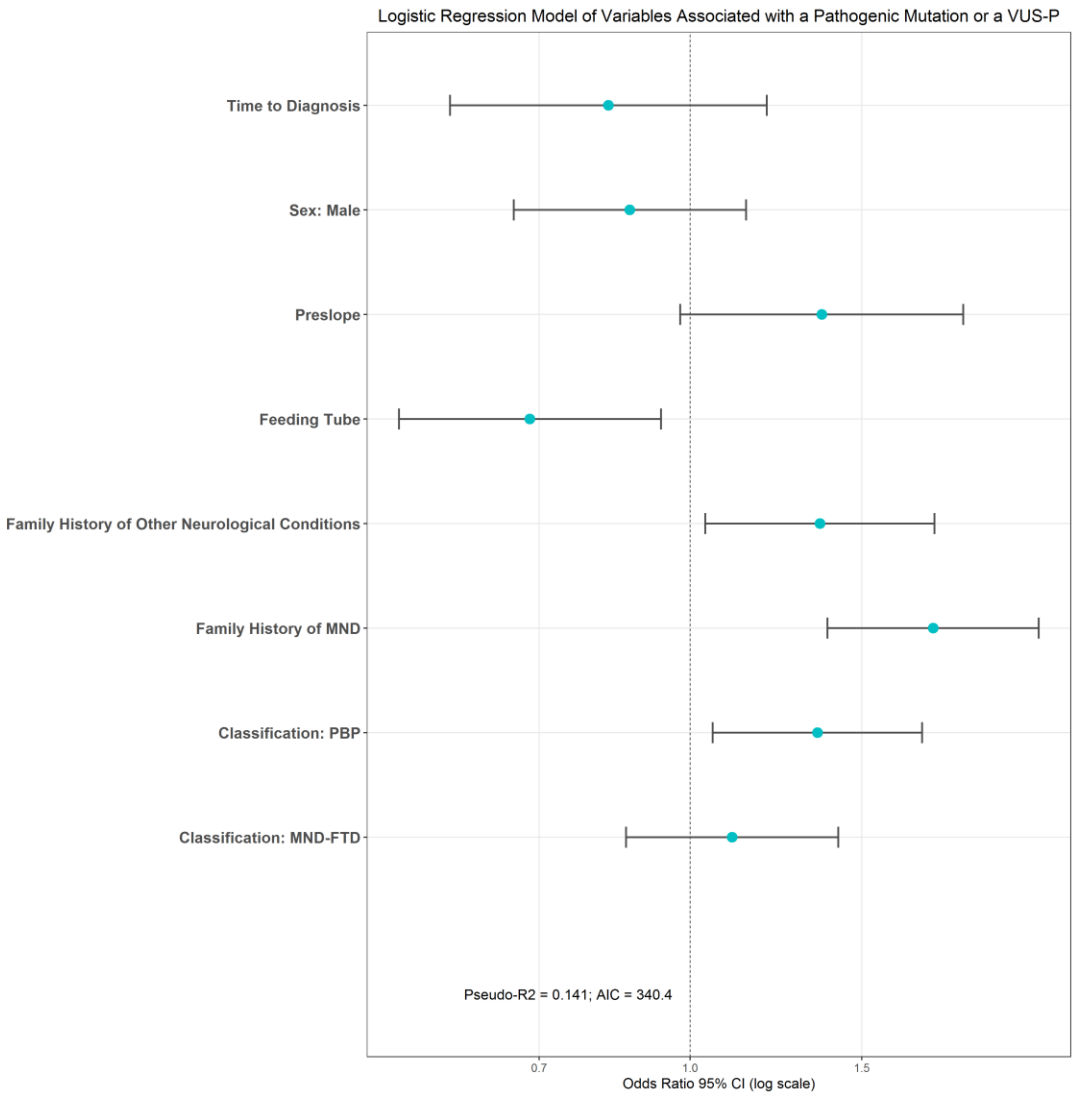


Figure 26 Logistic Regression Model of Variables Associated with a Pathogenic Variant, C9orf72 expansion or a VUS-P using regularised regression for feature selection

These results can be compared with those obtained using the more traditional method of selecting features from univariate analysis, as described in Chapter 5. Variables with univariate p-values ≤ 0.1 (Sex, Classification, Family History of MND, Family History of Other Neurological Conditions and Feeding Tube Insertion) were included in a second logistic regression model. After controlling for near-zero variance there were p=6 predictors. As before, Family History of MND (p<0.00001), Classification: PBP (p=0.0085) and Family History

of Other Neurological Conditions (p=0.034) were significantly associated with having a pathogenic mutation/VUS-P. In this model, there was no significant association with Feeding Tube Insertion (p=0.067). The pseudo-R^2 was 0.116 (95% CI 0.113-0.119) explaining 12% of the variance and the AIC was 345.5 (95% CI 344.4-346.6). These results are illustrated in Figure 27.

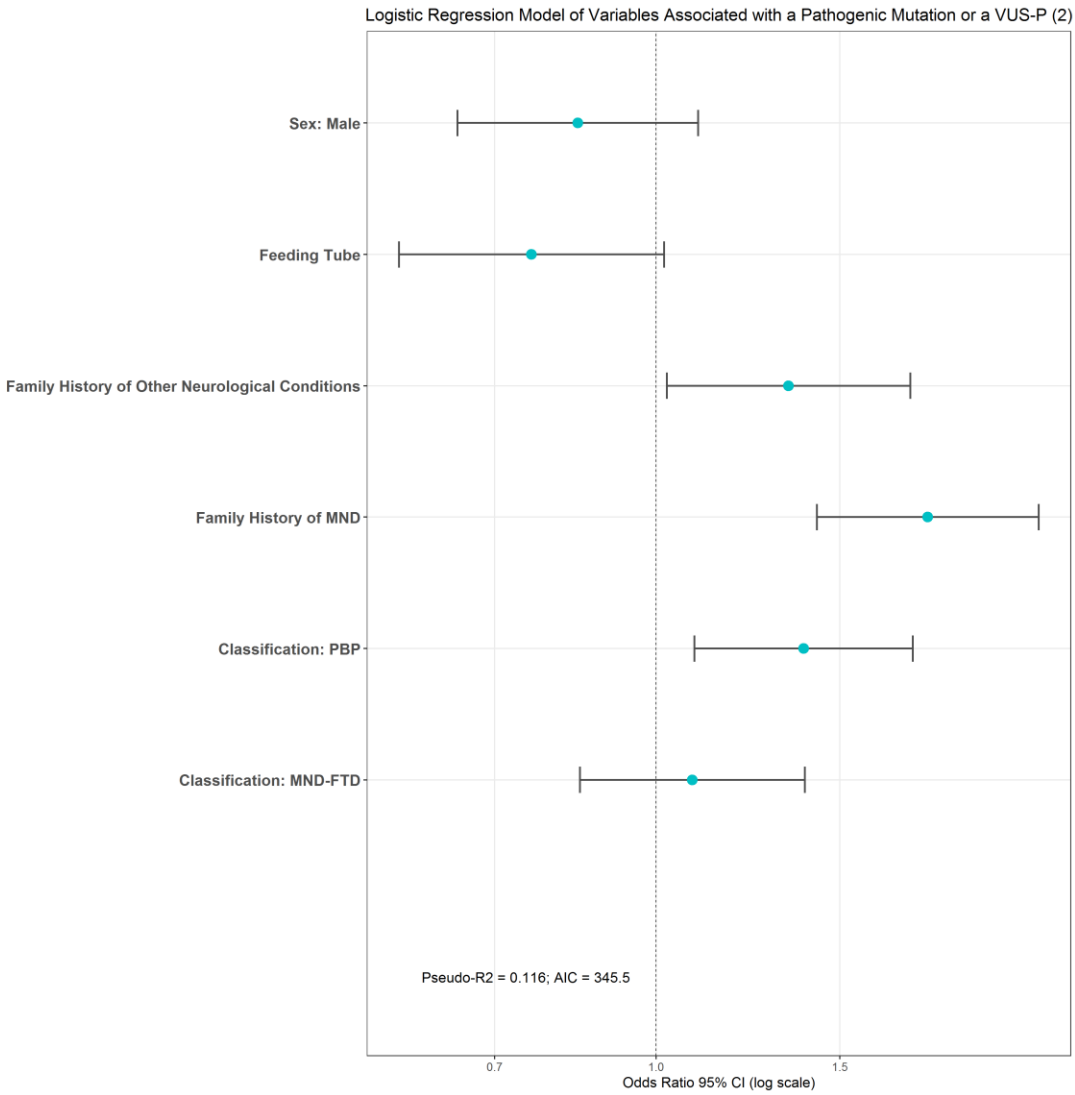


Figure 27 Logistic Regression Model of Variables Associated with a Pathogenic Variant, C9orf72 expansion or a VUS-P using univariate statistics for feature selection

Pathogenic Variant Carriers Only

Patients with convincing Pathogenic/disease-causing mutations were studied separately to see if increasing certainty of pathogenicity was associated with different phenotypic markers. This group included *C9orf72* expansion carriers (n=30), *SOD1* p.I114T carriers (n=9), *NEK1* pathogenic carriers (n=2) and the *VCP* pathogenic mutation carrier (n=1) (total n=41). The same methods were adopted to identify predictors. These individuals were compared against the rest of the cohort (n=297). However, to limit the number of predictors to fulfil the 'one in ten' rule, a 'tolerance' cut-off was applied to select the simplest model within a percentage of the model with the 'best' tuning parameters. The models were significant (p=0.004). The models generated twelve relevant predictors (Family History of MND, ALS Specific Score, Family History of Other Neurological Conditions, Sex, Time to Diagnosis, History of Head Injury, Ever Smoking, Use of NIV, Exercise Participation, Age of Onset, Feeding Tube Insertion and Past Medical History of Malignancy). When selected for logistic regression modelling, there were remained 12 predictors. Considering there were only 41 individuals with the outcome of interest, this violated the 'one in ten' rule and so the logistic regression model could not be used. Instead, univariate statistics with Bonferroni correction only could be tested, revealing that only a Family History of MND was associated with having a pathogenic mutation (Table 28).

C9orf72 expansion carriers

The same methods were again applied to carriers of *C9orf72* expansions only (n=30). These individuals were compared against the rest of the cohort (n=309). The models were significant (p<0.0001). The models generated seven significant features (Use of NIV, Family History of MND, ALS Specific Score, History of Blood Transfusion, Time to Diagnosis, Family History of Other Neurological Conditions and Site of Onset: Other). When selected for logistic regression modelling, there were eight predictors (taking into account different levels of Site of Onset). Considering there were only 30 individuals with the outcome of interest, this also violated the 'one in ten' rule. Univariate statistics with Bonferroni correction revealed that a Family History of MND and lower ECAS: ALS Specific Score were associated with having a *C9orf72* expansion (Table 28).

There were two individuals with intermediate-length repeat expansions. Both individuals were diagnosed in their early sixties with upper limb onset disease and had diagnoses of ALS with cognitive impairment identified via ECAS measurements (ECAS total score 73 in patient with 28 repeats, 79 in patient with 70 repeats). Neither had a family history of MND nor past medical or family history of psychiatric conditions. The individual with 28 repeats died 2.6 years after symptom onset whereas the individual with 70 repeats was alive at censorship (1.6 years after onset).

SOD1 carriers, *SOD1* I114T carriers

Due to the low number of *SOD1* variant carriers (n=12), multivariable analysis was not possible. However, univariate analysis showed a significant genotype-phenotype association between SOD1 carriers and having a Family History of MND (p=0.0001) (Table 28). SOD1 p.I114T carriers were studied separately and, on univariate testing, they were also significantly associated with having a Family History of MND (p=0.0003) (Table 28). Family history details of p.I114T carriers are described in Table 27.

Proband with SOD1 p.I114T variant	Age of Onset of Proband	Site of Onset of Proband	Number of Affected Relatives	Family History of MND Details	Other Family History
1	61	Limb	1	Father – limb onset age 63	Parental grandfather – diagnosed with MS
2	64	Limb	3	Sisters x2, parental cousin – disease site and onset unknown	
3	67	Limb	1	Father – site and onset unknown	Niece – diagnosed with MS
4	57	Limb	2	Father, Sister – limb onset	
5	48	Limb	1	Stepsister – site and onset unknown	
6	55	Limb	0	-	Parental grandfather – diagnosed with MS in 40s
7	67	Limb	0	-	
8	68	Limb	0	-	
9	64	Limb	0	-	

Table 27 Family histories of SOD1 p.I114T carriers; MS = multiple sclerosis

NEK1 carriers, Digenic carriers

There were no statistically significant genotype-phenotype associations for *NEK1* or digenic carriers (Table 28). The individual with both the *C9orf72* expansion and *SOD1* p.I114T variant was a male who had lower limb-onset ALS age 68 and who developed cognitive impairment

as assessed by ECAS. Interestingly, he had no family history of MND or other neurological conditions. He died 65 months (5.4 years) after symptom onset.

Phenotypic Characteristic	Missing Data (%)	Summary Statistic/ Test	All (n=339)	Nil Path/VUS-P (n=257)	Path/VUS-P (n=82)	Path Only (n=41)	*C9orf72* (n=30)	*SOD1* inc. I114T (n=12)	*SOD1* I114T (n=9)	*NEK1* (n=7)	Digenic (n=7)
Sex	0	*Male (%) Fisher's*	220 (64.9) -	175 (68.1) -	45 (54.9) p=0.03	21 (48.8) p=0.02	15 (50.0) p=0.1	9 (75.0) p=0.6	6 (66.7) p=1.0	2 (28.6) p=0.05	5 (71.4) p=1.0
Ethnicity	2.7	*Group A (%) Fisher's*	326 (98.8) -	247 (98.4) -	79 (100) p=0.6	40 (100) p=1.0	29 (100) p=1.0	12 (100) p=1.0	9 (100) p=1.0	7 (100) p=1.0	7 (100) p=1.0
Ever smoked	7.7	*Yes (%) Fisher's*	176 (56.2) -	137 (57.3) -	39 (52.7) p=0.5	19 (47.5) p=0.2	14 (48.3) p=0.4	6 (50.0) p=0.8	4 (44.4) p=0.5	4 (80.0) p=0.4	3 (42.9) p=0.7
Exercise participation	15.6	*Median (IQR) Wilcoxon*	Mod (Light-Mod) -	Mod (Light-Mod) -	Mod (Light-Mod) 0.6	Mod (Light-Mod)	Mod (Light-Mod) p=0.5	Mod (Mod-Heavy) p=0.3	Mod (Mod-Heavy) p=0.1	Mod (Light-Mod) p=0.5	Mod (Mod-Heavy) p=0.08
Heavy metal or Pesticide exposure	28.0	*Yes (%) Fisher's*	60 (24.6) -	44 (24.3) -	16 (25.4) p=0.9	6 (17.7) p=0.4	5 (20.8) p=0.8	2 (20.0) p=1.0	2 (25.0) p=1.0	1 (20.0) p=1.0	2 (40.0) p=0.6
PMH Cardiovascular disease	0	*Yes (%) Fisher's*	164 (48.4) -	127 (49.4) -	37 (45.1) p=0.5	17 (41.5) p=0.4	12 (40.0) p=0.3	6 (50.0) p=1.0	4 (44.4) p=1.0	5 (71.4) p=0.3	4 (57.1) p=0.7
PMH Autoimmune disease	0	*Yes (%) Fisher's*	53 (15.6) -	39 (15.2) -	14 (17.1) p=0.7	7 (17.1) p=0.8	6 (20.0) p=0.4	0 (0) p=0.2	0 (0) p=0.4	1 (14.3) p=1.0	1 (14.3) p=1.0
	0	*Yes (%)*	28 (8.3)	20 (7.8)	8 (9.8)	2 (4.9)	1 (3.3)	1 (8.3)	1 (11.1)	1 (14.3)	0 (0)

PMH Malignancy		*Fisher's*	-	-	p=0.6	p=0.6	p=0.5	p=1.0	p=0.5	p=0.5	p=1.0
PMH Psychiatric disease	0	*Yes %)* *Fisher's*	71 (20.9) -	52 (20.2) -	19 (23.2) p=0.6	7 (17.1) p=0.7	4 (13.3) p=0.4	4 (33.3) p=0.3	3 (33.3) p=0.4	0 (0) p=0.4	3 (42.9) p=0.2
History of Head Injury	26.8	*Yes (%)*	67 (27.0) -	54 (45.8) -	13 (21.7) p=0.3	5 (15.6) p=0.1	3 (13.0) p=0.1	3 (33.3) p=0.7	2 (28.6) p=1.0	1 (20.0) p=1.0	0 (0) p=0.2
History of Blood Transfusion	29.8	*Yes (%)*	25 (10.5) -	19 (10.6) -	6 (10.3) p=1.0	3 (9.7) p=1.0	0 (0) p=0.1	1 (11.1) p=1.0	1 (14.3) p=0.5	1 (20.0) p=0.4	0 (0) p=1.0
Family History of MND	1.8	*Yes (%)* *Fisher's*	28 (8.4) -	9 (3.6) -	**19 (23.2)** **p<0.00001**	**15 (36.6)** **p<0.00001**	**9 (30.0)** **p=0.0002**	**6 (50.0)** **p=0.0001**	**5 (55.6)** **p=0.0003**	0 (0) p=1.0	2 (28.6) p=0.1
Family History of Dementia	3.5	*Yes (%)* *Fisher's*	97 (29.7) -	74 (29.8) -	23 (29.1) p=1.0	13 (32.5) p=0.7	9 (30.0) p=1.0	5 (45.5) p=0.3	3 (37.5) p=0.7	1 (16.7) p=0.7	2 (28.6) p=1.0
Family History of Early-Onset Dementia	3.5	*Yes (%)* *Fisher's*	15 (4.6) -	10 (4.0) -	5 (6.3) p=0.4	3 (7.5) p=0.4	3 (10.0) p=0.1	1 (9.1) p=0.4	0 (0) p=1.0	0 (0) p=1.0	1 (14.3) p=0.3
Family History of Other Neurological Conditions	5.0	*Yes (%)* *Fisher's*	94 (29.2) -	63 (25.7) -	31 (40.3) p=0.02	18 (45.0) p=0.03	14 (46.7) p=0.03	7 (63.6) p=0.02	4 (50.0) p=0.2	1 (16.7) p=0.7	2 (28.6) p=1.0
Family History of Psychiatric Conditions	11.5	*Yes (%)* *Fisher's*	49 (16.3) -	37 (16.2) -	12 (16.7) p=1.0	7 (19.4) p=0.6	7 (26.9) p=0.2	2 (18.2) p=0.7	0 (0) p=0.4	0 (0) p=0.6	2 (33.3) p=0.3
	0.6	*Mean (SD)*	63.1 (10.8)	63.1 (10.7)	63.0 (11.1)	60.8 (8.8)	60.7 (7.8)	58.4 (7.8)	60.1 (6.5)	66.4 (12.5)	56.0 (10.9)

Age of onset (years)		*t-test*	-	-	p=0.9	p=0.08	p=0.1	p=0.06	p=0.2	p=0.5	0.1
Time to Diagnosis (months)		*Median (IQR)*	12.0 (8.0-23.0)	12.0 (8.0-24.0)	12.0 (7.0-22.8)	9.0 (7.0-18.0)	10.0 (7.0-18.0)	17.0 (8.5-35.8)	14.0 (7.0-35.0)	8.0 (5.5-20.5)	24.0 (16.0-35.0)
		Wilcoxon	-	-	p=0.2	p=0.09	p=0.08	p=0.3	p=0.8	p=0.3	0.04
Site of onset	0	*Bulbar %*	110 (32.5)	81 (31.5)	29 (35.4)	12 (29.3)	11 (36.7)	1 (8.3)	0 (0)	2 (28.7)	2 (28.7)
		Limb (%)	211 (62.2)	163 (63.4)	48 (58.5)	25 (61.0)	15 (50.0)	11 (91.7)	9 (100)	5 (71.4)	5 (71.4)
		Other (%)	18 (5.3)	13 (5.1)	5 (6.1)	4 (9.8)	4 (13.3)	0 (0)	0 (0)	0 (0)	0 (0)
		Fisher's	-	-	p=0.7	p=0.4	p=0.08	p=0.1	p=0.07	p=1.0	p=1.0
Classification	0	*ALS (%)*	261 (77.0)	198 (77.0)	63 (76.8)	34 (82.9)	23 (76.7)	11 (91.7)	9 (100)	7 (100)	6 (85.7)
		MND-FTD (%)	18 (5.3)	12 (4.7)	6 (7.3)	5 (12.2)	5 (16.7)	0 (0)	0 (0)	0 (0)	0 (0)
		PLS (%)	14 (4.1)	14 (5.4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		PMA (%)	16 (4.7)	13 (5.1)	3 (3.7)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		PBP (%)	20 (5.9)	11 (4.3)	9 (11.0)	2 (4.9)	2 (6.7)	0 (0)	0 (0)	0 (0)	1 (14.3)
		Other (%)	10 (3.0)	9 (3.5)	1 (1.2)	0 (0)	0 (0)	1 (8.3)	0 (0)	0 (0)	0 (0)
		Fisher's	-	-	p=0.04	p=0.1	p=0.08	p=0.7	p=1.0	p=1.0	p=0.8
ALSFRS-R Preslope	21.5	*Median (IQR)*	0.58 (0.28-1.00)	0.57 (0.28-1.00)	0.61 (0.25-1.12)	0.65 (0.30-1.11)	0.71 (0.32-1.11)	0.23 (0.18-0.58)	0.40 (0.14-0.62)	0.75 (0.48-2.25)	0.24 (0.20-0.28)
		Wilcoxon	-	-	p=0.5	p=0.5	p=0.3	p=0.03	p=0.1	p=0.4	p=0.004
ECAS ALS Specific Score	44.5	*Median (IQR)*	81.5 (70.8-87.0)	82.0 (71.8-88.0)	80.0 (68.8-86.2)	75.0 (65.5-82.8)	**69.0 (55.8-79.0)**	86 (76.5-87.0)	87.0 (78.0-87.0)	78.0 (77.5-86.5)	78.0 (74.8-80.5)
		Wilcoxon	-	-	p=0.5	p=0.02	**p=0.0005**	p=0.4	p=0.4	p=0.7	p=0.5
ECAS ALS Non-Specific Score	44.5	*Median (IQR)*	28.0 (24.0-31.0)	28.0 (24.0-31.0)	27.0 (24.8-30.2)	26.0 (24.0-29.5)	25.5 (24.0-29.0)	26.5 (26.0-30.5)	26.0 (25.0-30.0)	31.0 (29.0-31.5)	27.0 (26.2-27.2)
		Wilcoxon	-	-	p=0.3	p=0.4	p=0.3	p=0.9	p=1.0	p=0.2	p=0.5

Riluzole Use	0	*Yes (%)* *Fisher's*	138 (40.7) -	104 (40.5) -	34 (41.5) p=0.9	18 (43.9) p=0.7	14 (46.7) p=0.6	5 (41.7) p=1.0	3 (33.3) p=0.7	3 (42.9) p=1.0	1 (14.3) p=0.2
Feeding Tube Inserted	0.3	*Yes (%)* *Fisher's*	115 (34.0) -	93 (36.3) -	22 (26.8) p=0.1	(10 (24.4) p=0.2	10 (33.3) p=1.0	0 (0) p=0.01	0 (0) p=0.03	0 (0) p=0.1	1 (14.3) p=0.4
NIV Use	0.3	*Yes (%)* *Fisher's*	99 (29.3) -	79 (30.7) -	20 (24.7) p=0.3	9 (22.0) p=0.4	2 (6.7) p=0.003	8 (66.7) p=0.007	7 (77.8) p=0.003	2 (28.6) p=1.0	2 (28.7) p=1.0
Survival (Days)	0.6	*Median (IQR)* *Wilcoxon*	960 (692-1385) -	1010 (718-1414) -	866 (639-1201) p=0.04	794 (596-1050) p=0.005	772 (604-954) p=0.006	959 (692-1141) p=0.4	1020 (686-1310) p=0.8	958 (494-1158) p=0.4	1146 (1045-1903) p=0.1

Table 28 Descriptive statistics of phenotypic characteristics by genotype group. Bonferroni corrected significant values are highlighted in bold (p=0.0019)

Case studies

Key phenotypic characteristics of rare gene carriers are described in Table 29.

FUS, OPTN, TBK1 and VCP

Both patients with *FUS* mutations had young-onset ALS with short survival (20 months from onset, compared with 32 months for the genotyped cohort overall). Both also had family histories of MND, reinforcing the likely pathogenicity of these variants. Neither underwent cognitive assessment, perhaps related to their rapid rate of decline, but cognitive impairment was not a documented feature of disease. As such, both patients met expected phenotypic profiles for *FUS* carriers[79]. The pwMND with a variant in *OPTN* had a very young onset of disease (30 years) with bulbar-onset ALS but no family history. Patients with *OPTN* variants tend to have more slowly progressive disease and, indeed, this individual survived to point of censorship (33 months). The patient with a *TBK1* missense mutation had very rapidly progressive ALS (survival 10 months from onset). In view of this, it was not possible to undertake cognitive assessment. However, during the course of disease, she developed anxiety and low mood and was commenced on antidepressant therapy. She also had a premorbid history of depression, perhaps suggesting that this was a prodromal feature or disease feature of *TBK1* MND. This individual did not have a family history of MND; indeed, all previously described *TBK1* variants in Italian cohorts had sporadic disease[416]. A hypothesis is that *TBK1* variants alone have low disease penetrance; this is supported by reported digenic cases involving *TBK1*[416].

The individual with the Pathogenic *VCP* mutation had a two-generation family history with her mother being diagnosed with limb-onset MND (presumed ALS) and her maternal grandmother having had dementia. There was no documented history of Paget's disease or inclusion body myositis. This individual actually had relatively slowly progressive ALS, surviving 75 months (6.3 years) from onset at censorship.

New MND Genes: ANXA11 and CCNF

The patients with *ANXA11* and *CCNF* mutations had late age of onset of disease and no family history of MND. While they had typical ALS presentations, it is not possible to draw any further conclusions about the pathogenicity of these variants from the phenotypes.

Rare MND-syndrome Genes: DAO, DCTN1, ERBB4, MATR3, NEFH, PRPH, SETX, SQSTM1

People with *DAO* mutations are thought to have shorter survival[86]. However, our patient with a potentially pathogenic *DAO* variant had relatively long survival from onset (52 months) compared to the overall median for the cohort (32 months). Three *DCTN1* carriers were all males with lower limb-onset ALS. While *DCTN1* mutations in MND are rare, this perhaps hints at a link with a more classical ALS phenotype. No specific characteristics were linked with carriers of *ERBB4, MATR3, NEFH* or *PRPH* mutations. The individual with the *SETX* mutation had young onset (age 43 years) familial ALS. However, this patient also had a *SOD1* variant (p.A146N). The young age of onset might be related to the compound heterozygosity observed in this individual. Neither patient with *SQSTM1* variants had multisystem disease suggestive of Paget's disease or myopathy; however, one individual did have cognitive impairment.

FTD Genes: GRN, MAPT, PSEN1, PSEN2

When examining individuals carrying variants in FTD-associated genes (*GRN*, *MAPT, PSEN1, PSEN2),* we see that 3/5 individuals had evidence of some cognitive impairment; the remaining two did not have cognitive assessments. Of these five individuals, one had a family history of early-onset dementia, one a family history of MND and two had family histories of Parkinson's disease. As cognitive impairment is a feature of all these neurodegenerative disease, it is possible that the variants are contributing to an inheritable dementia phenotype.

Other Dementia Genes: APP, CSF1R, NOTCH3

Genes associated with other dementia processes (*APP, CSF1R, NOTCH3)* were also variably associated with cognitive impairment (2/6; 3/6 not assessed) and family histories of dementia (2/6) and Parkinson's disease (2/6).

The median Time to Diagnosis for these rare gene carriers was 16 months (IQR 8-28 months). This is longer than the overall median for the cohort (n=339; median 12 (IQR 8-23).

Phenotypic Characteristic	***VCP* (n=1)**	***ANXA11* (n=1)**	***APP* (n=1)**	***CCNF* (n=2)**	***CSF1R* (n=1)**	***DAO* (n=1)**	***DCTN1* (n=3)**	***ERBB4* (n=1)**
Sex	Female	Male	Male	Female Male	Female	Female	Male Male Male	Female
Ethnicity	A1A	A1A	A1A	A1A A1A	A1A	A1B	NA A1A A1A	A1A
Family History	MND (mother) Dementia (maternal grand-mother)	No	PD (father)	(i) No (ii) No	No	Dementia (brother), tremor (mother)	(i) No (ii) No (iii) No	No
Age of onset (years)	41	85	71	78 63	61	77	90 59 54	69
Time to Diagnosis (months)	44	16	9	1 7	3	29	4 37 24	11
Site of onset	Lower limb	Lower limb	Bulbar	Bulbar Lower limb	Mixed (limb)	Upper limb	Lower limb Lower limb Upper limb	Bulbar
Classification	ALS	ALS	PBP	ALS ALS	PMA	PMA	ALS ALS ALS	PBP
Cognitive Impairment	NA	NA	NA	NA No	Yes	Yes	No No Yes	No
Survival from Onset (Months)	75	27	25	7 20	36	52	9 76 47	18

Phenotypic Characteristic	*FUS* (n=2)	*GRN* (n=2)	*MAPT* (n=1)	*MATR3* (n=1)	*NEFH* (n=1)	*NOTCH3* (n=4)	*OPTN* (n=1)	*PRPH* (n=1)
Sex	Male Male	Female Female	Male	Male	Female	Male Female Female Female	Male	Female
Ethnicity	A1A A1A	A1A A1A	NA	A1A	A1B	A1B A1A A1Other A1A	A1A	A1A
Family History	(i) MND (maternal aunt) (ii) MND (father, sister)	(i) No (ii) EOD (sister), PD (father)	MND (unknown), Epilepsy (sister)	No	GBS (mother)	(i) Dementia (father), PD (paternal grandfather) (ii) No (iii) Dementia (mother) (iv) No	No	MS (maternal cousins x3, maternal aunt)
Age of onset (years)	47 41	75 68	77	56	76	74 76 62 51	30	66
Time to Diagnosis (months)	9 10	30 31	7	16	33	9 44 16 20	17	23
Site of onset	Bulbar Upper limb	Bulbar Bulbar	Bulbar	Upper limb	Bulbar	Bulbar Bulbar Bulbar Upper limb	Bulbar	Bulbar
Classification	ALS ALS	ALS PBP	PBP	ALS	PBP	ALS MND-FTD	ALS	ALS

						PBP ALS		
Cognitive Impairment	NA NA	NA Mild	NA	NA	No	NA Yes NA No	NA	NA
Survival from Onset (Months)	20 20	43 61	24	39	64	25 67 29 36	33	30

Phenotypic Characteristic	*PSEN1 (n=1)*	*PSEN2 (n=1)*	SETX (n=1)	SQSTM1 (n=2)	TBK1 (n=1)
Sex	Male	Male	Male	Male Male	Female
Ethnicity	A1Other	A1A	A1B	A1A A1A	A1A
Family History	No	PD (mother)	MND (mother, maternal grandmother) PD (father)	(i) Dementia (father) (ii) No	Depression (mother)
Age of onset (years)	60	78	43	57 56	64
Time to Diagnosis (months)	8	10	56	7 16	4
Site of onset	Cognition	Bulbar	Mixed (inc Bulbar)	Lower limb Upper limb	Upper limb
Classification	ALS	ALS	ALS	ALS ALS	ALS
Cognitive Impairment	NA	Yes	Yes	Yes NA	NA
Survival from Onset (Months)	24	15	70	29 39	10

Table 29 Key phenotypic characteristics of rare gene carriers. EOD = Early-onset Demetnia, GBS = Guillain Barre Syndrome, MS = Multiple Sclerosis, PD = Parkinson's Disease

7.4 PROGNOSTIC MODELLING: ADDING GENETIC DATA

The final process in the genotype-phenotype study was to assess the contribution of genetics to survival, in the context of other phenotypic variables. The "Pathogenic/VUS-P" group (n= 82) was considered the most informative genotype split as the most stringent statistical methods could be applied to this sample to obtain genotype-phenotype associations. This variable (Pathogenic Variant/VUS-P, Yes or No) was therefore as added to the prognostic model outlined in Chapter 4.3 with all missing data imputed (Figure 13 (iii)). Applying the same correlation matrix as previously, the only correlation was ECAS: ALS Non-Specific Score with ECAS: ALS Specific Score (0.61) and so the former was excluded. After excluding variables with near-zero variance and correlated variables, there remained 26 predictors. One hundred and seventy six individuals had the outcome of interest (death at follow-up). While this did not fulfil the 'one in ten' rule, there is certainly precedent for relaxing these rules in linear regression models[448]. Using all 26 predictors, a Family History of MND ($p=0.039$) and increasing ALSFRS-R Preslope ($p=0.0039$) predicted death whereas increasing Time to Diagnosis ($p<0.00001$), Classification: Other ($p=0.0032$) and a history of Ever Smoked ($p=0.047$) predicted survival (Figure 28). Having a Pathogenic Mutation or VUS-P did not influence survival ($p=0.12$). Further, Age of Onset ($p=0.081$) and Heavy Metal or Pesticide Exposure ($p=0.16$) were no longer significant. The overall R^2 for the model was 0.511 (95% CI 0.500-0.522) which was lower than the model which did not include genetic data, perhaps due to the higher number of predictors. Hazards ratios and confidence intervals for the genotype-phenotype prognostic model are summarised in Table 30, ranked by hazard ratio from most significant predictor of death to most significant predictor of survival.

