Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

The Environmental Epidemiology of Amyotrophic Lateral Sclerosis in Europe

A dissertation submitted to Trinity College Dublin in fulfilment for the award of Doctor of Philosophy (PhD)

James PK Rooney
B.Sc. , M.B. B.Ch. B.A.O. , M.Sc.

Supervisor: Professor Orla Hardiman

September 2018

The Environmental Epidemiology of Amyotrophic Lateral Sclerosis in Europe

By

James PK Rooney

Professor Julie Kelly
_____
Viva Voce Chair

Professor Joe Barry
_____
Internal Reviewer

Professor Neil Pearce
_____
External Reviewer

Professor Orla Hardiman
_____
Supervisor

Declaration

Declaration I declare that this thesis was not previously submitted as an exercise for a degree at this or any other university and, unless otherwise stated, it is entirely my own work. Where any of the content presented is the result of input or data from related collaborative research this is duly acknowledged in the text.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signed……
Date……/……/……..

Abstract

Amyotrophic Lateral Sclerosis (ALS) is a terminal neuro-degenerative disorder of adults, typically having rapid progression and involving both motor and cognitive function. ALS exhibits considerable heterogeneity of both symptom profile and progression. Nevertheless, median survival is typically under three years from disease onset. ALS is the most common motor neurone disease in adults, with an incidence of 2 to 3 per 100,000 in Europe. The cause of ALS is unknown, however current expert opinion is that the disease occurs as a result of the combined effects of genetic and environmental factors.

The primary aims of this thesis were to investigate environmental exposures as risk factors for ALS in Irish and European populations, and to investigate clinical and genetic prognostic factors in those populations. Data was obtained using the Irish ALS Register, and from Euro-MOTOR - an international, case-control study of Dutch, Irish and Italian ALS patients matched healthy controls.

Overall, the findings from this thesis provide evidence that occupational and environmental exposures have importance in ALS aetiology. Spatial epidemiological analysis of the Irish ALS cohort found two significant areas of low risk for ALS, and the Euro-MOTOR study revealed associations between oral contraceptive pill use, physical activity, occupational exposures and ALS risk. These findings indicate the need for large gene-environment studies in future ALS research. Furthermore, survival analyses reinforced previous evidence that attendance at an ALS multidisciplinary clinic is associated with improved survival, and provided fresh insight into the prognostic effect of the *C9orf72* expansion in ALS, indicating that stratification by *C9orf72* expansion status is important in epidemiological studies. Finally, longitudinal and prognostic characteristic of sub-scores of the ALS Functional Rating Scale (ALSFRS) were explored and were found to be independently associated with survival. This may help to inform future clinical trial design.

Acknowledgements

Primarily, I would like to thank and acknowledge the people with ALS, and their families, who allowed their information to be included in the Irish ALS Register, and who participated in the research studies upon which this thesis is based. This involved a sacrifice of time and energy when both were scarce, without which the research described in this thesis would not be possible. I would also like to thank the healthy volunteers who acted as control participants for these projects.

I thank Professor Hardiman for her advice, guidance and encouragement throughout this work. I would also like to thank Professor Anthony Staines, Dublin City University, who co-sponsored my HRB grant and freely shared his expertise in spatial epidemiology. I would like to acknowledge the Irish ALS research Team (Trinity College Dublin), in particular Mark Heverin who provided fantastic support for all of my work, and Tom Burke for providing feedback on this thesis.

I would like to thank my family for their continuing support, and for instilling in me the enthusiasm for learning that is necessary to complete a thesis such as this.

As part of my PhD I had the opportunity to work with ALS groups in different European countries and I would like to thank Professor Ammar Al-Chalabi (King's College London), Professor Leonard van den Berg and Professor Jan Veldink (University of Utrecht), and Professor Adriano Chiò (University of Turin) for inviting me to their respective institutions. Thanks also must go to Anna Kulka (King's College London), Fabrizio D'Ovidio (University of Turin), Ruben van Eijk, Anne Visser and Bastiaan Middelkoop (University of Utrecht) for welcoming me to those institutions and being excellent colleagues and collaborators.

Finally, I would like to thank the Health Research Board for sponsoring me to pursue my PhD thereby making this work possible (Grant number: HPF-2014-527).

Technical acknowledgements

I acknowledge Neil McCluskey, NCRI, and the National Cancer Registry of Ireland for their time, expertise, and the use of modified OSI shapefiles generated for the All Ireland Cancer Atlas. I acknowledge Ordnance Survey Ireland for granting us copyright to reproduce OSI shapefiles (Permit No. MP 0008613).

I thank Wayne Richter, New York State Department of Environmental Conservation for sharing R code to output ASCII data suitable for SaTScan, and Daniel Simpson and Herbert Rue of the Norwegian University of Science and Technology for support in using R-INLA.

I thank Trutz Haase for sharing data from the Pobal HP social deprivation index and his advice regarding use of the index. I would also like to thank Teagasc and NUI Galway (in association with the EPA) for granting license to access the National Soil Database to facilitate this research.

Table of Contents:

List of tables

List of Figures

List of publications

List of research articles published during this thesis:

1. Al-Chalabi A, Calvo A, Chio A, Colville S, Ellis CM, Hardiman O, Heverin M, Howard RS, Huisman MH, Keren N, Leigh PN, Mazzini L, Mora G, Orrell RW, **Rooney J**, Scott KM, Scotton WJ, Seelen M, Shaw CE, Sidle KS, Swingler R, Tsuda M, Veldink JH, Visser AE, van den Berg LH, Pearce N. Analysis of amyotrophic lateral sclerosis as a multistep process: a population-based modelling study. Lancet Neurol. 2014 Nov;13(11):1108–13. PMID: 25300936

2. Pupillo E, Messina P, Giussani G, Logroscino G, Zoccolella S, Chiò A, Calvo A, Corbo M, Lunetta C, Marin B, Mitchell D, Hardiman O, **Rooney J**, Stevic Z, Bandettini di Poggio M, Filosto M, Cotelli MS, Perini M, Riva N, Tremolizzo L, Vitelli E, Damiani D, Beghi E, EURALS Consortium. Physical activity and amyotrophic lateral sclerosis: a European population-based case-control study. Ann Neurol. 2014 May 75(5):708–16. PMID: 24706338

3. (Chapter 3.1) **Rooney J**, Heverin M, Vajda A, Crampsie A, Tobin K, Byrne S, et al. An Exploratory Spatial Analysis of ALS Incidence in Ireland over 17.5 Years (1995 – July 2013). Le W, editor. PLoS One. 2014 May 27;9(5):e96556. PMID: 24867594

4. (Chapter 4.1) **Rooney J**, Byrne S, Heverin M, Tobin K, Dick A, Donaghy C, et al. A multidisciplinary clinic approach improves survival in ALS: a comparative study of ALS in Ireland and Northern Ireland. J Neurol Neurosurg Psychiatry. 2015;86(5):496–501. PMID: 25550416

5. (Chapter 3.2) **Rooney J**, Vajda A, Heverin M, Elamin M, Crampsie A, McLaughlin R, Staines A, Hardiman O. Spatial cluster analysis of population amyotrophic lateral sclerosis risk in Ireland. Neurology. 2015;84(15):1537–44. PMID: 25770197

6. (Chapter 3.3) **Rooney JPK**, Tobin K, Crampsie A, Vajda A, Heverin M, McLaughlin R, Staines A, Hardiman O. Social deprivation and population density are not associated with small area risk of amyotrophic lateral sclerosis. Environ Res. Elsevier; 2015;142:141–7. PMID: 26142719

7. (Chapter 3.4) **Rooney J**, Vajda A, Heverin M, Crampsie A, Tobin K, McLaughlin R, Staines A, Hardiman O. No association between soil constituents and amyotrophic lateral sclerosis relative risk in Ireland. Environ Res. Elsevier; 2016;147:102–7. PMID: 26855128

8. Fogh I, Lin K, Tiloca C, **Rooney J**, Gellera C, Diekstra FP, et al. Association of a Locus in the CAMTA1 Gene With Survival in Patients With Sporadic Amyotrophic Lateral Sclerosis. JAMA Neurol. 2016;73(7):812–20. PMID: 27244217

9. Tobin K, Gilthorpe MS, **Rooney J**, Heverin M, Vajda A, Staines A, Hardiman O. Age-period-cohort analysis of trends in amyotrophic lateral sclerosis incidence. J Neurol; 2016: Jul 2;178(8):1265–71. PMID: 27372451

10. (Chapter 4.2) **Rooney J**, Heverin M, Vajda A, Burke T, Galvin M, Tobin K, Elamin M, Staines A, Hardiman O. Survival analysis of geospatial factors in the Irish ALS cohort. Amyotroph Lateral Scler Frontotemporal Degener. 2016;17(7–8):555–60. PMID: 27145090

11. (Chapter 4.4) **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;88(5):381–5. PMID: 27888187

12. Elamin M, Pinto-Grau M, Burke T, Bede P, **Rooney J**, O'Sullivan M, Lonergan K, Kirby E, Quinlan E, Breen N, Vajda A, Heverin M, Pender N, Hardiman O. Identifying behavioural changes in ALS: Validation of the Beaumont Behavioural Inventory (BBI). Amyotroph Lateral Scler Front Degener. 2016;8421(November):1–6. PMID: 27894191

13. (Chapter 2) Hardiman O, Chalabi A Al, Brayne C, Beghi E, Van den Berg LH, Chio A, Martin S, Logroscino G, **Rooney J**. The changing picture of amyotrophic lateral sclerosis: lessons from European registers. J Neurol Neurosurg Psychiatry. 2017; Jul;88(7):557-563. PMID: 28285264

14. (Chapter 4.3) **Rooney J**, Fogh I, Westeneng H, Vajda A, McLaughlin R, Heverin M, Jones A, van Eijk R, Calvo A, Mazzini L, Shaw C, Morrison K, Shaw PJ, Robberecht W, Van Damme P, Al-Chalabi A, van den Berg L, Chiò A, Veldink J, Hardiman O. C9orf72 expansion differentially affects males with spinal onset amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 2017 Apr 23;88(4):281. PMID: 27663272

15. (Chapter 5.1) D'Ovidio F[1], **Rooney JPK[1]**, Visser AE[1], Vermeulen RCH, Veldink JH, Van Den Berg LH, Hardiman O, Logroscino G, Chiò A, Beghi E; Euro-MOTOR Group. Critical issues in ALS case-control studies: the case of the Euro-MOTOR study. Amyotroph Lateral Scler Front Degener. 2017 Jul 3;18(5–6):411–8. PMID: 28594593

16. **Rooney JPK**, Brayne C, Tobin K, Logroscino G, Glymour MM. Benefits , pitfalls , and future design of population-based registers in neurodegenerative disease. 2017;1–9. PMID: 28515268

17. (Chapter 5.2) **Rooney JPK[1]**, Visser AE[1], D'Ovidio F [1], Vermoulen R, Beghi E, Chio A, Veldink JH, Logroscino G, Van den Berg LH, Hardiman O, for the Euro-MOTOR consortium. A case-control study of hormonal exposures as etiological factors for ALS in women: Euro-MOTOR. Neurology. 2017 Sep 19;89(12):1283-1290.

18. Pupillo E, Poloni M, Bianchi E, Giussani G, Logroscino G, Zoccolella S, Chiò A, Calvo A, Corbo M, Lunetta C, Marin B, Mitchell D, Hardiman O, **Rooney J**, Stevic Z, Bandettini di Poggio M, Filosto

M, Cotelli MS, Perini M, Riva N, Tremolizzo L, Vitelli E, Damiani D, Beghi E; EURALS Consortium. Trauma and amyotrophic lateral sclerosis: a European population-based case-control study from the EURALS consortium. Amyotroph Lateral Scler Frontotemporal Degener. 2018 Feb;19(1-2):118-125.

19. Visser AE, **Rooney JPK**, D'Ovidio F, Westeneng HJ, Vermeulen RCH, Beghi E, Chiò A, Logroscino G, Hardiman O, Veldink JH, Van den Berg LH, for the Euro-MOTOR consortium. A multicentre, cross-cultural, population-based case-control study of physical activity as risk factor for amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 2018 Aug;89(8):797-803.

20. Westeneng HJ, Debray TPA, Visser AE, van Eijk RPA, **Rooney JPK**, Calvo A, Martin S, McDermott CJ, Thompson AG, Pinto S, Koboleva X, Rosenbohm A, Stubendorff B, Sommer H, Middelkoop BM, Dekker AM, van Vugt JFA, van Rheenen W, Vajda A, Heverin M, Kazoka M, Hollinger H, Gromicho M, Körner S, Ringer TM, Rödiger A, Gunkel A, Shaw CE, Bredenoord AL, van Es MA, Corcia P, Couratier P, Weber M, Grosskreutz J, Ludolph AC, Petri S, de Carvalho M, Van Damme P, Talbot K, Turner MR, Shaw PJ, Al-Chalabi A, Chiò A, Hardiman O, Moons KGM, Veldink JH, van den Berg LH. Prediction of personalised prognosis in patients with amyotrophic lateral sclerosis: development and validation of a prediction model. Lancet Neurol. 2018 May;17(5):423-433.

Additional research articles published during the course of this thesis:

21. **Rooney JPK**. The retention time of inorganic mercury in the brain--a systematic review of the evidence. Toxicol Appl Pharmacol. Elsevier Inc.; 2014 Feb 1;274(3):425–35. PMID: 24368178

22. Hayes JP, **Rooney J**. The prevalence of respiratory symptoms among mushroom workers in Ireland. Occup Med (Lond). 2014 Oct;64(7):533–8. PMID: 25138013

Letters to the Editor published during the course of this thesis:

23. **Rooney JPK**, Dórea JG. Krakow's children cohort and long-term follow-up of thimerosal exposure—design and statistics. Eur J Pediatr. 2015;174(11):1555. PMID: 25994246

24. **Rooney JP**, Staines A, Hardiman O. Spatial cluster analysis of population amyotrophic lateral sclerosis risk in Ireland. Neurology. 2015 Nov 17;85(20):1822–3. PMID: 26574539

25. **Rooney JPK**, Frissen MN, Bass GA, Dórea JG. Dental amalgam fillings and Helicobacter pylori eradication rates: Wide-ranging implications. Eur J Gastroenterol Hepatol. 2015;27(10):1231. PMID: 26327504

26. Dórea JG, **Rooney J**. Comments on the "Effects of in utero exposure to polychlorinated biphenyls, methylmercury, and polyunsaturated fatty acids on birth size". Sci Total Environ; 2016 Feb 15;544:1136–7. PMID: 26857042

Glossary of Terms

AIC ……..Akaike Information Criterion

ALS ……Amyotrophic Lateral Sclerosis

ALSFRS ……Amyotrophic Lateral Sclerosis Functional Rating Scale

ALSFRS-R ……..Amyotrophic Lateral Sclerosis Functional Rating Scale - Revised

ANOVA ……Analysis of Variance

ASCII ……..American Standard Code for Information Interchange

BBI ……..Beaumont Behavioural Inventory

BIC ……..Bayesian Information Criterion

BMAA ……β-methylamino-L-alanine

BMI ……..Body Mass Index

BUGS ……Bayesian Inference Using Gibbs Sampling

BYM ……Besag-York-Mollié model

CAFS ……Combined Assessment of Function and Survival

CERA ……..Continuous Erythropoietin Receptor Activator

CFA ……Confirmatory Factor Analysis

CIDP ……..Chronic Inflammatory Demyelinating Polyradiculoneuropathy

CSF ……Cerebro-Spinal Fluid

CSO ……Central Statistics Office (Ireland)

DIC ……Deviance Information Criterion

DME ……Diesel Mineral Exhaust

DNA ……Deoxyribonucleic Acid

EFNS ……European Federation of Neurological Societies

ELF-EMF ……Extremely Low Frequency Electromagnetic Field

ENCALS ……European Network to Cure ALS

EPA ……..Environmental Protection Agency

EPH ……Ephrin (receptor)

EPO ……Erythropoietin

FAB ……Frontal Assessment Battery

FALS ……..Familial Amyotrophic Lateral Sclerosis

FDA ……..Food and Drug Administration (USA)

FTD ……Fronto-Temporal Dementia

FTLD ……..Frontotemporal Lobar Degeneration

FUS ……Fused in Sarcoma

FVC ……Forced Vital Capacity

GDPR……..General Data Protection Regulation

GIS ……Geographic Information System

GPS ……Global Positioning System

GWAS ……..Genome Wide Association Study

HIPE ……Hospital In-Patient Enquiry

HRT ……..Hormone Replacement Therapy

HSE ……Health Service Executive

IGF ……Insulin-like Growth Factor

INLA ……..Integrated Nested Laplace Approximation

IPD ……Individual Patient Data

IQR ……Inter-Quartile Range

ISCED ……..International Standard Classification of Education

ISCO ……International Standard Classification of Occupations

JEM ……Job Exposure Matrix

LAENALS ……Latin American Epidemiology Network of ALS

LMN ……..Lower Motor Neuron

MCMC ……..Markov Chain Monte Carlo

MDT ……Multi-Disciplinary Team

MESH ……Medical Subject Headings

MET ……Metabolic Equivalent of Task

MND ……Motor Neurone Disease

NEALS ……Northeast ALS Consortium

NICE ……The National Institute for Health and Care Excellence (UK)

NIPPV ……Non-Invasive Positive Pressure Ventilation

NIV ……Non-Invasive Ventilation

OCP ……Oral Contraceptive Pill

OSI ……Ordnance Survey Ireland

PAH ……Poly-Aromatic Hydrocarbons

PASW ……Predictive Analytics SoftWare (SPSS)

PCR ……Polymerase Chain Reaction

PDC ……..Parkinsonism-Dementia Complex

PDGF ……Platelet-Derived Growth Factor

PEG ……Percutaneous Endoscopic Gastrostomy

PLS ……Primary Lateral Sclerosis

PROACT ……Pooled Resource Open-Access ALS Clinical Trials Database

PSMA ……..Progressive Spinal Muscular Atrophy

RAN ……Repeat-associated Non-ATG

RCT ……..Randomised Controlled Trial

RIG ……Radiologically Inserted Gastrostomy

RNA ……Ribonucleic Acid

ROI ……Republic of Ireland

SBGH ……Sex Hormone Binding Globulin

SEALS ……South-East England ALS Register

SIR ……Standardized Incidence Ratios

SNIP ……..Sniff Nasal Inspiratory Pressure

SOD ……Superoxide Dismutase

TAR ……Trans-Activation Response element

TARDBP ……TAR DNA-binding protein

TIA ……Transient Ischaemic Attack

UMN ……Upper Motor Neuron

VEGF ……Vascular-Endothelial Growth Factor

VFI ……Verbal Fluency Index

VGDF ……..Vapours, Gasses, Dusts and Fumes

Chapter 1: Background

Chapter Outline:

1. Background

1.1. What is Amyotrophic Lateral Sclerosis ?

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disorder, most typically with onset in the middle period of life, and characterised by the progressive death of upper and lower motor-neurones leading to weakness, paralysis and eventually death. In recent years the disease has been increasingly recognised to include brain functions beyond the motor system – particularly executive and behavioural dysfunction[1,2], and ALS is known to have overlapping genetic aetiology with fronto-temporal dementia (FTD)[3,4]. The onset and progression of symptoms in ALS is heterogenous and prognosis ranges from several months to several decades[5].