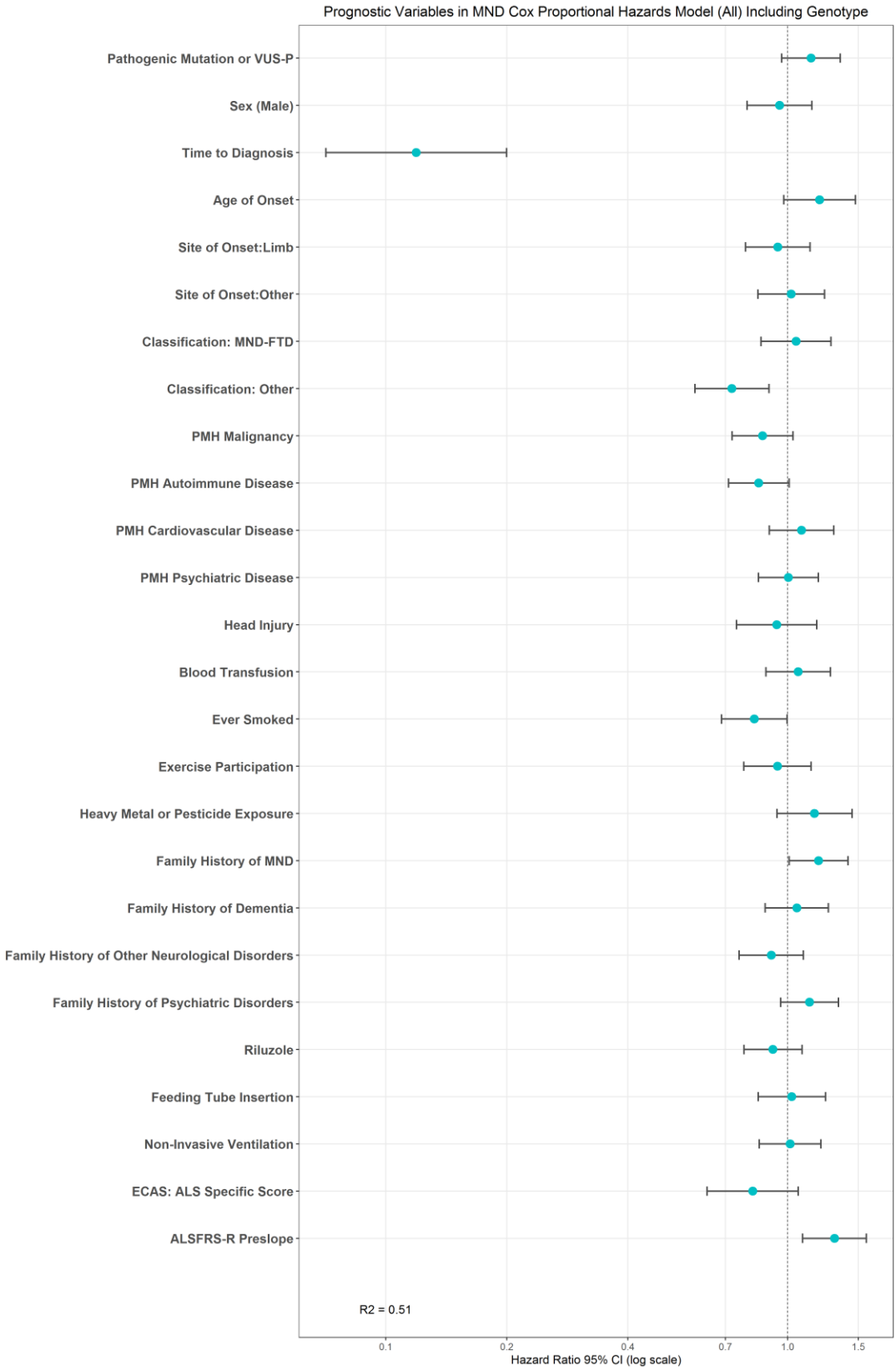


Figure 28 Cox Regression Model predicting survival with Pathogenic Mutation/VUS-P status (all missing data imputed)

Predictor	Hazard Ratio	95% CI
ALSFRS-R Preslope	**1.31**	**1.09-1.57**
Age of Onset	1.20	0.98-1.48
Family History of MND	**1.19**	**1.01-1.41**
Heavy Metal or Pesticide Exposure	1.17	0.94-1.45
Pathogenic Mutation or VUS-P	1.14	0.97-1.35
Family History of Psychiatric Disorders	1.13	0.96-1.34
PMH Cardiovascular Disease	1.08	0.90-1.30
Blood Transfusion	1.06	0.88-1.28
Family History of Dementia	1.05	0.88-1.26
Classification: MND-FTD	1.05	0.86-1.28
Feeding Tube Insertion	1.02	0.84-1.24
Site of Onset:Other	1.02	0.84-1.24
Non-Invasive Ventilation	1.01	0.85-1.21
PMH Psychiatric Disease	1.00	0.85-1.19
Sex (Male)	0.95	0.79-1.15
Site of Onset:Limb	0.95	0.79-1.14
Exercise Participation	0.94	0.78-1.14
Head Injury	0.94	0.75-1.18
Riluzole	0.92	0.78-1.09
Family History of Other Neurological Disorders	0.91	0.76-1.09
PMH Malignancy	0.87	0.73-1.03
PMH Autoimmune Disease	0.85	0.71-1.01
Ever Smoked	**0.83**	**0.68-1.00**
ECAS: ALS Specific Score	0.82	0.63-1.06
Classification: Other	**0.73**	**0.59-0.90**
Time to Diagnosis	**0.12**	**0.07-0.20**

Table 30 Hazard ratios of prognostic modelling predictors using genotype-phenotype model. ***bold*** *= significant predictor*

Finally, patients with pathogenic mutations only (including *C9orf72* expansions; n=41) were also studied within the same prognostic model. Again, 26 predictors were included in the model. Increasing ALSFRS-R Preslope (p=0.0011) predicted death. Family History of MND was no longer significant (p=0.081). Time to Diagnosis (p<0.00001), Classification: Other (p=0.0036) and a history of Ever Smoked (p=0.026) continued to predict survival (Figure 29).

Having a Pathogenic Mutation Only did not influence survival (p=0.087). The overall R^2 for the model was 0.517 (95% CI 0.509-0.525) which was higher than the model which included all potentially pathogenic mutations. Hazards ratios and confidence intervals for this model are summarised in Table 31.

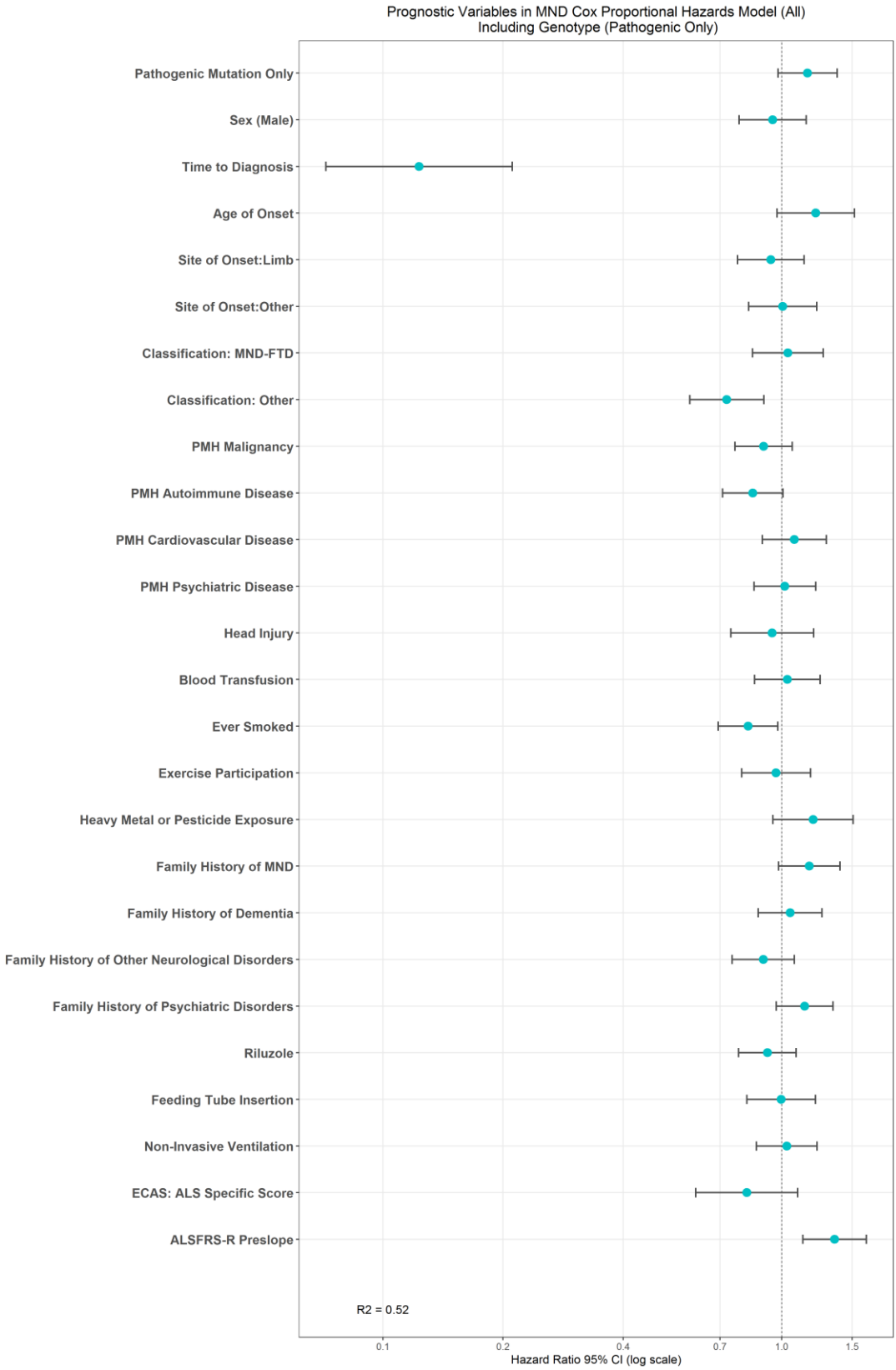


Figure 29 Cox Regression Model predicting survival with Pathogenic Mutation Only status (all missing data imputed)

Predictor	Hazard Ratio	95% CI
ALSFRS-R Preslope	**1.36**	**1.13-1.63**
Age of Onset	1.22	0.97-1.52
Heavy Metal or Pesticide Exposure	1.20	0.95-1.51
Family History of MND	1.17	0.98-1.40
Pathogenic Mutation Only	1.16	0.98-1.38
Family History of Psychiatric Disorders	1.14	0.97-1.35
PMH Cardiovascular Disease	1.08	0.89-1.29
Family History of Dementia	1.05	0.87-1.26
Classification: MND-FTD	1.04	0.85-1.27
Blood Transfusion	1.03	0.86-1.25
Non-Invasive Ventilation	1.03	0.86-1.23
PMH Psychiatric Disease	1.02	0.85-1.22
Site of Onset:Other	1.01	0.83-1.23
Feeding Tube Insertion	1.00	0.82-1.21
Exercise Participation	0.97	0.79-1.18
Sex (Male)	0.95	0.78-1.15
Head Injury	0.95	0.75-1.20
Site of Onset:Limb	0.94	0.77-1.14
Riluzole	0.92	0.78-1.09
PMH Malignancy	0.90	0.76-1.06
Family History of Other Neurological Disorders	0.90	0.75-1.08
PMH Autoimmune Disease	0.85	0.71-1.01
Ever Smoked	**0.82**	**0.69-0.98**
ECAS: ALS Specific Score	0.82	0.61-1.10
Classification: Other	**0.73**	**0.59-0.90**
Time to Diagnosis	**0.12**	**0.07-0.20**

Table 31 Hazard ratios of prognostic modelling predictors using genotype-phenotype model (pathogenic mutations only). bold = significant predictor

7.5 DISCUSSION

Genotype-phenotype associations

In this part of the analysis, I have used stringent statistical modelling with machine learning elements to identify genotype-phenotype associations. Machine learning approaches for feature selection were similar to univariate methods but the pseudo-R^2 and AIC were better. Similarity between the two methods suggests model stability and implies that we can make conclusions about the Scottish cohort based on the results.

With deeper phenotyping and genotyping, and an evolution in understanding of statistical methods, different predictors of having a genetic mutation were identified compared with our pilot analysis in Chapter 5, contradicting the hypothesis. In the latter analysis, Age of Onset and Female Sex loses significance. The male-to-female ratio in the gene carrying population remains low, although this did not reach Bonferroni-corrected significance on univariate testing (p=0.03). The differences between cohorts are likely related to more stringent statistical methods adopted latterly. A Family History of MND is strongly associated with having a genetic mutation, as predicted. However, a Family History of Other Neurological Conditions is also associated (p=0.027 in the logistic regression model). As described in Table 11, many patients in this category had family histories of Parkinson's Disease (PD) (34.6%) and multiple sclerosis (MS) (19.6%). There are also family histories of these other neurodegenerative diseases in rare gene carriers (Table 29). This may imply a shared genetic pathology. Indeed previous studies have observed higher incidences of PD in apparently sporadic MND and FTD kindreds[164,449,450].

Significantly fewer gene carriers undergo Feeding Tube Insertion, in spite of there being more people with PBP in this group. Factors which might impede gastrostomy placement include presence of cognitive impairment and rapidly progressive disease. More gene carriers than non-gene carriers were were cognitively impaired and had poorer survival and ALSFRS-R Preslopes, although this did not reach statistical significance. Individuals with PBP were significantly more likely to have a genetic mutation. As described in Chapter 1, PBP may be an artificial subtype of ALS: the majority of patients will progress to develop a more classical ALS phenotype[42]. Instead, the classification of PBP in this cohort may be a more accurate

measure of bulbar-onset disease. The Site of Onset variable is complicated by patient recall and 'mixed' onset presentation. By extension, we may hypothesise that gene-carriers are more likely to have bulbar-onset disease. This would support our previous link with *C9orf72* expansion carriers and bulbar-disease (the *C9orf72* expansion being the biggest monogenic contributor to Pathogenic/VUS-P cohort). However, these assumptions require validation using larger Scottish cohorts and an external validation cohort.

Univariate association testing of *C9orf72* expansion carriers showed that they have significantly poorer ALS-Specific ECAS scores, confirming the hypothesis. This finding parallels other studies and supports a link between *C9orf72* and MND-FTD spectrum disorders[75,128,132,133]. This highlights the utility and importance of early cognitive assessment using the ECAS assessment tool following diagnosis of MND. The penetrance for the *C9orf72* expansion is incomplete but is thought to be higher in MND than in pure FTD; it is also unaffected by prior family history of disease and increases with age[451]. Antisense oligonucleotide therapy trials are currently recruiting *C9orf72* expansion carriers[452]. Early identification of cognitive impairment will be crucial in the near future to guide appropriate genetic testing and ensure that patients are offered therapeutic intervention before they lose capacity.

While it did not meet Bonferroni-corrected threshold for significance, *C9orf72* expansion carriers in our population have poorer outcomes in terms of survival (p=0.006), as has been described in other studies[132,136]. Fewer people with *C9orf72* expansions were initiated on NIV, though, again, this did not meet the significance cut-off (p=0.003). *C9orf72* expansion carriers are thought to have fast respiratory decline[135]. From this we might infer that this population might not have had opportunity to commence on NIV due to inability to consent and comply with treatment (because of cognitive impairment) and because of rapidly progressive disease. Inclusion of Forced Vital Capacity respiratory measures in a longitudinal survival study might help to validate these findings. In the meantime, early assessment of *C9orf72* status and ECAS cognitive assessment in a clinical setting could guide intervention strategies and help to maximise patient access to available intervention. The male-to-female ratio in *C9orf72* expansion carriers was 50:50, ie. more females that would be expected in a typical MND cohort. In our study of a whole population cohort of patients in Chapter 3, we found that significantly fewer females than males were commenced on NIV (p<0.0001). This

was unexplained but, on reflection, *C9orf72* status might be a contributor. However, neither female sex nor NIV affects overall survival (Chapter 4).

We identified two patients with intermediate-length repeat expansions who both had classical ALS phenotypes and cognitive impairment. Intermediate repeats are more common in those with neuropsychiatric disease (including FTD) and our results provide further evidence of this phenotype.

In contrast to *C9orf72*, proportionally more *SOD1* carriers commenced NIV, though this did not reach Bonferroni-corrected significance (p=0.007). This may be due to their having more predictable limb-onset ALS disease. ECAS scores and survival of *SOD1* carriers were reflective of the population as a whole, suggesting that the mutation does not influence these factors. Patients with *SOD1* mutations tend not to have significant cognitive impairment[108] and so these findings were anticipated. Individuals who had the p.I114T mutation did not have any statistically significant phenotypic markers; however, all had limb-onset ALS suggesting that the variant may result in a more 'classical' form of MND. Indeed, none required gastrostomy insertion by the time of censorship, suggesting that bulbar disease is not a prominent feature. Family histories of individuals with the p.I114T variant revealed histories of limb-onset MND age similar ages to the proband, although details about disease site and onset are missing for some individuals. Family histories of MS were also apparent, perhaps implying either a shared genetic aetiology, or perhaps a misdiagnosis of phenotype.

While proportionally more patients with the *C9orf72* expansion had bulbar-onset disease and more patients with the *SOD1* mutation had limb-onset disease, this was not statistically significant. This contradicts predictions from the historical analysis in Chapter 5. This observation may relate to the more conservative statistical methods adopted in Chapter 7 and suggests that our initial findings are not necessarily stable or replicable.

Oligogenic patients also had no significantly different phenotypic markers. The sample size is very small and so this is unsurprising. Contrary perhaps to expectations, they appeared to have a slower rate of ALSFRS Preslope decline (p=0.004). However, this did not influence their overall survival.

Rare gene mutations were observed in this Scottish cohort, including a pathogenic mutation in *VCP.* These are rare in the UK population and are often associated with other diseases such as inclusion body myositis, FTD and Paget's disease[126]. Our patient had a family history of MND and dementia (there was insufficient information to determine if this was FTD) but no other diseases typical of the phenotype. The presence of rare potentially pathogenic mutations in MND and FTD/dementia-associated genes in the Scottish population suggests that these require further study. The median Time to Diagnosis for these rare gene carriers was 16 months (IQR 8-28 months). This is longer than the overall median for the cohort (n=339; median 12 (IQR 8-23). Early genetic testing of these individuals might have resulted in earlier diagnosis, better prognostication and more quality time for the patients and their families.

Prognostic Models

Rare variants have been shown to influence survival in a smaller study[453]. However, our larger study, which controls for a wide breadth of endogenous and exogenous variables, suggests that while genotype may influence disease onset, it is not an independent predictor of survival. A similar observation was seen in an Italian survival study which included *C9orf72*, *SOD1* and *TARDBP* status[178]. Clinical variables related to speed of progression of disease remained influential. It is possible that these factors overwhelm other biological hallmarks of disease once a patient has reached the critical threshold for disease manifestation. Age of Onset loses its significance in predicting death in the model that includes genetic data (HR 1.20 (95% CI 0.98-1.48) in model including genotype; HR 1.28 (95% CI 1.06-1.56) in model excluding genotype). We would expect increasing Age of Onset to be associated with worse prognosis and so the insignificance of this predictor is counter-intuitive. The confidence intervals are close to 1.0 and so this variable is vulnerable to instability when the number of predictors are high. Indeed, the R^2 of the model with genotyping is lower than the one without (0.51 versus 0.54) and the model does violate the 'one in ten' rule (26 predictors, 176 people with the outcome of interest). A solution would be to adopt a machine learning approach to regression modelling. This will be discussed further in Chapter 8.

Two groups of gene carriers were explored in the prognostic models: i) those with all potentially pathogenic gene mutations, and ii) those with clearly pathogenic/disease-causing mutations by strict ACMG-guided standards. Predictors of outcome remain relatively stable between the two models. However, Family History of MND loses its significance when combined with the Pathogenic Only gene variable. This reflects the raw data as definite pathogenic mutations are more highly correlated with family history than all potentially pathogenic mutations (correlation coefficients 0.38 and 0.30 respectively).

Limitations

The main constraint of these genotype-phenotype analyses are the small sample sizes, limiting the number of permitted multivariable analyses. Ways to overcome this would be i) to genotype more samples in the Scottish population, ii) to collaborate with other populations to increase sample sizes, and iii) to adopt machine learning methodology throughout the analysis, using methodology which might cope better with small numbers of event per variable. Options (i) and (iii) are achievable within the framework of CARE-MND and should be explored for future study. Option (ii) should also be explored, providing that the same framework for classification of variant pathogenicity is used between populations.

7.6 CONCLUSIONS

From this genotype-phenotype study using a wide variety of phenotypic variables and an extended gene panel, we can conclude that having pathogenic and VUS-P mutations may be associated with a shared pathological process with other neurodegenerative diseases. The *C9orf72* expansion is associated with cognitive impairment. Having a gene mutation did not influence survival when controlled for other phenotypic markers. Early clinical gene testing using an extended gene panel may help in the diagnostic process by shortening time to diagnosis. It may also guide management, either by prompting consideration of NIV prior to cognitive decline (*C9orf72*) or by preparing patients early for the likelihood of NIV (*SOD1*).

8. SUMMATIVE CONCLUSIONS

8.1 THESIS OBSERVATIONS

This thesis demonstrates what we can learn from a cohesive population-based study of a rare and life-limiting disease. Using the strengths of the Scottish healthcare system and Scottish government infrastructure, we have been able to create a harmonised platform for longitudinal assessment of pwMND. Through this, we have achieved 99% capture of incident pwMND over a three year period, allowing me to make conclusions about this population's phenotypes and genotypes though my PhD study. Unknowns relevant to Scottish pwMND and their care-providers have been elucidated and will now be summarised.

Question 1: Has the incidence of MND in Scotland changed over time?

The incidence of MND in Scotland is rising, in-keeping with anticipated global trajectories[21]. In the Scottish population, we suspect that this is due to better awareness of the broader phenotype of MND (including cognitive phenotypes), an increase in the number of MND clinical specialists and improved healthcare for competing diseases such as heart disease and stroke disease. The incidence of MND in Scotland is the highest reported standardised incidence of the disease globally[253]. This likely reflects our ascertainment processes and the infrastructure of CARE-MND. However, ongoing assessment of incidence post-2017 with comparison against up-to-date European population statistics will be required to see if these observations are sustained. An opposing argument is that they reflect recent public awareness of MND in Scotland, which has prompted a surge in diagnosis by healthcare providers.

The annual prevalence of MND ranged from 7.61-7.81/100,000 population. Compared with some other European populations, this is low (10.3 in the Netherlands 2008[266], 11.2 in Modena 2009[265], 10.5 in Piemonte and Valle D'Aosta 2014 (although this rose to 12.3 with the inclusion of 78 prevalent patients who had undergone tracheostomy)[257]). Survival from onset for the first two cohorts was 2.9 and 3.0 years respectively and in both cases

survival had improved over time, likely due to better multidisciplinary care[265,266]. In this thesis, I have observed a median survival from onset of 3.5 years in the historical cohort (diagnosed 1989-2014). However, in the incident cohort (mean follow-up time 1.9 years), 55% of patients were deceased. Longer follow-up of the 2015-17 incident cohort is required to study longer survivors in this group but our observations so far, combined with the low prevalence of disease, are discouraging and suggest that current interventions for MND do not have a significant survival benefit in Scotland (see Question 2).

Nevertheless, current awareness of incidence and prevalence will now allow us to anticipate service needs better. The MND Nurse Consultant for Scotland is able to use this information as evidence for economic and healthcare investment. Estimated annual incidence is important for power calculations for future clinical trials. Knowledge of epidemiology is important for long-term services planning, particularly in the climate of imminent drug therapies. The implementation of genetically targeted treatments, particularly those for *C9orf72* expansion carriers, is expected to result in a rise in prevalence of MND by up to 50% by 2066[454]. Such predictions should justify ongoing investment into MND care in Scotland.

Question 2: Can deep clinical phenotyping of the Scottish MND population reveal predictors of survival?

During this population study, I initially undertook limited phenotypic characterisation of an incident population cohort and a historical cohort of genotyped individuals with MND. These analyses revealed hypothesis generating questions which were explored further on deep clinical phenotyping of an incident research cohort of patients (n=437) using the CARE-MND platform. Using this information, I can summarise that Scottish individuals with MND are predominantly male and of White Scottish ethnicity, with a male-to-female ratio of 1.7:1. They tend to develop symptoms of disease age 64 years and receive a diagnosis age 66 years. The average time to diagnosis is 12 months. Lower limb is the most common site of disease onset and ALS the most typical manifestation. Six percent of the population are diagnosed with MND-FTD from the outset; however, cognitive impairment of some degree is present in 39% of patients. Uptake of interventions are variable, with 40% taking riluzole, 31% having

a feeding tube inserted and 27% being commenced on NIV at time of censorship (median follow-up 23 months).

I have shown that survival can be predicted at diagnosis and that models with multiple imputation for missing data perform better than those that exclude missing data. Using a fully imputed model, I found that steeper decline in the ALSFRS-R slope (ALSFRS-R performed within six months of diagnosis) and a shorter time to diagnosis predicts poorer survival. Median rate of decline is 0.6 points per month and median time to diagnosis 12 months. Patients with PLS, PMA and PBP have better outcomes. Older age at onset also predicts death. In summary, these findings suggest that increased burden of upper and lower motor neurone degeneration in older, vulnerable patients results in faster progression to death. A family history of MND suggests poorer outcome; however, having a pathogenic or potentially pathogenic mutation did not predict survival. This might suggest that there are, as yet, undiscovered genetic influences of disease. Another possibility may relate to the "gene-time-environment (GTE)" hypothesis as described in Chapter 1[160]. Once a patient with a genetic predisposition has reached a critical threshold, or accumulated the required risk burden for disease manifestation, there is a rapid cascade of decline related to the patient's innate vulnerability. In our models, post-onset factors may overwhelm the influence of premorbid characteristics.

Interventions such as riluzole, gastrostomy and NIV did not influence survival when controlled for other variables. The overall median survival from onset was 30 months (2.5 years); by two years, 49% of this research cohort were deceased. This is unchanged compared to a Scottish study published in 1993, in spite of the fact that uptake of riluzole and gastrostomy has increased substantially (5.6% vs 40.0% and 11.6% vs 31.3% respectively)[203]. Although longer follow-up time is required (as described above), it suggests that current interventions for MND in Scotland do not have significant survival benefit.

Exogenous variables - heavy metal/pesticide exposure and smoking – may negatively and positively influence survival, respectively. Case-controls studies in the Scottish population are required to replicate these findings. However, smoking remained protective of survival

in all models studied through this PhD. Potential theories underlying this observation relate to its effects on inflammation and the gut microbiome.

While the derived prognostic models do not violate the 'one-in-ten' rule, there are a high number of predictors in the context of a small sample size. As such, there is a high chance of overfitting of the data. Ways to overcome this will be explored in Chapter 8.3. The prediction observations are also only derived from the Scottish population; we have not yet validated them on an external cohort and so findings are not generalisable. However, close parallels have been seen in similar analyses from Italian registry studies[178]. This study examined the influence of similar phenotypic markers and clinical interventions and found that age of onset, diagnostic delay, ALS phenotype, presence/absence of dementia, riluzole use, BMI and El Escorial classification predicted survival using multivariable Cox regression. However, in this analysis, non-significant variables were 'dropped' from the model. Results must therefore be interpreted with caution as non-inclusion of non-significant variables may inflate the significance of remaining variable (or, reduce the suppression effect of these non-significant variables[190]).

Question 3: What is the genetic epidemiology of MND in Scotland? In particular, are there mutations unique to the Scottish population and are rare/newly described mutations present in this population?

Following a hypothesis-generating pilot study, we used an adapted MND-specific ACMG framework to test a comprehensive gene panel on an unselected cohort of incident pwMND in Scotland. To our knowledge, this is the first complete, manual application and assessment of utility of the ACMG guidelines for MND genomic data. This has highlighted classification criteria that are relevant and irrelevant to MND-specific genes, thereby identifying targets for improvement (eg. Focus on functional studies and sourcing of segregation data).

Using this framework, up to 22% of patients have a potentially causative gene mutation. This is higher than previously described estimates[77]. Two percent were oligogenic, in-keeping with other population estimates[94,378,442]. In the MND community, monogenic causes of MND are thought to be present in 70% of familial and 10% of apparently sporadic

individuals[455]. In this study, 68% of familial cases and 20% of apparently sporadic were thought to have potential pathogenic mutations. However, only 54% of familial cases and 10% of apparently sporadic cases had definite Pathogenic mutations. The *C9orf72* expansion is the biggest contributor, affecting 32% of familial and 6% of sporadic patients, matching current estimates of frequency (35% and 5% respectively)[455]. However, *SOD1* mutations are also important, particularly the p.I114T variant. *SOD1* variants were found in 2% of sporadic individuals and 18% of familial. These number are similar to quoted estimates[455] but, to our knowledge, Scotland has the highest frequency of p.I114T variant.

Between the historical and incident Scottish cohorts, six loss of function *NEK1* variants (absent from controls) were identified. As loss of function is the presumed mechanism of disease causation in this gene[90], our data provide evidence for association of this gene with MND in the Scottish population and suggest that it should be included in future clinical gene panels. Loss of function variants in *TBK1* were identified in the historical cohort but not in the incident cohort, where only one potentially pathogenic missense *TBK1* variant was identified in a patient with limb-onset ALS and short survival (10 months from onset). This collective information neither confirms nor discounts *TBK1* as a disease-causing gene in the Scottish population; we therefore recommend that clinical panels include this gene to further knowledge of its role and avoid missing a potentially inheritable cause of MND.

There remains considerable uncertainty regarding VUS-P and VUS in rare genes and genes associated with dementia phenotypes. Sharing of these data with public databases and collaboration with other population cohorts may help to clarify their significance within the wider MND community (see Future Direction).

Question 4: Can deep clinical phenotyping and extended genotyping of an incident MND cohort identify genotype-phenotype correlates (particularly with regard to *C9orf72* and *SOD1*)?

Question 5: Does genotype influence survival in a multivariable prognostic model of MND in Scotland?

Genotype-phenotype study of these individuals shows a significant relationship between having a mutation and having a family history of other neurological conditions, including Parkinson's disease and multiple sclerosis, perhaps suggesting a genetic spectrum across these neurodegenerative diseases. Having a *C9orf72* expansion is associated with an increased risk of cognitive impairment and, perhaps, poorer survival, and these factors may preclude NIV initiation. Alternatively, cognitive impairment provides a barrier to NIV and so hastens clinical deterioration. Early genetic testing of people with suspected MND may provide clues towards likely prognosis and development of cognitive impairment, and guide early intervention.

We have outlined the phenotypic characteristics of the largest series of patients with the *SOD1* p.I114T variant to our knowledge (n=27 across both pilot and incident cohorts; 3.5% of cases overall). All but one of these cases had limb-onset disease and the majority had a classical ALS phenotype. They were significantly associated with having a family history of MND, suggesting penetrance of this variant across generations. A previous study of this variant suggested phenotypic heterogeneity and variable penetrance within a family[144]. Our study alludes to an association of MS within families with this variant and it would be interesting to review the phenotypic characteristics of these family members.

Genotype-phenotype descriptions of rare variants are hypothesis generating. The high frequency of VUS-P associated with cognitive phenotypes/dementia supports the need for early cognitive assessment and exploration of family histories of dementia and neuropsychiatric symptoms. Genotype annotation of CARE-MND patients provides a valuable resource whereby individuals with specific phenotypes and genotypes can be selected for disease modelling, rather than retrospectively sourcing phenotypic information

for gene carriers. For example, patients with rare variants and cognitive impairment might be useful targets for pluripotent stem cell models.

Genetic results were not fed back to patients from this research study. Currently, clinical genetic testing in Scotland is only offered routinely to those with a family history of MND. Following the results of our pilot study (Chapter 5), a clinic was set-up in the West of Scotland to offer genetic counselling and testing to individuals with young-onset (<50 years) apparently sporadic disease. Similar approaches have been adopted by Northeast ALS Consortium members in the USA[171]. After spending time at Columbia University, I learned that genetic results, including those obtained from research studies, are routinely reported back to patients. Although there are concerns about patient anxiety in the clinic on receiving information about a variant of uncertain significance, variant classification multidisciplinary team meetings operate to ensure that patients are given appropriately full and clear information. Although this is not part of routine clinical practice in the UK, growing knowledge of genetics amongst clinicians and patients will likely require consideration of this in the near future. This would be the only way to learn more about the implications of rare gene variants in the Scottish population and would more easily permit trio studies of parents. Further, and crucially, a non-paternalistic transparent approach to gene testing would allow individuals to make more informed decisions regarding family planning and pre-implantation genetic diagnoses. Variable penetrance of mutations makes this area especially challenging but denying patients the relevant information is equally ethically questionable.

Finally, inclusion of genetic data in a multivariable survival model revealed that genetic mutation status was not a significant predictor of death. From this, we can hypothesise that genetic mutations predispose to, but do not necessarily perpetuate, disease. However, the number of predictors in this model was high and so validation is required with a bigger sample size and using alternative statistical methods.

The ultimate aim is that these observational studies will inform national guidelines of MND in Scotland and translate directly to the patients and families who have contributed. Already, we have been able to provide patients with answers to questions about the Scottish MND population (Figure 30).

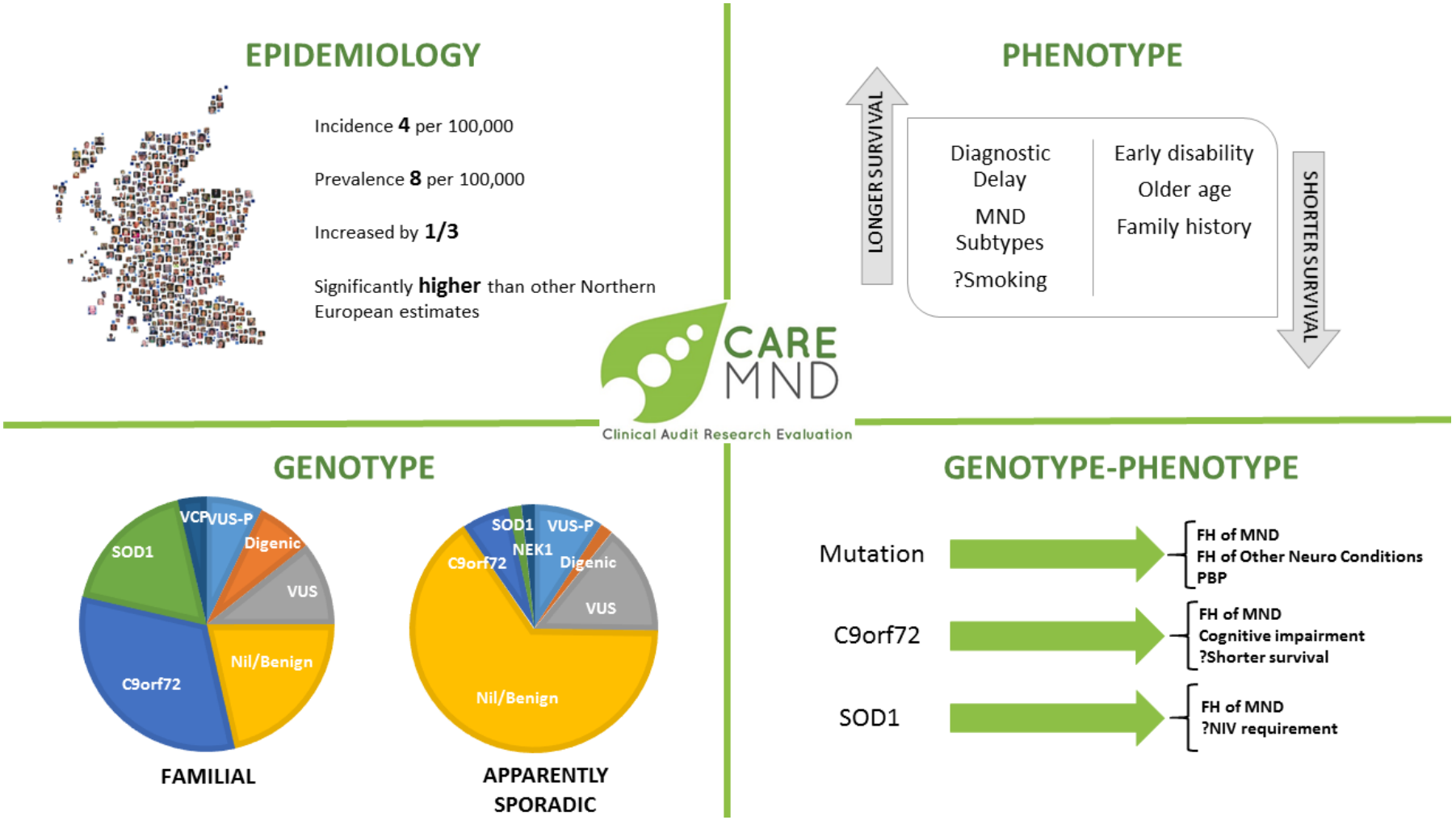


Figure 30 Summary of PhD findings for patient and public information. PLS = Primary Lateral Sclerosis; FH = Family History; VUS = Variant of Uncertain Significance

8.2 STUDY LIMITATIONS

This PhD benefits from a population cohort studied in 2015-17. Data prior to 2015 (1989-2014) were also analysed. However, a complete longitudinal picture is not possible due to differences in capture of phenotypic characteristics over time. As described in Chapter 3, the El Escorial classification was established during this time period and has evolved. Recognition of cognitive impairment in MND has also emerged over the past two decades.

Our models only include data from consenting patients on the CARE-MND Register. Patients without capacity at presentation are therefore excluded; we might expect such patients to have more rapid progress of disease. While research recruitment is addressed at the earliest opportunity in Scotland, such barriers are inevitable. There were also insufficient data to be able to include body mass index (BMI), electrophysiology, spirometry or imaging biomarkers in these analyses. BMI and post-diagnostic weight loss can be predictors of death[58]. Forced vital capacity (FVC) measurements have also been shown to be key predictors of early decline[6,179]. However, these variables are part of the CARE-MND platform proforma and should be included in future studies dedicated to specific exploration of each modality, whilst controlling for our predictive variables. One factor, which is not yet covered by CARE-MND, is a measurement of quality of life. This is independently related to survival outcome and could be incorporated into patient-centred questionnaires and future survival models[58].

This study provides a proof-of-concept of the viability of deep clinical phenotyping in a routinely-collected, population setting. However, a trade-off for a broad set of phenotypic variables is the lack of detail for each variable. For example, exogenous variables on CARE-MND exist as screening questions (for example, history of smoking: Yes/Ex/Never, with a free text box for comments). Answers are therefore crude and hypothesis generating for future case-control studies, which should include stratification of these factors (for example, pack year history, dates of starting and stopping smoking). Variables may also be subject to recall bias (for example, history of heavy metal and pesticide exposure, history of blood transfusion). Attempts to overcome this included a review of ISD hospital records and general practice information records but these are not exhaustive.

Throughout my PhD study period, my understanding of genetic variant classification and statistical modelling has evolved, corresponding with changes in the MND literature-base. Chapter 5 (genotype-phenotype study of a historical cohort) was hypothesis-generating and was published in 2016. Through this part of the PhD, I identified that methods for assessment of variant pathogenicity were arbitrary and inadequate, which prompted a systematic review of the ACMG guidelines and a more transparent approach to classification.

There has also been a move towards more sophisticated and systematic approaches for survival modelling, including machine learning[179,456]. During this PhD, univariate and then inclusion of near-significant variables in a multivariable model using SPSS was adopted in Chapter 5. However, R programming was adopted latterly to allow for more advanced and flexible analyses. This included imputation and machine learning approaches for feature selection. Missing data is an expected repercussion of routinely-collected population data. Exclusion of missing data risks losing individuals' valuable contributions to research. Although imputation resulted in clinically appropriate models, this method can come under criticism in medical research. One argument for imputation is the well-established and validated QRISK score for cardiovascular disease. In the development of this model, some variables included had up to 70% missing data, yet it is still considered an important clinical tool[194].

As described in Chapter 7, the numbers of individuals with genetic mutations were too small to allow for gene-specific genotype-phenotype study using multivariable models. We recognise that collaboration and contribution to larger datasets, such as Project MinE[457], will be a key application of this work.

Finally, the follow-up times for this project were short, constrained by the time frame of the PhD (mean follow-up time was 23.0 months (SD 9.2, range 6.0–39.0 months)). Longer follow-up may change our prognostic models and highlight as yet unseen benefits of NIV or gastrostomy.

8.3 FUTURE DIRECTION

The integrated MND care-research infrastructure will enable Scotland to become a leader in a stratified medical approach to individual patient care, resulting in better focused clinical trials. Parallels with other newly–launched national registries, such as the Swedish MND Quality Registry, the New Zealand MND Registry and the developing MND Register of England, Wales and Northern Ireland, will allow us to pursue extensive collaboration for patient centred-discovery science[458,459].

The findings in this thesis suggest that stratification of patients for clinical trials by site of onset is likely artificial. Clinical trial patients in Scotland should ideally be stratified by age of onset, family history of MND and El Escorial classification of disease. Time to diagnosis, ALSFRS-R preslope and smoking status should be controlled for when analysing outcome data. Register data in isolation cannot make conclusions about environmental risk factors, but combining data with control data from the same population in a case-control study would help to assess the impact of our environmental risk factors such as smoking and heavy metal/pesticide exposure[227]. Scotland is in a perfect position to perform such studies, with the availability of the Lothian Birth Cohorts and Scottish Government ISD data[460]. While this could provide important answers to questions about local environmental risk factors, perhaps the best use of CARE-MND data would be to collaborate internationally and use our data on a greater scale. Through CARE-MND we have endeavoured to make our data collection internationally compatible to streamline this process.

The prognostic models would benefit from longer follow-up time. This can be achieved through CARE-MND. I have demonstrated the use of machine learning methods for feature selection. However, other feature selection approaches could be explored, such as principle component analysis (PCA). PCA methods were not undertaken during this thesis as it is often difficult to make clinically-relevant assumptions from PCA groupings. Novel approaches include “concept-based” variable reduction whereby variables are grouped by clinical interaction[188]. It would be interesting to apply this to our Scottish dataset. Machine learning methodology also has the capability to perform cross-validated regression models. Examples include elastic net and random forest models. These have been explored

previously in ALS cohorts, with random forest models performing superiorly[179,188]. These models have the advantage of being able to cope with datasets where the number of predictors approaches the number of variables. Additional granularity could therefore be incorporated into our models, including FVC measures, a separating out of MND subtypes, and the inclusion of ALSFRS-R and ECAS subscores. ALSFRS-R motor and bulbar subscores have been found to provide additional information to recalled site of onset data.

Further dissection of our results could be achieved through a causal mediation analysis to elucidate the potential interactions between predictors in the multivariable models[461]. For example, there is clearly an interaction between family history of MND and the presence of pathogenic mutations but I have not yet been able to prove if the influence of family history on survival is mediated by results of our gene panel testing, or if there are other unknown genetic modifiers.

While the models described in this thesis are only relevant to the 2015-17 incident Scottish population, machine learning models would be more generalisable to other populations. Internal validation on a new Scottish dataset (eg. 2018-19 incident cohort) would be the next step. Gold-standard external validation on another population cohort would be ideal, perhaps through our collaborators at Columbia University or through other UK population databases such as the new MND Register for England and Wales. These analyses will be explored over the coming year. While the methods and results of the prognostic models are reported in this thesis with reference to TRIPOD guidelines, clear clinical applications of a model are not yet delineated, pending the above modifications[195,196]. The final model will ensure that full TRIPOD recommendations are met.

We recommend that the extended 49-gene panel be considered as a diagnostic panel for Scotland. Results from the research study will be communicated to the South-East Scotland laboratories to inform future variant interpretation. Ideally, Scottish variants would be uploaded into a centralised database so that we can learn more about the implications of rare variants in the Scottish population, determine heritability through family segregation or trio studies, and perhaps even identify new founder mutations. Scottish genetics laboratories are currently exploring this possibility and plan to work with our clinical neurology teams to develop multidisciplinary approaches to testing and variant classification.

By setting up this kind of infrastructure, anchored through CARE-MND, we would have access to a rich resource for selection of participants for disease modelling/stem cell studies and for genetically stratified drug trials. Ideally, this would also contribute to, or act as a stepping-stone, for a multi-site open-access database of annotated MND-associated gene variants through which the MND research and clinical community can continue to reassess variants, and identify missing classification criteria and targets for future research.

In conclusion, this PhD has set the scene for the future of MND discovery and treatment in Scotland. Through a population-wide model of participation, personalisation, prediction and prevention, we can move towards further advancement in the field and set precedent for other neurological diseases.

REFERENCES

1. Turner MR, Swash M. The expanding syndrome of amyotrophic lateral sclerosis: A clinical and molecular odyssey. J Neurol Neurosurg Psychiatry. 2015 Feb 2;86(6):667–73.
2. Logroscino G, Traynor BJ, Hardiman O, Chió A, Mitchell D, Swingler RJ, et al. Incidence of amyotrophic lateral sclerosis in Europe. J Neurol Neurosurg Psychiatry. 2010 Apr 1;81(4):385–90.
3. Forbes RB, Colville S, Parratt J, Swingler RJ. The incidence of motor nueron disease in Scotland. J Neurol. 2007 Jul;254(7):866–9.
4. Chiò A, Calvo A, Moglia C, Mazzini L, Mora G. Phenotypic heterogeneity of amyotrophic lateral sclerosis: a population based study. J Neurol Neurosurg & Psychiatry. 2011 Jul 1;82(7):740 LP-746.
5. Scialò C, Novi G, Bandettini di Poggio M, Canosa A, Sormani MP, Mandich P, et al. Clinical epidemiology of amyotrophic lateral sclerosis in Liguria, Italy: An update of LIGALS register. Amyotroph Lateral Scler Front Degener. 2016 Jul 11;17(7–8):535–42.
6. Westeneng HJ, Debray TPA, Visser AE, van Eijk RPA, Rooney JPK, Calvo A, et al. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. Lancet Neurol. 2018 May 1;17(5):423–33.
7. Rooney J, Byrne S, Heverin M, Corr B, Elamin M, Staines A, et al. Survival analysis of irish amyotrophic lateral sclerosis patients diagnosed from 1995-2010. PLoS One. 2013 Jan;8(9):e74733.
8. Strong MJ, Abrahams S, Goldstein LH, Woolley S, Mclaughlin P, Snowden J, et al. Amyotrophic lateral sclerosis - frontotemporal spectrum disorder (ALS-FTSD): Revised diagnostic criteria. Amyotroph Lateral Scler Front Degener. 2017 Apr 3;18(3–4):153–74.
9. National Institute for Health and Care Guidance. Motor neurone disease: assessment and management | Guidance and guidelines | NICE [Internet]. NICE guideline. NICE; 2016 [cited 2016 Oct 22]. Available from: https://www.nice.org.uk/guidance/ng42
10. Miller RG, Jackson CE, Kasarskis EJ, England JD, Forshew D, Johnston W, et al. Practice parameter update: The care of the patient with amyotrophic lateral sclerosis: Drug, nutritional, and respiratory therapies (an evidence-based review): Report of the quality standards subcommittee of the American academy of neurology. Neurology. 2009 Oct 13;73(15):1218–26.
11. Lacomblez L, Bensimon G, Leigh PN, Guillet P, Meininger V. Dose-ranging study of riluzole in amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis/Riluzole Study Group II. Lancet (London, England). 1996 May 25;347(9013):1425–31.
12. Bensimon G, Lacomblez L, Delumeau JC, Bejuit R, Truffinet P, Meininger V, et al. A study of riluzole in the treatment of advanced stage or elderly patients with amyotrophic lateral sclerosis. J Neurol. 2002 May 1;249(5):609–15.
13. Bensimon G, Lacomblez L, Meininger V. A Controlled Trial of Riluzole in Amyotrophic Lateral Sclerosis. N Engl J Med. 1994 Mar 3;330(9):585–91.
14. Writing Group K, Edaravone (MCI-186) ALS 19 Study Group M, Tsuji S, Itoyama Y, Sobue G, Togo M, et al. Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled

trial. Lancet Neurol. 2017 Jul 1;16(7):505–12.
15. Hardiman O, van den Berg LH. Edaravone: a new treatment for ALS on the horizon? Lancet Neurol. 2017 Jul 1;16(7):490–1.
16. Löser C, Aschl G, Hébuterne X, Mathus-Vliegen EMH, Muscaritoli M, Niv Y, et al. ESPEN guidelines on artificial enteral nutrition--percutaneous endoscopic gastrostomy (PEG). Clin Nutr. 2005 Oct;24(5):848–61.
17. Burkhardt C, Neuwirth C, Sommacal A, Andersen PM, Weber M. Is survival improved by the use of NIV and PEG in amyotrophic lateral sclerosis (ALS)? A post-mortem study of 80 ALS patients. Zhou R, editor. PLoS One. 2017 May 23;12(5):e0177555.
18. Bourke SC, Tomlinson M, Williams TL, Bullock RE, Shaw PJ, Gibson GJ. Effects of non-invasive ventilation on survival and quality of life in patients with amyotrophic lateral sclerosis: a randomised controlled trial. Lancet Neurol. 2006 Feb;5(2):140–7.
19. Aoun SM, Bentley B, Funk L, Toye C, Grande G, Stajduhar KJ. A 10-year literature review of family caregiving for motor neurone disease: Moving from caregiver burden studies to palliative care interventions. Palliat Med. 2013 May 20;27(5):437–46.
20. Chiò A, Logroscino G, Traynor BJ, Collins J, Simeone JC, Goldstein LA, et al. Global epidemiology of amyotrophic lateral sclerosis: A systematic review of the published literature. Vol. 41, Neuroepidemiology. NIH Public Access; 2013. p. 118–30.
21. Arthur KC, Calvo A, Price TR, Geiger JT, Chiò A, Traynor BJ. Projected increase in amyotrophic lateral sclerosis from 2015 to 2040. Nat Commun. 2016;7(12408).
22. Murphy M, Quinn S, Young J, Parkin P, Taylor B. Increasing incidence of ALS in Canterbury, New Zealand. Neurology. 2008 Dec 2;71(23):1889 LP-1895.
23. Alonso A, Logroscino G, Jick SS, Hernán MA. Incidence and lifetime risk of motor neuron disease in the United Kingdom: a population-based study. Eur J Neurol. 2009 Jun;16(6):745–51.
24. Connolly S, Heslin C, Mays I, Corr B, Normand C, Hardiman O. Health and social care costs of managing amyotrophic lateral sclerosis (ALS): An Irish perspective. Amyotroph Lateral Scler Front Degener. 2015 Mar 31;16(1–2):58–62.
25. Turner MR. The reunification of amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 2018 Oct 8;jnnp-2018-319470.
26. Brown RH, Al-Chalabi A. Amyotrophic Lateral Sclerosis. Longo DL, editor. N Engl J Med. 2017 Jul 13;377(2):162–72.
27. Brooks BR, Miller RG, Swash M, Munsat TL. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph Lateral Scler Other Motor Neuron Disord. 2000 Dec;1(5):293–9.
28. Brooks BR. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and th. J Neurol Sci. 1994 Jul;124 Suppl:96–107.
29. de Carvalho M, Dengler R, Eisen A, England JD, Kaji R, Kimura J, et al. Electrodiagnostic criteria for diagnosis of ALS. Clin Neurophysiol. 2008 Mar;119(3):497–503.
30. Costa J, Swash M, de Carvalho M. Awaji Criteria for the Diagnosis of Amyotrophic Lateral Sclerosis A Systematic Review. Arch Neurol. 2012;69(11):1410–6.
31. McCombe PA, Henderson RD. Effects of gender in amyotrophic lateral sclerosis. Gend Med. 2010 Dec;7(6):557–70.
32. Gautier G, Verschueren A, Monnier A, Attarian S, Salort-Campana E, Pouget J. ALS

with respiratory onset: clinical features and effects of non-invasive ventilation on the prognosis. Amyotroph Lateral Scler. 2010 Aug;11(4):379–82.
33. Moirangthem V, Micheal Ouseph M. Atypical Presentations of Amyotrophic Lateral Sclerosis: A Case Report Case Report. J Neuropsychiatry Clin Neurosci. 2011;23:3.
34. Charcot J. Sclerose des cordons lateraux de la moelle epinere chez femmehysterique atteinte de contracture peranemte des quatre membres. Bull Soc Med Hop Paris. 1965;2(2):24.
35. Gordon PH, Cheng B, Katz IB, Pinto M, Hays AP, Mitsumoto H, et al. The natural history of primary lateral sclerosis. Neurology. 2006 Mar 14;66(5):647–53.
36. Bruyn RPM, Koelman JHTM, Troost D, De Jong JMBV. Motor neuron disease (amyotrophic lateral sclerosis) arising from longstanding primary lateral sclerosis. J Neurol Neurosurg Psychiatry. 1995;58(6):742–4.
37. Gastaut JL, Bartolomei F. Mills' syndrome: ascending (or descending) progressive hemiplegia: a hemiplegic form of primary lateral sclerosis? J Neurol Neurosurg Psychiatry. 1994 Oct 1;57(10):1280–1.
38. Turner MR, Gerhard A, Al-Chalabi A, Shaw CE, Hughes RAC, Banati RB, et al. Mills' and other isolated upper motor neurone syndromes: in vivo study with 11C-(R)-PK11195 PET. J Neurol Neurosurg Psychiatry. 2005 Jun 1;76(6):871–4.
39. Kim WK, Liu X, Sandner J, Pasmantier M, Andrews J, Rowland LP, et al. Study of 962 patients indicates progressive muscular atrophy is a form of ALS. Neurology. 2009;73(20):1686–92.
40. Turner MR, Talbot K. Mimics and chameleons in motor neurone disease. Vol. 13, Practical Neurology. BMJ Publishing Group Ltd; 2013. p. 153–64.
41. Burrell JR, Vucic S, Kiernan MC. Isolated bulbar phenotype of amyotrophic lateral sclerosis. Amyotroph Lateral Scler. 2011 Jul 27;12(4):283–9.
42. Karam C, Scelsa SN, MacGowan DJL. The clinical course of progressive bulbar palsy. Amyotroph Lateral Scler. 2010 Aug 5;11(4):364–8.
43. Hu MT, Ellis CM, Al-Chalabi A, Leigh PN, Shaw CE. Flail arm syndrome: a distinctive variant of amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 1998 Dec;65(6):950–1.
44. Kiernan JA, Hudson AJ. Frontal lobe atrophy in motor neuron diseases. Brain. 1994 Aug;117 (Pt 4):747–57.
45. Gregory JM, McDade K, Bak TH, Pal S, Chandran S, Smith C, et al. Executive, language and fluency dysfunction are markers of localised TDP-43 cerebral pathology in non-demented ALS. J Neurol Neurosurg Psychiatry. 2019 Sep 12;jnnp-2019-320807.
46. Abrahams S, Leigh PN, Goldstein LH. Cognitive change in ALS: A prospective study. Neurology. 2005;64(7):1222–6.
47. Neary D, Snowden J., Mann DM. Cognitive change in motor neurone disease/amyotrophic lateral sclerosis (MND/ALS). J Neurol Sci. 2000 Nov 1;180(1–2):15–20.
48. Convery R, Mead S, Rohrer JD. Review: Clinical, genetic and neuroimaging features of frontotemporal dementia. Neuropathol Appl Neurobiol. 2019 Feb 1;45(1):6–18.
49. Grace GM, Orange JB, Rowe A, Findlater K, Freedman M, Strong MJ. Neuropsychological functioning in PLS: a comparison with ALS. Can J Neurol Sci. 2011 Jan;38(1):88–97.
50. de Vries BS, Rustemeijer LMM, van der Kooi AJ, Raaphorst J, Schröder CD, Nijboer TCW, et al. A case series of PLS patients with frontotemporal dementia and overview of the literature. Amyotroph Lateral Scler Front Degener. 2017 Oct 2;18(7–8):534–48.

51. Raaphorst J, de Visser M, van Tol M-J, Linssen WHJP, van der Kooi AJ, de Haan RJ, et al. Cognitive dysfunction in lower motor neuron disease: executive and memory deficits in progressive muscular atrophy. J Neurol Neurosurg Psychiatry. 2011 Feb 1;82(2):170–5.
52. Olney RK, Murphy J, Forshew D, Garwood E, Miller BL, Langmore S, et al. The effects of executive and behavioral dysfunction on the course of ALS. Neurology. 2005 Dec 13;65(11):1774–7.
53. Lillo P, Mioshi E, Hodges JR. Caregiver burden in amyotrophic lateral sclerosis is more dependent on patients' behavioral changes than physical disability: a comparative study. BMC Neurol. 2012 Dec 7;12(1):156.
54. Turner MR, Barnwell J, Al-Chalabi A, Eisen A. Young-onset amyotrophic lateral sclerosis: Historical and other observations. Brain. 2012;135(9):2883–91.
55. Pupillo E, Messina P, Logroscino G, Beghi E. Long-term survival in amyotrophic lateral sclerosis: a population-based study. Ann Neurol. 2014 Feb;75(2):287–97.
56. Turner MR, Parton MJ, Shaw CE, Leigh PN, Al-Chalabi A. Prolonged survival in motor neuron disease: a descriptive study of the King's database 1990-2002. J Neurol Neurosurg Psychiatry. 2003 Jul 1;74(7):995–7.
57. Mateen FJ, Carone M, Sorenson EJ. Patients who survive 5 years or more with ALS in Olmsted County, 1925-2004. J Neurol Neurosurg Psychiatry. 2010 Oct;81(10):1144–6.
58. Chiò A, Logroscino G, Hardiman O, Swingler R, Mitchell D, Beghi E, et al. Prognostic factors in ALS: A critical review. Amyotroph Lateral Scler. 2009;10(5–6):310–23.
59. Grad LI, Rouleau GA, Ravits J, Cashman NR. Clinical Spectrum of Amyotrophic Lateral Sclerosis (ALS). Cold Spring Harb Perspect Med. 2017 Aug 1;7(8):a024117.
60. Leigh PN, Anderton BH, Dodson A, Gallo J-M, Swash M, Power DM. Ubiquitin deposits in anterior horn cells in motor neurone disease. Neurosci Lett. 1988 Nov 11;93(2–3):197–203.
61. Arai T, Hasegawa M, Akiyama H, Ikeda K, Nonaka T, Mori H, et al. TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Biochem Biophys Res Commun. 2006 Dec 22;351(3):602–11.
62. Alami NH, Smith RB, Carrasco MA, Williams LA, Winborn CS, Han SSW, et al. Axonal transport of TDP-43 mRNA granules is impaired by ALS-causing mutations. Neuron. 2014 Feb 5;81(3):536–43.
63. Kanouchi T, Ohkubo T, Yokota T. Can regional spreading of amyotrophic lateral sclerosis motor symptoms be explained by prion-like propagation? J Neurol Neurosurg Psychiatry. 2012 Jul 1;83(7):739–45.
64. Polymenidou M, Cleveland DW. The seeds of neurodegeneration: Prion-like spreading in ALS. Vol. 147, Cell. Elsevier; 2011. p. 498–508.
65. Deng H, Hentati A, Tainer J, Iqbal Z, Cayabyab A, Hung W, et al. Amyotrophic lateral sclerosis and structural defects in Cu,Zn superoxide dismutase. Science (80-). 1993;261(5124).
66. Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature. 1993;362(6415):59–62.
67. Byrne S, Heverin M, Elamin M, Bede P, Lynch C, Kenna K, et al. Aggregation of neurologic and neuropsychiatric disease in amyotrophic lateral sclerosis kindreds: a population-based case-control cohort study of familial and sporadic amyotrophic lateral sclerosis. Ann Neurol. 2013 Nov;74(5):699–708.

68. Gibson SB, Figueroa KP, Bromberg MB, Pulst S-M, Cannon-Albright L. Familial clustering of ALS in a population-based resource. Neurology. 2014 Jan 7;82(1):17–22.
69. Pettersson E, Lundeberg J, Ahmadian A. Generations of sequencing technologies. Genomics. 2009 Feb 1;93(2):105–11.
70. Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) [Internet]. [cited 2018 Jul 9]. Available from: www.genome.gov/sequencingcostsdata.
71. Lill CM, Abel O, Bertram L, Al-Chalabi A. Keeping up with genetic discoveries in amyotrophic lateral sclerosis: the ALSoD and ALSGene databases. Amyotroph Lateral Scler. 2011 Jul 27;12(4):238–49.
72. Abel, Olubunmi; Al-Chalabi A. The Amyotrophic Lateral Sclerosis Online Genetics Database (ALSoD) [Internet]. Institute of Psychiatry, Psychology & Neuroscience, Kings College London. 2015. Available from: http://alsod.iop.kcl.ac.uk/
73. Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron. 2011 Oct 20;72(2):257–68.
74. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. 2011 Oct 20;72(2):245–56.
75. Rohrer JD, Isaacs AM, Mizlienska S, Mead S, Lashley T, Wray S, et al. C9orf72 expansions in frontotemporal dementia and amyotrophic lateral sclerosis. Lancet Neurol. 2015 Jan 28;
76. Ng ASL, Tan E-K. Intermediate C9orf72 alleles in neurological disorders: does size really matter? J Med Genet. 2017 Sep 1;54(9):591 LP-597.
77. Jiang T, Tan M-S, Tan L, Yu J-T. Application of next-generation sequencing technologies in Neurology. Ann Transl Med. 2014 Dec;2(12):125.
78. Kabashi E, Valdmanis PN, Dion P, Spiegelman D, McConkey BJ, Velde C Vande, et al. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat Genet. 2008 May 30;40(5):572–4.
79. Vance C, Rogelj B, Hortobágyi T, De Vos KJ, Nishimura AL, Sreedharan J, et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science. 2009 Feb 27;323(5918):1208–11.
80. Kwiatkowski TJ, Bosco DA, Leclerc AL, Tamrazian E, Vanderburg CR, Russ C, et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science. 2009 Feb 27;323(5918):1205–8.
81. Nishimura AL, Mitne-Neto M, Silva HCA, Richieri-Costa A, Middleton S, Cascio D, et al. A Mutation in the Vesicle-Trafficking Protein VAPB Causes Late-Onset Spinal Muscular Atrophy and Amyotrophic Lateral Sclerosis. Am J Hum Genet. 2004 Nov;75(5):822–31.
82. Maruyama H, Morino H, Ito H, Izumi Y, Kato H, Watanabe Y, et al. Mutations of optineurin in amyotrophic lateral sclerosis. Nature. 2010 May 28;465(7295):223–6.
83. Wu C-H, Fallini C, Ticozzi N, Keagle PJ, Sapp PC, Piotrowska K, et al. Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. Nature. 2012 Aug 15;488(7412):499–503.
84. Johnson JO, Mandrioli J, Benatar M, Abramzon Y, Van Deerlin VM, Trojanowski JQ, et al. Exome Sequencing Reveals VCP Mutations as a Cause of Familial ALS. Neuron. 2010 Dec 9;68(5):857–64.
85. Freischmidt A, Wieland T, Richter B, Ruf W, Schaeffer V, Müller K, et al.

Haploinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia. Nat Neurosci. 2015 May;18(5):631–6.
86. Cirulli ET, Lasseigne BN, Petrovski S, Sapp PC, Dion PA, Leblond CS, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science (80-). 2015 Feb 19;347(6229):1436–41.
87. Shu S, Li XL, Liu Q, Liu F, Cui B, Liu MS, et al. Screening of the TBK1 gene in familial and sporadic amyotrophic lateral sclerosis patients of Chinese origin. Amyotroph Lateral Scler Front Degener. 2016 Nov 16;17(7–8):605–7.
88. Kenna KP, van Doormaal PTC, Dekker AM, Ticozzi N, Kenna BJ, Diekstra FP, et al. NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. Nat Genet. 2016;48(9):1037–42.
89. Brenner D, Müller K, Wieland T, Weydt P, Böhm S, Lule D, et al. NEK1 mutations in familial amyotrophic lateral sclerosis. Vol. 139, Brain. 2016. p. e28.
90. Nguyen HP, Van Mossevelde S, Dillen L, De Bleecker JL, Moisse M, Van Damme P, et al. NEK1 genetic variability in a Belgian cohort of ALS and ALS-FTD patients. Neurobiol Aging. 2018 Jan 1;61:255.e1-255.e7.
91. Van Damme P, Van Den Bosch L, Robberecht W. Molecular mechanisms and therapeutics of amyotrophic lateral sclerosis. Nat Rev Neurosci. 2016;
92. Chiò A, Restagno G, Brunetti M, Ossola I, Calvo A, Canosa A, et al. ALS/FTD phenotype in two Sardinian families carrying both C9ORF72 and TARDBP mutations. J Neurol Neurosurg Psychiatry. 2012 Jul;83(7):730–3.
93. Lattante S, Ciura S, Rouleau GA, Kabashi E. Defining the genetic connection linking amyotrophic lateral sclerosis (ALS) with frontotemporal dementia (FTD). Trends Genet. 2015 Apr;31(5):263–73.
94. van Blitterswijk M, van Es MA, Hennekam EAM, Dooijes D, van Rheenen W, Medic J, et al. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. Hum Mol Genet. 2012 Sep 1;21(17):3776–84.
95. Kenna KP, McLaughlin RL, Byrne S, Elamin M, Heverin M, Kenny EM, et al. Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. J Med Genet. 2013 Nov;50(11):776–83.
96. Cady J, Allred P, Bali T, Pestronk A, Goate A, Miller TM, et al. Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. Ann Neurol. 2015 Jan;77(1):100–13.
97. Bury JJ, Highley JR, Cooper-Knock J, Goodall EF, Higginbottom A, McDermott CJ, et al. Oligogenic inheritance of optineurin (OPTN) and C9ORF72 mutations in ALS highlights localisation of OPTN in the TDP-43-negative inclusions of C9ORF72-ALS. Neuropathology. 2016 Apr;36(2):125–34.
98. Sproviero W, Shatunov A, Stahl D, Shoai M, van Rheenen W, Jones AR, et al. ATXN2 trinucleotide repeat length correlates with risk of ALS. Neurobiol Aging. 2017 Mar 1;51:178.e1-178.e9.
99. Mackenzie IRA, Bigio EH, Ince PG, Geser F, Neumann M, Cairns NJ, et al. Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. Ann Neurol. 2007 May;61(5):427–34.
100. Al-Chalabi A, Van Den Berg LH, Veldink J. Gene discovery in amyotrophic lateral sclerosis: implications for clinical management. Nat Rev Neurol. 2017 Feb 16;13(2):96–104.
101. Jones CT, Swingler RJ, Brock DJH. Identification of a novel S0D1 mutation in an apparently sporadic amyotrophic lateral sclerosis patient and the detection of lle113Thr in three others. Hum Mol Genet. 1994 Apr 1;3(4):649–50.

102. Renton AE, Chiò A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. Nat Neurosci. 2014 Jan;17(1):17–23.
103. Cruts M, Engelborghs S, van der Zee J, Van Broeckhoven C. C9orf72-Related Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. In: GeneReviews(R). University of Washington, Seattle; 2015.
104. Zou Z-Y, Zhou Z-R, Che C-H, Liu C-Y, He R-L, Huang H-P. Genetic epidemiology of amyotrophic lateral sclerosis: a systematic review and meta-analysis. J Neurol Neurosurg Psychiatry. 2017 Jul;88(7):540–9.
105. He J, Mangelsdorf M, Fan D, Bartlett P, Brown MA. Amyotrophic Lateral Sclerosis Genetic Studies: From Genome-wide Association Mapping to Genome Sequencing. Neuroscientist. 2014 Nov 5;
106. Al-Chalabi A, Fang F, Hanby MF, Leigh PN, Shaw CE, Ye W, et al. An estimate of amyotrophic lateral sclerosis heritability using twin data. J Neurol Neurosurg Psychiatry. 2010 Dec 1;81(12):1324–6.
107. Li HF, Wu ZY. Genotype-phenotype correlations of amyotrophic lateral sclerosis. Transl Neurodegener. 2016 Dec 3;5(1):3.
108. Yamashita S, Ando Y. Genotype-phenotype relationship in hereditary amyotrophic lateral sclerosis. Vol. 4, Translational Neurodegeneration. BioMed Central; 2015. p. 13.
109. Shimizu T, Kawata A, Kato S, Hayashi M, Takamoto K, Hayashi H, et al. Autonomic failure in ALS with a novel SOD1 gene mutation. Neurology. 2000 Apr 11;54(7):1534–7.
110. Andersen PM, Nilsson P, Ala-Hurula V, Keränen M-L, Tarvainen I, Haltia T, et al. Amyotrophic lateral sclerosis associated with homozygosity for an Asp90Ala mutation in CuZn-superoxide dismutase. Nat Genet. 1995 May 1;10(1):61–6.
111. van Es MA, Dahlberg C, Birve A, Veldink JH, van den Berg LH, Andersen PM. Large-scale SOD1 mutation screening provides evidence for genetic heterogeneity in amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 2010 May 1;81(5):562–6.
112. Zou Z-Y, Liu M-S, Li X-G, Cui L-Y. H46R SOD1 mutation is consistently associated with a relatively benign form of amyotrophic lateral sclerosis with slow progression. Amyotroph Lateral Scler Front Degener. 2016 Nov 16;17(7–8):610–3.
113. Cudkowicz ME, McKenna-Yasek D, Chen C, Hedley-Whyte ET, Brown RH. Limited corticospinal tract involvement in amyotrophic lateral sclerosis subjects with the A4V mutation in the copper/zinc superoxide dismutase gene. Ann Neurol. 1998 Jun;43(6):703–10.
114. Kato S, Hayashi H, Nakashima K, Nanba E, Kato M, Hirano A, et al. Pathological characterization of astrocytic hyaline inclusions in familial amyotrophic lateral sclerosis. Am J Pathol. 1997 Aug;151(2):611–20.
115. Georgoulopoulou E, Gellera C, Bragato C, Sola P, Chiari A, Bernabei C, et al. A novel SOD1 mutation in a young amyotrophic lateral sclerosis patient with a very slowly progressive clinical course. Muscle and Nerve. 2010 Aug 25;42(4):596–7.
116. Del Grande A, Conte A, Lattante S, Luigetti M, Marangi G, Zollino M, et al. D11Y SOD1 mutation and benign ALS: a consistent genotype-phenotype correlation. J Neurol Sci. 2011 Oct 15;309(1–2):31–3.
117. Yang Y, Hentati A, Deng HX, Dabbagh O, Sasaki T, Hirano M, et al. The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. Nat Genet. 2001 Oct;29(2):160–5.

118. Hadano S, Hand CK, Osuga H, Yanagisawa Y, Otomo A, Devon RS, et al. A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2. Nat Genet. 2001 Oct;29(2):166–73.
119. Siddiqi S, Foo JN, Vu A, Azim S, Silver DL, Mansoor A, et al. A Novel Splice-Site Mutation in ALS2 Establishes the Diagnosis of Juvenile Amyotrophic Lateral Sclerosis in a Family with Early Onset Anarthria and Generalized Dystonias. Raoul C, editor. PLoS One. 2014 Dec 4;9(12):e113258.
120. Marques VD, Barreira AA, Davis MB, Abou-Sleiman PM, Silva WA, Zago MA, et al. Expanding the phenotypes of the Pro56Ser VAPB mutation: Proximal SMA with dysautonomia. Muscle and Nerve. 2006 Dec;34(6):731–9.
121. Chiò A, Borghero G, Pugliatti M, Ticca A, Calvo A, Moglia C, et al. Large Proportion of Amyotrophic Lateral Sclerosis Cases in Sardinia Due to a Single Founder Mutation of the TARDBP Gene. Arch Neurol. 2011 May 1;68(5):594.
122. Benajiba L, Ber I Le, Camuzat A, Lacoste M, Thomas-Anterion C, Couratier P, et al. TARDBP mutations in motoneuron disease with frontotemporal lobar degeneration. Ann Neurol. 2009 Apr;65(4):470–4.
123. Hübers A, Just W, Rosenbohm A, Müller K, Marroquin N, Goebel I, et al. De novo FUS mutations are the most frequent genetic cause in early-onset German ALS patients. Neurobiol Aging. 2015 Aug 15;
124. Munoz DG, Neumann M, Kusaka H, Yokota O, Ishihara K, Terada S, et al. FUS pathology in basophilic inclusion body disease. Acta Neuropathol. 2009 Nov 15;118(5):617–27.
125. Yan J, Deng HX, Siddique N, Fecto F, Chen W, Yang Y, et al. Frameshift and novel mutations in FUS in familial amyotrophic lateral sclerosis and ALS/dementia. Neurology. 2010;75(9):807–14.
126. Kwok CT, Wang HY, Morris AG, Smith B, Shaw C, De Belleroche J. VCP mutations are not a major cause of familial amyotrophic lateral sclerosis in the UK. J Neurol Sci. 2015 Feb 15;349(1–2):209–13.
127. He J, Tang L, Benyamin B, Shah S, Hemani G, Liu R, et al. C9orf72 hexanucleotide repeat expansions in Chinese sporadic amyotrophic lateral sclerosis. Neurobiol Aging. 2015 Jun 9;36(9):2660.e1-8.
128. Vats A, Gourie-Devi M, Suroliya V, Verma S, Faruq M, Sharma A, et al. Analysis of C9orf72 repeat expansion in amyotrophic lateral sclerosis patients from North India. J Neurol Sci. 2017 Feb 15;373:55–7.
129. Smith BN, Newhouse S, Shatunov A, Vance C, Topp S, Johnson L, et al. The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. Eur J Hum Genet. 2013 Jan;21(1):102–8.
130. Mok K, Traynor BJ, Schymick J, Tienari PJ, Laaksovirta H, Peuralinna T, et al. The chromosome 9 ALS and FTD locus is probably derived from a single founder. Neurobiol Aging. 2012;33(1):209.e3-209.e8.
131. Patel AN, Sampson JB. Cognitive Profile of C9orf72 in Frontotemporal Dementia and Amyotrophic Lateral Sclerosis. Curr Neurol Neurosci Rep. 2015 Sep;15(9):582.
132. Byrne S, Elamin M, Bede P, Shatunov A, Walsh C, Corr B, et al. Cognitive and clinical characteristics of patients with amyotrophic lateral sclerosis carrying a C9orf72 repeat expansion: a population-based cohort study. Lancet Neurol. 2012 Mar;11(3):232–40.
133. Watson A, Pribadi M, Chowdari K, Clifton S, Joel Wood, Miller BL, et al. C9orf72 repeat expansions that cause frontotemporal dementia are detectable among patients with psychosis. Psychiatry Res. 2015 Dec 8;

134. Solje E, Aaltokallio H, Koivumaa-Honkanen H, Suhonen NM, Moilanen V, Kiviharju A, et al. The Phenotype of the C9ORF72 Expansion Carriers According to Revised Criteria for bvFTD. PLoS One. 2015 Jan;10(7):e0131817.
135. Miltenberger-Miltenyi G, Conceição VA, Gromicho M, Pronto-Laborinho AC, Pinto S, de Carvalho M. C9orf72 expansion is associated with accelerated decline of respiratory function and decreased survival in amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 2018 Apr 16;jnnp-2018-318032.
136. Rooney J, Fogh I, Westeneng HJ, Vajda A, McLaughlin R, Heverin M, et al. C9orf72 expansion differentially affects males with spinal onset amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 2017 Apr;88(4):295–300.
137. Dols-Icardo O, García-Redondo A, Rojas-García R, Sánchez-Valle R, Noguera A, Gómez-Tortosa E, et al. Characterization of the repeat expansion size in C9orf72 in amyotrophic lateral sclerosis and frontotemporal dementia. Hum Mol Genet. 2014 Feb 1;23(3):749–54.
138. Deng HX, Chen W, Hong ST, Boycott KM, Gorrie GH, Siddique N, et al. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. Nature. 2011 Sep 21;477(7363):211–5.
139. Harms MB, Baloh RH. Clinical neurogenetics: amyotrophic lateral sclerosis. Neurol Clin. 2013 Nov;31(4):929–50.
140. Bannwarth S, Ait-El-Mkadem S, Chaussenot A, Genin EC, Lacas-Gervais S, Fragaki K, et al. A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. Brain. 2014 Aug 1;137(8):2329–45.
141. Wong CH, Topp S, Gkazi AS, Troakes C, Miller JW, de Majo M, et al. The CHCHD10 P34S variant is not associated with ALS in a UK cohort of familial and sporadic patients. Neurobiol Aging. 2015 Oct 1;36(10):2908.e17-2908.e18.
142. Gijselinck I, Van Mossevelde S, Van Der Zee J, Sieben A, Philtjens S, Heeman B, et al. Loss of TBK1 is a frequent cause of frontotemporal dementia in a Belgian cohort. Neurology. 2015;85(24):2116–25.
143. Al-Chalabi A, Calvo A, Chio A, Colville S, Ellis CM, Hardiman O, et al. Analysis of amyotrophic lateral sclerosis as a multistep process: a population-based modelling study. Lancet Neurol. 2014 Nov;13(11):1108–13.
144. Lopate G, Baloh RH, Al-Lozi MT, Miller TM, Fernandes Filho JA, Ni O, et al. Familial ALS with extreme phenotypic variability due to the I113T SOD1 mutation. Amyotroph Lateral Scler. 2010;11(1–2):232–6.
145. Zoccolella S, Palagano G, Fraddosio A, Russo I, Ferrannini E, Serlenga L, et al. ALS-plus: 5 cases of concomitant amyotrophic lateral sclerosis and parkinsonism. Neurol Sci. 2002 Sep 1;23(0):s123–4.
146. Qureshi AI, Wilmot G, Dihenia B, Schneider JA, Krendel DA. Motor Neuron Disease With Parkinsonism. Arch Neurol. 1996 Oct 1;53(10):987–91.
147. Majounie E, Abramzon Y, Renton AE, Perry R, Bassett SS, Pletnikova O, et al. Repeat Expansion in *C9ORF72* in Alzheimer's Disease. N Engl J Med. 2012 Jan 19;366(3):283–4.
148. Lesage S, Le Ber I, Condroyer C, Broussolle E, Gabelle A, Thobois S, et al. C9orf72 repeat expansions are a rare genetic cause of parkinsonism. Brain. 2013 Feb 1;136(2):385–91.
149. Hensman DJ, Poulter M, Beck J, Hehir J, Polke JM, Campbell T, et al. C9orf72 expansions are the most common genetic cause of Huntington disease phenocopies. Neurology. 2014 Jan 28;82(4):292–9.
150. Karch CM, Wen N, Fan CC, Yokoyama JS, Kouri N, Ross OA, et al. Selective Genetic

Overlap Between Amyotrophic Lateral Sclerosis and Diseases of the Frontotemporal Dementia Spectrum. JAMA Neurol. 2018 Apr 9;
151. Therrien M, Dion PA, Rouleau GA. ALS: Recent Developments from Genetics Studies. Curr Neurol Neurosci Rep. 2016 Apr;16(6):59.
152. Tafuri F, Ronchi D, Magri F, Comi GP, Corti S. SOD1 misplacing and mitochondrial dysfunction in amyotrophic lateral sclerosis pathogenesis. Front Cell Neurosci. 2015 Aug 25;9:336.
153. Lagier-Tourenne C, Polymenidou M, Cleveland DW. TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. Hum Mol Genet. 2010 Apr 15;19(R1):R46–64.
154. Gendron TF, Rademakers R, Petrucelli L. TARDBP mutation analysis in TDP-43 proteinopathies and deciphering the toxicity of mutant TDP-43. J Alzheimers Dis. 2013;33 Suppl 1(Suppl 1):S35-45.
155. Solomon DA, Stepto A, Au WH, Adachi Y, Diaper DC, Hall R, et al. A feedback loop between dipeptide-repeat protein, TDP-43 and karyopherin-α mediates C9orf72-related neurodegeneration. Brain. 2018 Oct 1;141(10):2908–24.
156. Gellera C, Tiloca C, Del Bo R, Corrado L, Pensato V, Agostini J, et al. *Ubiquilin 2* mutations in Italian patients with amyotrophic lateral sclerosis and frontotemporal dementia. J Neurol Neurosurg Psychiatry. 2013 Feb;84(2):183–7.
157. van der Zee J, Gijselinck I, Van Mossevelde S, Perrone F, Dillen L, Heeman B, et al. *TBK1* Mutation Spectrum in an Extended European Patient Cohort with Frontotemporal Dementia and Amyotrophic Lateral Sclerosis. Hum Mutat. 2017 Mar 1;38(3):297–309.
158. Chia R, Chiò A, Traynor BJ. Novel genes associated with amyotrophic lateral sclerosis: diagnostic and clinical implications. Vol. 17, The Lancet Neurology. Elsevier; 2018. p. 94–102.
159. Cirulli ET, Lasseigne BN, Petrovski S, Sapp PC, Dion PA, Leblond CS, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science. 2015 Mar 27;347(6229):1436–41.
160. Al-Chalabi A, Hardiman O. The epidemiology of ALS: a conspiracy of genes, environment and time. Nat Rev Neurol. 2013 Oct 15;9(11):617–28.
161. Orsetti V, Pegoraro E, Cima V, D'Ascenzo C, Palmieri A, Querin G, et al. Genetic variation in KIFAP3 is associated with an upper motor neuron-predominant phenotype in amyotrophic lateral sclerosis. Neurodegener Dis. 2011 Jan;8(6):491–5.
162. Arning L, Epplen JT, Rahikkala E, Hendrich C, Ludolph AC, Sperfeld AD. The SETX missense variation spectrum as evaluated in patients with ALS4-like motor neuron diseases. Neurogenetics. 2013 Feb 6;14(1):53–61.
163. Chiò A, Borghero G, Restagno G, Mora G, Drepper C, Traynor BJ, et al. Clinical characteristics of patients with familial amyotrophic lateral sclerosis carrying the pathogenic GGGGCC hexanucleotide repeat expansion of C9ORF72. Brain. 2012 Mar 1;135(3):784–93.
164. Ferrari R, Wang Y, Vandrovcova J, Guelfi S, Witeolar A, Karch CM, et al. Genetic architecture of sporadic frontotemporal dementia and overlap with Alzheimer's and Parkinson's diseases. J Neurol Neurosurg Psychiatry. 2017 Feb 1;88(2):152–64.
165. Wilke C, Baets J, De Bleecker JL, Deconinck T, Biskup S, Hayer SN, et al. Beyond ALS and FTD: the phenotypic spectrum of TBK1 mutations includes PSP-like and cerebellar phenotypes. Neurobiol Aging. 2018 Feb;62:244.e9-244.e13.
166. Talbot K. Should all patients with ALS have genetic testing? J Neurol Neurosurg Psychiatry. 2014 May;85(5):475.

167. Wagner KN, Nagaraja H, Allain DC, Quick A, Kolb S, Roggenbuck J. Patients with Amyotrophic Lateral Sclerosis Have High Interest in and Limited Access to Genetic Testing. J Genet Couns. 2016 Oct 20;
168. Chiò A, Battistini S, Calvo A, Caponnetto C, Conforti FL, Corbo M, et al. Genetic counselling in ALS: facts, uncertainties and clinical suggestions. J Neurol Neurosurg Psychiatry. 2014 May;85(5):478–85.
169. Roggenbuck J, Quick A, Kolb SJ. Genetic testing and genetic counseling for amyotrophic lateral sclerosis: an update for clinicians. Genet Med. 2016 Aug 18;
170. National Guideline Clearinghouse (NGC). Guideline summary: EFNS guidelines on the clinical management of amyotrophic lateral sclerosis (MALS) — revised report of an EFNS task force. [Internet]. National Guideline Clearinghouse (NGC). 2012 [cited 2016 Oct 22]. Available from: https://www.guideline.gov/summaries/summary/38469
171. Klepek H, Nagaraja H, Goutman SA, Quick A, Kolb SJ, Roggenbuck J. Lack of consensus in ALS genetic testing practices and divergent views between ALS clinicians and patients. Amyotroph Lateral Scler Front Degener. 2019 Apr 1;1–6.
172. Benatar M, Stanislaw C, Reyes E, Hussain S, Cooley A, Fernandez MC, et al. Presymptomatic ALS genetic counseling and testing: Experience and recommendations. Neurology. 2016 Jun 14;86(24):2295–302.
173. Scarrott JM, Herranz-Martín S, Alrafiah AR, Shaw PJ, Azzouz M. Current developments in gene therapy for amyotrophic lateral sclerosis. Expert Opin Biol Ther. 2015 Jul;15(7):935–47.
174. Abdul Wahid SF, Law ZK, Ismail NA, Azman Ali R, Lai NM. Cell-based therapies for amyotrophic lateral sclerosis/motor neuron disease. In: Abdul Wahid SF, editor. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd; 2016.
175. Trias E, Ibarburu S, Barreto-Núñez R, Babdor J, Maciel TT, Guillo M, et al. Post-paralysis tyrosine kinase inhibition with masitinib abrogates neuroinflammation and slows disease progression in inherited amyotrophic lateral sclerosis. J Neuroinflammation. 2016 Jul 11;13(1):177.
176. Al-Chalabi A, Brown RH. Finding a Treatment for ALS — Will Gene Editing Cut It? Phimister EG, editor. N Engl J Med. 2018 Apr 12;378(15):1454–6.
177. Fournier C, Glass JD. Modeling the course of amyotrophic lateral sclerosis. Nat Biotechnol. 2015 Jan 9;33(1):45–7.
178. Calvo A, Moglia C, Lunetta C, Marinou K, Ticozzi N, Ferrante GD, et al. Factors predicting survival in ALS: a multicenter Italian study. J Neurol. 2017 Jan 1;264(1):54–63.
179. Taylor AA, Fournier C, Polak M, Wang L, Zach N, Keymer M, et al. Predicting disease progression in amyotrophic lateral sclerosis. Ann Clin Transl Neurol. 2016;3(11):866.
180. Beghi E, Chiò A, Couratier P, Esteban J, Hardiman O, Logroscino G, et al. The epidemiology and treatment of ALS: Focus on the heterogeneity of the disease and critical appraisal of therapeutic trials. Amyotroph Lateral Scler. 2011 Jan 11;12(1):1–10.
181. Gonzalez Calzada N, Prats Soro E, Mateu Gomez L, Giro Bulta E, Cordoba Izquierdo A, Povedano Panades M, et al. Factors predicting survival in amyotrophic lateral sclerosis patients on non-invasive ventilation. Amyotroph Lateral Scler Front Degener. 2016 Aug 17;17(5–6):337–42.
182. Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory

function. BDNF ALS Study Group (Phase III). J Neurol Sci. 1999 Oct 31;169(1–2):13–21.
183. Proudfoot M, Jones A, Talbot K, Al-Chalabi A, Turner MR. The ALSFRS as an outcome measure in therapeutic trials and its relationship to symptom onset. Amyotroph Lateral Scler Front Degener. 2016;17(5–6):414–25.
184. Kimura F, Fujimura C, Ishida S, Nakajima H, Furutama D, Uehara H, et al. Progression rate of ALSFRS-R at time of diagnosis predicts survival time in ALS. Neurology. 2006 Jan 24;66(2):265–7.
185. Elamin M, Bede P, Montuschi A, Pender N, Chio A, Hardiman O. Predicting prognosis in amyotrophic lateral sclerosis: a simple algorithm. J Neurol. 2015 Jun;262(6):1447–54.
186. Rooney J, Burke T, Vajda A, Heverin M, Hardiman O. What does the ALSFRS-R really measure? A longitudinal and survival analysis of functional dimension subscores in amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 2017 May 1;88(5):381–5.
187. Knibb JA, Keren N, Kulka A, Leigh PN, Martin S, Shaw CE, et al. A clinical tool for predicting survival in ALS. J Neurol Neurosurg Psychiatry. 2016 Dec 1;87(12):1361–7.
188. Pfohl SR, Kim RB, Coan GS, Mitchell CS. Unraveling the Complexity of Amyotrophic Lateral Sclerosis Survival Prediction. Front Neuroinform. 2018;12:36.
189. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. BMJ. 2016 Apr 27;352:i1981.
190. MacKinnon DP, Krull JL, Lockwood CM. Equivalence of the mediation, confounding and suppression effect. Prev Sci. 2000 Dec;1(4):173–81.
191. BEAULIEU-JONES BK, MOORE JH. MISSING DATA IMPUTATION IN THE ELECTRONIC HEALTH RECORD USING DEEPLY LEARNED AUTOENCODERS. In: Biocomputing 2017. WORLD SCIENTIFIC; 2017. p. 207–18.
192. Masconi KL, Matsha TE, Echouffo-Tcheugui JB, Erasmus RT, Kengne AP. Reporting and handling of missing data in predictive research for prevalent undiagnosed type 2 diabetes mellitus: a systematic review. EPMA J. 2015;6(1):7.
193. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res. 2011 Mar;20(1):40–9.
194. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. Vol. 339, BMJ (Online). British Medical Journal Publishing Group; 2009. p. 157–60.
195. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med. 2015 Jan 6;162(1):W1.
196. Collins, G., Reitsma H, Altman D. Reporting Guideline for Prediction Model Studies : TRIPOD T ransparent R eporting of a multivariable prediction model for I ndividual P rognosis O r D iagnosis. Ann Intern Med. 2015 Jan 6;162(1):55–63.
197. NHS Research Scotland. Research in Scotland | NHS Research Scotland | NHS Research Scotland [Internet]. 2018 [cited 2018 Jul 14]. Available from: http://www.nhsresearchscotland.org.uk/research-in-scotland
198. Mackenzie IS, Morant S V, Bloomfield GA, MacDonald TM, O'Riordan J. Incidence and prevalence of multiple sclerosis in the UK 1990-2010: a descriptive study in the General Practice Research Database. J Neurol Neurosurg Psychiatry. 2014 Jan 1;85(1):76–84.

199. Holloway SM, Emery AEH. The epidemiology of motor neuron disease in Scotland. Muscle Nerve. 1982 Feb 1;5(2):131–3.
200. Holloway SM, Mitchell JD. Motor neurone disease in the Lothian Region of Scotland 1961-81. J Epidemiol Community Health. 1986 Dec 1;40(4):344 LP-350.
201. Kahana E, Zilber N. Changes in the Incidence of Amyotrophic Lateral Sclerosis in Israel. Arch Neurol. 1984 Feb 1;41(2):157–60.
202. Hern JEC, Dundee RK, Davidson D, Forster A, Roberts R, Swingler RJ, et al. The Scottish motor neuron disease register: a prospective study of adult onset motor neuron disease in Scotland. Methodology, demography and clinical features of incident cases in 1989. J Neurol Neurosurg Psychiatry. 1992;55:536–41.
203. Chancellor AM, Slattery JM, Fraser H, Swingler RJ, Holloway SM, Warlow CP. The prognosis of adult-onset motor neuron disease: a prospective study based on the Scottish Motor Neuron Disease Register. J Neurol. 1993 Jun;240(6):339–46.
204. Forbes RB, Colville S, Swingler RJ. The epidemiology of amyotrophic lateral sclerosis (ALS/MND) in people aged 80 or over. Age Ageing. 2004 Mar;33(2):131–4.
205. Davenport RJ, Swingler RJ, Chancellor AM, Warlow CP. Avoiding false positive diagnoses of motor neuron disease: lessons from the Scottish Motor Neuron Disease Register. J Neurol Neurosurg Psychiatry. 1996 Feb;60(2):147–51.
206. Abrahams S, Newton J, Niven E, Foley J, Bak TH. Screening for cognition and behaviour changes in ALS. Amyotroph Lateral Scler Frontotemporal Degener. 2014 Mar;15(1–2):9–14.
207. Lulé D, Burkhardt C, Abdulla S, Böhm S, Kollewe K, Uttner I, et al. The Edinburgh Cognitive and Behavioural Amyotrophic Lateral Sclerosis Screen: A cross-sectional comparison of established screening tools in a German-Swiss population. Amyotroph Lateral Scler Frontotemporal Degener. 2014 Oct 8;(August):1–8.
208. Niven E, Newton J, Foley J, Colville S, Swingler R, Chandran S, et al. Validation of the Edinburgh Cognitive and Behavioural Amyotrophic Lateral Sclerosis Screen (ECAS): A cognitive tool for motor disorders. Amyotroph Lateral Scler Front Degener. 2015 Apr 27;16(3–4):172–9.
209. Crockford CJ, Kleynhans M, Wilton E, Radakovic R, Newton J, Niven EH, et al. ECAS A-B-C: alternate forms of the Edinburgh Cognitive and Behavioural ALS Screen. Amyotroph Lateral Scler Front Degener. 2018 Jan 2;19(1–2):57–64.
210. Poletti B, Solca F, Carelli L, Faini A, Madotto F, Lafronza A, et al. Cognitive-behavioral longitudinal assessment in ALS: the Italian Edinburgh Cognitive and Behavioral ALS screen (ECAS). Amyotroph Lateral Scler Front Degener. 2018 Jul 3;19(5–6):387–95.
211. Mora JS, Salas T, Fernández MC, Rodríguez-Castillo V, Marín S, Chaverri D, et al. Spanish adaptation of the edinburgh cognitive and behavioral amyotrophic lateral sclerosis screen (ECAS). Amyotroph Lateral Scler Front Degener. 2018 Jan 2;19(1–2):74–9.
212. Crockford C, Newton J, Lonergan K, Madden C, Mays I, O'Sullivan M, et al. Measuring reliable change in cognition using the Edinburgh Cognitive and Behavioural ALS Screen (ECAS). Amyotroph Lateral Scler Front Degener. 2018 Jan 2;19(1–2):65–73.
213. Crockford C, Newton J, Lonergan K, Chiwera T, Booth T, Chandran S, et al. ALS-specific cognitive and behavior changes associated with advancing disease stage in ALS. Neurology. 2018 Oct 9;91(15):e1370–80.
214. Jones CT, Swingler RJ, Simpson SA, Brock DJ. Superoxide dismutase mutations in an unselected cohort of Scottish amyotrophic lateral sclerosis patients. J Med Genet. 1995 Apr;32(4):290–2.
215. Cleary EM, Pal S, Azam T, Moore DJ, Swingler R, Gorrie G, et al. Improved PCR based

methods for detecting C9orf72 hexanucleotide repeat expansions. Mol Cell Probes. 2016 Aug;30(4):218–24.
216. Akimoto C, Volk AE, van Blitterswijk M, Van den Broeck M, Leblond CS, Lumbroso S, et al. A blinded international study on the reliability of genetic testing for GGGGCC-repeat expansions in C9orf72 reveals marked differences in results among 14 laboratories. J Med Genet. 2014 Jun 1;51(6):419–24.
217. Forbes RB, Colvile S, Cran GW, Swingler RJ. Unexpected decline in survival from amyotrophic lateral sclerosis/motor neurone disease. J Neurol Neurosurg Psychiatry. 2004 Dec 1;75(12):1753–5.
218. Forbes R, Colville S, Swingler R. Frequency, timing and outcome of gastrostomy tubes for amyotrophic lateral sclerosis/motor neurone disease. J Neurol. 2004 Jul;251(7):813–7.
219. Hayward C, Swingler RJ, Simpson SA, Brock DJ. A specific superoxide dismutase mutation is on the same genetic background in sporadic and familial cases of amyotrophic lateral sclerosis. Am J Hum Genet. 1996 Nov;59(5):1165–7.
220. Leighton D, Newton J, Colville S, Bethell A, Craig G, Cunningham L, et al. Clinical audit research and evaluation of motor neuron disease (CARE-MND): a national electronic platform for prospective, longitudinal monitoring of MND in Scotland. Amyotroph Lateral Scler Front Degener. 2019 Mar 20;1–9.
221. National Records of Scotland. Statistics and Data: Mid-Year Population Estimates [Internet]. National Records of Scotland. National Records of Scotland; 2017 [cited 2018 Feb 28]. Available from: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates
222. The Scottish Motor Neuron Disease Research Group. The Scottish Motor Neuron Disease Register: a prospective study of adult onset motor neuron disease in Scotland. Methodology, demography and clinical features of incident cases in 1989. J Neurol Neurosurg Psychiatry. 1992 Jul;55(7):536–41.
223. Quality Improvement Scotland. Neurological Health Services: Clinical Standards October 2009. 2009;
224. Health Improvement Scotland. Neurological health services in Scotland: Final Report. 2012;
225. Ferrie J, Robertson P, Rieck –, Watson N. Living With MND: An evaluation of care pathways available to adults with, and the families or carers of, adults with Motor Neurone Disease in Scotland [Internet]. 2013 [cited 2015 Mar 1]. Available from: https://www.euanmacdonaldcentre.com/wp-content/uploads/2015/02/Living-with-MND_Jo-Ferrie_report.pdf
226. Government S. Report on the Specialist Nursing and Care Fund. Scottish Gov. 2017;(September).
227. Hardiman O, Al-Chalabi A, Brayne C, Beghi E, Van Den Berg LH, Chio A, et al. The changing picture of amyotrophic lateral sclerosis: Lessons from European registers. Vol. 88, Journal of Neurology, Neurosurgery and Psychiatry. BMJ Publishing Group; 2017. p. 557–63.
228. Edgren G, Hjalgrim H, Rostgaard K, Lambert P, Wikman A, Norda R, et al. Transmission of neurodegenerative disorders through blood transfusion: A cohort study. Ann Intern Med. 2016 Sep 6;165(5):316–24.
229. Zheng Z, Sheng L, Shang H. Statins and amyotrophic lateral sclerosis: A systematic review and meta-analysis. Amyotroph Lateral Scler Front Degener. 2013 May 24;14(4):241–5.