By definition, ALS is the most common of a group of diseases collectively known as Motor Neurone Disease (MND). However the two terms are used somewhat interchangeably in Ireland and the UK, while ALS or the eponymous name "Lou Gehrig's disease" after the famous base-ball player, are more commonly used in the United States. Throughout this thesis I will use the term 'ALS' as the studies detailed herein included only patients who met the revised El-Escorial criteria for ALS[6].

1.1.1. Clinical Features

ALS is characterised by a history of progressive weakness and the presence of mixed upper (hyper-reflexia, clonus, spasticity) and lower (weakness, atrophy, fasciculations) motor neurone signs[6]. Although disease onset is typically said to follow the rule of thirds: "one third upper limb, one third lower limb and one third bulbar in onset", onset can be heterogeneous and initial presentation with respiratory onset, head-drop, weight-loss or even with fronto-temporal symptoms can occur [7]. Progression is also highly varied, and in some cases prolonged follow up is required to distinguish ALS from other MND's such as primary lateral sclerosis (PLS). A number of diseases associated with progressive motor weakness share some, though not all, of the features of ALS. Such mimic conditions include: chronic inflammatory demyelinating polyradiculoneuropathy (CIDP), inclusion

body myositis, Kennedy's syndrome, neuralgic amyotrophy, hereditary spastic parapesis, cervical myeloradiculopathy and primary progressive multiple sclerosis[7].

Diagnosis of ALS is in accordance with the clinical definition set forth by the revised El-Escorial criteria [6]. The criteria that stipulate the diagnosis of ALS are specified in Table 1-1. The criteria further classify ALS cases as definite, probable, probable (lab-supported), and possible ALS based on the distribution of symptoms [6]. (Note the suspected ALS category was removed when the criteria were revised [6]). Since there is not yet any definitive diagnostic test for ALS, the role of clinical investigations is primarily to exclude mimic disorders. Electrophysiological testing offers supportive evidence but is not diagnostic in its own right.

Table 1-1 The revised El-Escorial criteria

The diagnosis of ALS requires the presence of: • Evidence of lower motor neuron (LMN) degeneration by clinical, electrophysiological or neuropathological examination • Evidence of upper motor neuron (UMN) degeneration by clinical examination • Progressive spread of symptoms of signs within a region to other regions
Together with the absence of: • Electrophysiological or pathological evidence of other disease processes that might explain LMN signs and/or UMN degeneration • Neuroimaging evidence of other disease processes that might explain observed clinical and electrophysiological signs
Reference: Brooks *et al.*, 2000 [6]

1.1.2. Epidemiology

ALS is an uncommon disease in comparison to neurological disorders such as cerebro-vascular accidents or the dementias. The crude global incidence rate has been recently estimated as 1.75 (95%: 1.55–1.96) per 100,000 person-years, although data for some continents, particularly Africa, is lacking[8]. Heterogeneity across continents has been identified, with higher standardised incidence in Europe and North America compared to South or East Asia[8]. Due to several long running population based ALS registers the incidence of ALS in Europe is very well described (Table 1-2) and overall age-standardised incidence ranges from 2 to 4 per 100,000 person-years.

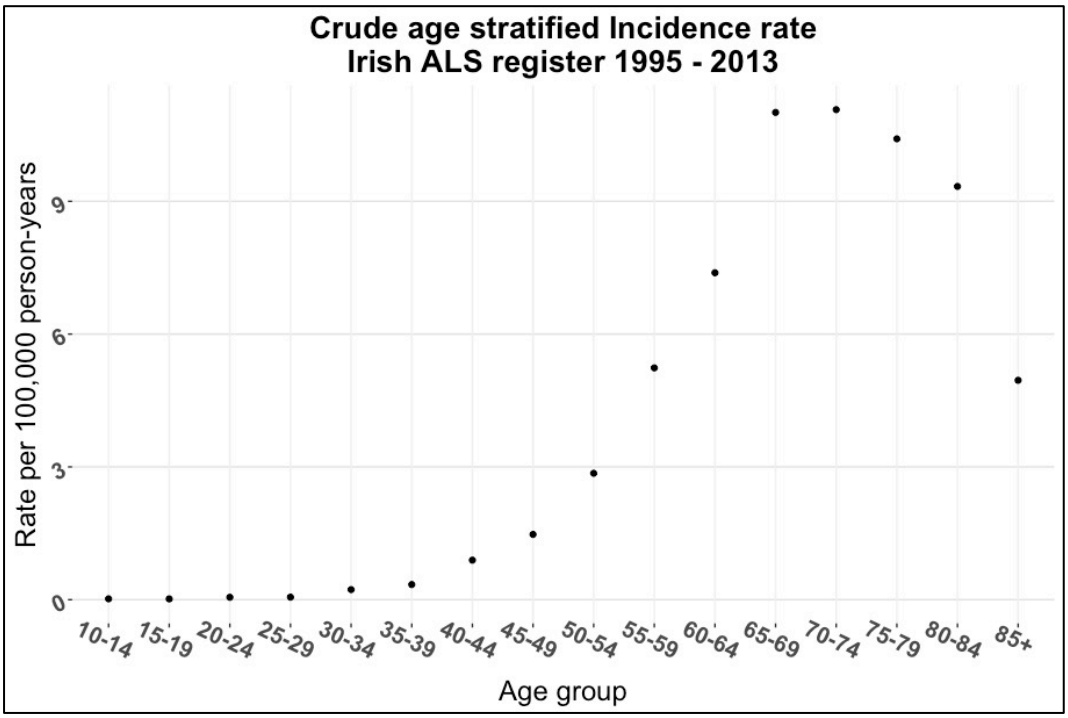
Table 1-2 Age standardised incidence of ALS in long running European population based registers

Country	Registration dates	Catchment population size	Number registered	Age-standardised incidence
England (SEALS)	2002 - 2009	4.023 million	438	2.19
Ireland	1995 - 2012	4.093 million	1,593	3.27
Italy (Piedmont)	1995 - 2004	4.398 million	1,260	3.19
Netherlands	2006 - 2012	16.546 million	1,757	2.16
Scotland	1989 - 1998	5.125 million	1,226	3.43

Legend: SEALS = South East England ALS register. Data sourced from Al-Chalabi *et al.*, 2014 [9]

There is a well-recognised excess of male cases over female, with a typical male: female ratio of 1.5: 1 (unadjusted for age) [10]. However females are more likely to be older at onset and to have bulbar onset disease [10–12]. Peak age of onset is typically in the middle period of life with a falling incidence at older ages (Figure 1-1).

Figure 1-1 Crude incidence of ALS in Ireland by age group



Familial disease was previously thought to make up around 5% of cases[13]. However, with familial aggregation studies of families we now know that in Ireland familial disease is at least 16% [14]. Survival is very heterogeneous, and while the majority of cases have a prognosis of under 3 years, some individuals survive for up to 10 years or longer [5,11,15]. Due to the poor prognosis, prevalence figures for ALS are not drastically higher than incidence, typically with point prevalence in the region of 6 to 10 per 100,000 persons [12,16,17].

Localised clustering of ALS cases have been reported in many studies. Early observations from Guam [18] and the Kii peninsula of Japan [19,20] of high incidences of ALS led to interest in environmental factors. More recently localised clusters of ALS cases have been reported in several European countries [21–23] and the United States [24–26]. Many such studies (though not all) suffer from methodological flaws such as the post-hoc investigation of purported clusters, and comparison across studies is difficult due to the use of diverse methods, therefore spatial methods have thus far failed to definitively identify local environmental

risk factors for ALS. In Ireland, crude analysis of incidence by county of cases incident from 1995 – 1997 suggested an increase of cases in county Donegal [27].

The full causal mechanisms of ALS are unknown. However, recent application of multistep models from cancer research to population based ALS cohorts indicates that a 6 step process, on average, is responsible [9]. Such a multi-step process is compatible with ALS as a oligo-genic process or as a process involving genes and environmental factors [9]. Recent progress has been made in unravelling the genetics of ALS, most notably with the discovery of the *C9orf72* expansion [3,4], responsible for up to 10% of cases in Northern Europe. A large number of environmental risk factors for ALS have been proposed. However, currently smoking is the only environmental factor established to increase ALS risk [28]. Chapter 1.2 summarises the most important genetic discoveries in ALS, while Chapter 1.3 provides a summary of proposed environmental risk factors.

1.2. The Genetics of ALS

As much as 20% of ALS cases in Ireland are now known to be familial, and approximately 10% of all cases are attributed to the pathological *C9orf72* expansion. The earliest known ALS gene, *SOD1*, which accounts for about 2% of ALS cases worldwide, has not been described in Ireland. The rate of discovery of pathogenic ALS genes has increased in recent years, consistent with a multi-step oligogenic, gene-environment process [9,29]. Over 30 genetic variants have been pathologically linked to ALS [29]. This section reviews the most important genes identified to date.

Superoxide dismutase 1 (SOD1)

SOD1 mutation as a pathological cause of ALS was discovered via genetic linkage analysis in 1993 [30,31], making it the first known ALS causing variant. It is responsible for up to 12% of familial cases and 1% of sporadic cases internationally, though has not been found in any Irish cases to date. Over 100 different mutations have been described, but distinct pathological pathways have been described for only a few[32]. *SOD1* ALS appears to have a distinct pathology from other forms of ALS, due to the absence of TDP-43 and/or FUS pathology which is seen in other forms of ALS (including *C9orf72* expansion ALS)[32,33]. It is thought that misfolded SOD1 protein aggregates are pathological, and may be able to propagate from affected cells to previously unaffected cells[34].

TAR DNA-binding protein 34 (TARDBP)

Although pathological studies had previously identified ubiquitinated TAR DNA binding protein (TDP-43) inclusions in tissue samples of ALS and FTD patients, mutations in the *TARDBP* gene were only firmly associated with ALS in two studies in 2008 [35,36]. While *TARDBP* mutations are themselves uncommon causes of ALS-FTD (1-2%), 97% of ALS cases exhibit TDP-43 inclusions[35–37], and they also occur in tau-negative FTD[35–37]. The TDP-43 protein plays a role in RNA spicing and stability, and is capable of nucleocytoplasmic shuttling[37].

Fused in sarcoma (FUS)

FUS is a nucleoprotein with functions including DNA and RNA binding, transcription and splicing, and mRNA transport [38]. A mutant form characterised by cytoplasmic FUS immuno-reactive inclusions and lower motor neurone predominant symptoms was associated with familial ALS 2011 [39]. These inclusions differ from the ubiquitin-positive and TDP-43-positive aggregates TDP-43 found in most ALS, but may imply that FUS and TDP-43 operate at different points on the same pathway [40].

Chromosome 9 open reading frame 72 (C9orf72)

Discovered in 2011 [3,4], the *C9orf72* hexanucleotide repeat expansion is responsible for 10% of ALS cases and 25% of FTD cases in Northern Europe [41]. Thus, it is the largest shared factor linking ALS & FTD as a clinical spectrum, and is the most important ALS gene discovered to date. *C9orf72* is thought to be involved in RNA processing, the pathological expansion may result in repeat-associated non-ATG (RAN) protein translation, and it has been linked to the production of toxic dipeptide repeats [42,43]. Furthermore, recent evidence has implicated the expansion in disrupting nucleocytoplasmic transport processes [44–47]. The phenotype of *C9orf72* expansions are not limited to ALS/FTD, and the expansion has been associated with parkinsonism in ALS/FTD, psychosis, and Huntington's Disease [48–51].

TANK Binding Kinase 1 (TBK1)

TBK1 is a kinase that plays an important role as a node protein for multiple signalling pathways [52]. It integrates upstream signals from a variety of proteins involved in pathogen detection, inflammation and ras-induced oncogenesis, and in turn modulates downstream targets including immune response/inflammation, autophagy, insulin signalling and proliferation/growth [52]. Mutations of TBK1 have recently been established as a causative pathway in ALS & FTD, supporting a role for autophagy and neuroinflammation in ALS [53].

1.3. Environmental risk factors

There is a large body of literature on the role of environmental risk factors in ALS. However with the exception of smoking, no purported risk factor has gained widespread acceptance as an established risk factor for ALS [28]. A MESH search of the Pubmed database published in the last 10 years found 20 systematic reviews and meta-analysis. Risk factors are now considered in turn.

1.3.1. Smoking

Considered as an established risk factor for ALS, the evidence for smoking in ALS has been assessed by three systematic reviews/meta-analyses [28]. In 2009, Armon performed a systematic review identifying 7 studies that met the inclusion criteria and on the strength of 3 class II studies and 1 class III study declared smoking to be a level A risk factor ('established risk factor') [28]. However, a subsequent meta-analysis of 18 cohort and case control studies in 2011 conflicted with Armon finding a pooled RR of ALS of 1.28 (95%CI: 0.97 to 1.68) for current versus never smokers and 1.12 (95%CI: 0.98 to 1.27) for ever versus never smokers[54]. The authors did note significant heterogeneity between studies and further found a RR of ALS in women of 1.66 (95% CI: 1.31 to 2.10) vs 0.86 (95% CI: 0.71 to 1.03) in men. Despite these conflicted studies smoking is generally accepted as an 'established risk factor' for ALS.

More recently, active smoking post-diagnosis has also been identified as a negative prognostic factor in ALS. Amongst 650 Italian ALS patients current smokers had a significantly shorter median survival (1.9 years, IQR 1.2–3.4) compared with former smokers (2.3 years, IQR 1.5–4.2) and never smokers (2.7 years, IQR 1.8–4.6)[55].

1.3.2. Physical activity

A topic of some controversy, there has long been interest in the role of physical activity in ALS. This interest grew in the 2000s, after reports of increased risk in Italian professional soccer players [56]. Nevertheless a systematic review by Armon in 2007 concluded that physical activity (and trauma) are "probably not"

risk factors for ALS, but acknowledged potential for outstanding confounding factors [57]. However a 2014 systematic review and meta-analysis of 37 studies again concluded that there was Armon class A evidence that physical activity was not a risk factor for ALS, but for occupation based physical activity the evidence was rated as class U (unknown), and for soccer the evidence was rated as class C (possible risk factor) [58]. Again the importance of confounding factors was recognised. Furthermore, it has been suggested that there might be a common factor underlying both physical fitness and ALS susceptibility, which may explain the observed heterogeneity of results [59].

1.3.3. Trauma

Having a history of trauma, in particular head trauma, has also been proposed as a risk factor for ALS. A case control study in 2007 incorporating a literature review and meta-analysis concluded that multiple head trauma was associated with increased risk of ALS (OR 3.1, 95%CI: 1.2 – 8.1), whereas injuries to other body parts were not associated with increased risk [60]. The authors noted heterogeneity between studies – possibly due to differing definitions of trauma between studies and in some cases study designs did not distinguish between general trauma and head trauma in particular [60]. Armon concluded in a review of evidence also in 2007 that trauma was "probably not" (level B evidence) a risk factor for ALS in a review of the literature at that time. However, in this review Armon did not specifically consider head trauma [57]. Therefore, in 2012 Armon & Nelson specifically reviewed the evidence pertaining to head trauma and based on 12 articles found the evidence to be inconclusive – i.e. the evidence for head trauma as a risk factor in ALS was level U by the Armon criteria [61]. In 2016 two general reviews of risk factors in ALS addressed the topic of trauma. Wang *et al.* performed a meta-analysis of 15 studies on trauma and found associations between trauma (OR = 1.73, 95% CI: 1.43 to 2.09, I^2 = 53%), and head-trauma as a sub-group (OR = 1.27, 95% CI: 1.02 to 1.57, I^2 = 9%) to be associated with ALS [62]. Belbasis *et al.* also performed meta-analysis of head-injury (OR = 1.65, 95% CI: 1.09 to 2.50), although they also calculated the 95% predictive interval (PI) which further accounts for between-study heterogeneity and on this basis concluded

there was not enough evidence to reach a definitive conclusion regarding head injury (95% PI 0.59 to 4.64) [63].

1.3.4. Pesticides

There has been considerable interest in the role of pesticides as risk factors and a number of systematic reviews and meta-analyses have been performed. In 2009, Sutedja *et al.* included pesticides in a systematic review of chemical risk factors for ALS applying the Armon classification of evidence and determined pesticides to be a 'possible' risk factor for ALS [64]. The issue was addressed again in two pesticide specific meta-analyses published in 2012. The study by Kamel *et al.* included new cohort data and a meta-analysis of 8 previous studies [65]. The meta-analysis found that exposure to pesticides as a group was associated with ALS (OR 1.9, 95% CI: 1.1 to 3.1), but did not find significant associations for categories of pesticides in the cohort study [65]. Malek *et al.* performed meta-analysis of 6 previous studies and found pesticides exposure to be significant in men (OR = 1.88, 95% CI: 1.36 to 2.61) but not women (OR = 1.31, 95% CI: 0.69 to 2.47) [66]. Malek *et al.* noted that the chemical category of pesticides and duration/frequency of exposures were not specified in most studies making meta-analysis of exposure to specific pesticides and cumulative exposure analysis impossible [66]. Kang *et al.'s* 2014 systematic review had a slightly different focus – assessing risk of ALS with respect to rural residence, farmers and pesticides exposures [67]. Meta-analysis supported an association between pesticides exposure and ALS (OR, 1.44; 95% CI, 1.22-1.70), with farming occupation (OR, 1.42; 95% CI, 1.17 to 1.73) and a higher OR (1.96) for pesticides exposure in men was noted [67]. Conversely, Belbasis *et al.* found inconclusive evidence for pesticides as a risk factor for ALS in their 2016 meta-analysis based on an assessment of predicted interval (PI) (OR = 1.44, 95% CI: 1.22 to 1.70, 95% PI: 0.94 to 2.20) [63].

1.3.5. Heavy Metals

Another category of risk factor that has been often suggested and studied numerous times is that of heavy metal exposure. There has been particular interest in lead, mercury and selenium in part due to the well-recognised clinical

neuro-motor toxicity of these metals with high exposure and also case reports of mimic syndromes of ALS after acute exposures [68]. Studies have also investigated associations between aluminum, cadmium, chromium, manganese, metals unspecified and ALS [64]. However, Sutedja *et al.* included heavy metals in their 2009 systematic review of chemical risk factors and found only three out of 50 studies to be of high quality and, due to heterogeneous study designs, they were unable to perform a meta-analysis for metal exposures [64]. Lead exposure has been more extensively studied than other metals, and in 2014 Wang *et al.* were able to perform a meta-analysis of studies assessing lead exposure as an ALS risk factor [69]. The meta-analysis included 9 studies and found an association between lead exposure and ALS (OR of 1.81, 95% CI 1.39 to 2.36) with low heterogeneity in estimates across studies [69]. In 2016 Belbasis *et al.* extended the analysis of Wang *et al.* [69] to calculate a predictive interval, PI = 1.14 – 2.28, concluding that lead was the only exposure they assessed that presented robust evidence for a convincing association with ALS. In a 2016 meta-analysis of multiple risk factors, Wang *et al.* found that exposure to heavy metals in general was associated with higher risk of ALS (OR 1.69, 95% CI: 1.13 to 2.52), an estimate in approximate agreement with a separate 2016 meta-analysis from Hersi *et al.*[70] (OR 2.13, 95% CI: 1.33 – 3.42), although it is more difficult to interpret such findings due to the non-specific classification of heavy metals as a single group.