230. Seelen M, van Doormaal PTC, Visser AE, Huisman MHB, Roozekrans MHJ, de Jong SW, et al. Prior medical conditions and the risk of amyotrophic lateral sclerosis. J Neurol. 2014 Oct 25;261(1):1949–56.
231. Belbasis L, Bellou V, Evangelou E. Environmental Risk Factors and Amyotrophic Lateral Sclerosis: An Umbrella Review and Critical Assessment of Current Evidence from Systematic Reviews and Meta-Analyses of Observational Studies. Neuroepidemiology. 2016 Jan 6;46(2):96–105.
232. Malek AM, Barchowsky A, Bowser R, Youk A, Talbott EO. Pesticide exposure as a risk factor for amyotrophic lateral sclerosis: A meta-analysis of epidemiological studies. Environ Res. 2012 Aug;117:112–9.
233. Sutedja NA, Veldink JH, Fischer K, Kromhout H, Heederik D, Huisman MHB, et al. Exposure to chemicals and metals and risk of amyotrophic lateral sclerosis: A systematic review. Vol. 10, Amyotrophic Lateral Sclerosis. 2009. p. 302–9.
234. Ingre C, Roos PM, Piehl F, Kamel F, Fang F. Risk factors for amyotrophic lateral sclerosis. Clin Epidemiol. 2015 Jan;7:181–93.
235. Alonso A, Logroscino G, Hernán MA. Smoking and the risk of amyotrophic lateral sclerosis: A systematic review and meta-analysis. Vol. 81, Journal of Neurology, Neurosurgery and Psychiatry. 2010. p. 1249–52.
236. Gallo V, Bueno-De-Mesquita HB, Vermeulen R, Andersen PM, Kyrozis A, Linseisen J, et al. Smoking and risk for amyotrophic lateral sclerosis: Analysis of the EPIC cohort. Ann Neurol. 2009 Apr;65(4):378–85.
237. Gallo V, Vanacore N, Bueno-de-Mesquita HB, Vermeulen R, Brayne C, Pearce N, et al. Physical activity and risk of Amyotrophic Lateral Sclerosis in a prospective cohort study. Eur J Epidemiol. 2016 Mar 11;31(3):255–66.
238. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep. 1991 Dec;14(6):540–5.
239. Healthcare Improvement Scotland. Neurological Health Services - NHSScotland local reports [Internet]. NHS Scotland . 2012 [cited 2019 Jun 29]. Available from: http://www.healthcareimprovementscotland.org/our_work/long_term_conditions/neurological_health_services/peer_review_evaluation.aspx
240. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienne, Austria: R Foundation for Statistical Computing, Vienna, Austria; 2017.
241. Swingler R. An audit of the health of people with motor neurone disease in Scotland: The Scottish Motor Neurone Disease Audit, Research and Trials (SMART) Study. Unpublished; 2015.
242. Leighton DJ, Newton J, Stephenson LJ, Colville S, Davenport R, Gorrie G, et al. Changing epidemiology of motor neurone disease in Scotland. J Neurol. 2019 Apr 25;266(4):817–25.
243. Black HA, Leighton DJ, Cleary EM, Rose E, Stephenson L, Colville S, et al. Genetic epidemiology of motor neuron disease-associated variants in the Scottish population. Neurobiol Aging. 2017 Mar;51:178.e11-178.e20.
244. Radakovic R, Stephenson L, Colville S, Swingler R, Chandran S, Abrahams S. Multidimensional apathy in ALS: validation of the Dimensional Apathy Scale. J Neurol Neurosurg Psychiatry. 2016 Jun;87(6):663–9.
245. Henstridge CM, Sideris DI, Carroll E, Rotariu S, Salomonsson S, Tzioras M, et al. Synapse loss in the prefrontal cortex is associated with cognitive decline in amyotrophic lateral sclerosis. Acta Neuropathol. 2018 Feb 22;135(2):213–26.
246. McGeachan AJ, Hobson E V., Al-Chalabi A, Stephenson J, Chandran S, Crawley F, et

al. A multicentre evaluation of oropharyngeal secretion management practices in amyotrophic lateral sclerosis. Amyotroph Lateral Scler Front Degener. 2017 Jan 2;18(1–2):1–9.
247. McLaughlin RL, Schijven D, van Rheenen W, van Eijk KR, O'Brien M, Kahn RS, et al. Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. Nat Commun. 2017 Mar 21;8:14774.
248. van Rheenen W, Shatunov A, Dekker AM, McLaughlin RL, Diekstra FP, Pulit SL, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nat Genet. 2016 Jul 25;48(9):1043–8.
249. MND Scotland. Figures show "Aikman" impact on MND Care [Internet]. MND Scotland Website. 2018 [cited 2018 Nov 21]. Available from: https://www.mndscotland.org.uk/latest/news/figures-show-aikman-impact-on-mnd-care/
250. Gorrie GH, Chandran S, Colville S, Newton J, Leighton D, Mcdonald M, et al. Improved survival and 30-day mortality after gastrostomy in Scottish motor neurone disease patients: evidence from a national retrospective cohort study using STROBE criteria. Amyotroph Lateral Scler Front Degener. 2019 Mar 5;1–7.
251. Elliott E, Newton J, Rewaj P, Gregory JM, Tomarelli L, Colville S, et al. An epidemiological profile of dysarthria incidence and assistive technology use in the living population of people with MND in Scotland. Amyotroph Lateral Scler Front Degener. 2019 Oct 9;1–7.
252. Marin B, Logroscino G, Boumédiene F, Labrunie A, Couratier P, Babron MC, et al. Clinical and demographic factors and outcome of amyotrophic lateral sclerosis in relation to population ancestral origin. Vol. 31, European Journal of Epidemiology. Springer Netherlands; 2016. p. 229–45.
253. Marin B, Boumédiene F, Logroscino G, Couratier P, Babron M-C, Leutenegger AL, et al. Variation in worldwide incidence of amyotrophic lateral sclerosis: a meta-analysis. Int J Epidemiol. 2016 May 16;46(1):57–74.
254. United States Census Bureau 2010. 2010 Census Data Products: United States [Internet]. United States Census Bureau. 2010. Available from: https://www.census.gov/population/www/cen2010/glance/
255. Lilienfeld DE, Stolley PD, Lilienfeld AM. Foundations of epidemiology. Oxford University Press; 1994. 371 p.
256. Schoenberg BS. Calculating Confidence Intervals for Rates and Ratios. Neuroepidemiology. 1983;2(3–4):257–65.
257. Chiò A, Mora G, Moglia C, Manera U, Canosa A, Cammarosano S, et al. Secular trends of amyotrophic lateral sclerosis: The Piemonte and Valle d'Aosta register. JAMA Neurol. 2017 Sep 1;74(9):1097–104.
258. Donaghy C, Clarke J, Patterson C, Kee F, Hardiman O, Patterson V. The epidemiology of motor neuron disease in Northern Ireland using capture-recapture methodology. Amyotroph Lateral Scler. 2010 Aug 16;11(4):374–8.
259. Rosenbohm A, Peter RS, Erhardt S, Lulé D, Rothenbacher D, Ludolph AC, et al. Epidemiology of amyotrophic lateral sclerosis in Southern Germany. J Neurol. 2017 Apr 20;264(4):749–57.
260. Horrocks S, Wilkinson T, Schnier C, Ly A, Woodfield R, Rannikmäe K, et al. Accuracy of routinely-collected healthcare data for identifying motor neurone disease cases: A systematic review. Le W, editor. PLoS One. 2017 Feb 28;12(2):e0172639.
261. Doyle P, Brown A, Beral V, Reeves G, Green J. Incidence of and risk factors for motor neurone disease in UK women: a prospective study. BMC Neurol. 2012 May 6;12:25.

262. Maxwell R, Wells C, Verne J. Under-reporting of progressive supranuclear palsy. Lancet (London, England). 2010 Dec 18;376(9758):2072.
263. Demetriou CA, Hadjivasiliou PM, Kleopa KA, Christou YP, Leonidou E, Kyriakides T, et al. Epidemiology of Amyotrophic Lateral Sclerosis in the Republic of Cyprus: A 25-Year Retrospective Study. Neuroepidemiology. 2017;48(1–2):79–85.
264. Joensen P. Incidence of amyotrophic lateral sclerosis in the Faroe Islands. Acta Neurol Scand. 2012 Jul;126(1):62–6.
265. Georgoulopoulou E, Vinceti M, Bonvicini F, Sola P, Goldoni CA, Girolamo G De, et al. Changing incidence and subtypes of ALS in Modena, Italy: A 10-years prospective study. Amyotroph Lateral Scler. 2011 Nov 6;12(6):451–7.
266. Huisman MHB, De Jong SW, Van Doormaal PTC, Weinreich SS, Schelhaas HJ, Van Der Kooi AJ, et al. Population based epidemiology of amyotrophic lateral sclerosis using capture-recapture methodology. J Neurol Neurosurg Psychiatry. 2011 Oct 1;82(10):1165–70.
267. Chiò A, Mora G, Calvo A, Mazzini L, Bottacchi E, Mutani R. Epidemiology of ALS in Italy. Neurology. 2009 Feb 24;72(8):725 LP-731.
268. O'Toole O, Traynor BJ, Brennan P, Sheehan C, Frost E, Corr B, et al. Epidemiology and clinical features of amyotrophic lateral sclerosis in Ireland between 1995 and 2004. J Neurol Neurosurg Psychiatry. 2008 Jan 1;79(1):30–2.
269. Marin B, Fontana A, Arcuti S, Copetti M, Boumédiene F, Couratier P, et al. Age-specific ALS incidence: a dose–response meta-analysis. Eur J Epidemiol. 2018 Jul 23;33(7):621–34.
270. Information Services Division (ISD) S. Annual Trends in Consultant-led Outpatient Activity [Internet]. ISD Scotland. 2017 [cited 2018 May 16]. Available from: http://www.isdscotland.org/Health-Topics/Hospital-Care/Outpatient-Activity/
271. Beghi E, Logroscino G, Chiò A, Hardiman O, Mitchell D, Swingler R, et al. The epidemiology of ALS and the role of population-based registries. Vol. 1762, Biochimica et Biophysica Acta - Molecular Basis of Disease. 2006. p. 1150–7.
272. Rooney JPK, Tobin K, Crampsie A, Vajda A, Heverin M, McLaughlin R, et al. Social deprivation and population density are not associated with small area risk of amyotrophic lateral sclerosis. Environ Res. 2015;142:141–7.
273. Roberts AL, Johnson NJ, Chen JT, Cudkowicz ME, Weisskopf MG. Race/ethnicity, socioeconomic status, and ALS mortality in the United States. Neurology. 2016 Nov 29;87(22):2300–8.
274. Caslake R, Taylor K, Scott N, Gordon J, Harris C, Wilde K, et al. Age-, gender-, and socioeconomic status-specific incidence of Parkinson's disease and parkinsonism in North East Scotland: The PINE study. Parkinsonism Relat Disord. 2013 Mar 5;19(5):515–21.
275. Chancellor AM, Slattery JM, Fraser H, Warlow CP. Risk factors for motor neuron disease: a case-control study based on patients from the Scottish Motor Neuron Disease Register. J Neurol Neurosurg Psychiatry. 1993 Nov;56(11):1200–6.
276. Information Services Division. Scottish Heart Disease Statistics. Natl Stat Publ Scotl. 2019;
277. Scottish Government. Scottish Government: Evidence Finder: Summary: Ethnicity [Internet]. Scottish Government. 2015. Available from: http://www.gov.scot/Topics/People/Equality/Equalities/DataGrid/Ethnicity
278. Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, et al. Proportionally more deleterious genetic variation in European than in African populations. Nature. 2008 Feb 21;451:994.

279. Tobin K, Gilthorpe MS, Rooney J, Heverin M, Vajda A, Staines A, et al. Age-period-cohort analysis of trends in amyotrophic lateral sclerosis incidence. J Neurol. 2016 Oct 2;263(10):1919–26.
280. Moody CJ, Mitchell D, Kiser G, Aarsland D, Berg D, Brayne C, et al. Maximizing the potential of longitudinal cohorts for research in neurodegenerative diseases: A community perspective. Front Neurosci. 2017 Aug 29;11(AUG):467.
281. Turner MR, Goldacre R, Ramagopalan S, Talbot K, Goldacre MJ. Autoimmune disease preceding amyotrophic lateral sclerosis: An epidemiologic study. Neurology. 2013 Oct 1;81(14):1222–5.
282. Miller ZA, Sturm VE, Camsari GB, Karydas A, Yokoyama JS, Grinberg LT, et al. Increased prevalence of autoimmune disease within C9 and FTD/MND cohorts Completing the picture. Neurol Neuroimmunol NeuroInflammation. 2016 Dec;3(6):e301.
283. Freedman DM, Curtis RE, Daugherty SE, Goedert JJ, Kuncl RW, Tucker MA. The association between cancer and amyotrophic lateral sclerosis. Cancer Causes Control. 2013 Jan 23;24(1):55–60.
284. Freedman DM, Wu J, Daugherty SE, Kuncl RW, Enewold LR, Pfeiffer RM. The risk of amyotrophic lateral sclerosis after cancer in U.S. elderly adults: A population-based prospective study. Vol. 135, International Journal of Cancer. 2014. p. 1745–50.
285. Turner MR, Goldacre R, Talbot K, Goldacre MJ. Cerebrovascular injury as a risk factor for amyotrophic lateral sclerosis: Table 1. J Neurol Neurosurg Psychiatry. 2016 Mar;87(3):244–6.
286. Chen H, Richard M, Sandler DP, Umbach DM, Kamel F. Head injury and amyotrophic lateral sclerosis. Am J Epidemiol. 2007 Oct 1;166(7):810–6.
287. McKee AC, Gavett BE, Stern RA, Nowinski CJ, Cantu RC, Kowall NW, et al. TDP-43 proteinopathy and motor neuron disease in chronic traumatic encephalopathy. J Neuropathol Exp Neurol. 2010 Sep 1;69(9):918–29.
288. Walt GS, Burris HM, Brady CB, Spencer KR, Alvarez VE, Huber BR, et al. Chronic Traumatic Encephalopathy Within an Amyotrophic Lateral Sclerosis Brain Bank Cohort. J Neuropathol Exp Neurol. 2018 Dec 1;77(12):1091–100.
289. Turner MR, Abisgold J, Yeates DGR, Talbot K, Goldacre MJ. Head and other physical trauma requiring hospitalisation is not a significant risk factor in the development of ALS. J Neurol Sci. 2010 Jan 15;288(1–2):45–8.
290. Belbasis L, Bellou V, Evangelou E. Environmental risk factors and amyotrophic lateral sclerosis: An umbrella review and critical assessment of current evidence from systematic reviews and meta-analyses of observational studies. Neuroepidemiology. 2016 Jan 6;46(2):96–105.
291. Peters TL, Fang F, Weibull CE, Sandler DP, Kamel F, Ye W. Severe head injury and amyotrophic lateral sclerosis. Amyotroph Lateral Scler Front Degener. 2013 May 4;14(4):267–72.
292. Hamidou B, Couratier P, Besançon C, Nicol M, Preux PM, Marin B. Epidemiological evidence that physical activity is not a risk factor for ALS. Vol. 29, European Journal of Epidemiology. 2014. p. 459–75.
293. Huisman MHB, Seelen M, de Jong SW, Dorresteijn KRIS, van Doormaal PTC, van der Kooi AJ, et al. Lifetime physical activity and the risk of amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 2013 Sep;84(9):976–81.
294. Eaglehouse YL, Talbott EO, Chang Y, Kuller LH. Participation in physical activity and risk for amyotrophic lateral sclerosis mortality among postmenopausal women. JAMA Neurol. 2016 Mar 1;73(3):329–36.

295. McGuire V, Longstreth WT, Nelson LM, Koepsell TD, Checkoway H, Morgan MS, et al. Occupational exposures and amyotrophic lateral sclerosis: A population- based case-control study. Am J Epidemiol. 1997 Jun 15;145(12):1076–88.
296. Weisskopf MG, Morozova N, O'Reilly EJ, McCullough ML, Calle EE, Thun MJ, et al. Prospective study of chemical exposures and amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 2009 Apr 9;80(5):558–61.
297. Vinceti M, Filippini T, Violi F, Rothman KJ, Costanzini S, Malagoli C, et al. Pesticide exposure assessed through agricultural crop proximity and risk of amyotrophic lateral sclerosis. Environ Heal A Glob Access Sci Source. 2017 Dec 29;16(1):91.
298. Kamel F, Umbach DM, Stallone L, Richards M, Hu H, Sandler DP. Association of lead exposure with survival in amyotrophic lateral sclerosis. Environ Health Perspect. 2008 Jul;116(7):943–7.
299. Santurtún A, Villar A, Delgado-Alvarado M, Riancho J. Trends in motor neuron disease: association with latitude and air lead levels in Spain. Neurol Sci. 2016 Aug 20;37(8):1271–5.
300. Johnson FO, Atchison WD. The role of environmental mercury, lead and pesticide exposure in development of amyotrophic lateral sclerosis. Neurotoxicology. 2009 Sep 1;30(5):761–5.
301. Wills A-M, Landers JE, Zhang H, Richter RJ, Caraganis AJ, Cudkowicz ME, et al. Paraoxonase 1 (PON1) organophosphate hydrolysis is not reduced in ALS. Neurology. 2008 Mar 18;70(12):929–34.
302. D'Ovidio F, d'Errico A, Calvo A, Costa G, Chiò A. Occupations and amyotrophic lateral sclerosis: are jobs exposed to the general public at higher risk? Eur J Public Health. 2017 Aug 1;27(4):643–7.
303. Bergman BP, Mackay DF, Pell JP. Motor neurone disease and military service: evidence from the Scottish Veterans Health Study. Occup Environ Med. 2015 Dec 1;72(12):877–9.
304. Golomb BA, Verden A, Messner AK, Koslik HJ, Hoffman KB. Amyotrophic Lateral Sclerosis Associated with Statin Use: A Disproportionality Analysis of the FDA's Adverse Event Reporting System. Drug Saf. 2018 Apr;41(4):403–13.
305. Drory VE, Bronipolsky T, Artamonov I, Nefussy B. Influence of statins treatment on survival in patients with amyotrophic lateral sclerosis. J Neurol Sci. 2008 Oct 15;273(1–2):81–3.
306. Zinman L, Sadeghi R, Gawel M, Patton D, Kiss A. Are statin medications safe in patients with ALS? Amyotroph Lateral Scler. 2008 Jan 10;9(4):223–8.
307. Armon C. Smoking may be considered an established risk factor for sporadic ALS. Vol. 73, Neurology. American Academy of Neurology; 2009. p. 1693–8.
308. Nelson LM, McGuire V, Longstreth WT, Matkin C. Population-based case-control study of amyotrophic lateral sclerosis in western Washington State. I. Cigarette smoking and alcohol consumption. Am J Epidemiol. 2000 Jan 15;151(2):156–63.
309. Uma K MH. Data Collection Methods and Data Pre- processing Techniques for Healthcare Data Using Data Mining. Int J Sci Eng Res. 2017;8(6):1131–6.
310. Information Services Division (ISD) Scotland. ISD Scotland | Information Services Division | Data Dictionary | Ethnicity Code [Internet]. NHS National Services. 2019 [cited 2019 Jun 19]. Available from: https://www.ndc.scot.nhs.uk/Dictionary-A-Z/Definitions/index.asp?Search=E&ID=243&Title=Ethnicity Code
311. Spataro R, Volanti P, Lo Coco D, La Bella V. Marital status is a prognostic factor in amyotrophic lateral sclerosis. Acta Neurol Scand. 2017 Dec 1;136(6):624–30.
312. Office for National Statistics. ONS Occupation Coding Tool [Internet]. Office for

National Statistics. 2016 [cited 2019 Jun 23]. Available from: https://onsdigital.github.io/dp-classification-tools/standard-occupational-classification/ONS_SOC_occupation_coding_tool.html
313. National Institute for Health and Care Excellence. Overview | Alcohol-use disorders: prevention | Guidance | NICE [Internet]. National Institute for Health and Care Excellence. NICE; 2010 [cited 2019 Jun 23]. Available from: https://www.nice.org.uk/guidance/ph24
314. Czaplinski A, Yen AA, Appel SH. Amyotrophic lateral sclerosis: Early predictors of prolonged survival. J Neurol. 2006 Nov 13;253(11):1428–36.
315. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography (Cop). 2013 Jan 1;36(1):27–46.
316. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996 Dec;49(12):1373–9.
317. Marshall A, Altman DG, Holder RL. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. BMC Med Res Methodol. 2010 Dec 31;10(1):112.
318. Rubin DB, Wiley J, York N, Brisbane C, Singapore T. Multiple Imputation for Nonresponse in Surveys. 1987.
319. National Records of Scotland. Welcome to Scotland's Census | Scotland's Census [Internet]. [cited 2019 Aug 18]. Available from: https://www.scotlandscensus.gov.uk/
320. Mclean J, Christie S, Hinchliffe S, Gray L, Bardsley D, Dean L, et al. A National Statistics Publication for Scotland. 2017;1.
321. Turner MR, Goldacre R, Talbot K, Goldacre MJ. Psychiatric disorders prior to amyotrophic lateral sclerosis. Ann Neurol. 2016 Oct 19;80(6):935–8.
322. Leighton D, Manson L, McHutchison C, Sherlock L, Newton J, Abrahams S, et al. 11 Premorbid neuropsychiatric disease in patients with motor neurone disease in scotland. J Neurol Neurosurg Psychiatry. 2017 Aug 13;88(8):A16.3-A17.
323. Calvo A, Canosa A, Bertuzzo D, Cugnasco P, Solero L, Clerico M, et al. Influence of cigarette smoking on ALS outcome: a population-based study. J Neurol Neurosurg Psychiatry. 2016 Nov 1;87(11):1229–33.
324. Wolf J, Safer A, Wöhrle JC, Palm F, Nix WA, Maschke M, et al. Factors predicting one-year mortality in amyotrophic lateral sclerosis patients - data from a population-based registry. BMC Neurol. 2014;14(1):197.
325. Li X, Li W, Liu G, Shen X, Tang Y. Association between cigarette smoking and Parkinson's disease: A meta-analysis. Arch Gerontol Geriatr. 2015 Nov 1;61(3):510–6.
326. Derkinderen P, Shannon KM, Brundin P. Gut feelings about smoking and coffee in Parkinson's disease. Mov Disord. 2014 Jul 1;29(8):976–9.
327. Brenner D, Hiergeist A, Adis C, Mayer B, Gessner A, Ludolph AC, et al. The fecal microbiome of ALS patients. Neurobiol Aging. 2018 Jan 1;61:132–7.
328. Wright ML, Fournier C, Houser MC, Tansey M, Glass J, Hertzberg VS. Potential Role of the Gut Microbiome in ALS: A Systematic Review. Biol Res Nurs. 2018 Oct 20;20(5):513–21.
329. Mazzini L, Mogna L, De Marchi F, Amoruso A, Pane M, Aloisio I, et al. Potential Role of Gut Microbiota in ALS Pathogenesis and Possible Novel Therapeutic Strategies. J Clin Gastroenterol. 2018;52:S68–70.

330. Thompson AG, Turner MR. Untangling neuroinflammation in amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 2019 Jul 11;jnnp-2019-321242.
331. Peters S, Visser AE, D'Ovidio F, Vlaanderen J, Portengen L, Beghi E, et al. Effect modification of the association between total cigarette smoking and ALS risk by intensity, duration and time-since-quitting: Euro-MOTOR. J Neurol Neurosurg Psychiatry. 2019 Aug 21;jnnp-2019-320986.
332. Langkammer C, Enzinger C, Quasthoff S, Grafenauer P, Soellinger M, Fazekas F, et al. Mapping of iron deposition in conjunction with assessment of nerve fiber tract integrity in amyotrophic lateral sclerosis. J Magn Reson Imaging. 2010 May 24;31(6):1339–45.
333. Kwan JY, Jeong SY, van Gelderen P, Deng HX, Quezado MM, Danielian LE, et al. Iron accumulation in deep cortical layers accounts for MRI signal abnormalities in ALS: Correlating 7 tesla MRI and pathology. Ashizawa T, editor. PLoS One. 2012 Apr 17;7(4):e35241.
334. Abel O, Shatunov A, Jones AR, Andersen PM, Powell JF, Al-Chalabi A. Development of a Smartphone App for a Genetics Website: The Amyotrophic Lateral Sclerosis Online Genetics Database (ALSoD). JMIR mHealth uHealth. 2013;1(2):e18.
335. Abel, O. ALSoD (6.0) [Internet]. 2016. Available from: http://alsod.iop.kcl.ac.uk/
336. Jones CT, Brock DJ, Chancellor AM, Warlow CP, Swingler RJ. Cu/Zn superoxide dismutase (SOD1) mutations and sporadic amyotrophic lateral sclerosis. Lancet (London, England). 1993 Oct 23;342(8878):1050–1.
337. Kerr SM, Liewald DC, Campbell A, Taylor K, Wild SH, Newby D, et al. Generation Scotland: Donor DNA Databank; A control DNA resource. BMC Med Genet. 2010 Dec 23;11(1):166.
338. Andrews S. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. 2012 [cited 2018 Jul 17]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
339. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011 May 2;17(1):10.
340. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013 Mar 16;
341. Broad Institute. Picard tools. Broad Institute. 2016.
342. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep 1;20(9):1297–303.
343. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010 Sep 1;38(16):e164–e164.
344. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015 Oct 1;526(7571):68–74.
345. Lek M, Karczewski KJ, Minikel E V, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016 Aug 18;536(7616):285–91.
346. Stenson PD, Mort M, Ball E V., Evans K, Hayden M, Heywood S, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Vol. 136, Human Genetics. 2017. p. 665–77.
347. Brunak S, Engelbrecht J, Knudsen S. Prediction of human mRNA donor and acceptor sites from the DNA sequence. J Mol Biol. 1991 Jul 5;220(1):49–65.