1.3.6. Alcohol

Perhaps surprisingly, there is some evidence that alcohol may be a protective factor against ALS. This has been suggested by the results of a number of studies and in 2016 E *et al.* published a meta-analysis of four case-control studies and one cohort study on the topic [71]. The meta-analysis found an OR for the association between alcohol consumption and ALS of 0.57 (95% CI: 0.51 to 0.64), but they were unable to explore a dose-response relationship as most of the included studies did not report the necessary data [71]. The reasons why alcohol might be protective against ALS are not clear, although E *et al.* noted that there have been a few studies showing alcohol to be a possible protective factor in multiple sclerosis and systemic lupus erythematosus [71].

1.3.7. Diesel exhaust, solvents and others

A number of other potential risk factors have been assessed in ALS. Exposure to oil or diesel fumes has been of particular interest. Sutedja *et al.* found inconclusive evidence for exposure to oil including petroleum and machine fuels in 2009 [64]. More recent systematic reviews have not focussed on this exposure. However, as numerous occupations with exposure to diesel exhausts such as driving and military occupations have been linked to ALS [72], interest in this exposure remains high. Solvents including aromatic or chlorinated solvents have been of interest and again evidence in 2009 was inconclusive [64]. While earlier studies were not specific regarding type of solvents, by 2014 more studies did specify particular solvent families and some reports of increased OR were reported [62]. However, studies remained small and overall evidence remains inconclusive [62].

A 2006 review of electrical injury as an ALS risk factor concluded that the evidence did not support a causal link [73]. However, Wang *et al.* performed a meta-analysis of 5 case-control studies and found evidence to suggest electrical injury might be a risk factor (OR=3.27 (1.87-5.73)) [62]. A systematic review on the association between diabetes mellitus and ALS included seven studies, however, due to methodological variability and changing diagnostic definitions for both diseases across time periods, no firm conclusions were reached [74].

1.4. Prognostic Factors

Prognostic factors have been studied extensively. Amongst demographic characteristics age at disease onset is an important survival predictor, while gender is not associated with survival [11]. Change of BMI after disease onset is an important predictor of survival with more rapidly progressing patients losing BMI more quickly and having shorter survival [75]. Diagnostic delay (time from disease onset to diagnosis) is a very strong predictor of survival, with shorter diagnostic delay being associated with shorter survival [11]. This is likely because more rapidly progressing patients seek medical attention more quickly. Patients with bulbar onset disease progress more rapidly than spinal onset [11], as do patients with executive function deficits [1].

With regard to therapeutic interventions, riluzole, licensed for treatment of ALS since the 1990's, is associated with survival benefit of approximately 2 to 3 months [76]. In clinical trials non-invasive ventilation was associated with improved survival [77]. Observational studies typically don't show this survival benefit [11,78], although this is likely due to selection bias of patients who receive NIV. Gastrostomy placement, either radiologically inserted gastrostomy (RIG) or percutaneous endoscopic gastrostomy (PEG), is often used for par-enteral feeding when bulbar symptoms create mastication and swallowing difficulties. Currently there is no strong evidence that this improves survival [79], although several studies indicate the RIG is associated with less complications when compared to PEG [79–81]. Attendance at a specialist ALS multidisciplinary clinic (MDT) has been associated with improved survival in observational studies in a number of different countries [11,82–85]. Recently, the drug Edaravone has been licensed by the FDA for use in ALS patients in the United States based on results of a recent phase 3 randomised trial [86]. However, the drug has yet to be licensed in Europe and ENCALS has invited the manufacturer, MT Pharma, to conduct further clinical trials in Europe to resolve unanswered questions, particularly to provide clarity regarding any survival benefits [87].

Only a few genes have been robustly identified that modify prognosis in ALS. However, as the most frequently identified causative mutation in ALS, a number of studies have associated *C9orf72* with a negative prognosis in cohorts in Europe and the United States with hazard ratios ranging from 1.1 (95% CI: 08. – 1.5) to 1.9 (95% CI:1.1 – 3.7) [88–94]. In addition, polymorphisms of Unc-13 homolog A (*UNC13A*), which is involved in the regulation of neurotransmitter release, has been associated with survival in a number of studies [95–97] with carriers having median survival of 1 year lower than non-carriers in one study of 500 patients in Italy [95]. A GWAS study of 4,256 patients identified two loci associated with survival – one located in the insulin-degrading enzyme (IDE) gene (HR 1.61; 95%CI: 1.38-1.89) and another in the calmodulin-binding transcription activator 1 gene (*CAMTA1*) (HR 1.17; 95%CI: 1.11-1.24) which has been previously linked with cerebellar ataxia and variations in human episodic memory [98]. Early reports of associations between mutations in the Kinesin-associated protein 3 gene (*KIFAP3*)[99] and the EPH receptor A4 gene (*EPHA4*) [100] have not been born out in replication studies [97,98,101].

In general, few studies have examined the role of environmental/behavioural factors and survival in ALS. However, recently it has been shown that smoking is an important prognostic factor [55]. In a population based cohort of 610 Italian patients, survival was shorter in current smokers (1.9 years, IQR: 1.2 – 3.4) compared to ex-smokers (2.3 years, IQR 1.5 to 4.2) or never smokers (2.7 years, IQR: 1.8 – 4.6), and there was evidence of a dose-response effect [55]. A study from the South East of England found that London residency was associated with poorer prognosis, whilst coastal residency was associated with improved prognosis, and authors theorized that this might be due to pollution in London or socioeconomic factors [102]. In the Netherlands it was found that amongst female ALS patients, a longer reproductive life-span (HR = 0.96, 95% CI: 0.92 - 0.99) and a longer lifetime endogenous oestrogen exposure (HR = 0.90, 95% CI: 0.84 - 0.98) were both associated with longer survival [103].

1.5. Aims of this thesis

On this background, this thesis aims to shed new light on the epidemiology of ALS using data from the Irish ALS register, and from European ALS registers through the ENCALS organisation. Specific aims of this thesis are as follows:

1. To review the role of population based registers of ALS research in Europe.
2. Spatial epidemiology:
    a. To describe the overall spatial pattern of ALS incidence in Ireland.
    b. To formally test for cluster of ALS cases in Ireland.
    c. To examine associations between spatially structured factors such as population density and social deprivation and ALS incidence.
3. Survival and progression:
    a. To quantify the survival benefit of attending the ALS multidisciplinary clinic in patients from Ireland and Northern Ireland.
    b. To determine the impact on survival of spatially structured factors such as population density and social deprivation.
    c. To quantify the effect on survival of the *C9orf72* expansion using data from the Irish and other European ALS registers.
    d. To examine the prognostic and longitudinal characteristics of ALSFRS sub-scores.
4. Environmental epidemiology - The Euro-MOTOR study:
    a. To examine the association between lifetime physical activity and ALS risk.
    b. To characterise lifetime BMI in ALS patients vs controls.
    c. To examine the association between hormonal factors and ALS risk in females.
    d. To examine the association between occupational exposures and ALS risk.

1.6. Thesis Organisation

This thesis follows the "Thesis by publication" format. Chapters or sub-chapters which have been published are marked via footnote including the publication details. As such, each chapter/sub-chapter is written in publication format and is intended as a stand-alone piece of work. Chapters are organised by theme as follows: Chapter 2 reviews the literature regarding the role of population based registers in ALS research in Europe. Chapter 3 includes studies on the spatial epidemiology of ALS in Ireland. Chapter 4 focusses on the survival of ALS patients and attempts to delineate important prognostic factors. Chapter 5 includes results from the Euro-MOTOR study – a large multinational case-control study that attempts to identify environmental risk factors for ALS. Finally, Chapter 6 considers the broader implications of these findings within the ALS field.

Chapter 2: Population based ALS Registers – The building blocks of ALS research in Europe

Chapter Outline:

2. Population based ALS Registers – The building blocks of ALS research in Europe[1]

2.1. Introduction

Prospectively designed population based registers with complete ascertainment of all affected individuals in a defined geographic catchment area are of increasing value in epidemiologic research, as evidenced by the success of European population-based registers for Amyotrophic Lateral Sclerosis (ALS) [104,105].

Although a rare disease with an incidence in Europe of 2-3/100,000 person-years [104,105], ALS provides an excellent model to study neurodegenerative diseases. ALS has a rapid progression and is uniformly fatal, and there is high clinico-pathological correlation (autopsy diagnosis matches clinical diagnosis), rendering in vivo diagnosis accurate in the vast majority of cases. As is the case with other neurodegenerations, it is now recognized that ALS is a heterogeneous condition associated with more than one pathogenic mechanism and with different clinical manifestations and trajectories. With evolving recognition of disease heterogeneity as an important factor in a precision medicine approach towards the development of new therapeutics, population based datasets can play a crucial role in defining the full range of disease phenotype and demography.

ALS population based Registers capture all cases regardless of age, health or socioeconomic status, and can thus provide a wealth of information about disease incidence, prevalence, spatial distribution, heterogeneity in clinical phenotype, outcome and analysis of risk. The Irish and Italian ALS Registers have been in

[1] Chapter 2 has been published as:

Hardiman O, Chalabi AA, Brayne C, Beghi E, van den Berg LH, Chio A, Martin S, Logroscino G, **Rooney J**. The changing picture of amyotrophic lateral sclerosis: lessons from European registers. J Neurol Neurosurg Psychiatry. 2017 Jul(8);557 - 653

continuous operation for over 2 decades, and the Dutch and English for over a decade, during which insights into the patho-physiology, epidemiology and genetics of ALS have developed greatly. Each of these registers has provided valuable country-specific data regarding epidemiology, disease progression, clinical phenotyping and genetics, health service planning, quality of life assessment and mapping of the patient journey [9,105–109].

Some of the older European ALS Registers have combined their data as part of the EURALS consortium. This has enabled harmonized collaboration across European Registers, providing insights that would not be otherwise possible. For example, combined appraisal of Registers has indicated the presence of population based variability, has provided evidence that ALS is a multistep process [9], and has facilitated unbiased studies of spatial epidemiology [110–112].

The overall aims of this review is to provide a detailed analysis of what has been learned from existing ALS Registers, to identify and recognize hidden biases and confounders, and to explore future challenges to population based Registers and how these might be met.

2.2. Incidence and Prevalence of ALS

Comparative analysis of ALS across European Registers has provided incidence rates of 2.6 per 100,000 person-years and prevalence rates of 7-9 per 100,000 persons, with a mean life expectancy of 30 months from first symptoms [11,12,105]. European ALS Registers are advantaged by the presence of stable populations with defined geographic borders which ensure accuracy of incident and prevalent figures [105,113]. The core principle of population based Registers is that they capture all patients, nobody is too old, too poor or too sick, and a stable population structure with limited mobility also reduces the risks of loss to follow up.

In general, multiple data sources provide the best mechanism for accurate data capture [104,113]**.** Depending on the type of health system, European ALS

Registers have ascertained patients by a combination of unique patient identifiers (UK, Italy), referrals through networks of clinical professionals and death certification (Ireland), self-reporting and coding, and face to face or telephone based interviews with self-reported questionnaires (Netherlands). To assess the degree of under-ascertainment when two or more independent sources are used, a capture-recapture system is often used [12,17], although this is less accurate if data sources are linked in any way as has been shown by the Irish Register [114].

2.3. How hidden bias can affect estimates of incidence

A number of studies have suggested that the incidence of ALS is increasing [115–117], but this assertion is not fully supported by data from long running population based European Registers [113]. Careful evaluation of data held within European Registers suggests that apparent increases in incidence are more likely explained by the increased recognition of phenotypes that might in the past have been excluded from collection, including in some patients with "possible ALS" as per the EL Escorial diagnostic criteria and those with predominant features of fronto-temporal dementia (FTD), and an associated motor degeneration. This can lead to subtle shifts in ascertainment, as Registers evolve to include patients who might otherwise have escaped ascertainment. In the case of ALS, European population based Registers suggest that this is likely to account in part for the observed upward shift in disease incidence of ALS, particularly in later life [21]. This apparent increase in cases is contemporaneous with the evolving recognition that ALS patients can experience a range of cognitive changes which in some cases may be the presenting symptom [33,88]. Additionally, all Registers, no matter how well they are designed, will miss some cases at the beginning. The development of a national Register, or indeed the presence of a specialist service within a particular region, can have a secondary effect of increasing awareness of the condition, which also improves ascertainment as time goes on. Thirdly, the demographics of European countries are such that an increasing proportion of the population has entered the age range associated with increased risk of developing ALS. In this instance, although the total number of cases of ALS might increase within a population, the standardized rates may not have changed.

Failure to recognize the inherent biases of subtle changes in disease recognition, coupled with improvements in ascertainment strategies and changes in population demography can drive assumptions about increasing frequencies of ALS that are not fully supported by appropriately adjusted data.

2.4. How biases in datasets can affect survival estimates

Population based Registers have demonstrated that up to 70% of patients with ALS die within 3 years of first symptoms, and that approximately 5-10% can survive for more than 8 years [11,78]. Some clinical trial-based datasets (e.g. ProACT: https://nctu.partners.org/ProACT) have suggested that survival of ALS patients has increased over time [118], an effect that has been attributed in part to increased use of interventions such as non-invasive ventilation. However, while there is evidence that non-invasive ventilation may affect survival in some patients, data from European population based Registers has not identified a significant overall effect on survival at population level since the inception of NIV as a standard of care in ALS management [11,78]. Moreover, detailed analysis of population based registers suggests that the apparent increase in survival described in some clinical trial based datasets is more likely to be a function of systematic bias. European population based Registers have demonstrated that those patients who participate in clinical trials, regardless of whether the therapeutic agent has been efficacious, have a different disease trajectory compared to the disease trajectory of the overall population based cohort [119]. Table 2-1 and Figure 2-1 exemplify this by comparing participants of the "Lithium in patients with amyotrophic lateral sclerosis" (LiCALS) trial to the parent register population, i.e. the South-East England ALS (SEALS) register [120]. (NB. The El-Escorial criteria has been shown to have poor inter-rater reliability [121], and for that reason is typically not treated as an explanatory or prognostic variable when comparing cohorts or to define trial entry criteria beyond the exclusion of the suspected category, which is not a part of the revised criteria [6]).

Table 2-1 Demographic and clinical characteristics of patients enrolled in the LiCALS trial versus the SEALS population based register[2]

Variable		Seals (n = 296)	LiCALS (n = 217)	P value
Mean age at onset (yrs)		63.5	58.0	< 0.001 (t)
Median diagnostic delay (months)		12.5	9.8	<0.001 (MW)
Sex (M:F %)		52:48	70:30	<0.001 (Chi^2)
Site of onset (%)	Bulbar	29.3	21.7	
	Spinal	66.9	78.3	
	Not recorded	3.7	0	0.004 (Chi^2)
El Escorial (5)	Definite	27.7	38.2	
	Probable	48.6	37.3	
	Probable – lab supported	1.7	18	
	Possible	11.8	6.5	
	Suspected	2.7	0	
	Not recorded	7.4	0	<0.001 (Chi^2)

t = Student's t Test ; MW = Mann-Whitney U test ; Chi^2 = Chi squared test

Trial datasets usually comprise prevalent patients who attend specialist clinics and who are sufficiently well to meet trial entry criteria and sufficiently motivated to enrol in clinical trials. It has been long recognized that differences in disease trajectories between incident and prevalent cohorts can bias survival analyses, and that this bias underpins the differences in clinical characteristics between

[2] Table 2-1 data prepared by Sarah Martin, King's College London

clinic-based (primarily prevalent) and population-based (incident and prevalent) cohorts [16].

Figure 2-1 Kaplan-Meier curve of LiCALS vs SEALS data[3]

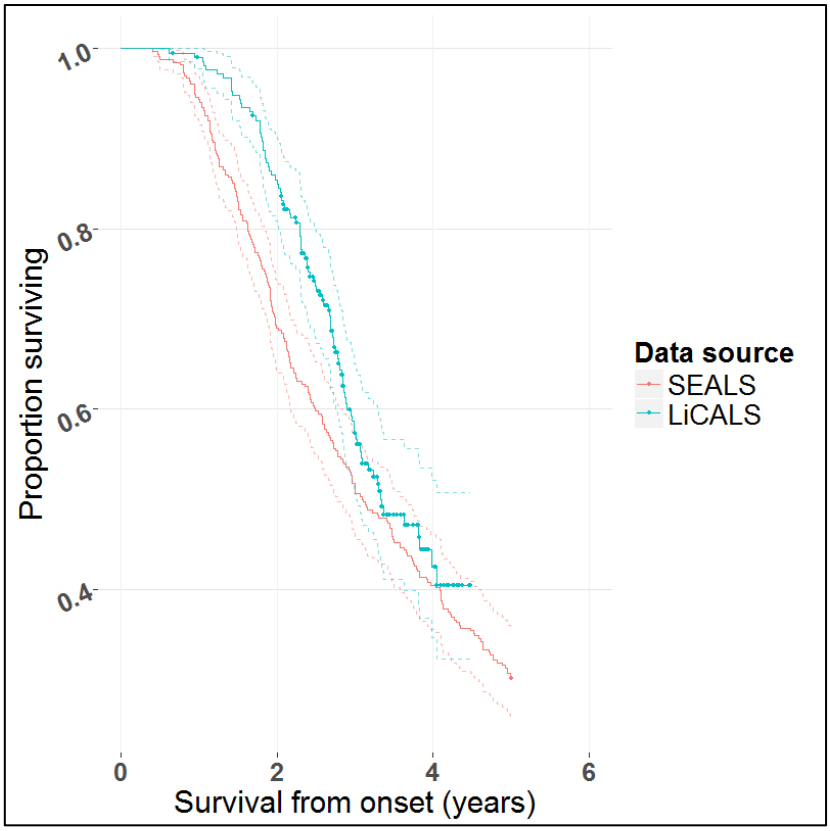


Figure 2-1 Legend: Kaplan-Meier survival curves for patients enrolled in the LiCALS trial (blue) versus the population based SEALS register from which participants were recruited. Characteristics of these groups are described in Table 2-1.

It is therefore more likely that improvements in survival in data repositories such as ProACT are due to subtle differences in the composition of disease cohorts rather than a true shift in disease behaviour. However, it is also the case that direct comparison of datasets (including Registers) across different epochs as a means of determining changes in disease outcome can lead to unintentional bias due to cohort effects. This is one of the reasons why the use of historical controls for comparative purposes is therefore not recommended, even when using

[3] Figure 2-1 prepared by Sarah Martin, King's College London

population based data from Registers. European ALS Registers have shown that data captured in the early years of Registers is unlikely to be of the same complexity or quality of subsequently captured data. For example, in the Irish ALS Register, a subtle shift in age profile has been ascertained within the study cohort [122]. This is most likely to reflect a transition in awareness of disease within the elderly population among referring practitioners. Maturing Registers eventually shift detection from a mixture of prevalent and incidence cases in the earlier Register years towards ascertainment of incident cases as the Register became established (Figure 2-2). Comparative analyses within Registers and across different periods must therefore take these potential confounders into account in drawing conclusions about disease behaviour.

Figure 2-2 Proportion of illness prior to diagnosis versus year of death[4]

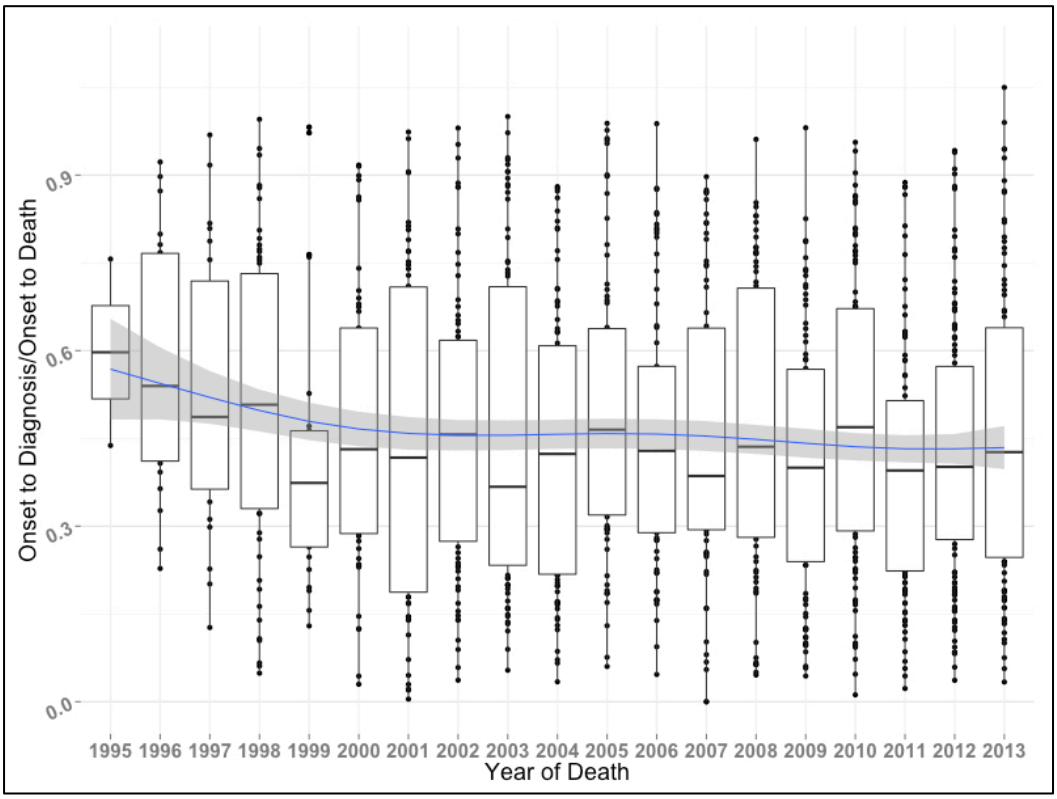


Figure 2-2 Legend: The above graph shows time for onset to diagnosis as a proportion of time from onset to death by year of death for incident cases on the Irish ALS register from 1995 - 2013. In the early years of the register this proportion appears higher before stabilizing to a more consistent ratio after the year 2000. This could be representative of improving survival during the early years of the register, however, this would be at odds with our previous survival analysis [11]**. Alternatively, this could be explained by the detection of 'legacy' prevalent cases being detected in the early years of a population disease register. Note that this graph is generated only from cases diagnosed from 1995 and onwards. If this trend is explained by pre-existing 'legacy' cases in the population, who may have developed the disease some years previously, the effect represents a 'start-up bias' in the early years of the disease register, as such cases are by definition likely to be long survivors.**

[4] Figure 2-2 prepared by Katy Tobin, Trinity College Dublin

2.5. Expanding the ALS Phenotype

Italian and Irish ALS Registers provide detailed analyses of disease sub-phenotypes that have helped to characterize the clinical and cognitive changes associated with ALS [1,123–125], demonstrating how Registers provide new insights that can help to accurately classify patients into different clinical and prognostic subgroups. This can be helpful for clinical trial stratification. For example, Irish and Italian Registers have shown that cognitive and behavioural changes are intrinsic features of ALS, affecting up to 50% of patients [124], and are associated with significant prognostic implications [1,125]. Interrogation of individual datasets can also provide important differentiating features that allow exclusion of possible mimic syndromes [126].

Registers have also been helpful in characterizing the presence of possible endo-phenotypes, among probands and their extended family members. Such population-based observations can in turn lead to novel and previously unrecognized pathogenic mechanisms. For example, a recent population based study from Ireland has shown higher rates of psychosis and suicide in first and second degree relatives of ALS patients compared to controls [14]. This observed family aggregation of neuropsychiatric disorders in ALS kindred's provided the necessary hypothesis to undertake a combined summary statistics GWAS Analysis of ALS and schizophrenia which has revealed a hitherto unrecognized 14% polygenic overlap between ALS and schizophrenia, suggesting the presence of shared pathogenic mechanisms between these two clinically divergent disorders [14].

2.6. Ascertaining Environmental Risk

While Registers of themselves, cannot define risk, well established population based prospective Registers such as the Dutch ALS Register can support detailed population based case control studies aimed to assess environmental risk, including physical activity, BMI, consumption of alcohol and fat, smoking and other exposures [59,108,127,128]. Because ALS is a rare disease, large case control studies require extensive collaboration between different centres and across different geographic regions. Subtle differences in ascertainment and disease

definition can introduce bias unless the data collection has been standardized with respect to ascertainment and characterization. The recent Euro-MOTOR study comprising cases and controls from countries with population based Registers demonstrates that combined ascertainment can significantly enhance power for risk assessment [129]. The Euro-MOTOR project has now established a repository of over 1500 population based incident cases and 3000 matched controls with extensive phenotype, environmental, and genomic characterization [129]. The Euro-MOTOR design has since been exported to other regions, most recently to 3 Latin American countries (Cuba, Chile and Uruguay) which have formed the Latin American Epidemiology Network for ALS (LAENALS).

Registers can also demonstrate how non-uniformity of access to health systems can bias analyses of risk. This is exemplified by conflicting observations relating to the association of disease risk and socioeconomic status. For example, an association between ALS with area-based socio-economic status was reported in New Jersey [130], where access to health services is not uniform and those in lower socioeconomic groups may not be captured. This contrasts with recent analyses from the European Registers, which have not demonstrated any association between social deprivation and ALS incidence [110–112].

Registers can additionally be used to address claims of disease clustering. Many studies have been published suggesting increased rates of ALS in regions thought to be associated with specific risk. This is exemplified by the suggestion that environmental pollutants or cyanotoxin exposure are associated with ALS. A geographic ascertainment bias (the so called “Texas sharp-shooter phenomenon”) is generated by examination of reported clusters in the absence of complete surveillance data [131]. By contrast careful population based analysis using well established Registers has to date failed to identify a spatial association between specific environmental pollutants and disease risk, as exemplified most recently by the negative evidence for clustering in a heavily polluted region of Italy [132]. Indeed, to date other than Guam and the Kii peninsula of Japan, no reproducible areas of clustering have been noted. A region of reduced incidence (“cold spot”) has recently been reported by the Irish Register [133] (Chapter 3), The reasons for

this are unclear, but may be related to subtle historical differences in local population structure.

2.7. Making Sense of Genetics

2.7.1. Defining Familial Disease

Prior to the recognition of the importance of the C9orf72 repeat expansion as an important causative gene in ALS, familial ALS (FALS) has been reported to account for 5% of cases [13]. More detailed analysis of family history and genotyping of at least one population based Register (Ireland) now suggests that the true proportion of FALS is closer to 16-20% of all ALS cases [14]. Low reported rates of familial disease are most likely a function of biased study design that do not collect within a population-based setting. It is also the case that incident patients may not be aware of a family history, or may not recognize the link between the proband and a family history of progressive neurological decline. Longer running Registers can provide important insights in this regard, as new patients are ascertained within kindred's that had previously been classified as "sporadic".

It must also be noted that genetic studies that do not recognize the presence of variations in population structure can confound analysis, as they make assumptions of uniform prevalence of gene variants and clinical phenotype [91,134]. Important variation in the prevalence of at risk genes is known to be the case for at least two major ALS genes - the frequency of the C9orf72 repeat expansion is high in populations of European extraction and low in the Asian population [41], while variants in SOD1 account for 13% of familial ALS in Italy, but are not found in Ireland and are rare in Holland [135]. The presence of population isolates can also affect the genetic epidemiology of ALS. For example, higher rates of ALS have been identified in a population isolate in The Netherlands, leading to the discovery of ALS associated variants in the NEK1 gene [106]. Similarly, higher rates of familial ALS have been noted in Sardinia due to founder effects with respect to TDP43 and C9orf72 [91].

It is important to note that Registers can not only help to identify new genes as in the case of NEK1, they can also limit the impact of referral bias on genetic studies [136]. Many genetics studies are of necessity clinic based, and since it is known that clinic based cohorts are phenotypically distinct from population based cohorts (Table 2-1), it can be assumed that reports of the prevalence of at risk genotypes are also biased.

2.7.2. Complex Genetics, Ancestral Origin and Disease Risk

Interrogation of incidence, prevalence and risk genetically admixed populations is of increasing interest. As noted, there is now considerable evidence that the incidence of ALS varies significantly across countries [8,137] and the phenotype and outcome of the disease vary in relation to population ancestral origin [138]. South American populations of mixed ancestral origin may have lower rates of ALS compared with those reported in Europe [137,139]. A population based mortality study from Cuba has reported different rates of ALS in different ancestral populations, with higher rates in those of European origin and lower rates in the admixed population (which corresponds to the "Latino" population in the US) [137]. Population based Registers that ascertain within a region of mixed ancestral origin are therefore of particular interest from genetic and environmental perspectives. Differences in ancestral risk may be significant sources of bias in the generation of Registers in countries such as the US, where differential access to and utilization of health services is linked to race, ethnicity, language, rurality, and socioeconomic status [140,141]. The design of European population based ALS Registers can counteract such biases by capturing all patients using multiple different sources and care pathways [105].

2.8. Health Services

Population based Registers can inform health services. The availability of precise incident, prevalent and clinical trajectory data can permit detailed service planning, and can enable projection of future societal needs. While Registers cannot of themselves provide sufficiently rich datasets to inform the entire patient journey, well-constructed and compatible Registers within different jurisdictions

(Table 2-2) can also permit comparative analyses of different types of services, as has been demonstrated in the island of Ireland [142]. Comparison of survival outcome between Registers in the Republic of Ireland and Northern Ireland, which have similar population structures but which provide different types of specialist care for ALS patients, have shown that multidisciplinary care within a single clinic is superior to devolved care provided within a defined “hub and spoke” model of care [142]. Registers also permit high level comparative studies of different interventions within individual geographic regions using outcomes such as hospitalizations and survival. In Puglia, the model of care is such that there is no survival difference between patients attending local neurologists and those receiving care in a specialist multidisciplinary clinic [84]. Similarly, nested work by Dutch researchers shows that the additional availability of a regional care worker does not improve quality of life among carers [83], although analysis of outcome using population based datasets in the Netherlands and Lombardy (Italy) show that multidisciplinary clinics are also better value for money, reduce hospitalizations and enhance quality of life of patients [143,144].

Table 2-2 Abridged ENCALS – ALS core clinical dataset

Personal data	Clinical Data	Genetics	ALSFRS	ECAS Cognitive and Behavioural data
Date of birth	Diagnosis	Which ALS genes tested?	Date of ALS-FRS	ECAS additional clinical data
Date of death	Date of diagnosis	Mutation in ALS gene ?	ALS-FRS 12 subscores	ECAS cognitive data
Gender	Co-morbidities	Free text notes	ALS-FRS total scores	ECAS behavioural data and carer interview
Patient/control	El Escorial category	Free text to add other genes	Rater name	
ID numbers	Date of disease onset			
Optional identifying data	Site of disease onset			
Free text notes	Other first symptoms (e.g. weight loss, cognitive/ behavioural)			
	Forced Vital Capacity / Sniff nasal inspiratory pressure			
	Family history			
	Endpoints – date of death, date of tracheostomy, date of NIV > 23hrs/day, date of latest endpoint check			

See full ENCALS ALS core clinical data guidelines on the ENCALS website.

2.8.1. Health Services, Clinical Trials and "Real World Data"

There is an increasing recognition that clinical trials by necessity select patients that are not representative of the true population, rendering decisions regarding the generalization of trial findings challenging from a health policy perspective [145]. While the best study design for assessment of treatment effectiveness is the randomized clinical trial, therapeutic effectiveness can also be assessed in more generalizable context using prospective observational studies nested within Registers, as has been demonstrated in the case of Riluzole [146]. Prospective cohort studies using population based Registers in which the outcomes are collected after exposure, or intervention in patients can provide valuable "real world" information regarding the longer term effect of a therapeutic intervention.

2.8.2. Sustainability of Population Based Registers.

Registers are difficult to fund as they are often viewed as infrastructure by research bodies, and as research initiatives by health services. Many Registers rely on the energy and interest of a single founder, and are challenged at the time of retirement of the key principal investigator, as occurred in the case of the Scottish Register on the retirement of the founder clinician, Dr. Robert Swingler. Fortunately, recent recognition of the value of the population-based Register for ALS by the Scottish health authorities has enabled the re-establishment of this valuable resource with the provision of ring fenced funding.

Long term sustainability of Registers can also be eroded by limitations on the types of data disease Registers are permitted to record. While issues of privacy and data protection must be clearly addressed in Registers as part of an over-arching governance structure, recent changes in European legislation are of potential concern to the operation of true population based Registers. For example, inclusion of data relating to living patients without their expressed informed consent is now in breach of European data protection laws. As institutional review boards are taking an increasingly stringent position regarding patient's autonomy with the introduction of "consent to contact" requirements, there is now a real risk of under-ascertainment of cases. Without derogation based on the principles

of public health benefit, it becomes increasingly difficult to create and sustain accurate population based Registers for most conditions.

On 25th May, 2018, the EU's General Data Protection Regulation (GDPR) came into effect, replacing the EU's Data Protection Directive 95/46/EC. Although many of the regulations in GDPR overlap with those of the previous directive, a number of changes of importance to medical research have been introduced. Changes of note in medical research include a broader definition of personal data and Special Category of Personal Data with additional conditions of use (note this now includes genetic and biometric data) [147]. Also of relevance is the concept of "Privacy by design and default" which has now been explicitly defined in the GDPR. Under this concept, data privacy is to be considered from the very outset of a study's design, with end-to-end privacy designed in for the lifetime of the project, including what happens to the data at the end of the project [147]. Transparency is also important, patients included in a study (or registry), should be able to request a copy of their data in a readable format at any time on request [147]. Similarly, the concept of consent has been expanded within GDPR. Consent must be unbundled (e.g. treatment not conditional on partaking in research) and freely given, it must be specific and granular (i.e. generic consent not sufficient, consent for each data use must be given specifically) and it must be easy to withdraw at any time [147]. GDRP also places new obligations on the organisation collecting data – including naming any third party organisations and individuals with whom data will be shared, obligatory reporting of data protection breaches and increased penalties in the case of breaches. There is also an obligation to have an end of project plan regarding the ultimate fate of the data, for example deletion, archival for a period of time or full anonymisation (note that the GDPR does not apply to fully anonymised data, or data from deceased individuals). These changes may present some challenges to medical research and disease research, particular with respect to data sharing. Specific sub-disciplines within medicine may have special concerns, for example observational critical care research has particular concerns regarding the new consent requirements, since the majority of their patients are unconscious [148]. However, Bill Davidson, the joint head of policy at

the Heath Research Authority in the UK is of the opinion that other sectors will be affected more than the medical research sector [149].

In the context of disease registers, the main impact of GDPR is likely to centre around the retention of pseudonymised data long term. Since GDPR defines pseudonymised data as personal data and states that it should be deleted when a research project is over, this is somewhat at odds with the aims of disease registers. It may also be at odds with the growing movement towards open reproducible research – the concept that research data files should be deposited in a data archive for future researchers to examine and reanalyse [150]. Such concerns can be partly addressed by fully anonymising data, however in the case of genetic data which is considered personal data by default within GDPR this may not be possible. It is advised that medical researchers and disease registry managers seek expertise through their university Data Protection Officer.

Legislation providing derogation will require an understanding and recognition by the public of the important potential societal benefits of population-based epidemiological research, and in particular the potential public health benefit of identifying and communicating data regarding regional variations in disease incidence, prevalence and survival. This principle is implicit in the case of notifiable infectious and communicable diseases and is of particular import in the case of rare neurodegeneration. This has been recognized in some jurisdictions for some types of Registers (e.g. the Irish Cancer Register, and the US ALS Registry), and by the International Rare Disease Research Consortium (http://www.irdirc.org/). However, there remains a disappointingly limited recognition within the Europe legislature of the significant benefits of population based Registers.

2.9. Conclusions

Prospective population-based disease registers are invaluable in patient oriented research of rare diseases. As exemplified by the success of European ALS Registers, population based databases can identify and address biases that are intrinsic to other types of data. While some biases cannot be completely eliminated, their

recognition can provide the necessary caution in data interpretation. Notwithstanding their limitations, Registers can provide unique and often unexpected insights into disease epidemiology and pathobiology, and can inform the types of healthcare that are of greatest benefit to patients. Table 2-3 contains our summary recommendations for the operation of a successful population based register.

It is imperative that funding agencies, healthcare providers and institution review bodies recognize the value of these types of Registers, particularly in the case of rare disease such as ALS, and that forthcoming data protection legislation, while well intentioned and appropriate in many ways, does not compromise our ability to fully understand disease heterogeneity, and to continue to improve the lives of patients with ALS and related neurodegenerations.

Table 2-3 Core recommendations for the operation of a population based register

• Clearly defined case definitions should be used including inclusion and exclusion criteria
• Include a clearly labelled register subsection for cases that should be tracked but do not fulfil the formal inclusion criteria
• Register variables should be carefully selected and a "core content" paradigm should be agreed in advance
• International collaborative efforts and/or national merger of data in large countries using multiple registers to cover different regions are advisable for rare diseases.
• Dedicated staff time to ensure effective set-up and maintenance of the register
• Defined Capture methodology including multiple sources
• Regular comparison of ascertainment rates and patient demographics
• Investigation of "ascertainment holes"
• Employment of careful statistical analyses of data collected in the first 3-5 years to account for *"start-up bias"*
• Exclude the most recent 1-2 years of data capture, particularly for survival analyses
• Security is paramount for system software, yet flexibility to accommodate a shifting knowledge base is essential
• Including population-based controls in a register enables valuable case-control studies for studying environment/lifestyle/genetic risk factors

Chapter 3: Spatial epidemiology of ALS

Chapter Outline:

3. Spatial epidemiology of ALS

3.1. An exploratory spatial analysis of ALS incidence in Ireland over 17.5 years[5]

3.1.1. Introduction:

Amyotrophic Lateral Sclerosis (ALS) is a terminal neurodegenerative condition of complex genetic origin, with an annual incidence in Ireland of approximately 2.6 per 100,000 and a lifetime risk of 1:300 [16]. At least 18 Mendelian inherited genes are known to be important in ALS pathogenesis, the most prevalent of which is a hexa-nucleotide repeat expansion in *C9orf72* which accounts for up to 11% of all ALS in Ireland [135]. A larger number of possible susceptibility genes have also been suggested, but many are yet to be confirmed [151]. As ALS is a disease that manifests later in life, genetic factors present at birth are likely to interact with later-life environmental factors. Early observations from Guam [18] and the Kii peninsula of Japan [19,20] of localised higher incidence rates of ALS generated interest in possible environmental causes, leading to interest in the spatial analysis of ALS incidence [21,25,110,111]. Since then, the methods of spatial analysis have evolved along with advances in technology, the development of new statistical approaches and with the emergence of geospatial information systems (GIS) [152].