348. Del Bo R, Tiloca C, Pensato V, Corrado L, Ratti A, Ticozzi N, et al. Novel optineurin mutations in patients with familial and sporadic amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 2011 Nov 1;82(11):1239–43.
349. Williams KL, McCann EP, Fifita JA, Zhang K, Duncan EL, Leo PJ, et al. Novel TBK1 truncating mutation in a familial amyotrophic lateral sclerosis patient of Chinese origin. Neurobiol Aging. 2015 Dec;36(12):3334.e1-3334.e5.
350. Gijselinck I, Van Mossevelde S, van der Zee J, Sieben A, Philtjens S, Heeman B, et al. Loss of TBK1 is a frequent cause of frontotemporal dementia in a Belgian cohort. Neurology. 2015 Nov 18;
351. Borghero G, Pugliatti M, Marrosu F, Marrosu MG, Murru MR, Floris G, et al. TBK1 is associated with ALS and ALS-FTD in Sardinian patients. Neurobiol Aging. 2015 Jul;43:180.e1-180.e5.
352. Exome Aggregate Consortium. ExAC Browser [Internet]. Online. 2016 [cited 2018 Nov 2]. Available from: http://exac.broadinstitute.org/
353. Le Ber I, De Septenville A, Millecamps S, Camuzat A, Caroppo P, Couratier P, et al. TBK1 mutation frequencies in French frontotemporal dementia and amyotrophic lateral sclerosis cohorts. Neurobiol Aging. 2015 Nov;36(11):3116.e5-3116.e8.
354. Pottier C, Bieniek KF, Finch NC, van de Vorst M, Baker M, Perkersen R, et al. Whole-genome sequencing reveals important role for TBK1 and OPTN mutations in frontotemporal lobar degeneration without motor neuron disease. Acta Neuropathol. 2015 Jul 6;130(1):77–92.
355. Tsai P-C, Liu Y-C, Lin K-P, Liu Y-T, Liao Y-C, Hsiao C-T, et al. Mutational analysis of TBK1 in Taiwanese patients with amyotrophic lateral sclerosis. Neurobiol Aging. 2016 Apr;40:191.e11-191.e16.
356. Lattante S, Conte A, Zollino M, Luigetti M, Del Grande A, Marangi G, et al. Contribution of major amyotrophic lateral sclerosis genes to the etiology of sporadic disease. Neurology. 2012 Jul 3;79(1):66–72.
357. Li TM, Alberman E, Swash M. Clinical features and associations of 560 cases of motor neuron disease. J Neurol Neurosurg Psychiatry. 1990 Dec 1;53(12).
358. Logroscino G, Traynor BJ, Hardiman O, Chio' A, Couratier P, Mitchell JD, et al. Descriptive epidemiology of amyotrophic lateral sclerosis: new evidence and unsolved issues. J Neurol Neurosurg Psychiatry. 2008 Jan 1;79(1):6–11.
359. Traynor BJ, Codd MB, Corr B, Forde C, Frost E, Hardiman OM. Clinical features of amyotrophic lateral sclerosis according to the El Escorial and Airlie House diagnostic criteria: A population-based study. Arch Neurol. 2000 Aug 1;57(8):1171–6.
360. Andersen PM. Amyotrophic lateral sclerosis associated with mutations in the CuZn superoxide dismutase gene. Vol. 6, Current Neurology and Neuroscience Reports. Current Medicine Group; 2006. p. 37–46.
361. Johnson L, Miller JW, Gkazi AS, Vance C, Topp SD, Newhouse SJ, et al. Screening for OPTN mutations in a cohort of British amyotrophic lateral sclerosis patients. Neurobiol Aging. 2012 Dec;33(12):2948.e15-2948.e17.
362. Shu S, Li XGL, Liu Q, Liu F, Cui B, Liu MS, et al. Screening of the TBK1 gene in familial and sporadic amyotrophic lateral sclerosis patients of Chinese origin. Amyotroph Lateral Scler Front Degener. 2016 Nov 16;17(7–8):605–7.
363. FALCONER DS. The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann Hum Genet. 1965 Aug 1;29(1):51–76.
364. Millecamps S, Boillée S, Le Ber I, Seilhean D, Teyssou E, Giraudeau M, et al. Phenotype difference between ALS patients with expanded repeats in C9ORF72 and patients with mutations in other ALS-related genes. J Med Genet. 2012

Apr;49(4):258–63.
365. Millecamps S, Salachas F, Cazeneuve C, Gordon P, Bricka B, Camuzat A, et al. SOD1, ANG, VAPB, TARDBP, and FUS mutations in familial amyotrophic lateral sclerosis: genotype-phenotype correlations. J Med Genet. 2010 Aug 1;47(8):554–60.
366. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May;17(5):405–24.
367. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. Nature. 2014 Apr 24;508(7497):469–76.
368. Cheon JY, Mozersky J, Cook-deegan R. Variants of uncertain significance in BRCA : a harbinger of ethical and policy issues to come ? 2014;1–10.
369. Murray ML, Cerrato F, Bennett RL, Jarvik GP. Follow-up of carriers of BRCA1 and BRCA2 variants of unknown significance: variant reclassification and surgical decisions. Genet Med. 2011 Dec;13(12):998–1005.
370. Ly C V., Miller TM. Emerging antisense oligonucleotide and viral therapies for amyotrophic lateral sclerosis. Vol. 31, Current opinion in neurology. 2018. p. 648–54.
371. Marangi G, Traynor BJ. Genetic causes of amyotrophic lateral sclerosis: New genetic analysis methodologies entailing new opportunities and challenges. Brain Res. 2015 May 14;1607:75–93.
372. Dekker AM, Seelen M, van Doormaal PTC, van Rheenen W, Bothof RJP, van Riessen T, et al. Large-scale screening in sporadic amyotrophic lateral sclerosis identifies genetic modifiers in C9orf72 repeat carriers. Neurobiol Aging. 2016 Mar 1;39:220.e9-220.e15.
373. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science. 2012 Jul 6;337(6090):64–9.
374. Vrabec K, Koritnik B, Leonardis L, Dolenc-Grošelj L, Zidar J, Smith B, et al. Genetic analysis of amyotrophic lateral sclerosis in the Slovenian population. Neurobiol Aging. 2015 Mar 1;36(3):1601.e17-1601.e20.
375. Özoğuz A, Uyan Ö, Birdal G, Iskender C, Kartal E, Lahut S, et al. The distinct genetic pattern of ALS in Turkey and novel mutations. Neurobiol Aging. 2015 Apr 1;36(4):1764.e9-1764.e18.
376. Kim H-J, Oh K-W, Kwon M-J, Oh S-I, Park J-S, Kim Y-E, et al. Identification of mutations in Korean patients with amyotrophic lateral sclerosis using multigene panel testing. Neurobiol Aging. 2015 Sep 30;37:209.e9-209.e16.
377. Nakamura R, Sone J, Atsuta N, Tohnai G, Watanabe H, Yokoi D, et al. Next-generation sequencing of 28 ALS-related genes in a Japanese ALS cohort. Neurobiol Aging. 2016 Mar 1;39:219.e1-219.e8.
378. Morgan S, Shatunov A, Sproviero W, Jones AR, Shoai M, Hughes D, et al. A comprehensive analysis of rare genetic variation in amyotrophic lateral sclerosis in the UK. Brain. 2017 Jun 1;140(6):1611–8.
379. Keogh MJ, Wei W, Wilson I, Coxhead J, Ryan S, Rollinson S, et al. Genetic compendium of 1511 human brains available through the UK Medical Research Council Brain Banks Network Resource. Genome Res. 2017 Jan;27(1):165–73.
380. UK Genetic Testing Network. London Institute of Neurology - UK Genetic Testing Network [Internet]. NHS UK Genetic Testing Network. [cited 2019 Jul 17]. Available from: https://ukgtn.nhs.uk/find-a-test/search-by-laboratory/laboratory/london-

institute-of-neurology-47/
381. UK Genetic Testing Network. Sheffield RGC - UK Genetic Testing Network [Internet]. [cited 2019 Jul 17]. Available from: https://ukgtn.nhs.uk/find-a-test/search-by-laboratory/laboratory/sheffield-rgc-40/
382. National Centre for Biotechnology Information: Gene Testing Registry. ALS panel - Tests - GTR - NCBI [Internet]. 2017 [cited 2019 Jul 17]. Available from: https://www.ncbi.nlm.nih.gov/gtr/tests/503152/
383. Hamosh A, Scott AF, Amberger J, Valle D, McKusick VA. Online Mendelian Inheritance in Man (OMIM). Hum Mutat. 2000;15(1):57–61.
384. ClinVar. ClinVar [Internet]. National Center for Biotechnology Information, U.S. National Library of Medicine. 2016 [cited 2018 Nov 2]. Available from: https://www.ncbi.nlm.nih.gov/clinvar/
385. University of California Santa Cruz. UCSC Genome Browser. UcscEdu. 2016;(May):3–5.
386. Galaxy Developer Team. Galaxy Project Page [Internet]. Online. 2010 [cited 2018 Nov 2]. Available from: https://usegalaxy.org/
387. SeattleSeq Annotation. SeattleSeq Variation Annotation [Internet]. University of Washington and Hudson-Alpha Institute for Biotechnology. [cited 2018 Nov 2]. Available from: http://snp.gs.washington.edu/SeattleSeqAnnotation138/
388. Computational Biology Center | Memorial Sloan Kettering Cancer Center. MutationAssessor.org /// functional impact of protein mutations [Internet]. 2015 [cited 2018 Nov 2]. Available from: http://mutationassessor.org/r3/
389. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino Acid Substitutions and Indels. de Brevern AG, editor. PLoS One. 2012 Oct 8;7(10):e46688.
390. ANNOVAR. ANNOVAR Documentation [Internet]. 2018 [cited 2018 Nov 2]. Available from: http://annovar.openbioinformatics.org/en/latest/
391. Fox BI, Hollingsworth JC, Gray MD, Hollingsworth ML, Gao J, Hansen RA. Developing an expert panel process to refine health outcome definitions in observational data. J Biomed Inform. 2013 Oct;46(5):795–804.
392. Thompson BA, Spurdle AB, Plazzer J-P, Greenblatt MS, Akagi K, Al-Mulla F, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat Genet. 2014 Feb;46(2):107–15.
393. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet. 2015 Apr 15;24(8):2125–37.
394. Krippendorff K. Agreement and Information in the Reliability of Coding. Commun Methods Meas. 2011 Apr 1;5(2):93–112.
395. Picher-Martel V, Valdmanis PN, Gould P V., Julien J-P, Dupré N. From animal models to human disease: a genetic approach for personalized medicine in ALS. Acta Neuropathol Commun. 2016 Jul 11;4(1):70.
396. Jarvik GP, Browning BL. Consideration of Cosegregation in the Pathogenicity Classification of Genomic Variants. Am J Hum Genet. 2016 Jun 2;98(6):1077–81.
397. Amendola LM, Jarvik GP, Leo MC, Mclaughlin HM, Akkari Y, Amaral MD, et al. Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. 2016;
398. Nalbandian A, Llewellyn KJ, Badadani M, Yin HZ, Nguyen C, Katheria V, et al. A progressive translational mouse model of human valosin-containing protein disease:

the VCP(R155H/+) mouse. Muscle Nerve. 2013 Feb;47(2):260–70.
399. Philips T, Rothstein JD. Rodent Models of Amyotrophic Lateral Sclerosis. In: Current Protocols in Pharmacology. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2015. p. 5.67.1-5.67.21.
400. Williams KL, Topp S, Yang S, Smith B, Fifita JA, Warraich ST, et al. CCNF mutations in amyotrophic lateral sclerosis and frontotemporal dementia. Nat Commun. 2016 Apr 15;7:11253.
401. Dormann D, Rodde R, Edbauer D, Bentmann E, Fischer I, Hruscha A, et al. ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. EMBO J. 2010 Aug 18;29(16):2841–57.
402. Vance C, Scotter EL, Nishimura AL, Troakes C, Mitchell JC, Kathe C, et al. ALS mutant FUS disrupts nuclear localization and sequesters wild-type FUS within cytoplasmic stress granules. Hum Mol Genet. 2013 Jul 1;22(13):2676–88.
403. Yang C, Danielson EW, Qiao T, Metterville J, Brown RH, Landers JE, et al. Mutant PFN1 causes ALS phenotypes and progressive motor neuron degeneration in mice by a gain of toxicity. Proc Natl Acad Sci U S A. 2016 Oct 11;113(41):E6209–18.
404. Tanaka Y, Hasegawa M. Profilin 1 mutants form aggregates that induce accumulation of prion-like TDP-43. Prion. 2016 Jul 3;10(4):283–9.
405. Daoud H, Valdmanis PN, Kabashi E, Dion P, Dupré N, Camu W, et al. Contribution of TARDBP mutations to sporadic amyotrophic lateral sclerosis. J Med Genet. 2009 Feb 1;46(2):112 LP-114.
406. Chio A, Calvo A, Mazzini L, Cantello R, Mora G, Moglia C, et al. Extensive genetics of ALS : A population-based study in Italy. Neurology. 2012 Nov 6;79(19):1983–9.
407. Renaud L, Picher-Martel V, Codron P, Julien J-P. Key role of UBQLN2 in pathogenesis of amyotrophic lateral sclerosis and frontotemporal dementia. Acta Neuropathol Commun. 2019 Dec 18;7(1):103.
408. Bersano A, Del Bo R, Lamperti C, Ghezzi S, Fagiolari G, Fortunato F, et al. Inclusion body myopathy and frontotemporal dementia caused by a novel VCP mutation. Neurobiol Aging. 2009 May;30(5):752–8.
409. Golden Helix I. VarSeq [Internet]. Golden Helix, Inc, Bozeman, MT. 2018 [cited 2018 Nov 11]. Available from: http://goldenhelix.com
410. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. Hum Mutat. 2016 Mar;37(3):235–41.
411. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016 Dec 6;17(1):122.
412. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res. 2014 Dec 16;42(22):13534–44.
413. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. 2019 Jan 30;531210.
414. Akimoto C, Morita M, Atsuta N, Sobue G, Nakano I. High-Resolution Melting (HRM) Analysis of the Cu/Zn Superoxide Dismutase (SOD1) Gene in Japanese Sporadic Amyotrophic Lateral Sclerosis (SALS) Patients. Neurol Res Int. 2011;2011:1–8.
415. Stewart HG, Mackenzie IR, Eisen A, Brännström T, Marklund SL, Andersen PM. Clinicopathological phenotype of ALS with a novel G72C SOD1 gene mutation mimicking a myopathy. Muscle Nerve. 2006 May;33(5):701–6.
416. Lattante S, Doronzio PN, Marangi G, Conte A, Bisogni G, Bernardo D, et al.

Coexistence of variants in TBK1 and in other ALS-related genes elucidates an oligogenic model of pathogenesis in sporadic ALS. Neurobiol Aging. 2019;
417. Watts GDJ, Wymer J, Kovach MJ, Mehta SG, Mumm S, Darvish D, et al. Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. Nat Genet. 2004 Apr 21;36(4):377–81.
418. Kim E-J, Park Y-E, Kim D-S, Ahn B-Y, Kim H-S, Chang YH, et al. Inclusion Body Myopathy With Paget Disease of Bone and Frontotemporal Dementia Linked to VCP p.Arg155Cys in a Korean Family. Arch Neurol. 2011 Jun 1;68(6):787–96.
419. Topp SD, Fallini C, Shibata H, Chen HJ, Troakes C, King A, et al. Mutations in the vesicular trafficking protein Annexin A11 are associated with amyotrophic lateral sclerosis Bradley N. Smith. Sci Transl Med. 2017 May 3;9(388):eaad9157.
420. Lee A, Rayner SL, Gwee SSL, De Luca A, Shahheydari H, Sundaramoorthy V, et al. Pathogenic mutation in the ALS/FTD gene, CCNF, causes elevated Lys48-linked ubiquitylation and defective autophagy. Cell Mol Life Sci. 2018 Jan 29;75(2):335–54.
421. Liu X, Wu C, He J, Zhang N, Fan D. Two rare variants of the ANXA11 gene identified in Chinese patients with amyotrophic lateral sclerosis. Neurobiol Aging. 2019 Feb;74:235.e9-235.e12.
422. Tsai P-C, Liao Y-C, Jih K-Y, Soong B-W, Lin K-P, Lee Y-C. Genetic analysis of ANXA11 variants in a Han Chinese cohort with amyotrophic lateral sclerosis in Taiwan. Neurobiol Aging. 2018 Dec 1;72:188.e1-188.e2.
423. Mitchell J, Paul P, Chen H-J, Morris A, Payling M, Falchi M, et al. Familial amyotrophic lateral sclerosis is associated with a mutation in D-amino acid oxidase. Proc Natl Acad Sci U S A. 2010 Apr 20;107(16):7556–61.
424. Vilarino-Guell C, Wider C, Soto-Ortolaza AI, Cobb SA, Kachergus JM, Keeling BH, et al. Characterization of DCTN1 genetic variability in neurodegeneration. Neurology. 2009 Jun 9;72(23):2024–8.
425. Takahashi Y, Fukuda Y, Yoshimura J, Toyoda A, Kurppa K, Moritoyo H, et al. ERBB4 Mutations that Disrupt the Neuregulin-ErbB4 Pathway Cause Amyotrophic Lateral Sclerosis Type 19. Am J Hum Genet. 2013 Nov 7;93(5):900–5.
426. Takahashi Y, Uchino A, Shioya A, Sano T, Matsumoto C, Numata-Uematsu Y, et al. Altered immunoreactivity of ErbB4, a causative gene product for ALS19, in the spinal cord of patients with sporadic ALS. Neuropathology. 2019 May 24;neup.12558.
427. Boehringer A, Garcia-Mansfield K, Singh G, Bakkar N, Pirrotte P, Bowser R. ALS Associated Mutations in Matrin 3 Alter Protein-Protein Interactions and Impede mRNA Nuclear Export. Sci Rep. 2017 Dec 6;7(1):14529.
428. Leblond CS, Gan-Or Z, Spiegelman D, Laurent SB, Szuto A, Hodgkinson A, et al. Replication study of MATR3 in familial and sporadic amyotrophic lateral sclerosis. Neurobiol Aging. 2016 Jan 1;37:209.e17-209.e21.
429. Johnson JO, Pioro EP, Boehringer A, Chia R, Feit H, Renton AE, et al. Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. Nat Neurosci. 2014 May 30;17(5):664–6.
430. Al-Chalabi A, Andersen PM, Nilsson P, Chioza B, Andersson JL, Russ C, et al. Deletions of the heavy neurofilament subunit tail in amyotrophic lateral sclerosis. Hum Mol Genet. 1999 Feb 1;8(2):157–64.
431. Chen Y-Z, Bennett CL, Huynh HM, Blair IP, Puls I, Irobi J, et al. DNA/RNA Helicase Gene Mutations in a Form of Juvenile Amyotrophic Lateral Sclerosis (ALS4). Am J Hum Genet. 2004 Jun;74(6):1128–35.
432. Fecto F, Yan J, Vemula SP, Liu E, Yang Y, Chen W, et al. SQSTM1 mutations in familial

and sporadic amyotrophic lateral sclerosis. Arch Neurol. 2011 Nov 1;68(11):1440–6.
433. Nykamp K, Anderson M, Powers M, Garcia J, Herrera B, Ho Y-Y, et al. Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria. Genet Med. 2017 Oct 11;19(10):1105–17.
434. Nguyen HP, Van Broeckhoven C, van der Zee J. ALS Genes in the Genomic Era and their Implications for FTD. Vol. 34, Trends in Genetics. Elsevier Current Trends; 2018. p. 404–23.
435. Barber IS, Braae A, Clement N, Patel T, Guetta-Baranes T, Brookes K, et al. Mutation analysis of sporadic early-onset Alzheimer's disease using the NeuroX array. Neurobiol Aging. 2017;49:215.e1-215.e8.
436. Swerdlow RH, Miller BB, Lopes MBS, Mandell JW, Wooten GF, Damgaard P, et al. Autosomal dominant subcortical gliosis presenting as frontotemporal dementia. Neurology. 2009 Jan 20;72(3):260–7.
437. Xi Z, van Blitterswijk M, Zhang M, McGoldrick P, McLean JR, Yunusova Y, et al. Jump from Pre-mutation to Pathologic Expansion in C9orf72. Am J Hum Genet. 2015 May 20;96(6):962–70.
438. Iacoangeli A, Al Khleifat A, Jones AR, Sproviero W, Shatunov A, Opie-Martin S, et al. C9orf72 intermediate expansions of 24–30 repeats are associated with ALS. Acta Neuropathol Commun. 2019 Dec 17;7(1):115.
439. Majounie E, Renton AE, Mok K, Dopper EG, Waite A, Rollinson S, et al. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. Lancet Neurol. 2012 Apr;11(4):323–30.
440. Chiò A, Mazzini L, D'Alfonso S, Corrado L, Canosa A, Moglia C, et al. The multistep hypothesis of ALS revisited. Neurology. 2018 Jul 25;91(7):e635–42.
441. McCampbell A, Cole T, Wegener AJ, Tomassy GS, Setnicka A, Farley BJ, et al. Antisense oligonucleotides extend survival and reverse decrement in muscle response in ALS models. J Clin Invest. 2018 Aug 1;128(8):3558–67.
442. Keogh MJ, Wei W, Aryaman J, Wilson I, Talbot K, Turner MR, et al. Oligogenic genetic variation of neurodegenerative disease genes in 980 postmortem human brains. J Neurol Neurosurg Psychiatry. 2018 Jan 13;89(8):813–6.
443. Shashi V, McConkie-Rosell A, Schoch K, Kasturi V, Rehder C, Jiang YH, et al. Practical considerations in the clinical application of whole-exome sequencing. Clin Genet. 2016 Feb;89(2):173–81.
444. Mu W, Lu H-M, Chen J, Li S, Elliott AM. Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. J Mol Diagnostics. 2016 Nov;18(6):923–32.
445. Wang H, Peng J, Wang B, Lu X, Zheng JZ, Wang K, et al. Inconsistency Between Univariate and Multiple Logistic Regressions. Shanghai Arch psychiatry. 2017 Apr 25;29(2):124–8.
446. Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008 Nov 10;28(5):1–26.
447. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016 Jan;69:245–7.
448. Vittinghoff E, McCulloch CE. Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression. Am J Epidemiol. 2007 Jan 12;165(6):710–8.
449. Fallis BA, Hardiman O. Aggregation of neurodegenerative disease in ALS kindreds. Amyotroph Lateral Scler. 2009 Jan 10;10(2):95–8.
450. Majoor-Krakauer D, Ottman R, Johnson WG, Rowland LP. Familial aggregation of

amyotrophic lateral sclerosis, dementia, and Parkinson's disease: evidence of shared genetic susceptibility. Neurology. 1994 Oct 1;44(10):1872–7.
451. Murphy NA, Arthur KC, Tienari PJ, Houlden H, Chiò A, Traynor BJ. Age-related penetrance of the C9orf72 repeat expansion. Sci Rep. 2017 Dec 18;7(1):2116.
452. Klim JR, Vance C, Scotter EL. Antisense oligonucleotide therapies for Amyotrophic Lateral Sclerosis: Existing and emerging targets. Int J Biochem Cell Biol. 2019 May 1;110:149–53.
453. Pang SYY, Hsu JS, Teo KC, Li Y, Kung MHW, Cheah KSE, et al. Burden of rare variants in ALS genes influences survival in familial and sporadic ALS. Neurobiol Aging. 2017;58:238.e9-238.e15.
454. Gowland A, Opie-Martin S, Scott KM, Jones AR, Mehta PR, Batts CJ, et al. Predicting the future of ALS: the impact of demographic change and potential new treatments on the prevalence of ALS in the United Kingdom, 2020–2116. Amyotroph Lateral Scler Front Degener. 2019 Apr 3;20(3–4):264–74.
455. Volk AE, Weishaupt JH, Andersen PM, Ludolph AC, Kubisch C. Current knowledge and recent insights into the genetic basis of amyotrophic lateral sclerosis. Vol. 30, Medizinische Genetik. Springer; 2018. p. 252–8.
456. Grollemund V, Pradat PF, Querin G, Delbot F, Le Chat G, Pradat-Peyre JF, et al. Machine learning in amyotrophic lateral sclerosis: Achievements, pitfalls, and future directions. Vol. 13, Frontiers in Neuroscience. 2019. p. 135.
457. Project MinE. Home - Project MinE [Internet]. [cited 2019 Oct 8]. Available from: https://www.projectmine.com/
458. Longinetti E, Regodón Wallin A, Samuelsson K, Press R, Zachau A, Ronnevi L-O, et al. The Swedish motor neuron disease quality registry. Amyotroph Lateral Scler Front Degener. 2018 Oct 2;19(7–8):528–37.
459. Walker KL, Rodrigues MJ, Watson B, Reilly C, Scotter EL, Brunton H, et al. Establishment and 12-month progress of the New Zealand Motor Neurone Disease Registry. J Clin Neurosci. 2019 Feb 1;60:7–11.
460. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: The lothian birth cohorts of 1921 and 1936. Int J Epidemiol. 2012 Dec 1;41(6):1576–84.
461. Lee H, Herbert RD, McAuley JH. Mediation Analysis. JAMA. 2019 Feb 19;321(7):697.

Appendix 1: CARE-MND Paper Proforma Version 6 (Final), 29/01/2016

Motor Neurone Disease (MND) Clinical Specialist Assessment

Patient Details

Family Name		First Name		Addressograph ***or*** *Name DoB Unit number CHI*
Maiden Name		Known As		
Address 1		Telephone		
Address 2		Mobile		
Town		Email		
Postcode				

DoB		CHI		Gender	
Hospital 1		Marital Status		Religion	
Ethnic Origin		Occupational History			
Military Service					

- Patient has given permission for the following patient alert to be placed on their electronic Trak record
  Yes ☐ No ☐ N/A☐ (e.g. other hosp)
- Patient has given permission for data to be shared with other healthcare professionals
  Yes ☐ No ☐
- Patient has given permission for data to be shared with voluntary organisations
  Yes ☐ No ☐
- Patient has agreed to be contacted by a member of a research team
  Yes ☐ No ☐

Care Advisor (CA) Name		Date of Initial Assessment	
Date of Symptom Onset	Date of Diagnosis	Date of Referral to CA	Referring Agent to CA
Date of 1st Contact with CA	Location of Initial Assessment	Date of 1st Home Visit	Date of Initial Letter to GP by CA

MND Consultant Name:	MND Clinic: Date	MND Clinic: Location	**Method of follow-up** *(circle):*
Date Referral to Neurology	Date 1st Seen in Neurology	Teleneurology? Y / N	*Home / Clinic / Other - please state:*

Next of Kin

Next of Kin: Name: Relationship:	e-mail: Phone: Mob: Address: Postcode
Emergency contact (or same as Next of kin ☐): Name: Relationship:	e-mail: Phone: Mob: Address: Postcode:
Permission to share information Yes ☐ No ☐	Date Permission Given

Addressograph ***or***
Name
DoB
Unit number
CHI

Please select the Health Board area where the patient normally resides: (please circle)

Ayrshire & Arran	Borders	Dumfries & Galloway	Fife	Forth Valley	Grampian	Greater Glasgow & Clyde	Highland	Lanark -shire	Lothian	Orkney	Shetland	Tayside	Western Isles

Please select the Local Authority area where the patient normally resides: (please circle)

Aberdeen City	Aberdeenshire	Angus	Argyll & Bute	Clackmannanshire	Dumfries & Galloway	Dundee City
East Ayrshire	East Dunbartonshire	East Lothian	East Renfrewshire	City of Edinburgh	Falkirk	Fife
Glasgow City	Highland	Inverclyde	Midlothian	Moray	North Ayrshire	North Lanarkshire
Orkney Islands	Perth & Kinross	Renfrewshire	Scottish Borders	Shetland Island	South Ayrshire	South Lanarkshire
Stirling Council	West Dunbartonshire	West Lothian	Western Isles			

Support Network

	Contact name	Address	Tel:	Mob:	E-mail:
Consultant Neurologist					
Other Consultant					
GP					
District Nurse					
OT					
Physio					
SLT					
Dietician					
Care Manager/ Social Worker					
Home Care Co-ordinator					
Palliative Care					
Respiratory Service					
Respiratory Consultant					
Rehabilitation					
Private Care Agency					
Welfare & Benefits Officer					

Presenting Symptoms/Clinical History

Addressograph ***or***
Name
DoB
Unit number
CHI

Source of Clinical History and Information:

Handedness (circle):
R L

Presenting Symptoms	**Bulbar**	**Upper Limb**	**Lower Limb**	**Other**
Region of onset				
Regions currently affected				

Investigations

		Date	Comments
Imaging	Yes ☐ No ☐		
Electrophysiology	Yes ☐ No ☐		
Lumbar Puncture	Yes ☐ No ☐		
Other	Yes ☐ No ☐		

Classification (circle)					
Amyotrophic Lateral Sclerosis (ALS) (El Escorial possible/probable/definite)	Progressive Muscular Atrophy (PMA)	Primary Lateral Sclerosis (PLS)	Progressive Bulbar Palsy (PBP)	MND with Fronto-temporal dementia (MND-FTD)	Other (specify)

Addressograph ***or***
Name
DoB
Unit number
CHI

Past Medical and Surgical History	Date

History of Head Injury: Yes ☐ No ☐

If Yes, Date and Description:

History of Blood Transfusion: Yes ☐ No ☐

If Yes, Dates and Description:

Social History	**Exercise Participation:**
Smoking: Alcohol: Drugs:	Light ☐ Moderate ☐ Heavy ☐ Describe:

History of heavy metal exposure: Yes ☐ No ☐

History of pesticide exposure: Yes ☐ No ☐

If Yes, Dates and Descriptions:

Known Allergies	Metal implants
Special Risks	Pacemaker Yes ☐ No ☐
	Medical Alert Badge Yes ☐ No ☐

Addressograph **or**
Name
DoB
Unit number
CHI

Known Familial Cases of:	
MND	Dementia
Other neurological conditions	Psychiatric conditions

Genogram	
Information Provided By:	Date Provided:
Referred for Genetic counselling Yes ☐ No ☐	Date Referred

Riluzole Discussed Yes ☐ No ☐	Date Discussed	Patient Decision Yes ☐ No ☐ Undecided ☐
Date shared Protocol info sent to GP	Bloods Done Yes ☐ No ☐	Date Riluzole Discontinued
Reason for discontinuation:		

CURRENT MEDICATION (including non-prescription medications)	Dates

Addressograph ***or***
Name
DoB
Unit number
CHI

Research

	Date
Scottish MND Register registration/Patient consents to be contacted: Yes ☐ No ☐	
Scottish MND Register registration: Yes ☐ No ☐	
Scottish Regenerative Neurology Tissue Bank discussed: Yes ☐ No ☐	
Scottish Regenerative Neurology Tissue Bank consen/tdonation: Yes ☐ No ☐	
Brain and Spinal Cord Donation Discussed: Yes ☐ No ☐	
Brain and Spinal Cord Donation Consent: Yes ☐ No ☐	
Other Research Discussed:	
Other Comments	

King's Staging (For Research)	
Date	Stage (1 – 4)

Addressograph **or**
Name
DoB
Unit number
CHI

Assessment

Home Environment

Access to House	House Layout
House Ownership	Mobility/Falls Risk
Lives Alone Yes ☐ No ☐	Lives with:

Understanding of MND	Patient	
	Family/Carer	
	Patient Goals	
Social Support		
Psychological Support		
Spiritual Support		
Falls Assessment	Date	Comment
Emotional Lability		
Cognitive Assessment		
Referral to Neuro-psychologist	Date	Neuropsychologist Details
Cognitive Screening/ ECAS	Date	Comments

Addressograph ***or***
Name
DoB
Unit number
CHI

Social Work

Referral of Patient	Date:	Details
Carer Assessment Referral	Date:	Details

Nutrition Assessment

Date	Weight (kg)	Height (cm)	BMI	Healthy Weight	Weight Loss %	Comments

Nutrition

Referral to Dietician	Date	Details
Feeding Tube Discussed Date:	Comments	
Does Patient wish a feeding tube? Yes ☐ No ☐ Undecided ☐ Date:	Date of GI Referral	GI Team Details
Date of GI Assessment	GI Outcome	
Date Gastrostomy/ Nasogastric Inserted	Gastrostomy/ Nasogastric Hospital	
Care & Use of Feeding Tube	Patient ☐ Family ☐ Carer ☐ Social Work ☐	

Speech and Language Therapy

Referral to SLT	Date	Details
Discussion of Communication Options	Date	Outcome
Sialorrhea Yes ☐ No ☐	Comments	
Augmentative and alternative communication (AAC)		

Addressograph ***or***
Name
DoB
Unit number
CHI

Speech Assessment

Date	Outcome

Swallowing Assessment

Date	Outcome

INDICATIONS FOR NON-INVASIVE VENTILATION IN MND

1 Symptoms related to respiratory muscle weakness (at least one of the following must be present)

Symptom	**Results**	**Results**	**Results**	**Results**	**Results**	**Results**	**Results**
Date							
Cough							
Dyspnoea							
Orthopnoea							
Disturbed sleep							
Morning headaches							
Poor concentration							
Daytime sleepiness							
Recurrent Chest Infections							
Epworth score >9 *							
Anorexia							
Spot tc CO_2 (If appropriate)							

Addressograph ***or***
Name
DoB
Unit number
CHI

2 *Evidence of respiratory muscle weakness – Pulmonary Function Testing*

Reminder **Height:** cms latest weight recorded:

		Erect (Sitting)		Supine (Lying)		SNIP
Date	**SpO2/Pulse**	**FVC (%)**	**FEV1 (%)**	**FVC (%)**	**FEV1 (%)**	**Cm H2O**

Is there an Indication for NIV? Yes ☐ No ☐	Does Patient wish ventilator Support? Yes ☐ No ☐ Undecided ☐	Date of Referral to Respiratory Team
Hospital	Consultant	Date of Ventilation

Palliative Care

Discussion on Advanced Care Planning		Date
Discussion on Palliative Care Options		Date
Power of Attorney Yes ☐ No ☐	Details	
Does Patient wish Specialist Palliative Care Input Yes ☐ No ☐		Palliative Care Contact
DNACPR Yes ☐ No ☐	DS1500	Anticipatory drugs/Just in Case Kit

Information & Discussion

	Date		Date
MND Scotland's Role		Patient Info Folder	
Carer Information		Counselling	
Family Info Evenings		Support Groups	
Comp Therapies		Small Grant	
Mailing List		Blue Badge	
Benefits		DVLA Informed	
Car Insurance Informed		GP Pack	
Comments			

Print extra pages for subsequent visits if required

Addressograph ***or***
Name
DoB
Unit number
CHI

Care and Management Plan (Follow-up details/action plan):		
Signature:	Printed:	
Designation:	Date:	Time:

Care and Management Plan (Follow-up details/action plan):		
Signature:	Printed:	
Designation:	Date:	Time:

Care and Management Plan (Follow-up details/action plan):		
Signature:	Printed:	
Designation:	Date:	Time:

Addressograph ***or***
Name
DoB
Unit number
CHI

The Amyotrophic Lateral Sclerosis Functional Rating Scale – Revised (ALSFRS - R)
Cedarbaum et al J Neurol Sci. 1999 Oct 31;169(1 - 2):13-21

I Comparisons are made with the patient's status prior to the onset of the disease, not with status of the last visit.

II Patient's response (on a 5-point scale) is recorded in relation to the question: "How are you doing at (…)?" for each of 12 functions listed in the ALSFRS-R.