A number of different approaches have been taken in spatial analysis of ALS and these can be crudely divided into two groups (although some studies have characteristics of both groups). The first group includes studies that have analyzed area based incidence rates – largely as an exploratory hypothesis-generating tool. The second group of studies include those which have implemented formal cluster

[5] Chapter 3.1 has been published as:

Rooney J, Heverin M, Vajda A, Crampsie A, Tobin K, Byrne S, et al. An Exploratory Spatial Analysis of ALS Incidence in Ireland over 17.5 Years (1995 – July 2013). PLoS One. 2014 May 27;9(5):e96556

analysis – either as an exploratory tool, or alternatively to investigate purported clusters. Findings have been conflicting, partly because many of the reported studies have been based on mortality data [21,22,153–155], which may be subject to incomplete case ascertainment thereby introducing a spatial bias. These data are therefore considered to be less reliable than those derived from prospective data collection, for example via population based register. Only two studies, one from Italy, one from the South-East of England are truly population based and have used a prospective design and case ascertainment, although numbers from the latter study are relatively small [110,111].

Whilst most cluster analysis studies have used one or several spatial scan statistics, few studies have employed spatial smoothing. Just two of the studies used Bayesian spatial smoothing [111,153]. This computationally intensive method, which has become practical with advances in computer hardware and software, has been used extensively to map cancer incidence in Ireland [153]. This method has the advantage of being able to allow for global and local random effects within a conditional auto-regression model [156].

Earlier analysis (1999) of the epidemiology of ALS in Ireland yielded a suggestion of an area of high incidence in the North-West of the country – encompassing the counties of Donegal, Sligo and Leitrim [27]. This finding was significant using crude incidence rates ($P = 0.017$). However, the observation did not meet criteria for statistical significance after standardization ($P = 0.07$).

This study aims to employ modern spatial analysis methods to produce a high-quality map and perform a formal spatial analysis of ALS incidence in Ireland. We have taken a step-by-step approach to employ modern GIS techniques to map ALS in Ireland, focusing in this paper on an exploratory areal visualization of Bayesian smoothed Standardized Incidence Ratios (SIRs) using a population based cohort from the Irish ALS register.

3.1.2. Methods:

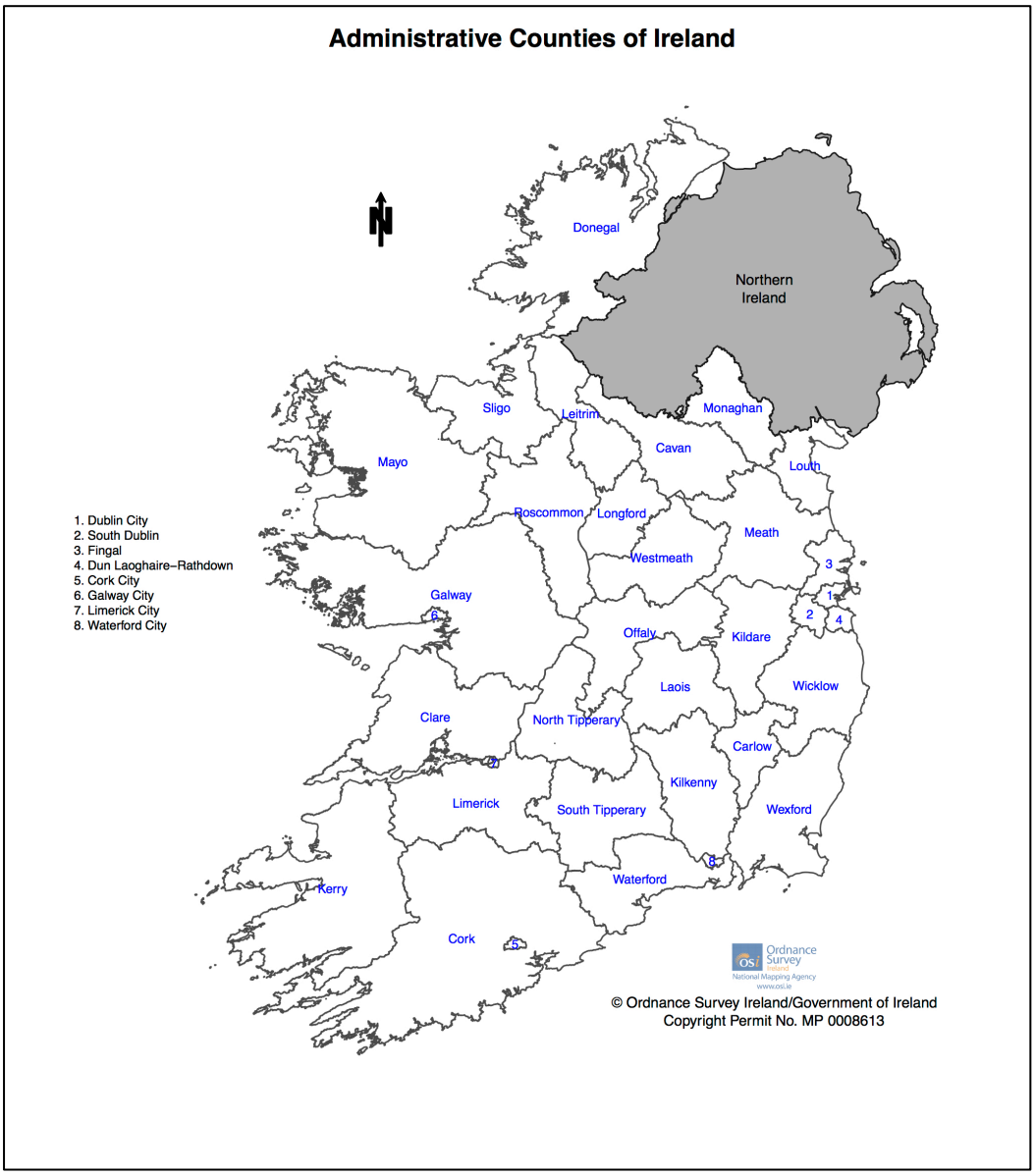
Data Sources:

Ireland is divided into 34 administrative areas – depicted in Figure 3-1. These larger administrative areas, are further subdivided into 3,409 Electoral Divisions (for the 2011 Census). Boundary files for the Electoral Divisions were obtained from the Ordnance Survey of Ireland [157]. Data on the Irish population was obtained from the Irish Central Statistics Office (CSO). Census data by Electoral Division were available online for the years 2002, 2006 and 2011 [158–160]. For the census year 1996, 5-year age group data was only available for larger areas. However, total male and female data was available by electoral division [161] and these were used to apportion the larger area 5-year age group data to electoral divisions on a proportional basis. Average population per Electoral Division by gender per 5 year age group across the four censuses was then calculated.

Patient data was obtained from the Irish ALS register which has been extensively described [16,27]. All eligible cases (diagnosed between 1st January 1995 and 31st July 2013) were geocoded using their address at time of diagnosis, (X-Y co-ordinates of latitude and longitude, World Geodetic System 84). As Ireland is one of the few developed countries in the world that does not use postcodes[6], geocodes were produced manually using the Ordnance Survey Ireland website [162]. Some addresses were more suited to precise geocoding than others - when the address contains a house-name or street number a relatively precise geocode could be generated compared to an address that referred only to a townland. In these cases the central point of the townland was taken as the address. However, even rural addresses can be geocoded to an accuracy of within a few hundred metres of the house, depending on the size of the townland. Each search result was inspected visually to check that the parameters of the address (townland, town, county) were consistent with the search result.

[6] This work was completed before the introduction of Eircode in 2015

Figure 3-1 Administrative Counties of Ireland



Calculation of Standardized Incidence Rates:

Indirect standardization by 5 year age group and sex was employed. Nationwide incidence rates per 5 year age group aged 20 and over, and by gender were calculated. These nationwide incidence rates were then applied to the average population data per Electoral Division to calculate the expected number of cases over the period of the study. For any cases missing age at diagnosis a correction factor was used to distribute each case across all age groups according to the overall age distribution. This ensured that observed and expected case totals were equal. This was also preferable to excluding those cases as that would have led to

large localized errors when calculating ED SIRs – since each ED has only a few cases the exclusion of one case could have a large local effect.

The GPS coordinates for cases were transformed to the Irish Grid coordinate system, overlaid onto the boundary files using R 2.15 statistical software [163] and additional packages for spatial analysis and mapping [156,163–177]. The number of cases per Electoral Division was then determined. Observed cases were then divided by the expected cases based on census data to calculate crude SIRs per electoral division. Note that over various years of the Census, the boundaries of some Electoral Divisions have been redefined, and this necessitated the combining of some Electoral Divisions – this has been previously described by the All Ireland Cancer Atlas and we have followed the same combinations of Electoral Divisions as outlined their report[178]. There were 3,355 EDs after combinations.

Next the *poly2nb()* command of R's *spdep* package [169] was used to determine neighbour relations between Electoral Divisions. Again as outlined in the All Ireland Cancer Atlas, we created artificial neighbour relations between islands, headlands and peninsulas [178]. In addition, we created 2 additional artificial neighbour links (between ED33008 -> ED28030 and between ED33011 -> ED33016). This was because no data were available for Northern Island and as a result Donegal was connected to the rest of the island by only one neighbour link in two separate locations. True neighbours were given a weight of 1.0 whilst artificial neighbours were given a weight of 0.5.

Spatial Smoothing Strategy:

The data were input into OpenBUGS 3.2.2 [179] and Bayesian conditional autoregression was implemented using the Besag-York-Mollié model [111,156,180,181]. This model assumes a Poisson model for spatial distribution of cases, allowing for global and local random effects as follows:

$$O_i \sim Poisson(E_i\theta_i)$$
$$\log(\theta_i) = \alpha + u_i + v_i + (\beta * d_i)$$

where:

O_i = observed cases in area i

E_i = expected cases in area i

θ_i = estimated Relative Risk for area i

α = regression intercept term

u_i = global random effect

v_i = spatially structured local random effect

$(\beta * d_i)$ = optional covariate term where β is the regression coefficient and d_i is the population density for area i

The model was compiled in OpenBugs and data on observed and expected cases loaded. After initializing, OpenBugs was given a burn-in run of 1,000,000 iterations before a further production run of 1,000,000 iterations across two chains - which were thinned by 100 to yield 20,000 estimates for the Relative Risk (RR) of each ED. Convergence of the chains was assessed by graphical inspection of CODA output.

Post smoothing analysis

After Bayesian smoothing, R was used to plot graphs of smoothed RR's for the entire cohort, and sub-cohorts of particular interest. Due to the skewed distribution of population density across EDs, the mean RR across EDs was 0.96. Therefore, dividing by the mean was used to centre the RR's. To further examine the relationship between population density and RR, log-normalisation of the

census-averaged population density per ED was attempted and uni-variate linear regression of RR's versus the natural log of population density was performed for total cases and also for stratified male and female cases. The linear regression was performed separately for both smoothed RR, and smoothed RR weighted by expected cases per electoral division.

3.1.3. Results:

The total number of cases meeting the inclusion criteria was 1,645. Of those only 7 could not be geocoded (due to incomplete addresses) leaving 1,638 geocoded cases. Table 3-1 shows the clinical characteristics of the cohort. Only 61 cases had no known age at diagnosis (due either to a missing birthdate or a missing diagnosis date). The population density per ED, after averaging population across census years, is shown in Figure 3-2.

Table 3-1 Basic clinical characteristics of Irish ALS patients from Jan 1995 to July 2013 (N = 1,645)

Variable	Strata	N (%)	Missing (%)
Geocoded ?		1638 (99.6 %)	7 (0.4 %)
Sex	Male	934 (56.8 %)	
	Female	711 (43.2%)	0
Escorial	Definite	876 (57.9%)	
	Probable	436 (28.8%)	
	Possible	183 (12.1%)	
	Suspected	17 (1.1%)	133 (8%)
Site Of Onset	Limb	923 (59.1%)	
	Other	638 (40.9%)	84 (5.1%)
Age at diagnosis	Median (range)	66.3 (21 - 94)	61 (3.7%)

Figure 3-2 Average Population Distribution across 4 census years

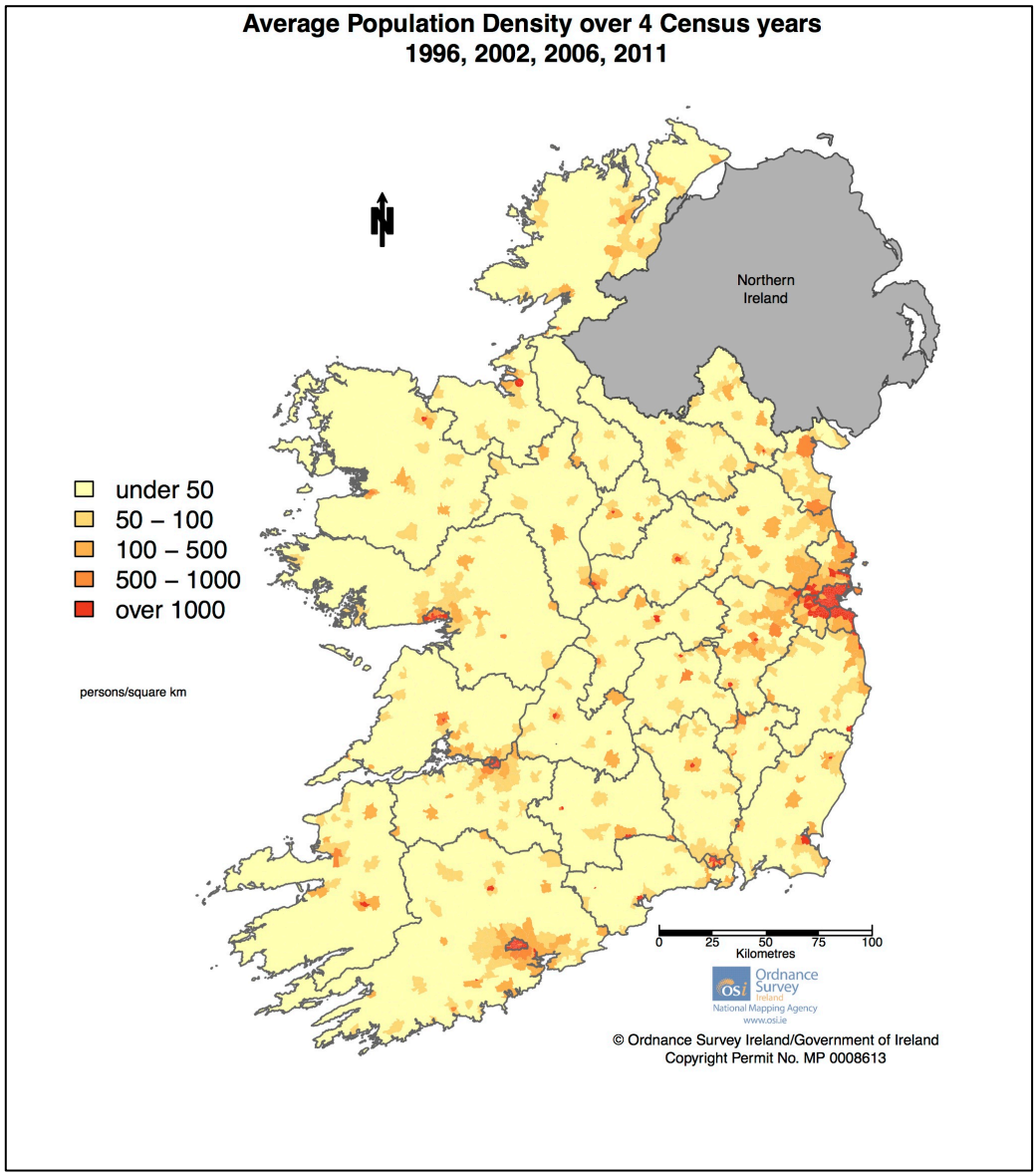


Figure 3-3 displays a histogram of total population per ED after census averages were calculated. As can be seen the distribution is highly asymmetric with a very large number of EDs having a population less than 2,000, and a small number of EDs with population into the tens of thousands. Figure 3-4 displays crude SIRs for all cases of ALS in the Republic of Ireland between January 1995 and July 2013. Due to random variation it is difficult to discern any overall geographical pattern in the map.

Figure 3-3 Distribution of Population per Electoral Division

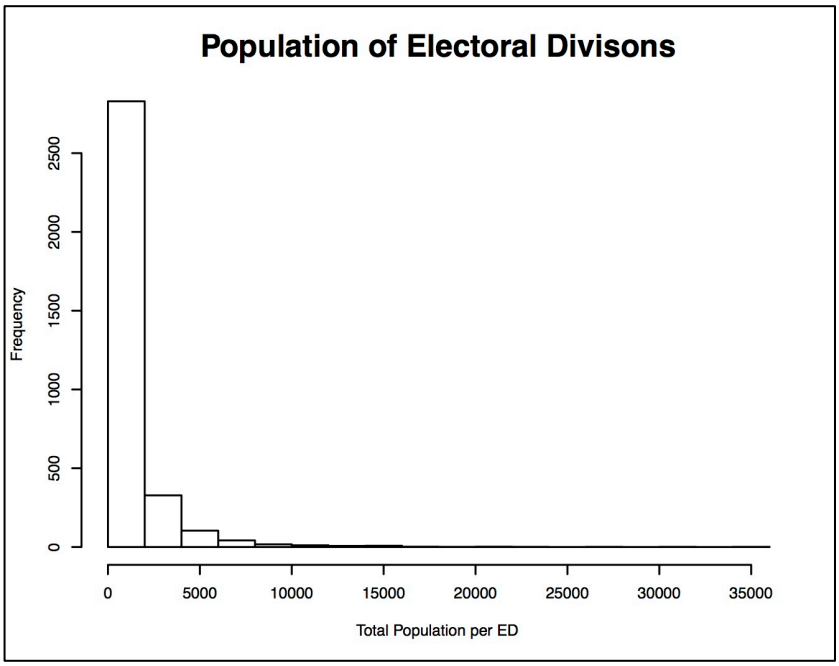
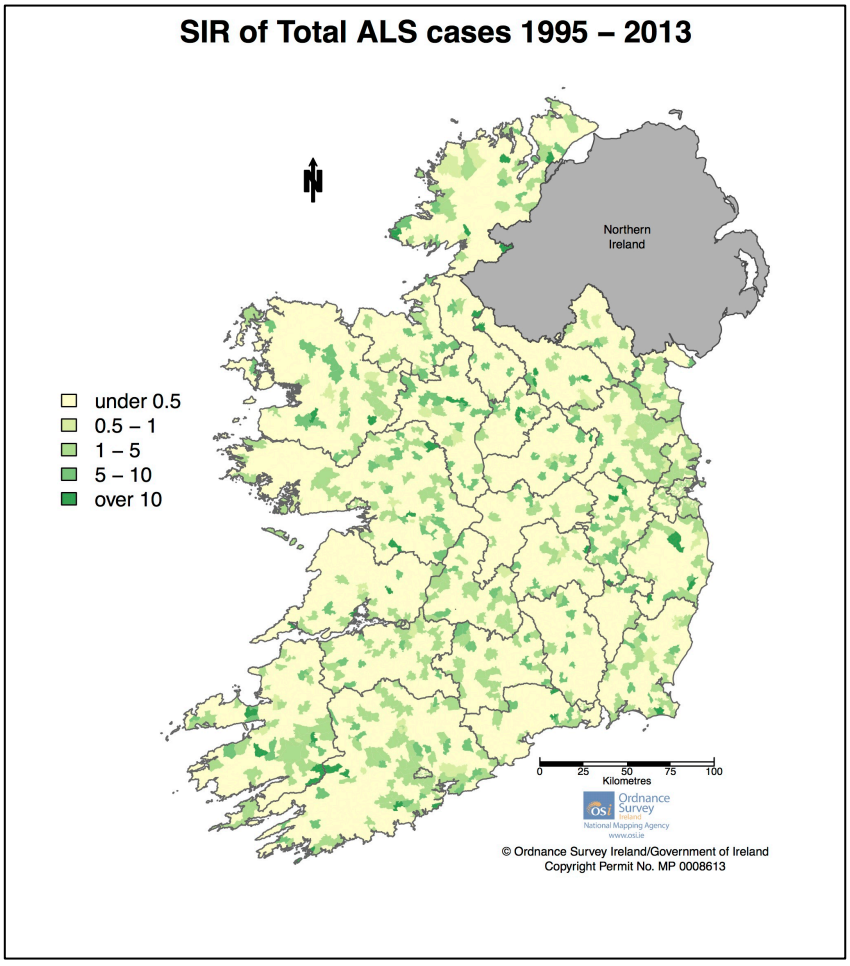
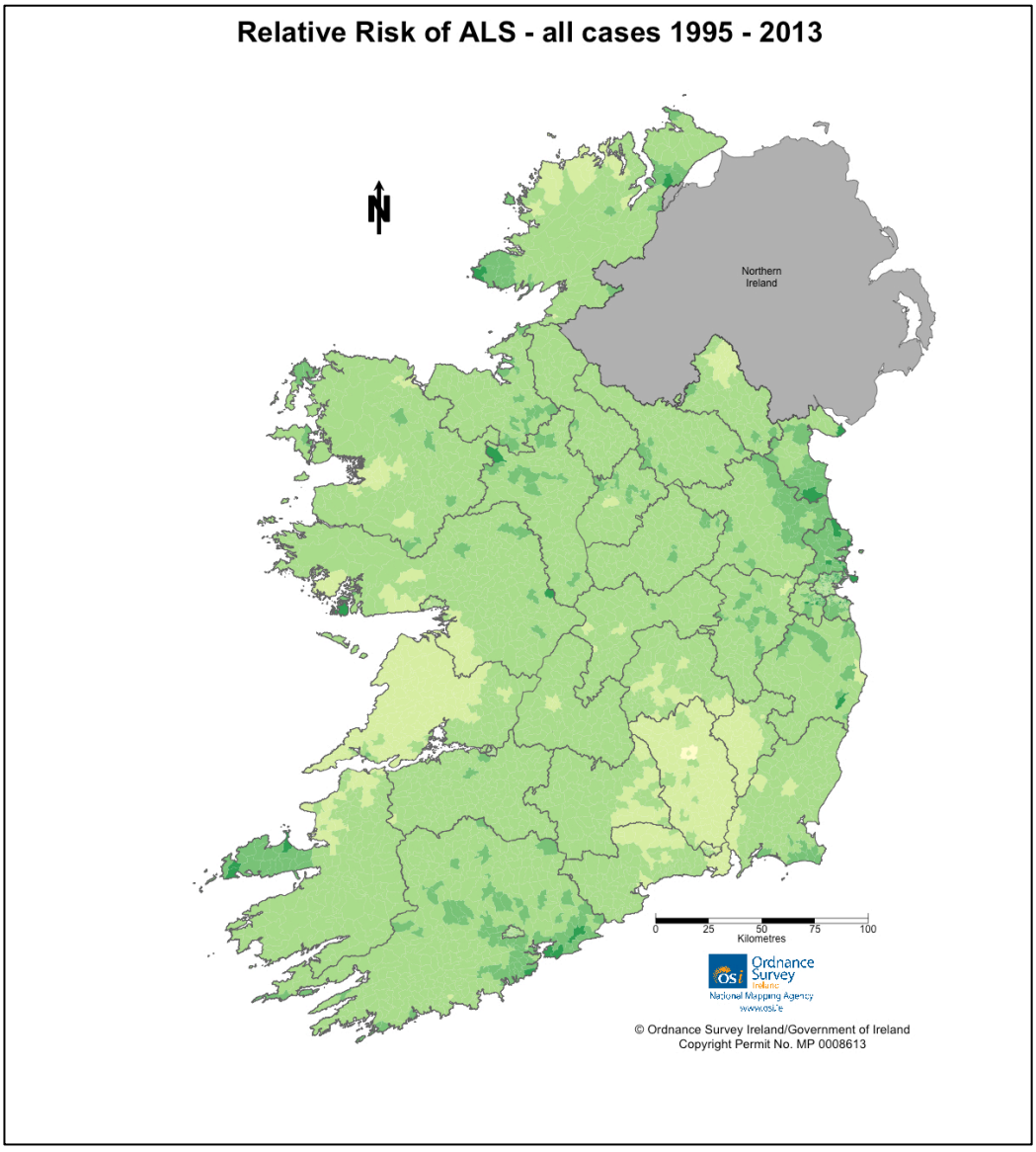


Figure 3-4 Crude (unsmoothed) standardized incidence ratio (SIR) per ED for all cases of ALS in Ireland 1995 – July 2013



Initial Bayesian smoothing of all cases including total population density revealed a coefficient for population density that did not differ significantly from zero (mean: -1.60×10^{-5}; 95% credible interval: -4.73×10^{-5} to 1.36×10^{-5}). As this result indicated that population density was not associated with ALS risk, this was therefore omitted from further models. Figure 3-5 displays smoothed relative risks for all cases of ALS after Bayesian smoothing without including a term for population density.

Figure 3-5 Smoothed Relative Risk for all cases of ALS in Ireland from 1995 to July 2013



As can be seen in Figure 3-5 there is a variation in RR across the country with RRs ranging from 0.69 – 1.58. Several areas appeared to have slightly higher RR – the North-East coast (Counties Louth, Meath and North County Dublin), Cork city, the Dingle peninsula in Kerry, and the Western part of Donegal. County Kilkenny appears to have lower RR, as does the Western part of County Clare.

Figure 3-6 displays smoothed relative risks separately for male and female cases. Both graphs show only mild spatial variation in general. However, there are some localized areas of interest. For females, parts of Kerry, Cork and Louth/Meath show higher RRs, whilst for males, the Dingle peninsula and the Western-most section of Donegal show higher RR.

Figure 3-7 displays smoothed relative risks separately for those aged under 55 and those over 55 at diagnosis (early onset ALS is more likely to be of genetic origin [135]). Those diagnosed under the age of 55 (n = 349) show a pattern of raised RR in the Cork/Kerry area, whereas those diagnosed over the age of 55 (n = 1,289) shows a raised RR in the Louth/Meath/Dublin area.

Figure 3-6 Smoothed Relative Risks for male and female Irish ALS cases 1995 – July 2013

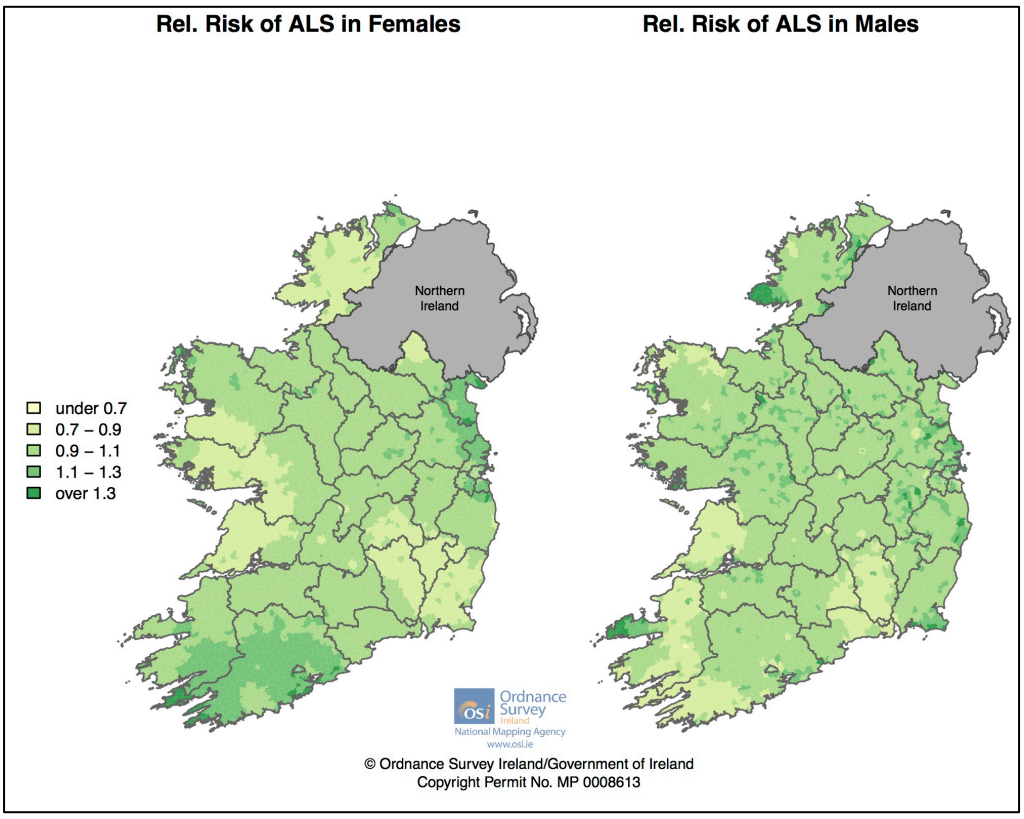
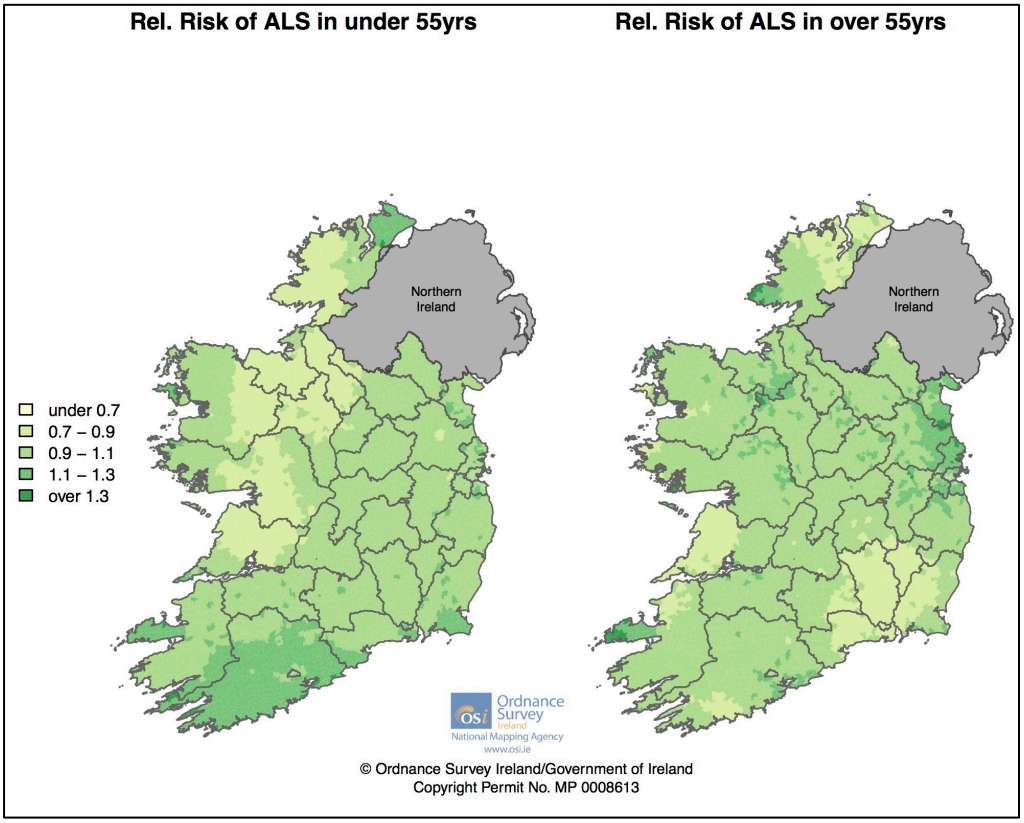


Figure 3-7 Smoothed Relative Risks for Irish ALS cases aged under 55 and over 55 at diagnosis between 1995 – July 2013



Graphs of RR weighted by expected cases for all cases, males and females versus natural log of population density are shown in Figure 3-8. As can be seen there is a statistically significant relationship between the smoothed weighted RR of all cases and ln(pop. density). For the un-weighted regressions, we found a statistically significant relationship between RR and ln(pop. density) with R^2 of 0.10, 0.11 & 0.04 for all cases, females and males respectively.

Figure 3-8 Linear Regression of post Bayesian smoothing RR x expected cases for all cases, males and females versus ln(population density)

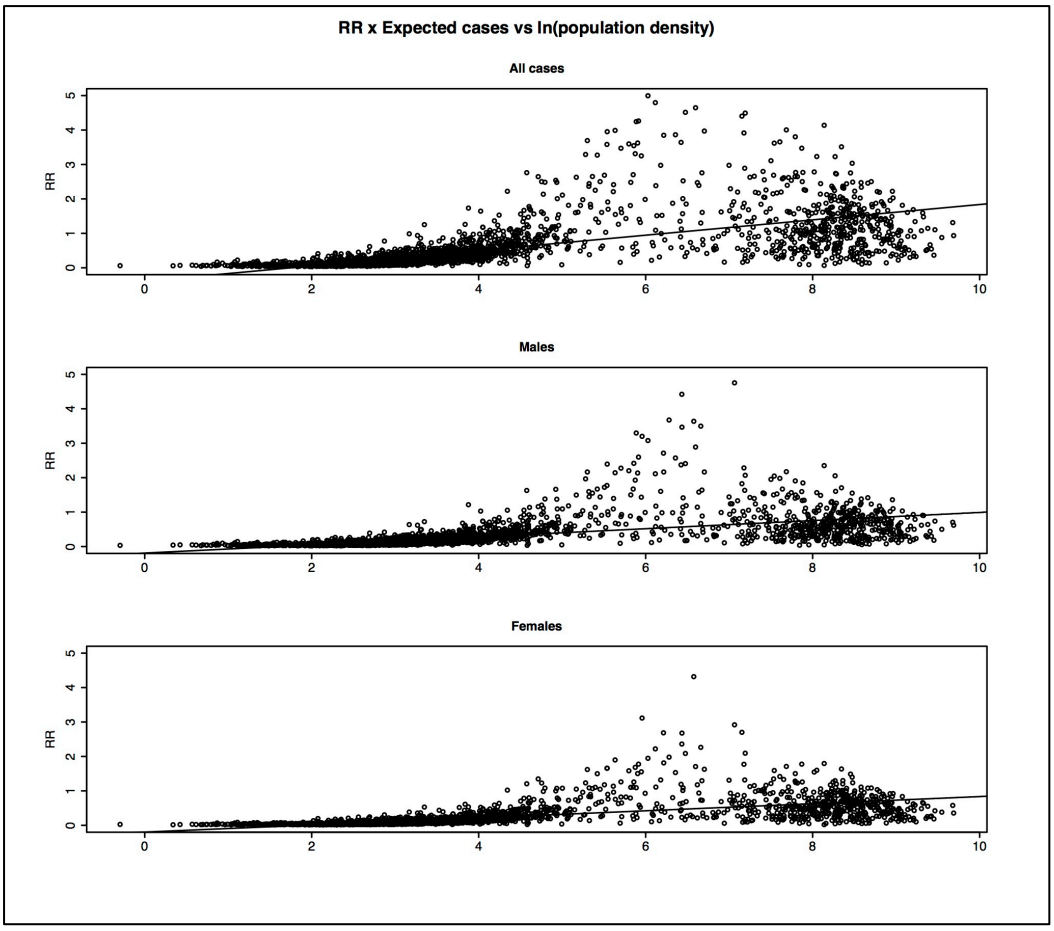


Figure 3-8 **Legend:**

Weighted regression coefficients:

All cases: Coeff. = 0.22; 95%CI(0.211, 0.234); $P < 2 \times 10^{-16}$; $R^2 = 0.32$

Males only: Coeff. = 0.12; 95%CI(0.112, 0.124); $P < 2 \times 10^{-16}$; $R^2 = 0.31$

Females only: Coeff. = 0.10; 95%CI(0.098, 0.108); $P < 2 \times 10^{-16}$; $R^2 = 0.34$

3.1.4. Discussion

To the best of our knowledge, this is the first areal mapping of ALS incidence based upon a national prospective population based ALS register. The overall pattern of RRs for all cases displays mild variation throughout the country. No nationwide North-South pattern (as reported by some earlier studies [153,154,182]) could be discerned (Figure 3-5). Areas of moderately higher risk appear in North Dublin, Louth & Meath, Cork, the Dingle peninsula in Kerry, and the area on the south-west coast of Donegal. Of these areas, North Dublin, Louth, Meath and Cork are relatively developed and of higher population density (Figure 3-2), whilst the area in Kerry and Donegal includes primarily rural areas and small traditional fishing with populations that have experienced little inward immigration. These factors may suggest the possibility of local genetic founder effects in some regions – an observation supported by the increased RR for those diagnosed aged under 55 years in the Cork/Kerry area.

Our findings contrast those of other groups. A 2009 Swedish study utilizing a hospital inpatient register with partial patient histories identified 3,390 incident cases of ALS between 1991 and 2005 [182]. Subsequent analysis of areal standardized incidence rates suggested an increasing South to North risk gradient – although a test for linear trend was not significant [182]. Similarly, a geographical and temporal analysis of ALS mortality between 1990 – 2005 in Spain identified 9,475 deaths –with higher rate in the North [153]. This finding may reflect the population genetics of Spain with higher degree of admixture in the South, as a previous population based mortality study in Cuba suggests that admixed populations exhibit a lower rate of ALS [137].

A study of age adjusted ALS mortality rates from 1969 – 1998 that divided the US into 12 geographical areas identified 105,318 deaths due to ALS according to death certificates [154]. This study also found a broad geographic pattern with declining rates in a North-Western to South-Eastern distribution, both in the total population and also in non-Hispanic whites [154]. As the population substructure in the US differs with respect to ethnicity, this distribution may reflect subtle differences in genetic risk.

Results from Bayesian spatial auto-regression vs post-Bayesian linear regression analysis of the association of population density and RR were conflicting. This discrepancy may be routed in the asymmetrical population (Figure 3-3) and population density distributions across electoral divisions (Figure 3-2), combined with different mathematical relationships employed by the two approaches: i.e. the Bayesian approach specifies the natural log of RR as a function of population density (and other variables), whereas in the linear regression we examined the relationship between RR and ln(population density) – a more symmetrical distribution. This demonstrates the mathematical problems that can be caused in areal spatial analysis by an asymmetric population distribution across areas.