			Scoring			
		Date				
		Initials				
		Time				
a).Speech	Normal speech processes	4				
	Detectable speech disturbances	3				
	Intelligible with repeating	2				
	Speech combined with non-vocal communication	1				
	Loss of useful speech	0				
b).Salivation	Normal	4				
	Slight but def excess of saliva in mouth;may drool	3				
	Moderately/excessive saliva, min drooling	2				
	Marked excess with some drooling	1				
	Marked drooling	0				
c).Swallowing	Normal eating habits	4				
	Early eating problems, occasional choking	3				
	Dietary consistency changes	2				
	Needs supplemental tube feeding	1				
	NPO (exclusively parenteral or enteral feeding)	0				
d).Handwriting	Normal	4				
	Slow or sloppy; all words are legible	3				
	Not all words are legible	2				
	Able to grip pen but unable to write	1				
	Unable to grip pen	0				
e). Cutting Food and Handling Utensils (patients without gastrostomy)	Normal	4				
	Somewhat slow and clumsy, but no help needed	3				
	Can cut most foods, although clumsy and slow; some help needed	2				
	Food must be cut by someone, but can still feed slowly	1				
	Needs to be fed	0				
e). Cutting Food and Handling Utensils (patients with gastrostomy)	Normal	4				
	Clumsy but able to perform all manipulations independently	3				
	Some help needed with closures and fasteners	2				
	Provides minimal assistance to caregiver	1				
	Unable to perform any aspect of task	0				
f).Dressing and Hygiene	Normal function	4				
	Independent and complete self-care with effort of decreased efficiency	3				
	Intermittent assistance or substitute methods	2				
	Needs attendant for self-care	1				
	Total dependence	0				

Addressograph ***or***
Name
DoB
Unit number
CHI

g).Turning in Bed and Adjusting Bed Clothes	Normal	4				
	Somewhat slow and clumsy, but no help needed	3				
	Can turn alone or adjust bed sheets, but with great difficulty	2				
	Can initiate, but not turn or adjust sheets alone	1				
	Helpless	0				
h). Walking	Normal	4				
	Early ambulation difficulties	3				
	Walks with assistance	2				
	Non-ambulatory functional movement	1				
	No purposeful left movement	0				
i). Climbing stairs	Normal	4				
	Slow	3				
	Mild unsteadiness or fatigue	2				
	Needs assistance	1				
	Cannot do	0				
j). Dyspnoea	None	4				
	Occurs when walking	3				
	Occurs with one or more of the following: eating, bathing, dressing (ADL)	2				
	Occurs at rest, difficulty when either sitting or lying	1				
	Significant difficulty, considering using mechanical respiratory support	0				
k). Orthopnoea	None	4				
	Some difficulty sleeping at night due to shortness of breath, does not routinely use more than 2 pillows	3				
	Needs extra pillows in order to sleep (more than two)	2				
	Can only sleep sitting up	1				
	Unable to sleep	0				
l). Respiratory insufficiency	None	4				
	Intermittent use of BiPAP	3				
	Continuous use of BiPAP during the night	2				
	Continuous use of BiPAP during the night and day	1				
	Invasive mechanical ventilation by intubation or tracheostomy	0				
		Total Score				

Addressograph ***or***
Name
DoB
Unit number
CHI

Epworth Sleepiness Scale (ESS) Score

In contrast to just feeling tired, how likely are you to fall asleep in the following circumstances.
This relates to your usual way of life at the moment. Even if you have not done any of these recently, try to work out how they would have affected you.
By using the following scale choose the most appropriate number for each situation.

0 - would ***never*** doze
1 - ***slight*** chance of dozing
2 - ***moderate*** chance of dozing
3 - ***high*** chance of dozing

SITUATION	CHANCES OF DOZING							
Date								
Sitting and reading								
Watching television								
Sitting inactive in a public place (ie Drs, meeting, cinema)								
As a passenger in a car for an hour without a break								
Lying down in the afternoon – circumstances permitting								
Sitting and talking to someone								
Sitting quietly after a lunch without alcohol								
In a car; while stopped for a few minutes in the traffic								
***TOTALS**								

Addressograph **or**
Name
DoB
Unit number
CHI

EDINBURGH COGNITIVE AND BEHAVIOURAL ALS SCREEN – ECAS SUMMARY

COGNITIVE SCORES

COMPONENT	PATIENT SCORE	CUT-OFF
Language	/28	26
Verbal Fluency	/24	14
Executive	/48	33
ALS SPECIFIC SCORE	/100	77
Memory	*/24*	*13*
Visuospatial	*/12*	*10*
ALS NON-SPECIFIC SCORE	*/36*	*24*
ECAS TOTAL SCORE	/136	105

SCORES EQUAL TO OR BELOW CUT-OFF = IMPAIRMENT

Addressogpraph **or**

Name:
DoB:
Unit Number:
CHI:

	Tick if present
Behavioural disinhibition	
Apathy or inertia	
Loss of sympathy or empathy	
Perseverative, stereotyped, compulsive or ritualistic	
Hyperorality or altered food	
TOTAL BEHAVIOUR	/5
PSYCHOSIS	/3

Summary of ECAS:

Information given to patient ☐ and carer* ☐ allied health professionals ☐

Care and Management Plan (F/up details/action plan):

Signature: Printed:

Addressograph ***or***
Name
DoB
Unit number
CHI

Death

Notified by:			Date of Death:			
Place of Death:	Home ☐	Hospital ☐	Care Home ☐	Hospice ☐	Other ☐	
Cause of Death:						
Date of Funeral:			Consent to notify others:	Yes ☐	No ☐	
Post Mortem:						

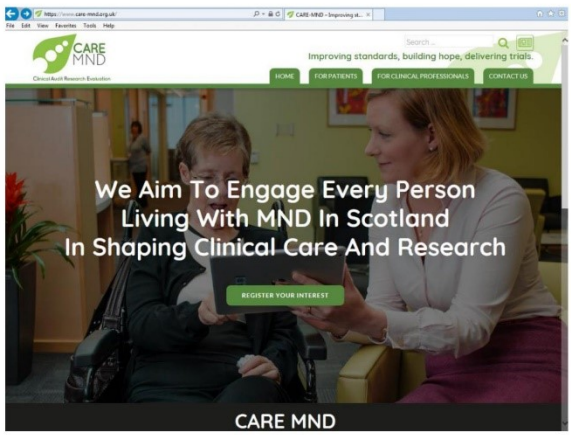
Appendix 2: CARE-MND Platform Standard Operating Procedures (SOP)/User Guide Version 2 28/06/2017

Clinical Audit Research Evaluation

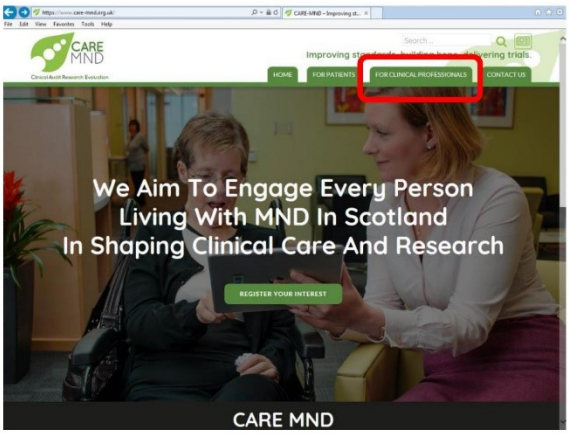
CARE-MND Platform User Guide

Step 1: Logging on to CARE-MND

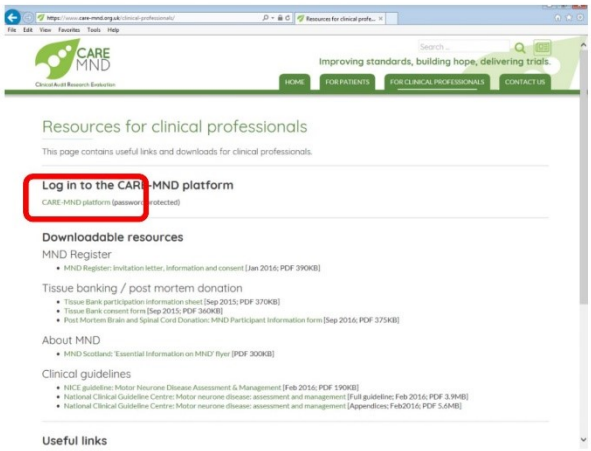
1. **Go to CARE-MND website: www.care-mnd.org.uk**



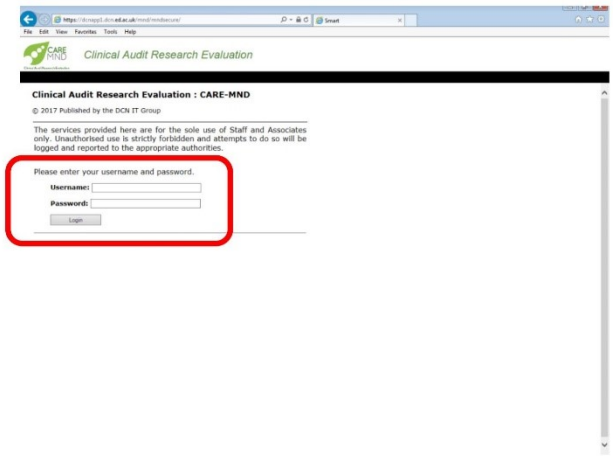
2. **Click on “For Clinical Professionals” tab**



3. **Click on the link to the “CARE-MND platform”**



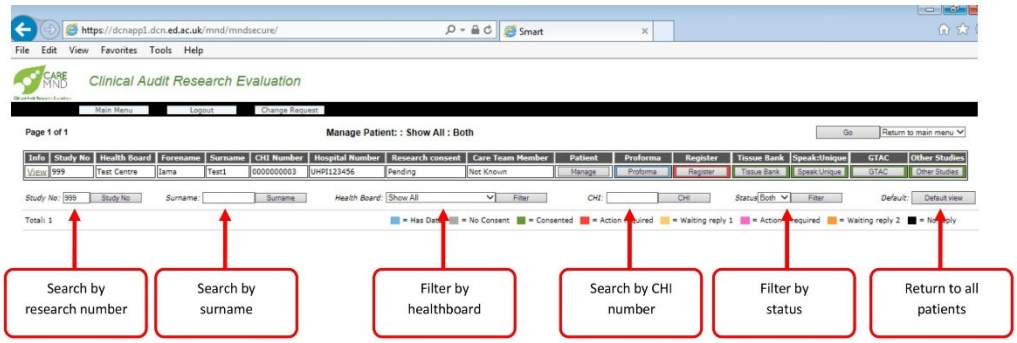
4. **Log-in using your secure personal details**



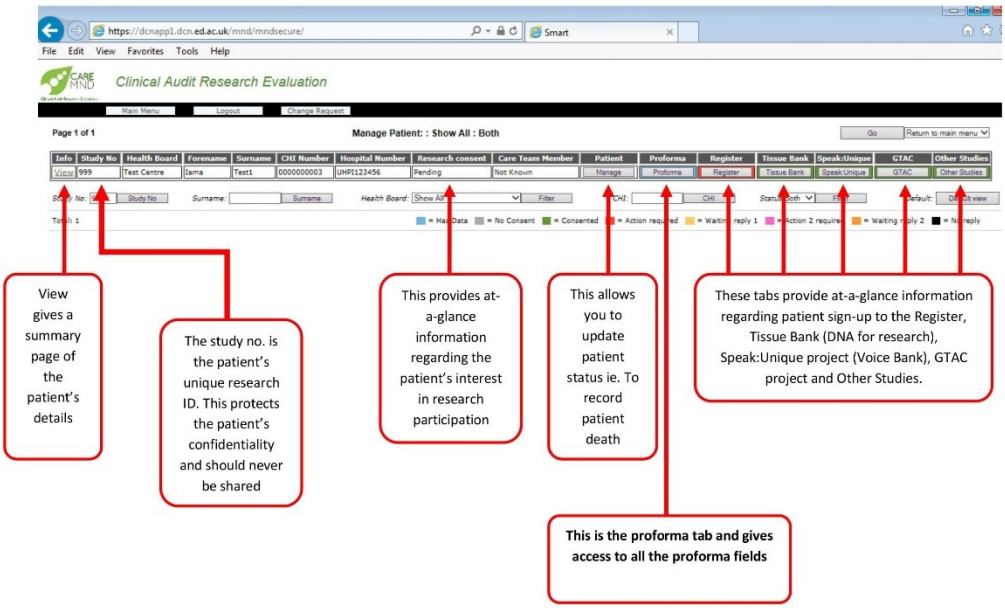
Step 2: Visualising your patients

The website will display all the patients under your care (including those that are deceased). The following steps show you how to navigate your patient list using a test patient ("Iama Test1").

1. **The following picture shows how you can search your patient list**

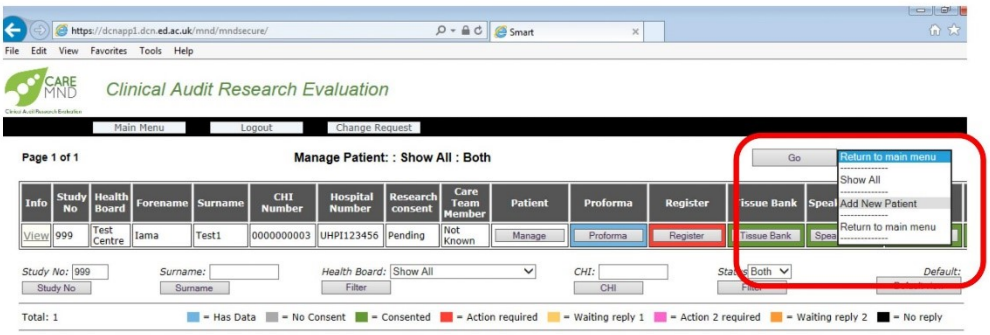


2. **The following picture shows what each patient item means**

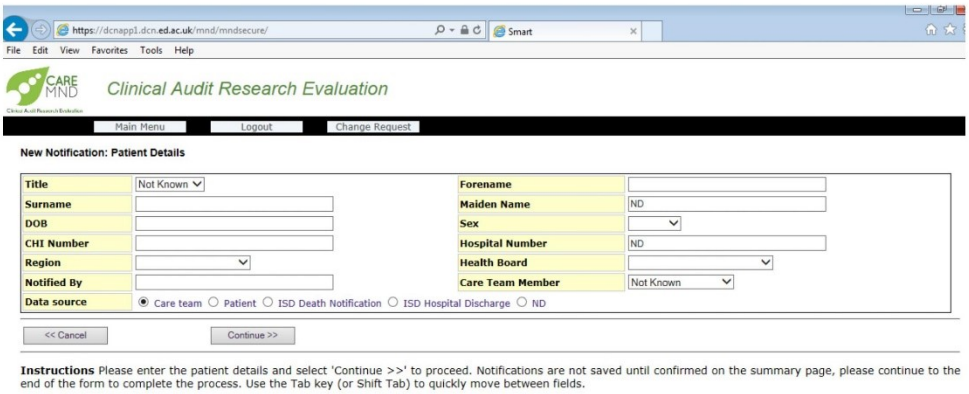


Step 3: Entering a new patient

1. **To enter a new patient, go to the drop-down tab at the top right of the screen and select "Add New Patient"**



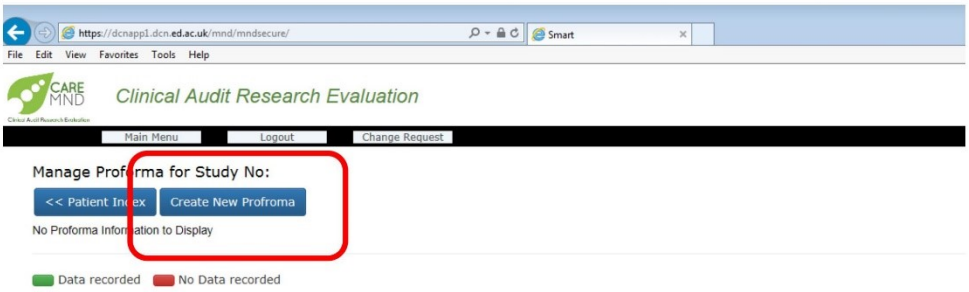
2. **You will be taken to an online form – please make sure to fill in all fields where possible**



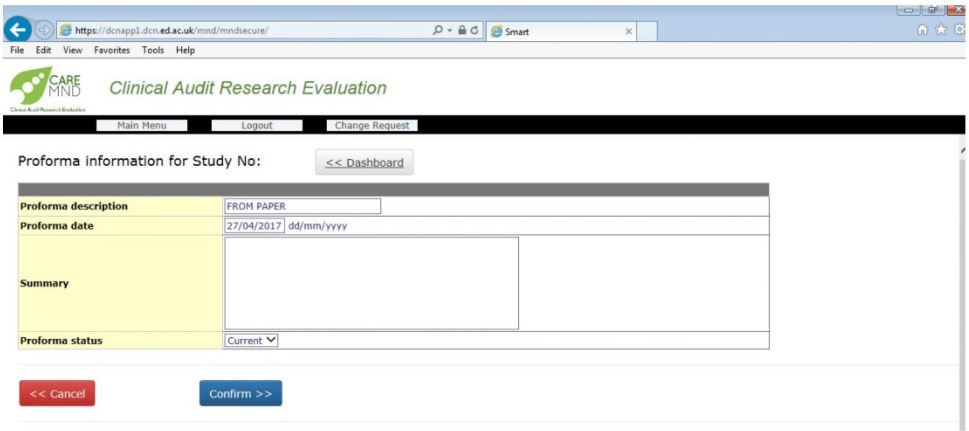
3. **Once you have completed this form, your patient will be on the database. They will be assigned a new Study Number. You can search for your patient using the Study Number, Surname or CHI number.**

Step 4: Starting a new proforma

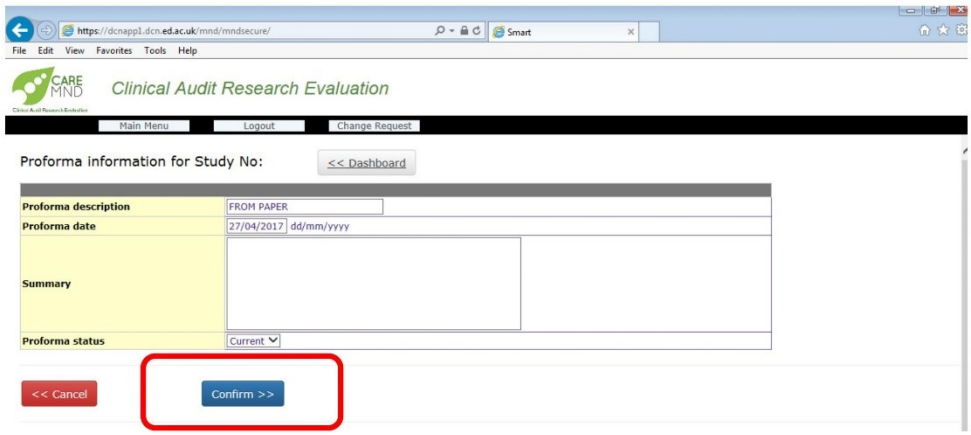
1. **To start a proforma for a new patient, click on the "Proforma" button for that patient. If the button is white, a proforma has not been started. If the button is blue, a proforma has already been started. Then click on the "Create New Proforma" button.**



2. **The first page of the proforma entry is shown below. You should indicate here where the data is coming from (ie. Copied from a paper proforma or new data entry) and the date of entry. In the "Summary" box you should include the dates of when the proforma is updated.**

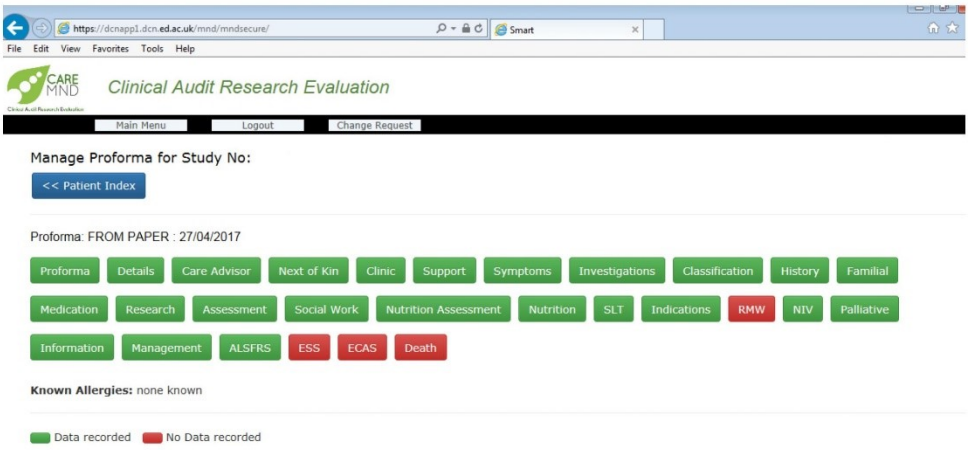


3. **You should then complete the proforma as you would the paper version. <u>Please remember that you need to click the blue "Confirm" box to save any changes that you make.</u>**



Step 5: Updating an existing proforma

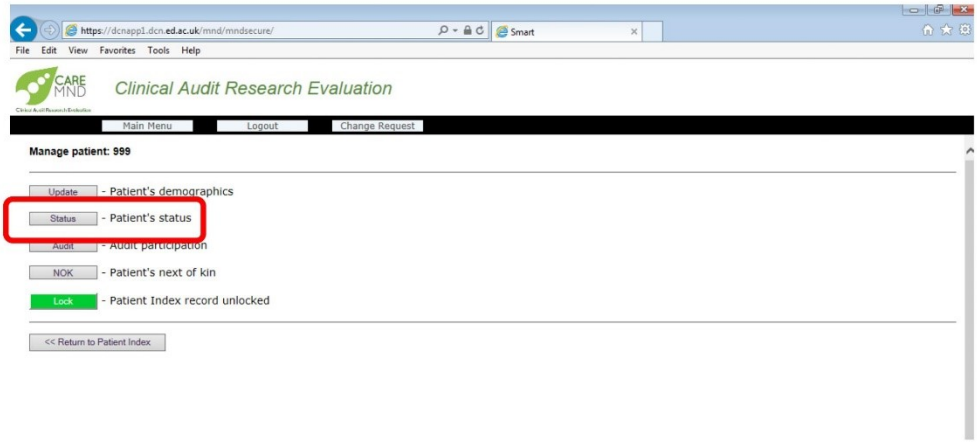
1. **If a proforma is in progress then the "Proforma" button will be blue. When you click on this, it will take you to the proforma "Dashboard". The "Dashboard" shows you the different pages of the proforma. If the relevant page is green that means that there is some data there – however, it may not be complete yet. If the page is red, there is no data. Click on any button (green or red) to update.**



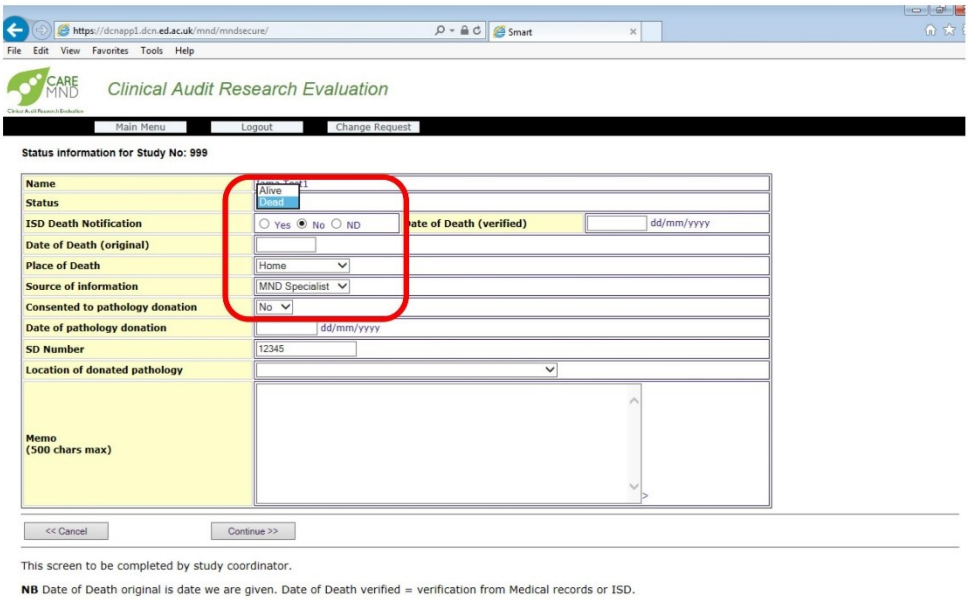
2. **Once a patient has died, it is important to check that their information is up to date. Please review each proforma page to make sure all relevant data is complete.**

Step 6: Updating patient status – recording a patient's death

1. **If one of your patients passes away, it is important to record this on the database so they are removed from your active case-load and so they don't receive any further research information. To do this click on the "Manage" button next to the patient's name. Then click on the "Status" button.**



2. **You can then update the patient's "Status" to "Dead" and add relevant details. There is also space on this page to record "ISD" information (National Records data) and "Pathology Donation" (Post-Mortem) but this will be updated by the relevant research teams.**



For more information or trouble-shooting, please contact Judith Newton, National Nursing Lead for MND, or the CARE-MND team at info@smart-mnd.org.

Data entry assistance is provided by Giuilia Melchiorre and Harry Gordon, Anne Rowling Regenerative Neurology Clinic, University of Edinburgh.

The CARE-MND platform a secure system and is managed by David Buchanan, IT Programmer, University of Edinburgh.

Appendix 3: Register Invitation Letter Version 3 11/01/2016

Version 3
11/01/2016

MND Register
Euan MacDonald Centre for
MND Research
University of Edinburgh
Chancellor's Building
49 Little France Crescent
Edinburgh
EH16 4SB

Tel: 0131-242-7985
e-mail: info@smart-mnd.org

We would like to invite you to participate in the Scottish MND Register.
The purpose is twofold:

1. To record basic information about diagnosis and care journey of the individuals with MND. This is necessary to improve the quality and timeliness of care provision.
2. To offer people living with MND the opportunity to participate in research, including clinical trials to test potential new drugs.

Research and audit findings from the register have already resulted in some changes to practice and have helped local care teams deliver higher standards of care (eg. more timely interventions to improve quality of life, an increase in the number of specialist care team members) Ongoing audit through the register means we can ensure continuous measurement and improvement of care provision for people living with MND.

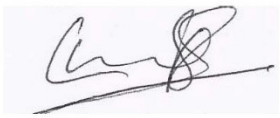
It is important to emphasise that volunteering to be on the Register does not in any way influence your routine care. Similarly you can be on the register and not participate in any research studies. You do not need to give any reasons for your decisions. There is no obligation whatsoever to participate in research. If you do agree to receive research information, it is important to note that some research projects can be undertaken at your home, without the need to travel.

Please read the information sheet carefully and do contact us if you have any questions. Once you are satisfied that your questions have been answered please complete the form with the relevant option ticked, along with the consent form. If you would like more time to consider your involvement, or to ask further questions, please take your time, and return the form in the enclosed envelope.

The Register is part of the Scottish Motor neurone disease Audit, Research and Trials (SMART) project which is based at Edinburgh University (www.smart-mnd.org).

Thank you for your consideration.

Yours faithfully,



Shuna Colville
Research Project Manager

Suvankar Pal
Consultant Neurologist

Siddharthan Chandran
Consultant Neurologist

Encl: Information sheet and freepost envelope

NHS Lothian

THE EUAN MACDONALD CENTRE FOR MOTOR NEURONE DISEASE RESEARCH

MND Scotland

1

MND Register

Full name of the person who has MND: _____

Maiden name _____

Date of Birth _____
day month year

current address of the person with MND: _____

_____

_____

Postcode _____

Telephone _____

Hospital attended: _____

Consultant Name: _____

Who is the GP of the person named above?

Name: _____

Address: _____

_____

Please tick one

I would like to be on the Register and receive information on research studies, including clinical trials that may be of interest to me: ☐
I would like to be on the Register but NOT receive information on research studies: ☐
I do not wish to be on the Register: ☐

Thank you for taking the time to complete this form. Please return this along with the consent form in the envelope provided. Please keep the information sheet provided.

For more information please contact:

Laura Stephenson
MND Scotland Research Nurse
Euan MacDonald Centre for MND Research
University of Edinburgh
Chancellor's Building
49 Little France Crescent
Edinburgh
EH16 4SB

Tel: 0131-242-7985
Email: info@smart-mnd.org .

Shuna Colville
Research Project Manager
Anne Rowling Clinic
University of Edinburgh
Chancellor's Building
49 Little France Crescent
Edinburgh

0131-465-9520
Website: www.smart-mnd.org

Version 3
11/01/2016

MND Register - Consent form

Name of Researchers: Prof Siddharthan Chandran, Dr Suvankar Pal, Shuna Colville

Please initial box

1) I confirm that I have read and understood the information sheet (V3, 11/01/2016) for the above study and have the opportunity to ask members of the Research Team questions at any time. ☐

2) I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason. Should I withdraw, my medical care or legal rights will not be affected. ☐

3) I understand that my medical notes will be examined over time, Information processed and stored securely by the responsible individuals from the research team or their regulators. **No identifiable information will be given to third parties.** ☐

4) I understand that I will not benefit financially from taking part in this study. ☐

5) I agree to take part in the above study. ☐

6) I agree that my GP can be informed that I am taking part in this study. ☐

7) I agree to receiving information about future research projects which may include clinical trials. (This does not commit you to participating in the future research it is purely informational). ☐

OR Complete only if you do not wish research information

8) I agree to my name being placed on the Register but do not wish to be contacted regarding research studies. ☐

Name of Patient: _____ Date:

Signature: _____

Where the patient is physically unable to sign a proxy can complete the form and sign on their behalf as long as the proxy is satisfied that the patient has understood the information sheet and the consent form.

Name of witness: _____ Date:

Signature of witness: _____

Name of Care Team member: _____ Date:
(if present)
Signature _____

NHS Lothian THE EUAN MACDONALD CENTRE FOR MOTOR NEURONE DISEASE RESEARCH MND Scotland 3

MND Register - Participant information sheet

We would like to invite you to join MND Register. There are three options to this register:

1) You are willing to have your name placed on the Register and we (The MND research Team) can tell you about future studies, including clinical trials, that you may be interested in. receive information on research projects as they become available
2) You are willing to have your name placed on Register but do NOT wish information on research studies
3) You do not wish to have your name on Register and do not wish to receive information on research studies

Please ask if there is anything that is not clear or if there is an area that you would like more information.

What is the purpose of the Register?
The purpose of the Register is to improve quality and equity of care across Scotland and to enable research and clinical trials. The register was established in response to many approaches from patients and health care professionals enquiring about standards of care and crucially a strong desire to participate in research studies and clinical trials. A national register is the standard and best method of ensuring these clinical and research needs are met. Scotland is also ideally placed to be a leader in this area due to the infrastructural strengths of NHS Scotland and the network of dedicated MND care advisors and nurses.

Why have I been approached?
The clinical care teams approach all people who have been diagnosed with motor neurone disease. If you require any further information we would be happy to answer your questions, our contact details are at the end of the document.

Do I have to take part?
No, and if you choose not to be included in the register this will not have any influence on your clinical management. You are also free to withdraw from the register at any time without providing a reason.

What do I have to do?
You need to read this information sheet carefully and ask any questions that you feel you need to (our contact details are below). When you are satisfied that your questions have been answered you can complete the consent form and information form and return them in the envelope provided.

Are there any risks or disadvantages of taking part?
There are no risks or disadvantages to being included in the register.

What are the possible benefits of taking part?
You will receive information about research projects as they arise. You can then decide whether or not to participate in any of them. There are no financial benefits to being on the Register.

Will my taking part in the Register be kept confidential?
Your medical records will be viewed by the research team to classify diagnosis and to research quality of care. All data / records are anonymised by a unique study number. Personal or identifiable data will never be disclosed to anyone outside the core research team based at the University of Edinburgh. All details will be kept securely on an NHS approved system within Edinburgh University and situated at the Western General Hospital, Edinburgh. If you agree, your GP would be informed of your consent to being on the Register. If you agree, we can tell you about future research studies including clinical trials that you may be interested in taking part in.

What if there are any clinically relevant findings form the research project?
If any clinically relevant findings become apparent as a result of this research (for example something that can be treated and has so far gone unnoticed, or there are questions about your diagnosis) this will be fed back to your clinical care team for them to action if relevant.

Do other researchers access my identifiable information?
Some studies may include completion of anonymised questionnaires ie no identifiable data is involved. Other studies may involve face to face interviews or tests which do require the external researcher to know who you are. Even here it is important to highlight that no identifiable information will ever be in the public domain or provided to third parties. Nonetheless to ensure that you are in control at all times of whether you participate in research we have put in place the following process.