Notwithstanding, our finding of a relationship between rates of ALS and population density is in keeping with 2009 findings from a register based study in the South-East of England which used Poisson spatial regression and SaTScan to find a cluster of cases in greater London [110]. The findings contrast with a 2013 Italian register based analysis performed using data from the Piedmont and Aosta Valley Register for ALS [111]. The study was methodologically similar, and identified 1,216 patients diagnosed between 1995 and 2004 before using Bayesian smoothing to create maps of RR's for subgroups of the cohort. The authors found rural areas of low population density and high ALS risk. They theorized that areas of higher incidence post smoothing might be related to the use of agricultural chemicals or clusters of familial disease in rural and mountainous areas prone to genetic isolation [111] – factors which may subtly differ between countries due to differences in geography and agricultural practices. Converse to the Italian findings, our finding of increased risk in higher population density areas points towards urban environmental effects or possibly urban socio-economic factors as risk factors.

The strengths of this study are that it uses a national prospective ALS register as the case-source. It also has a very high level of data integrity with complete data on gender, just 7 cases (0.4%) lacking geocodes and 61 (3.7%) lacking a precise age at diagnosis – therefore bias due to missing data is unlikely. We have used high quality census and spatial data obtained from the CSO and the OSI, and we have

implemented a robust smoothing strategy which has previously been implemented in Irish cancer epidemiology [178].

There are, however, some weaknesses to the study – primarily the asymmetric distribution of population amongst Electoral Divisions. We have also performed the analysis inclusive of known familial/genetic cases of ALS that may have led to higher SIRs & RRs in certain areas. Finally, we have not included a formal cluster analysis at this point – although this was not our aim in this phase of the study.

In conclusion, our spatial mapping of population based incident ALS cases from 1995 to 2013 in Ireland revealed several localized areas of higher relative risk for ALS. While we found conflicting results after using two different approaches to determine if relative risk was related to population density, an approach of linear regression of post Bayesian risk estimates found evidence of a weak relationship, suggesting the possibility of increased risk in urban rather than rural areas. Further analysis of the Irish ALS population is currently underway, including detailed cluster analysis and genetic epidemiologic mapping.

3.2. Spatial cluster analysis of population ALS risk in Ireland[7]

3.2.1. Introduction

Recently we performed Bayesian risk mapping of prospective ALS cases in Ireland using nationwide population data from the Irish ALS register including all cases in Ireland from 1st January 1995 to 31st July 2013 (Chapter 3.1). Our results indicated localised areas of increased risk, some in areas of high population density and others in more remote coastal areas. However, whilst Bayesian smoothing approaches to mapping are excellent for visualizing the spatial distribution of risk, they do not allow the hypothesis testing that is needed to identify clusters. Therefore, we have now performed a follow-up formal cluster analysis of the same Irish ALS cohort, and describe a novel method to perform a spatio-temporal audit of case ascertainment of the Irish ALS register.

3.2.2. Methods:

Data Sources and Data preparation

Our inclusion and exclusion criteria, and age and gender standardisation calculations were as previously reported in Chapter 3.1. However this analysis includes cases from 1st January 1995 to 31st December 2013. Briefly, all incident cases from those time periods were included and geocoded for address at the time of diagnosis. Boundary files for 3,409 Electoral Divisions (EDs) in Ireland were obtained from Ordnance Survey Ireland [157] and combined with Census data from 4 census years (1996, 2002, 2006, 2011) [158–161]. This data was then used to produce an average population per ED for the overall study period grouped by sex and 5 year age group. After necessary joins of certain neighbouring EDs due to census changes over the years, 3,355 EDs (Chapter 3.1) were included in the final

[7] Chapter 3.2 has been published as:

Rooney J, Vajda A, Heverin M, Elamin M, Crampsie A, McLaughlin R, et al. Spatial cluster analysis of population amyotrophic lateral sclerosis risk in Ireland. Neurology. 2015;84(15):1537–44

analysis. The number of observed cases per ED was then determined. Indirect standardisation using the population data was used to calculate age and sex standardised expected cases per ED (Chapter 3.1).

Cluster Analysis Methods

SaTScan [183] is a well-established cluster analysis tool that allows for a variety of spatio-temporal cluster analyses based on various probability models, and can identify areas of high risk, low risk or both [183,184]. However it is limited to detecting circular and elliptical clusters [183–185]. This has led to the development of an alternative cluster analysis program known as FleXScan [185–187]. FleXScan, although limited to detecting areas of high risk and not low risk, has the advantage of being able to identify clusters of irregular shapes [185–187]. Both SaTScan and FleXScan have been previously used to investigate ALS clustering [22,24,110,155]. As our previous Bayesian smoothing analysis found areas of localised increased risk on the highly irregular Irish western coastline, and areas of low risk in the south and southwest (Chapter 3.1), we applied both SaTScan and FleXScan to our dataset.

The prepared data were exported to ASCII files for use with SaTScan and FleXScan. SaTScan was run using a discrete scan statistic based on a Poisson probability model, scanning for a cluster of statistically high risk, with a maximal cluster size of 50% of the population (recommended default in SaTScan user guide [184]). In addition to scanning for circular clusters, SaTScan also searched for ellipsoid shaped clusters using a medium non-compactness penalty. Finally, SaTScan was rerun to identify statistically significant areas of low risk, with all other settings remaining the same. FlexScan was also run using a Poisson probability model searching for clusters of flexible size with ED neighbour relations supplied via ASCII file. It was run with the restricted likelihood ratio statistic enabled, to allow detection of clusters larger than 30 with the restriction parameter (α) set 0.1, 0.2 & 0.3 used in successive runs to improve detection of small area high risk, medium area moderate risk, & large area low risk clusters respectively [185]. A maximal spatial cluster size of 1000 EDs was used for all FlexScan runs.

Spatio-temporal audit of case ascertainment

Clusters, particularly areas of significant low risk, could be caused by localised variation in case ascertainment. Since the register is long running (almost 20 years), we used a number of graphical and statistical approaches to audit case ascertainment of the register. This approach assumes that any deficit in regional case ascertainment is unlikely to have remained constant in given regions over extended time periods.

Time period spatial smoothing

We applied Bayesian smoothing to cases identified during specific time periods only – namely 1995 – 1999 (n = 338), 2000 – 2004 (n = 411), 2005 – 2009 (n = 441), 2010 – 2013 (n = 433). We then plotted maps of the spatially smoothed risks for each time period, and graphically assessed ALS risk in each map searching for any areas of decreased incidence in each time period. Integrated Nested Laplace Approximation (INLA) was used to implement Bayesian smoothing via the R-INLA package [188,189].

Administrative area statistics

We determined incidence rates for the 34 administrative areas in our dataset (shown in Figure 3-1). For each administrative area we calculated observed and expected case numbers based on an average population from four different censuses. We then applied a Poisson test to determine if any of the observed case counts were significantly different from expected cases in each area. Furthermore, we applied the same technique to selected periods of the overall time that the register has been in operation in order to determine if there were temporary local failures in case ascertainment. Due to the use of multiple tests a Bonferroni correction was applied to determine a threshold of statistical significance of $P < 0.0015$. (Note – both cluster analysis programs already incorporate correction for multiple comparisons, therefore a Bonferroni correction was not applied to cluster analysis results).

Software

R 3.1 statistical software [190] was used to perform data preparation, routine statistics, and ALS risk mapping as described previously (Chapter 3.1). R was also used to prepare ASCII files for SaTScan v9.3 [183,184] and FleXScan v3.1.2 [185–187].

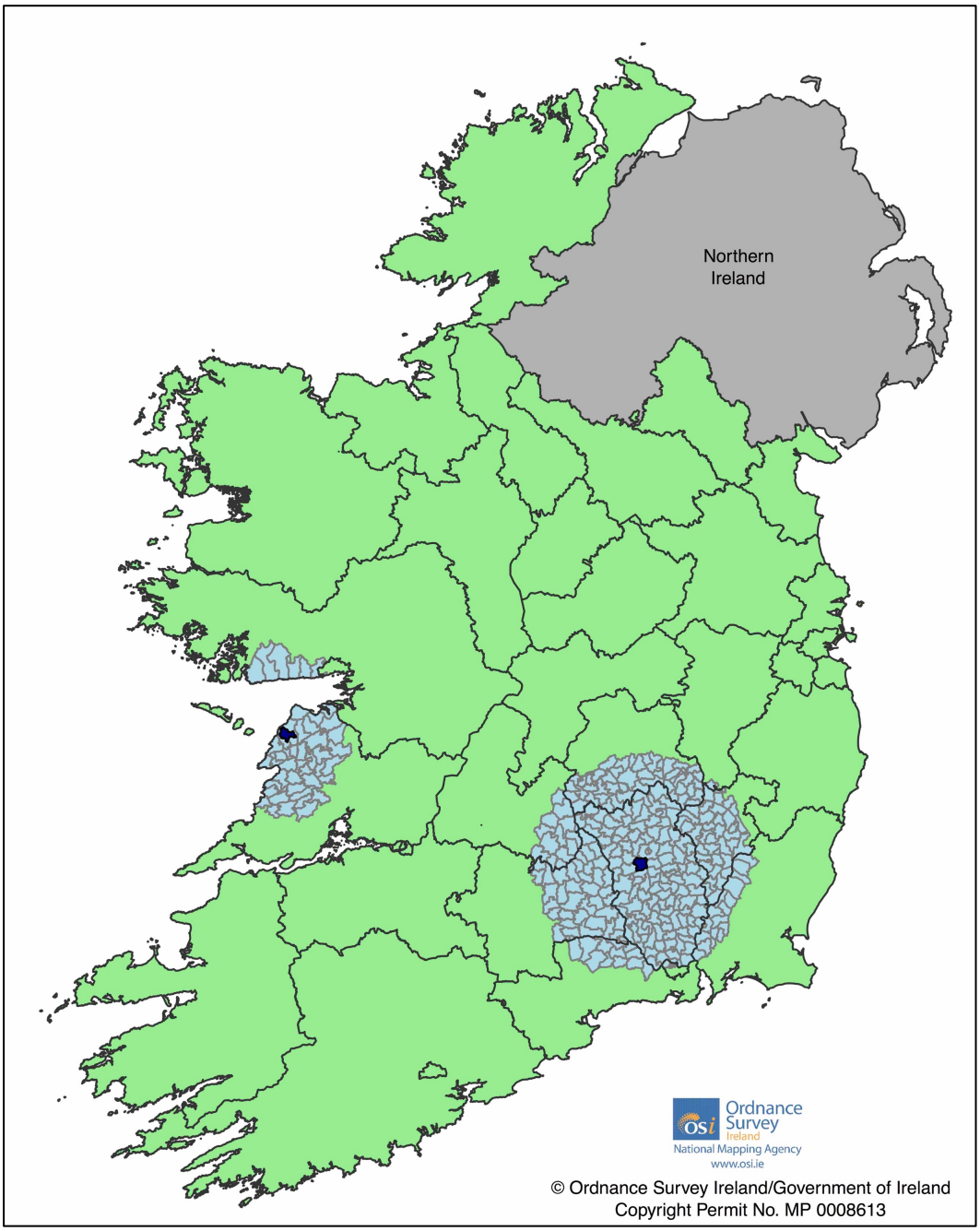
3.2.3. Results:

The total number of cases meeting the inclusion criteria was 1,684. The population density of electoral divisions and the geographic spread of population have been described in Chapter 3.1.

SaTScan

SaTScan did not identify any statistically significant circular or ellipsoid clusters of higher relative risk across any of the parameters used. However, SaTScan highlighted two low risk clusters of statistical significance. The larger area was located in the south of the country, corresponding to the counties of Carlow and Kilkenny. This cluster encompassed an area of 302 EDs with a total population of 240,710. Only 57 cases were observed where 105 were expected and SaTScan calculated a relative risk of 0.53 for this cluster with a P-value = 0.016. A smaller low risk region was located in the west and south-west, corresponding to the counties of Clare and Galway, and encompasses 49 EDs, with a total population of 27,409. No cases were detected where 13 were expected (relative risk = 0, P = 0.028). The location of both clusters is shown in Figure 3-9. FleXScan failed to find any statistically significant flexible shaped clusters of higher relative risk across varied alpha settings. A number of potential clusters were identified, but the lowest P value was = 0.588 indicating that these regions represented random background variation and did not represent true high-risk clusters.

Figure 3-9 Location of significant low risk area of ALS risk in Ireland

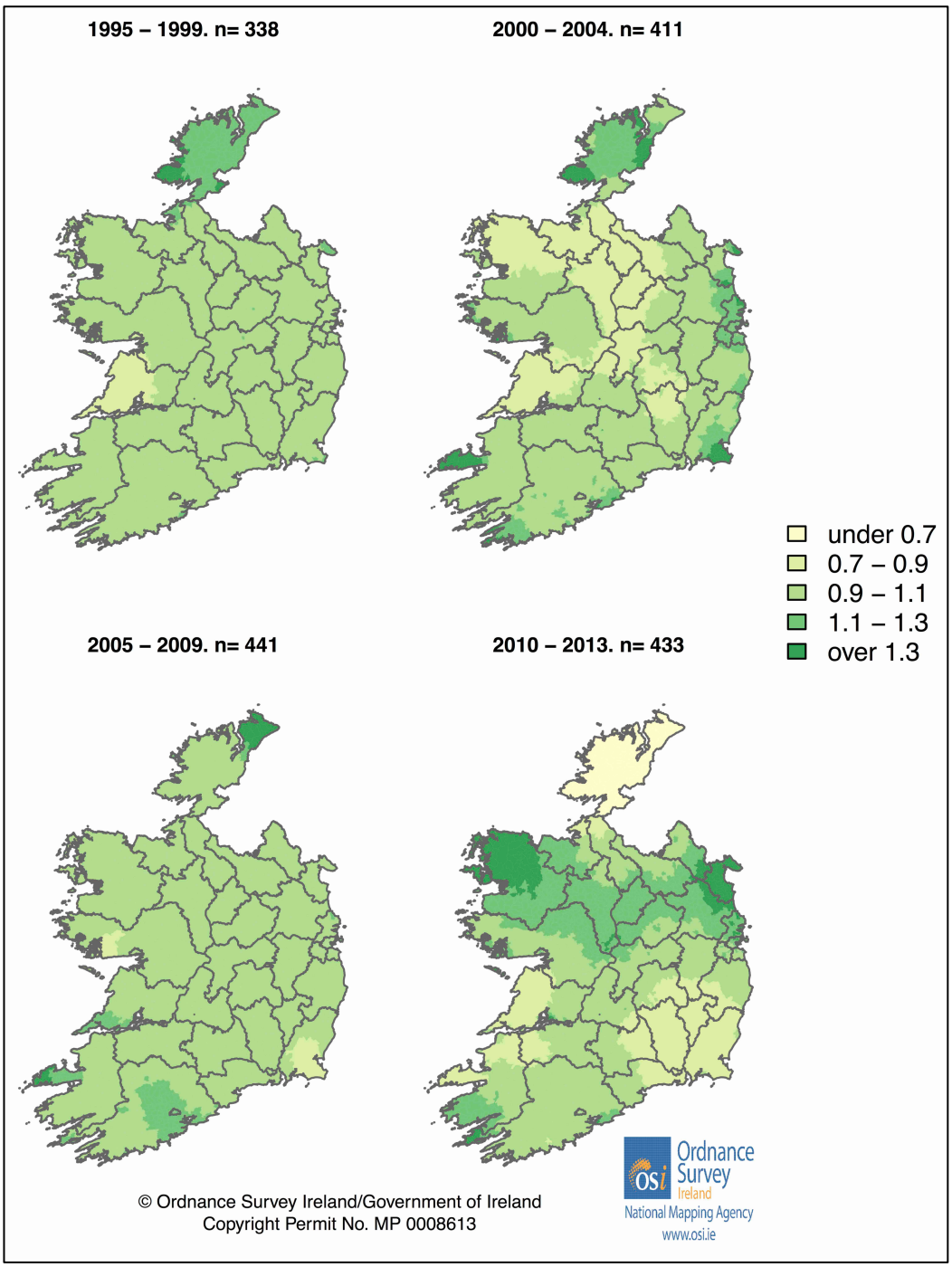


Audit of case ascertainment

Bayesian smoothed risk of ALS per ED for incident cases diagnosed during the given time periods are shown in Figure 3-10. There are different patterns between each time period. The latter period Jan 2010 to December 2013 shows a degree of variation with lower risk of ALS depicted in the entirety of Donegal county (< 0.7), and a band of higher risk across the middle of the country. The low risk area at Carlow / Kilkenny, identified via cluster analysis, shows average risk in two out of

four time periods and mildly reduced risk (RR 0.7 – 0.9) in other time periods, while the low risk area of Clare / Galway shows mildly reduced risk (RR 0.7 – 0.9) in three time periods.

Figure 3-10 Bayesian smoothed RR of ALS in Ireland by electoral division stratified by time period



A summary of the administrative area Poisson tests of observed versus expected cases is shown in Appendix A, Tables A1 & A2. Two tests met the threshold for overall significance. The test for all cases from 1995 – December 2013 in Kilkenny County (observed 18; expected 37), and the test for all cases between January 2010 and December 2013 for Donegal County (observed 4, expected 17). The findings from Donegal are explained by the relatively short follow-up time from 2013 to our data analysis (May 2014). Indeed, a further 5 cases have presented from County Donegal in the intervening time (to September 2014), and we have learned of 2 separate patients who had opted for treatment in Northern Ireland. This increases the observed cases during this time period to 7 (excluding those under treatment in Northern Ireland) – and we recalculated the Poisson statistic to obtain a new P = 0.006 - which no longer meets the significance threshold of P < 0.0015.

3.2.4. Discussion

The absence of significant high risk clusters in the Republic of Ireland despite using both SaTScan and FleXScan contrasts with our expectations, given our previous Bayesian analysis, which had shown several areas of higher risk in West Donegal, the Dingle peninsula, Cork and the East coast (Chapter 3.1). However the low risk areas in our Bayesian study are more robust than any of the small high risk areas that tend to draw focus, and in fact correspond with the two areas identified by SaTScan as significantly low risk areas. Thus, we view the two approaches as complementary in nature, with parallel Bayesian mapping and formal cluster analysis offering advantages over either method alone.

These results also contrast with previous findings of significant high-risk clusters in other countries, summarized in Table 3-2. A 2003 Finnish study included 1,000 cases identified from death certificates, aggregated them by municipalities, and used the mathematical tool SaTScan [183] to identify two clusters of relative risk (RR) 1.79 (P < 0.0001) and 1.32 (P = 0.013) respectively, which the authors attributed to one or more gene variants that were unknown at the time [21]. Similarly, analysis of national mortality rates in Italy using SaTScan [183] identified

16 clusters of high risk (Table 3-2) spread across the country for which no cause was identified [22]. A smaller 2009 study from the South East of England used high quality prospective register data and SaTScan [183] found a significant cluster in Greater London (RR 1.7, P = 0.012), as well as smaller clusters in urban areas outside London [110]. Three significant clusters amongst 12,173 deaths from ALS in Japan [155] were identified using a flexible spatial scan statistic. A 2011 study of the Limousin region in France used multiple techniques to identify three clusters (Table 3-2) and determined that out of 50 environmental factors, paper paste and water treatment plants had proximity to the cluster areas [23].