External researchers will **not** approach you directly.
- All requests to use the Register to identify potential research participants are first evaluated by the Register Management Committee comprising Professor Siddharthan Chandran, Dr Suvankar Pal and Shuna Colville to ensure that the research is appropriate and ethically approved.
- If deemed suitable study information (project invitation letters and information sheets) will then be sent by the MND Research team to potential participants who have already indicated that they wish to receive information about future studies.
Confidential details are not released by the MND research team to other researchers without your permission.

What happens to the results of the Register?
The results of individual research projects will be presented at academic meetings and published in academic medical journals. This will always (see above) be anonymised. In other words your personal details will never be identifiable. A summary of the results, when available, will be published in the quarterly MND Scotland publication, Aware. If you would like to receive a summary of the results please let us know by contacting us (address supplied below).

Who is organising the research?
The Register is organised by a team of researchers based in Scotland and is part of the Scottish Motor neurone disease Audit, Research and Trials (SMART) project. The research currently has 3 years funding from MND Scotland and does not involve medical or pharmaceutical companies. The researchers are not paid to recruit patients into the Register.

Who has reviewed the study?
The study has been reviewed by Scotland A Research Ethics Committee and externally peer reviewed by MND Scotland's research panel.

What if there is a problem?
All core research staff are covered by NHS contracts and the NHS public liability insurance. We do not anticipate any problems with this study but if you do have a complaint, please report this using the standard NHS Lothian Complaints Procedure (0131 558 3681).

Is there anyone I can speak to about this study who can offer me independent advice?
If you have any questions needing independent advice, please contact: Dr Peter Connick (Consultant Neurologist), 0131-465-9500, email: pconnick@exseed.ed.ac.uk

Contact for further Information:

Shuna Colville
Research Project Manager
Anne Rowling Clinic
Chancellor's Building
49 Little France Crescent
Edinburgh
EH16 4SB

Tel: 0131-465-9520
Email: info@smart-mnd.org

Laura Stephenson
MND Scotland Research Nurse
Euan MacDonald Centre for MND Research
Chancellor's Building
49 Little France Crescent
Edinburgh
EH16 4SB

0131-242-7985
website: www.smart-mnd.org

NHS Lothian THE EUAN MACDONALD CENTRE FOR MOTOR NEURONE DISEASE RESEARCH MND Scotland

Appendix 4: R Code for Statistical Analyses (truncated)

```
#install packages
install.packages("readr")
install.packages("caret")
install.packages("doParallel")
install.packages("psych")
install.packages("RANN")
install.packages("pROC")
install.packages("clinfun")
install.packages("combinat")
install.packages("gtools")
install.packages("DescTools")
install.packages("plyr")
install.packages("car")
install.packages("glmnet")
install.packages("ggplot2")
install.packages("ModelMetrics")
install.packages("Metrics")
install.packages("DMwR")
install.packages("survival")
install.packages("VIM")
install.packages("mice")
install.packages("corrplot")
install.packages("mitools")


preproc_pheno = read_csv("PhenoGeno1517_Simplified3.csv")


#Only include columns of interest (including outcome)
pheno = preproc_pheno[,c(2:33)]


#Visualise and count missing data
aggr_plot = aggr(pheno, numbers=TRUE, sortVars = T, labels = names(data),
cex.axis = .7, gap=3, ylab=c("Histogram of missing data", "Pattern"))


#dummyCode (for categorical variables)
phenoPreproc = pheno[,!(colnames(pheno) %in% c("Censor_or_Death",
"Survival_Onset_Days", "Less2year_survival_onset"))]
```

```
dummies = dummyVars(~.,data=phenoPreproc, fullRank=T)
phenoPreproc = data.frame(predict(dummies,newdata = phenoPreproc))

#Remove variables with 0 variance
nzv_cols = nearZeroVar(phenoPreproc)
if(length(nzv_cols) >0) phenoPreproc = phenoPreproc[,-nzv_cols]

#Standardise
preprocvalues = preProcess(phenoPreproc,method=c("center", "scale"))
phenoPreproc = predict(preprocvalues, phenoPreproc)

# Identify correlated variables
corrMatrix = corr.test(phenoPreproc)
corrMatrix005 = corrMatrix
corrMatrix005$r[corrMatrix005$p>0.05] = 0
png(file="corr.png", res=300, width=4500, height=4500)
corrplot(corrMatrix005$r, method="color", type="lower", addCoef.col =
"white", number.cex = 0.5)
dev.off()

# Remove correlated variables

#Cox regression
#Remove columns with <=25% missing data
pheno25 = phenoPreproc[, colMeans(is.na(phenoPreproc)) <= .25]

#Add outcome back in
pheno25cox = pheno25
pheno25cox$Survival = pheno$Survival_Onset_Days
pheno25cox$Censor = pheno$Censor_or_Death

#Create "temporary" imputed dataset. m=10 = 10 sets of imputed data
temppheno25cox = mice(pheno25cox, m=10, maxit=5, method = 'pmm', seed=987)
View(pheno25cox)

#Cox model
cox25model = with(temppheno25cox,
coxph(Surv(Survival,Censor)~Sex.M+Diagnostic_Delay+Age_Onset+Site_Onset.Lim
b+Classification.MND.FTD+Classification.Other+PMH_Malignancy.Yes+PMH_AI.Yes
```

```
+PMH_CV.Yes+PMH_Psych.Yes+Smoking_EverOrNever.Yes+Exercise_participation+FH
_MND.Yes+FH_Dementia.Yes+FH_Other_neurological_conditions.Yes+FH_Psychiatri
c_conditions.Yes+Riluzole.Yes+Feeding_tube.Yes+NIV.Yes+Pre_Slope))

summary(pool(cox25model))

r2_cox25 = NULL

for(i in seq(1:10))

{

  r2_cox25[i] = summary(cox25model$analyses[[i]])$rsq[[1]]

}

MeanCI(r2_cox25)


#Code for box/whiskers

boxLabels_ep = c("Sex (Male)","Time to Diagnosis", "Age of Onset","Site of
Onset:Limb", "Classification: MND-FTD", "Classification: Other", "PMH
Malignancy", "PMH Autoimmune Disease", "PMH Cardiovascular Disease", "PMH
Psychiatric Disease", "Ever Smoked", "Exercise Participation", "Family
History of MND", "Family History of Dementia", "Family History of Other
Neurological Disorders", "Family History of Psychiatric Disorders",
"Riluzole", "Feeding Tube Insertion", "Non-Invasive Ventilation", "ALSFRS-R
Preslope")


#manually extract each column exp(coef) which is the OR (not the intercept)

boxOdds_ep = NULL

boxCILow_ep = NULL

boxCIHigh_ep = NULL


for(i in seq(1:20))

{

  boxOdds_ep = c(boxOdds_ep, exp(summary(pool(cox25model))[,1][[(i)]]))

  boxCILow_ep = c(boxCILow_ep, exp((summary(pool(cox25model))[,1][[(i)]])-
(1.96*(summary(pool(cox25model))[,2][[(i)]]))))

  boxCIHigh_ep= c(boxCIHigh_ep,
exp((summary(pool(cox25model))[,1][[(i)]])+(1.96*(summary(pool(cox25model))
[,2][[(i)]]))))

}


df_ep <- data.frame(boxLabels_ep, boxOdds_ep, boxCILow_ep, boxCIHigh_ep)


png("hr_coefs_cox_25_2.png",res = 300, width = 30, height = 45, units =
'cm')

(p <- ggplot(df_ep, aes(x = boxOdds_ep, y = df_ep$boxLabels_ep)) +

    geom_vline(aes(xintercept = 1), size = .25, linetype = 'dashed') +

    geom_errorbarh(aes(xmax = boxCIHigh_ep, xmin = boxCILow_ep), size = .7,
height =
```

```
                    .2, color = 'gray30') +
    geom_point(size = 3.5, color = '#00BFC4') +
    theme_bw() +
    theme(axis.text.y = element_text(face = "bold", color = "gray30", size
= 12)) +
    theme(panel.grid.minor = element_blank()) +
    scale_x_continuous(breaks = c(0.05,0.1,0.2,0.4,0.7,1,1.5,3,6,12,24) ) +
    coord_trans(x = 'log') +
    ylab('') +
    xlab('Hazard Ratio 95% CI (log scale)') +
    labs(title = "i) Prognostic Variables in MND Cox Proportional Hazards
Model (<=25% Missing Data)")
  + annotate("text", x=0.05, y=-0.5, label = "R2 = 0.51")
  + theme(plot.title = element_text(hjust = 0.5))
  + scale_y_discrete(limits = rev(df_ep$boxLabels_ep))
)
dev.off()


# Repeat above for <=50% missing data and 0 missing data


#Adding in Genetic data


preproc_geno = read_csv("PhenoGeno_GenotypeOnly_Simplified2.csv")


#Univariate Analyses by Mutation Type (not shown)
#Logistic regression model with imputation for genotyping
#PathVUSP vs no PathVUSP
#select variables using regularisation


#Only include columns of interest (including outcome)
genoPath = preproc_geno[,c(4, 54, 55, 57, 59:83)]
#Visualise and count missing data
#dummyCode (for categorical variables)
#Remove variables with 0 variance
#Add outcome back in
#Control (within 1 SE)
control <- trainControl(method="cv", number=10, classProbs=TRUE,
summaryFunction=twoClassSummary, selectionFunction ="oneSE")
```

```
#Path_VUSP Outcome and Model

results_Path = list()

mods_Path = list()


#15 sites

for(i in seq(1:15))

{

  set.seed(987)


  #set up leave one site out cv

  test_set = genoPathglm[ which(genoPathglm$Health_Board ==
levels(genoPathglm$Health_Board)[i]), ]

  train_set = genoPathglm[ -which(genoPathglm$Health_Board ==
levels(genoPathglm$Health_Board)[i]), ]


  #remove site column

  test_set = test_set[ ,!(colnames(test_set) %in% c("Health_Board"))]

  train_set = train_set[ ,!(colnames(train_set) %in% c("Health_Board"))]


  #Get the observed outcome classes for this test set

  result = data.frame(obs=test_set$Path_VUSP)


  #train model over grid of 10 by 10 lambda alpha, knn impute, standardise

  mod <- train(Path_VUSP ~ ., data=train_set, method="glmnet",
metric="ROC", tuneLength = 10, preProc = c("center", "scale","knnImpute"),


trControl=control,na.action = na.pass)

  result$pred = predict(mod, test_set, type = "prob", na.action = na.pass)

  result$row = as.numeric(rownames(test_set))


  results_Path[[i]] = result

  mods_Path[[i]] = mod

}


#Collecting results

results_Path_seq = NULL

results_Path_seq$pred =
c(results_Path[[1]]$pred$Y,results_Path[[2]]$pred$Y,
results_Path[[3]]$pred$Y,
```

```
results_Path[[4]]$pred$Y,results_Path[[5]]$pred$Y,
results_Path[[6]]$pred$Y,


results_Path[[7]]$pred$Y,results_Path[[8]]$pred$Y,results_Path[[9]]$pred$Y,


results_Path[[10]]$pred$Y,results_Path[[11]]$pred$Y,results_Path[[12]]$pred
$Y,


results_Path[[13]]$pred$Y,results_Path[[14]]$pred$Y,results_Path[[15]]$pred
$Y)

results_Path_seq$obs =
factor(c(as.character(results_Path[[1]]$obs),as.character(results_Path[[2]]
$obs),


as.character(results_Path[[3]]$obs),as.character(results_Path[[4]]$obs),


as.character(results_Path[[5]]$obs),as.character(results_Path[[6]]$obs),


as.character(results_Path[[7]]$obs),as.character(results_Path[[8]]$obs),


as.character(results_Path[[9]]$obs),as.character(results_Path[[10]]$obs),


as.character(results_Path[[11]]$obs),as.character(results_Path[[12]]$obs),


as.character(results_Path[[13]]$obs),as.character(results_Path[[14]]$obs),a
s.character(results_Path[[15]]$obs)))

results_Path_seq$row = c(results_Path[[1]]$row,results_Path[[2]]$row,
results_Path[[3]]$row,

                         results_Path[[4]]$row,results_Path[[5]]$row,
results_Path[[6]]$row,


results_Path[[7]]$row,results_Path[[8]]$row,results_Path[[9]]$row,


results_Path[[10]]$row,results_Path[[11]]$row,results_Path[[12]]$row,


results_Path[[13]]$row,results_Path[[14]]$row,results_Path[[15]]$row)



auc_Path_seq = roc(predictor = results_Path_seq$pred, response =
results_Path_seq$obs)$auc

roc_Path_seq = roc(predictor = results_Path_seq$pred, response =
results_Path_seq$obs)



plot(roc_Path_seq)

#PSI = (PPV+NPV)-1

#LR+ = sens/(1-spec)

#LR- = (1-sens)/spec

set.seed(987)
```

```
ci.coords(roc_Path_seq,"best", best.method = "closest.topleft", ret =
c("specificity", "sensitivity", "accuracy", "tn", "tp", "fn", "fp", "npv",
"ppv"))



set.seed(987)

Path_auc_null = NULL

for(i in seq (1:10001))

{

  Path_perm = permute(results_Path_seq$obs)

  Path_auc_null = c(Path_auc_null, roc(predictor = results_Path_seq$pred,
response = Path_perm)$auc)

}



#get p value by taking proportion of permutated values greater or equal to
the actual value

(1+sum(Path_auc_null >= auc_Path_seq))/10001




#https://github.com/nogueirs/JMLR2018

getStability <- function(X,alpha=0.05) {

  ## the input X is a binary matrix of size M*d where:

  ## M is the number of bootstrap replicates

  ## d is the total number of features

  ## alpha is the level of significance (e.g. if alpha=0.05, we will get
95% confidence intervals)

  ## it's an optional argument and is set to 5% by default

  ### first we compute the stability



  M<-nrow(X)

  d<-ncol(X)

  hatPF<-colMeans(X)

  kbar<-sum(hatPF)

  v_rand=(kbar/d)*(1-kbar/d)

  stability<-1-(M/(M-1))*mean(hatPF*(1-hatPF))/v_rand ## this is the
stability estimate



  ## then we compute the variance of the estimate

  ki<-rowSums(X)

  phi_i<-rep(0,M)

  for(i in 1:M){
```

```
    phi_i[i]<-(1/v_rand)*((1/d)*sum(X[i,]*hatPF)-(ki[i]*kbar)/d^2-
(stability/2)*((2*kbar*ki[i])/d^2-ki[i]/d-kbar/d+1))

  }

  phi_bar=mean(phi_i)

  var_stab=(4/M^2)*sum((phi_i-phi_bar)^2) ## this is the variance of the
stability estimate



  ## then we calculate lower and upper limits of the confidence intervals

  z<-qnorm(1-alpha/2) # this is the standard normal cumulative inverse at a
level 1-alpha/2

  upper<-stability+z*sqrt(var_stab) ## the upper bound of the (1-alpha)
confidence interval

  lower<-stability-z*sqrt(var_stab) ## the lower bound of the (1-alpha)
confidence interval



return(list("stability"=stability,"variance"=var_stab,"lower"=lower,"upper"
=upper))



}



coefs_Path = NULL

for (i in seq(1:15))

{

  coefs_Path = c(coefs_Path, coef(mods_Path[[i]]$finalModel,
mods_Path[[i]]$bestTune$lambda))

}



#just get numbers

coefs_Path_extract = NULL

for(i in seq(1:15))

{

  coefs_Path_extract = rbind(coefs_Path_extract, coefs_Path[[i]][1:28])

}



#get matrix of coefficients presence (1) or absence (0)

#Presence or absence of predictors across all 15 LOSOCV models

coefs_Path_presence = NULL

coefs_Path_presence = coefs_Path_extract[1:15,2:28]

coefs_Path_presence[coefs_Path_presence != 0] <- 1
```

```
#stability of feature selection http://jmlr.org/papers/volume18/17-514/17-514.pdf
getStability(coefs_Path_presence)


#work out number of predictors shared across all models
table(colMeans(coefs_Path_presence))


#get rank of coef by importance as in sports ranking
coefs_Path_baseline_rank = NULL


for(i in seq(c(1:15)))
{
  #rank absolute value excluding the intercept for each model
  coefs_Path_baseline_rank = rbind(coefs_Path_baseline_rank, rank(abs(coefs_Path_extract[i,2:28]), ties.method = "min"))
}


# rank the mean ranks of each column across all models
coefs_Path_baseline_rank_mean = colMeans(coefs_Path_baseline_rank)


#Invert order of rank to identify top models
coefs_Path_baseline_order = rank(-coefs_Path_baseline_rank_mean)


#Get the column names (not the intercept)
coef_Path_names = dimnames(coefs_Path[[1]])[[1]][2:28]


coefs_Path_means = colMeans(coefs_Path_extract)[2:28]


df_Path_names = data.frame(coef_Path_names, coefs_Path_baseline_order, coefs_Path_means, colMeans(coefs_Path_presence))


#View the best predictor variables by absolute value in order (lower is better)
View(df_Path_names)


##Run the LR model with top predictors


#Only include columns of interest (including outcome)
genoPath = preproc_geno[,c(4, 54, 57, 62, 73, 76, 79, 83)]
```

```
#Visualise and count missing data

#dummyCode (for categorical variables)

#Remove variables with 0 variance

#Standardise

#Add outcome back in

#See which variables are include


#Create "temporary" imputed dataset. m=10 = 10 sets of imputed data

tempgenoPathglm = mice(genoPathglm, m=10, maxit=5, method = 'pmm',
seed=987)


#Glm model

glmPathmodel = with(tempgenoPathglm,
glm(Path_VUSP~Sex.M+Diagnostic_Delay+Classification.MND.with.Fronto.tempora
l.dementia..MND.FTD.+Classification.Progressive.Bulbar.Palsy..PBP.+FH_MND.Y
es+FH_Other_neurological_conditions.Yes+FH_MND.Yes+Feeding_tube_inserted.Ye
s+Pre_slope, family="binomial"))

summary(pool(glmPathmodel))


#get ORs

exp(summary(pool(glmPathmodel))[,1][[3]])


#Excluding individuals with missing outcome data (n=0)

glmPathmodelnull = genoPathglm[which(!is.na(genoPathglm$Path_VUSP)),]

nullgenoPathglmmod = glm(glmPathmodelnull$Path_VUSP~1,family = "binomial")


#pseudo R2 and AIC for each iteration of mice (10) (in a sequence, 10x)

pseudor2_genoPathglm = NULL

AIC_genoPathglm = NULL

for(i in seq(1:10))

{

  pseudor2_genoPathglm[i] = (1-logLik(getfit(glmPathmodel,
i))/logLik(nullgenoPathglmmod))[1]

  AIC_genoPathglm[i] = getfit(glmPathmodel,i)$aic

}


MeanCI(pseudor2_genoPathglm)

MeanCI(AIC_genoPathglm)
```

```
#Code for box/whiskers
##Run the LR model as above with univariate predictors instead to compare
```

Appendix 5: Scottish Regenerative Neurology Tissue Bank Information Sheet and Consent Form, Version 1.1 25/09/2014

Scottish Regenerative Neurology Tissue Bank
Participant Information Sheet

Name of researchers: Professor Siddharthan Chandran, Dr Suvankar Pal, Shuna Colville

Please ask if anything is unclear or if you would like further information.

Studying genes that may contribute to neurological conditions

Participant information sheet
You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with friends, relatives and your GP if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of this study?
We research disorders that damage the development or function of nerve cells (neurons) in the brain and spinal cord. These disorders are called neurological conditions, and include motor neurone disease, multiple sclerosis, Parkinson's disease, Fragile X syndrome, dementias and stroke.

You are being asked to consider providing consent to allow some of your blood, saliva (or sample taken through clinical care(spinal fluid or blood sample)), to be stored in the form of CSF/serum/plasma and/or DNA. We hope to create a comprehensive bank of samples from patients with neurological conditions. If you give your consent, we will use the sample to study the causes and mechanisms of these conditions. The samples will be valuable for understanding why these conditions affect certain people and also for the development and evaluation of new diagnostic tests.

How will I be involved?
If you have had a sample taken as part of your clinical care you will be asked to sign a consent form to allow researchers to use any spare sample for research purposes. If a clinical sample is not available you will be asked for a blood sample (usually approximately 9 mls or two teaspoons but may vary depending on the study). Blood is made up of different components (red cells, white cells, plasma), which we will extract and store for research purposes. It may be possible to take this blood sample at the same time that you are having other routine blood tests. If taking a blood sample is not possible then an alternative is to donate some saliva, from which the DNA can also be extracted.

What will happen to the sample I donate?
The different components, including DNA, will be extracted from the sample and stored in a dedicated and secure tissue bank in the Wellcome Trust Clinical Research Facility at the Western General Hospital, Edinburgh.

Will the sample and information obtained be confidential?
Your sample will be coded so that no personal information is available to staff outside the research team. The Scottish Regenerative Neurology Tissue Bank Steering Group is required to approve any studies involving the use of these samples. Researchers who are given permission to analyse the coded samples will be unable to access your personal data. Access to the link between the code number for your sample and your name will be restricted to the core members of the research team. No identifiable information will be given to a third party.

The questions that we ask you with regards to your symptoms, previous illness and family history will be stored in a secure NHS-approved computer.

Why have I been chosen?
You have been chosen as you are being investigated for, or have, a neurological condition that we are interested in studying.

Do I have to take part?
No. Participation in this study is entirely up to you. If you decide not to participate, this will not affect your care or treatment. If you decide to take part you will be given this information sheet to keep and be asked to sign a consent

form. If you decide to participate you are free to withdraw your sample from the Tissue Bank at any time and without giving a reason. If you request, your anonymised data can be withdrawn from any future analysis and, at your request, your sample will be destroyed.

Use of samples in future projects

Current knowledge of the genetics and disease-causing factors of neurological conditions is limited, but as this understanding increases there may be new studies that we can undertake using the donated samples. For this reason we are seeking your permission to use your stored sample in future research projects that arise as our understanding increases, which could include whole-genome analysis.

Our future studies may involve collaboration with high-quality international research centres that are linked in collaborative networks to share non-identifiable data. The best research increasingly requires partnerships with other research centres, and in working together one maximises the value of the precious samples that have been donated. Ultimately we hope that the research will lead to the involvement of pharmaceutical companies in the development of new therapies. We will grant access by pharmaceutical companies to the anonymised samples if the Regenerative Neurology Tissue Bank Steering Group agrees that this is the best way forward.

Are there any benefits or disadvantages to taking part in the studies?

You are unlikely to personally benefit directly from the results of this study, but the results may be of benefit to people with a neurological condition in the future. It is anticipated that our research will lead to a better understanding of these disorders and could help in producing new and better treatments, and ultimately, in preventing neurological conditions.

The only disadvantage to taking part in the study is the need to give up the time to have an ordinary blood test, which can be a little uncomfortable and/or leave some bruising.

What will happen to the results of the studies?

The results of any research may be presented at academic meetings and published in academic medical journals online and/or in print. All study reports will use anonymised data, so you will not be identified in any presentation or publication. Unfortunately it is not possible for us to feedback individual results on anonymised research samples.

What if there is a problem?

We do not anticipate any problems with this study but if you do have a problem/complaint, please report this using the standard NHS Lothian Complaints Procedure (0131-558-3681) or discuss the problem with the Principal Investigator, Professor Siddharthan Chandran, or Dr Suvankar Pal using the address below. The NHS indemnity scheme will cover participants in the conduct and management of the research in the unlikely event of them being harmed.

Who is organising the research?

The study is part of the Regenerative Neurology Research that is led by Professor Siddharthan Chandran, MacDonald Professor of Neurology, at the Anne Rowling Regenerative Neurology Clinic, University of Edinburgh.

Who has reviewed the study?

The Chief Scientist Office for Scotland has reviewed this study.

Contact for further information?

If you would like to know more about the science behind these studies please contact Professor Mary Porteous, Consultant Clinical Geneticist, South East of Scotland Genetics Service, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, Tel: 0131-537-1116

You may wish to discuss whether to participate with your local medical team.

Thank you very much for taking the time to read this document and for considering taking part in this study. If you decide to participate you will be given a copy of this information sheet and a signed consent form to keep.

Contact for further Information:

Laura Stephenson
MND Scotland Research Nurse
Euan Macdonald Centre for MND Research
Chancellor's Building, 49 Little France Crescent, Edinburgh, EH16 4SB, **0131 242 7985**, Email: **info@smart-mnd.org**

EUAN MACDONALD CENTRE
FOR MOTOR NEURONE DISEASE RESEARCH

THE UNIVERSITY of EDINBURGH

Scottish Regenerative Neurology Tissue Bank

PARTICIPANT CONSENT FORM

Name of researchers: Prof Siddharthan Chandran, Dr Suvankar Pal, Shuna Colville

Please initial box

1.	I confirm that I have read and understood the information sheet (dated Sept 2014) for the above study and have the opportunity to ask members of the research team questions at any time.	☐
2.	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason. Should I withdraw, my medical care and legal rights will not be affected.	☐
3.	I understand that my medical notes may be examined by responsible individuals from the research team or their regulators. **No identifiable information will be given to third parties.**	☐
4.	I agree to storage, processing and transfer of the data in the way described in the information sheet.	☐
5.	I agree to DNA analysis of cell material derived from my blood (or saliva) sample.	☐
6.	I understand that the sample is for research purpose and that **no results will be made routinely available to me.**	☐
7.	I understand my sample/DNA will be stored and may be used for future research with appropriate ethical approvals.	☐
8.	I agree to storage of my genetic material for research use in future.	☐
9.	I agree to my genetic material being used for genome scanning in future studies.	☐
10.	I understand that I will not benefit financially from taking part in this study even if the research results in the development of new drug therapies.	☐
11.	I agree to take part in the above study by allowing my sample collected either specifically for research or clinically to be used for research purposes.	☐
12.	I agree that my GP or family doctor can be informed that I am taking part in this study.	☐

Name of Patient: _____ Date:

Signature: _____

Name of Researcher: _____ Date:

Signature _____

THE UNIVERSITY of EDINBURGH

Anne Rowling Regenerative Neurology Clinic

Appendix 6: Consensus Methodology for American College of Medical Genetics Classification of MND Genomes

Abbreviations

PVS = Pathogenic Very Strong

PS = Pathogenic Strong

PM = Pathogenic Moderate

PP = Pathogenic Supporting

BA = Benign Stand-alone

BS = Benign Strong

BP = Benign Supporting

VUS = Variant of Uncertain Significance

PVS1: *Null variant (nonsense, frameshift, canonical +/- 1 or 2 splice sites, initiation codon, single or multiexon deletion) where loss of function (LoF) is a known mechanism of disease* - LoF known mechanism of disease only applicable for *ALS2*, *OPTN* and *TBK1* (probability of being loss-of-function intolerant *(*pLI*)* scores extracted from gnomAD). Functional assays for splice site variants only – not needed for frameshifts or truncations.

PS1: *The amino acid change is seen in a previously established pathogenic variant (but with a different nucleotide change)* - Ensure previous established variants meets ACMG criteria for Pathogenic.

PS2: *De novo variant (both maternity and paternity confirmed) in patient with the disease and no family history* - Review of literature including supplementary materials.

PS3: *Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product* - Rodent model of a given mutation accepted as supportive if the rodent had motor or cognitive symptoms, there was impact on survival, or proven loss of motor neurons pathologically[395,398,399]. Pathologically similar findings without neurodegeneration not accepted (eg. presence of DP43 mis-localization, dipeptide repeats, *C9of72* foci). The only accepted *in vitro* assay was cytoplasmic localization/nuclear exclusion of a *FUS* variant.

PS4: *Prevalence in affected individuals is significantly higher than in controls* - Relative Risk or Odd's Ratio >5.0 as obtained from case-control studies and confidence interval does not include 1.0

PM1: *Located in a mutational hot spot and/or critical and well-established functional domain (without any benign variants)* - *SOD1*, *VAPB*, *PFN1*, *CHCHD10* and *ALS2* – no hot spot; *FUS* – NLS (exon 15); *TARDBP* – exon 6; *VCP* – exons 3,4 and 5 or N-terminal domain; *UBQLN2* - PXX domain

PM2: *Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium* – Most complete database at time of analysis (ExAC database) used only. Extremely low frequency if recessive taken to be 0.01% MAF.

PM3: *If recessive, detected in trans with a pathogenic variant (if parental or offspring samples available* - Review of literature including supplementary materials.

PM4: *Protein length changes as a result of in-frame deletions/insertions in a non-repeat region OR stop-loss variants (stop codon is changed to an amino acid)* – Amino acid sequences visualised using UCSC and Ensembl Genome Browsers.

PM5: *Novel missense change at the same site as a different pathogenic amino acid missense change (eg. Trp38Ser and Tryp38Leu)* - Assessed with knowledge of pathogenicity of other variants – if one of the variants is pathogenic, this will impact on all other variants.

PM6: *Assumed de novo but paternity and maternity not confirmed.*

PP1: *Cosegregation with disease in multiple affected family members in gene definitively known to cause the disease* - Co-segregation calculated as per methods described in parallel publication[396].

PP2: *A novel missense variant in a gene that has a low rate of benign missense variants, and in which missense variants are a common mechanism of disease* - Z-score of ≥2.00 in ExAC accepted ensuring that the gene would be the top 5% of missense variant intolerant genes. Z-scores for missense variants: *ALS2* = 0.80; *CCNF* 0.22; *CHCHD10* 0.99; *FUS* = 2.60; *OPTN* = 0.06; *PFN1* = 2.97; *SOD1* = 2.34; *TBK1* = 1.17; *TARDBP* = 4.33; *UBQLN2* = 1.56; *VAPB* = 1.13; *VCP* = 6.47.

PP3: *Multiple lines of computational evidence supports a deleterious effect on the gene or gene product (eg conservation, evolutionary or splicing impact etc) (NB PP3 can only be used once)* - Review of seven *in silico* prediction algorithms. Rules for pathogenic: PhastCons >0.9; GERP > 4.0; CADD >20; Grantham >100; SIFT <0.05; Mutation Assessor >0.65/High; Polyphen probably-damaging. All algorithms must agree in order to achieve criteria. In view of high number of algorithms used, a 'leave-one-out' approach was adopted if there was

disagreement. If the remaining algorithms agreed, the criterion was accepted. Meta-SVM and meta-LR methods were also used (see Methods).

PP4: *Phenotype or family history is highly specific for a disease with a single genetic aetiology* - Due to phenotypic and genotypic heterogeneity of ALS, this was only thought to be relevant for patients with *VCP* who have typical disease syndrome (inclusion body myositis (IBM), frontotemporal dementia (FTD) and Paget's disease) and *ALS2* for infantile ascending hereditary spastic paraplegia (IAHSP).

PP5: *Reputable source (eg. clinical laboratory) recently reports variant as pathogenic but the evidence is not available for independent evaluation* - If a variant was reported clinically in ClinVar as pathogenic this criterion was accepted (evidence of clinical genotyping, not literature only).

BA1: *Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium* - ExAC database only used.

BS1: *Allele frequency is greater than expected for disorder (in any population (eg African American or European American)* - ExAC database only used; MAF cut-off = 0.001

BS2: *Disease is fully penetrant at an early age and the variant is observed in a well-documented healthy adult individual for a recessive (homozygous), dominant (heterozygous) or X-linked (hemizygous) disorder* - Only applicable for *ALS2* (juvenile disease)

BS3: *Well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing*

BS4: *Lack of segregation in affected members of a family* - Only applicable if complete/early penetrance therefore not applicable for these genes

BP1: *Missense variant in a gene for which primarily truncating variants are known to cause disease* - Only truly applicable for *ALS2*

BP2: *Observed in trans with a pathogenic variant for a fully penetrant/dominant gene/disorder, OR, observed in cis with a pathogenic variant in any inheritance*

BP3: *In-frame deletions/insertions in a repetitive region without a known function*

BP4: *Multiple lines of computational evidence suggest no impact on gene or gene product* – Methods as per PP3

BP5: *Variant found in a case with an alternate molecular basis for disease (dominant variants)* – *C9orf72* expansions were considered alternative molecular explanations for disease.

BP6: *Reputable source recently reports variant as benign but evidence is not available for independent evaluation* - As per PP5

BP7: *A synonymous variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site, AND, the nucleotide is not highly conserved* - Splicing algorithms – MaxEntScan >7 for splice impact and dbscsnv >0.8 for splice impact.

VUS: *Does not meet any of the above criteria or conflicting evidence (pathogenic and benign criteria met)*