A large number of studies have used cluster analysis methods to test specific hypotheses. SaTScan [183] was used to investigate a purported cluster surrounding a lead smelter in Jefferson County, Missouri [24]. Thirty-six patients were identified from hospital records, primary care providers and other sources and a small cluster (n = 3) of borderline significance (RR 6.4, P = 0.04) was identified by the authors. However, case ascertainment included only the Jefferson County in which the smelter was located [24].

β-methylamino-L-alanine (BMAA), a neurotoxic protein produced by cyanobacteria, has been linked to increased rates of amyotrophic lateral sclerosis/Parkinsonism-dementia complex (ALS/PDC) amongst the population of Guam [25], and in recent years a series of cluster analysis studies have attempted to detect ALS clusters in proximity to naturally occurring cyanobacteria blooms in lakes prone to such blooms. Of 278 cases identified in New Hampshire from Hospital records and 'community databases', 9 cases that occurred with higher than expected frequency in close proximity to such a lake were labeled a cluster, and the odds ratio of having lived within 0.5 miles of a lake with current or past cyanobacterial blooms was calculated at 2.32 (95% CI 1.42–3.80)[25]. An expansion of this study to three New England states used the Anselin Local Moran's I test identified 11 significant clusters (Table 3-2) amongst 688 cases and the authors postulated a role for cyanobacteria blooms and methyl-mercury contamination, although they also noted that they lacked complete case ascertainment [26]. Clustering of 381 patients diagnosed between 1994 & 2009 at

the centre in Montpellier, Hérault district, France was analysed using SaTScan [183] and a significant cluster was identified in close proximity to a lagoon prone to cyanobacteria blooms (RR 2.24, P = 0.0024) [191]. High BMAA levels were identified in locally caught shellfish although a direct link could not be proven [191].

With the exception of the London group, these studies have not used prospective population based ALS register data [21–26,155,191], and many of them have been limited to regional areas [23–26,110,191], which may exaggerate the significance of clusters [131]. During the course of our analysis, we identified two patients from the Northwest of Ireland who had opted for treatment across the border in Northern Ireland – illustrating one type of 'border problem' phenomenon that can occur. Our finding of significant low risk clusters has not been previously reported for ALS. While it is possible that the reduced frequency in these areas is a function of lower case ascertainment, we have implemented a novel approach to audit case ascertainment, based on the assumption that regional ascertainment pathways are unlikely to have remained constant over the 20 years lifetime of the register. Moreover, during this period the numbers of neurologists in Ireland trebled, with no change in case ascertainment. Furthermore, the Carlow / Kilkenny area did not have significant results on Poisson testing of county estimates stratified by time periods, and only showed mildly reduced risk on Bayesian smoothing of two out of four time strata. County Clare also did not have significant results on Poisson testing of county estimates stratified by time periods, and had mildly reduced risk estimates on Bayesian smoothing of three time periods.

The Carlow / Kilkenny region of significantly low risk roughly aligns with the Nore and Barrow River valleys. These rivers were historically significant trading routes from earliest times and contain several historic towns as well as Kilkenny City. This region has a complex settlement history. There is archaeological evidence of prehistoric settlement, the area roughly aligns with the historic kingdom of Ossory, Waterford at the mouth of the river system was initially established as a Viking centre and the area was also settled by the Anglo-Normans in the 12th century forming part of the Earldom of Ormond. It is therefore possible that the

population genetic admixture of this region differs from surrounding areas. The low risk region of Clare / Galway on the other hand centres on the unique karst landscape of the Burren. This environmentally distinct region had a much more sheltered settlement history and lower population densities until the mid-nineteenth century. Interestingly, however, this low risk area does loosely coincide with an area that became home to Roman Catholics forcibly transplanted from other areas of Munster and South Leinster by Cromwell [192]. It is possible that some population movement from the Carlow / Kilkenny area to this region could have resulted in similar genetic admixture. However, given the alignment of this region with the karst landscape it is more likely that some complex environmental factors are at play which will require further examination.

Notably, our study is of the exploratory design using nationwide prospective register based data, and has the advantage that it covers an entire national healthcare system where cross border treatment is uncommon due to public and private system entitlements and limitations. This is a particular advantage considering many previous ALS cluster analysis studies have sought clusters in limited regions defined as potentially containing a cluster (Table 3-2), an approach known to exaggerate the likelihood of identifying significant clusters [131]. Moreover, in our study, case ascertainment includes the majority of the physical island of Ireland. We have also demonstrated that failure of case ascertainment is an unlikely explanation for our finding of significant low risk areas. Despite these advantages, the failure to find significant high-risk clusters may be a function of lack of power to detect clusters in low population areas, despite our large cohort size.

We have used a novel methodology to detect clustering of ALS using the Irish Population based register, which has been in operation for over 18 years. We have not identified any significant ALS clusters. Conversely, we have identified 2 regions of consistently low incidence, and, as there are no obvious environmental similarities to explain both of these low risk areas, we postulate that they may reflect either differences in the population substructure in these areas or complex

environmental factors unique to each area. Further work is required to confirm this hypothesis.

Table 3-2 Summary of selected ALS cluster analysis studies

Year [Ref.]	Country / design	No. of spatial units	No. of ALS cases	No. of clusters found	Obs. cases	Exp. cases	Relative Risk	P value
2003 [21]	Finland Ret. (M) Exp.	452	1000	3¶	120	66.99	1.79	0.00001
					229	174.13	1.32	0.013
					227	148.04	1.53	0.00001
2007 [22]	Italy Ret. (M) Exp.	8,099	n/a	16	41	0.65	63.03	0.001
					35	1.12	31.24	0.001
					47	3.17	14.85	0.001
					78	14.79	5.30	0.001
					35	2.90	12.09	0.001
					11	0.36	30.23	0.001
					10	0.35	28.20	0.001
					9	0.33	27.67	0.001
					19	2.79	6.81	0.001
					38	11.68	3.25	0.001
					12	1.22	9.85	0.001
					62	26.95	2.31	0.001
					149	91.82	1.63	0.003
					17	3.32	5.13	0.003
					54	24.12	2.24	0.003
					17	4.06	4.19	0.025
2008 [24]	United States Pro. Conf.	n/a	36	1	3	0.47	6.4	0.044
2009 [25] **	United States Ret. Conf.	n/a	278	1	9	n/a	n/a	n/a
2009 [110]	England Pro. (Reg) Exp.	420	472	1	97	n/a	1.7	0.012
				4	n/a	n/a	n/a	> 2.5 SDs from mean
2010 [155]	Japan Ret. (M) Exp.	359	12,173	3	181 (Males)	115.7 (Males)	1.56	<0.00001
					254 (Males)	182.16 (Males)	1.39	0.03
					178 (Females)	118.67 (Females)	1.50	0.02
2011 [23]	France Pro. Exp.	747	199	3	10	2.55	3.93 (95% CI:	n/a

							1.88 - 7.22	
					9	2.30	3.91 (95% CI: 1.79 – 7.43)	n/a
					6	1.24	4.84 (95% CI: 1.77 – 10.54	n/a
2013 [26]	United States Ret. Conf.	n/a	688	11	9	1.49	n/a	<0.009
					8	1.42	n/a	<0.001
					9	1.44	n/a	<0.009
					29	5.63	n/a	<0.003
					25	4.13	n/a	<0.01
					7	1.14	n/a	<0.001
					8	1.34	n/a	<0.005
					7	0.91	n/a	<0.001
					12	2.04	n/a	<0.001
					8	0.66	n/a	<0.001
					6	0.79	n/a	<0.003
2013 [191]	France Pro. Con.	n/a	381	1	68	33.7	2.24	0.0024

*** Correction for multiple comparisons performed.**

**** Although this study reported a 'potential cluster' it did not clearly implement a defined cluster test statistic**

¶ 2 clusters at time of death, 1 at time of birth.

Ret. = retrospective ; Pro. = prospective ; Exp. = exploratory ; Conf. = confirmatory

(M) = based on mortality statistics.

(Reg) = ALS disease register was used for ascertainment.

Exploratory studies did not assume the presence of a cluster, whilst confirmatory studies were designed to specifically assess a supposed cluster

3.3. Social deprivation and population density are not associated with risk of ALS[8]

3.3.1. Introduction

Recently we characterized spatial relative risk and performed formal cluster analysis for ALS in Ireland based upon prospective incident ALS cases occurring in Ireland between 1995 and 2013 (Chapters 3.1 and 3.2). We found weak evidence of an association between area ALS risk and population density (Chapter 3.1) – a finding also reported by others [193]. We also performed formal cluster analysis, finding no significant areas of high risk, and two highly significant areas of low risk, one centring in Co. Kilkenny and the other centring in Co. Clare (Chapter 3.2).

Both of our previous analyses were performed using 3,355 historical electoral divisions (EDs), as areal definitions for our analysis. As we noted previously, the EDs have a highly skewed distribution of populations and this may have influenced the analysis (Chapter 3.1). In 2011, the Central Statistics Office used new areal units of smaller size (and thus higher spatial resolution) for census reporting including approximately 18,500 small areas (SAs) [160].

We aim to examine ALS spatial incidence in Ireland using the SAs and to compare this analysis with our previous ED based analysis in order to determine the influence of spatial scale and normalized population distribution on our results, and to inspect for associations between small area social deprivation and ALS incident risk.

---

[8] Chapter 3.3 has been published as:

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

3.3.2. Methods

Data Sources

Boundary files for the SAs in 2011 were obtained from the Ordnance Survey of Ireland [194]. Small areas include a minimum of 50 households and a maximum of 200 households, with an average of approximately 100 households, resulting in areas that generally equate to the housing development or street level [195]. 2006 census data on the Irish population and the social deprivation index defined for these small areas was obtained from Haase and Pratschke [195]. Social deprivation was quantified using the PobalHP index, developed by Haase and Pratschke over successive years of analysis of the Irish census [195]. This index was developed using confirmatory factor analysis (CFA) to quantify social deprivation using multiple parameters falling under three core dimensions: Demographic Profile, Social Class Composition and Labour Market Situation quantified for each small area [195]. We used relative PobalHP index values from 2006 as a midpoint estimate of social deprivation for our cohort, with mean 0, standard deviation 10 and values ranging from approx. -40 (most deprived) to +40 (most affluent) [195]. The categories for PobalHP scores are shown in Table 3-3.

Patient data was obtained from the Irish ALS register which has been extensively described previously [16,27]. All eligible cases (diagnosed between 1st January 1995 and 31st December 2013 – data extracted 1st October 2014) were geocoded as described previously (Chapter 3.1). As the boundary data was from the 2011 census and the population data was from 2006, a small number of polygon joins were necessary to ensure compatibility. Therefore 18,222 resulting SA's were included in the analysis.

Table 3-3 Distribution of Pobal HP Index Relative Index

Relative Index Score	Standard Deviation	Label
Over 30	> 3	Extremely affluent
20 to 30	2 to 3	Very affluent
10 to 20	1 to 2	Affluent
0 to 10	0 to 1	Marginally above average
-10 to 0	-1 to 0	Marginally below average
-20 to -10	-2 to -1	Disadvantaged
-30 to -20	-3 to -2	Very disadvantaged
Under -30	< -3	Extremely disadvantaged

Source: Haase and Pratschke 2012 [195]

Calculation of Standardised Incidence Rates

Indirect standardization by five-year age group and sex was applied to each small area population using the total national cases and population data to define the standard population, and expected cases per SA were calculated, analogous to our previous indirect standardization at ED level (Chapter 3.1). Expected cases were calculated separately for males, females, patients under 55 yrs and patients over 55yrs, limb onset and non-limb onset patients. Limb and non-limb categories were used to stratify site of onset as there were not enough 'other' onset cases (e.g. respiratory) to generate separate maps for those categories.

Bayesian Smoothing

The *poly2nb()* command of R's spdep package [169] was used to define a neighbourhood matrix for SA's. We created artificial neighbour relations between islands and the mainland. Spatial smoothing was achieved by implementing the Besag-York-Mollié model [156,180] via the R-INLA package [188,189]. A model was built for all patients, and stratified models were built for males, females, age of onset under or over 55 years, site of onset and time periods. Further models were built including social deprivation and population density as Bayesian covariates. Maps were then drawn showing smoothed relative risk.

Post Smoothing Analysis

Post smoothing analysis of our previous ED based analysis revealed a weak association between ALS risk and population density, which was in contrast with our results found by including population density as a Bayesian covariate (Chapter 3.1). The asymmetrical distribution of population across EDs may have contributed to this discrepancy (Chapter 3.1). However, the SAs were designed with statistical considerations in mind in order to have a much less skewed population distribution. Therefore, linear regression of population density onto post smoothing ALS risk was performed to further explore this relationship. Methods used were the same as previously described (Chapter 3.1).

Cluster analysis

We previously used SaTScan [183] to identify two areas of low risk for ALS on the ED geography level that were statistically significant (Chapter 3.2). However the geographical size and shape of the low risk areas may have been in part determined by the shapes of the EDs. Therefore there is value in comparing our previous results with cluster analysis on the SA level. SaTScan was again used to identify high or low risk clusters of ALS risk amongst the SAs using the same methods employed previously (Chapter 3.2).

Finally, mean values for population density and social deprivation were compared between areas falling into identified clusters and the rest of the country using the Students t-test.

Software

R 3.1 statistical software [163] was used to perform data preparation, routine statistics, and ALS risk mapping as described previously (Chapter 3.1 and 3.2). R was also used to prepare ASCII files for SaTScan v9.3 [183,184].

3.3.3. Results

The total number of cases meeting the inclusion criteria was 1,701. Table 3-4 shows the clinical characteristics of the cohort. Only 61 cases had no known age at diagnosis (due either to a missing birthdate or a missing diagnosis date). 92 cases were missing a site of onset.

Table 3-4 Descriptive statistics of cohort. N=1,701

Variable		Male (n = 959)	Female (n =742)	P value
Age (mean)		64.5 (12.0)	66.4 (11.2)	0.0007 ¶
Site of onset	Limb	597 (62.3%)	341 (46.0%)	
	Non-limb	308 (32.1%)	365 (49.2%)	
	Missing value	54 (5.6%)	36 (4.8%)	<0.0001†
Familial		57 (5.9%)	54 (7.3%)	0.2776 †

¶ Student's t-test **† Fishers exact test**

Bayesian Smoothed Risks

Figure 3-11 shows the Bayesian smoothed risk for SAs for all cases. Similar to our previous analysis at ED level (Figure 3-5), the SA map's most striking features are the low risk areas in counties Kilkenny and Clare. However, on the SA map the low risk area at the West coast of county Clare, and the low risk area in Kilkenny are reduced in size relative to the ED map. Again, matching our ED analysis, we see areas of mildly elevated RR in counties Louth / Meath / Dublin, Cork, the Dingle peninsula and the south-western part of county Donegal. The SA map of risk in males and females (Figure 3-12) were also in close agreement with the ED level maps for gender (Figure 3-6).

The map for limb onset disease shows an area of elevated risk in Louth / Meath / Dublin and low risk areas in counties Kilkenny, Clare and Donegal (Figure 3-13). The map for non-limb onset show high-risk areas in the North-West, the north western corner of Mayo, Dingle, Cork and Wexford, with a more intense low risk area on the county Clare coastline than seen in other maps. It was noted that limb

vs non-limb maps were distinct from gender stratified maps despite the association between limb onset and male gender (Table 3-4).

Figure 3-11 Map of Bayesian smoothed SIR per Census small area in Ireland

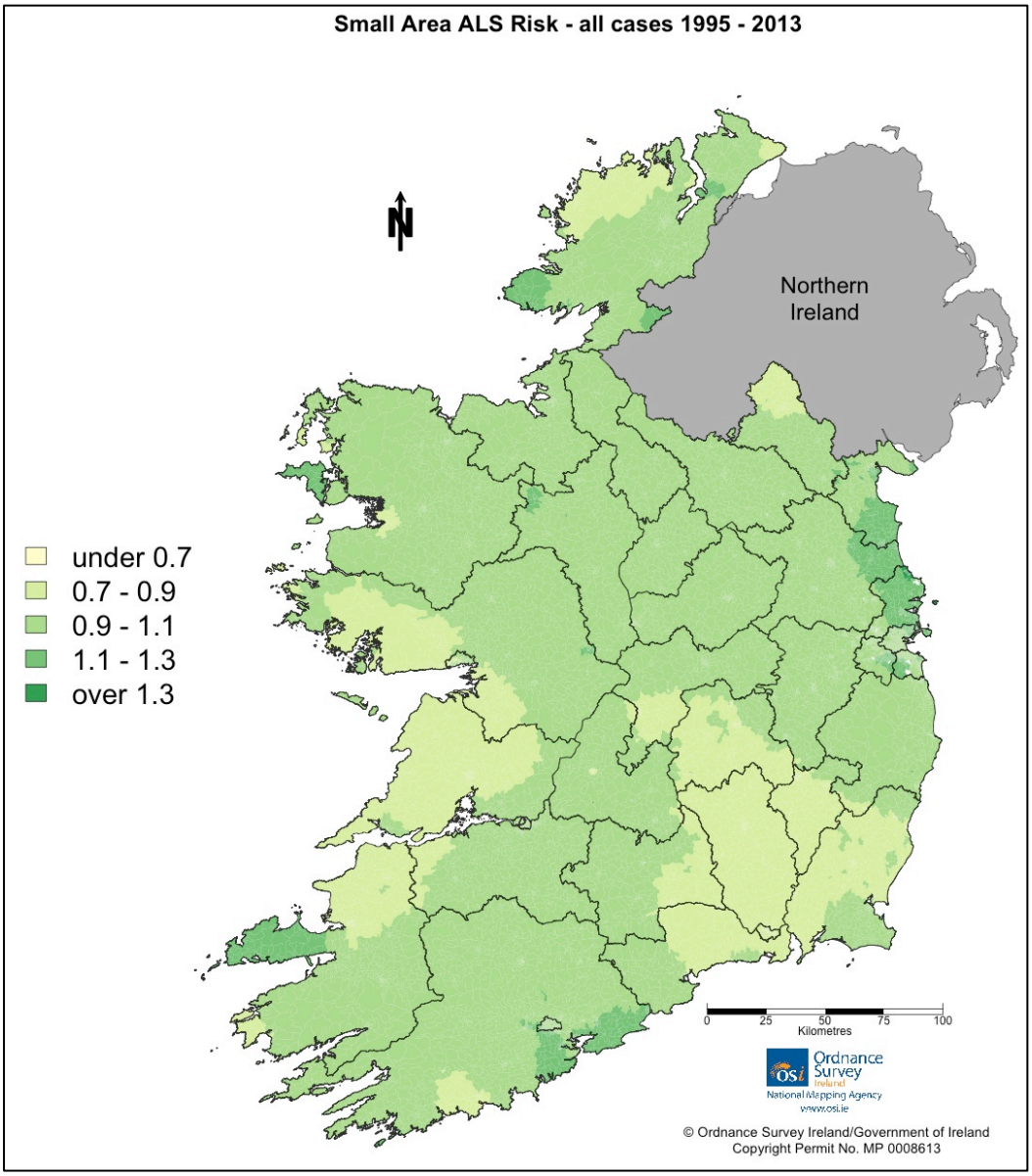


Figure 3-12 Bayesian smoothed SIR per Census small area in Ireland stratified by gender

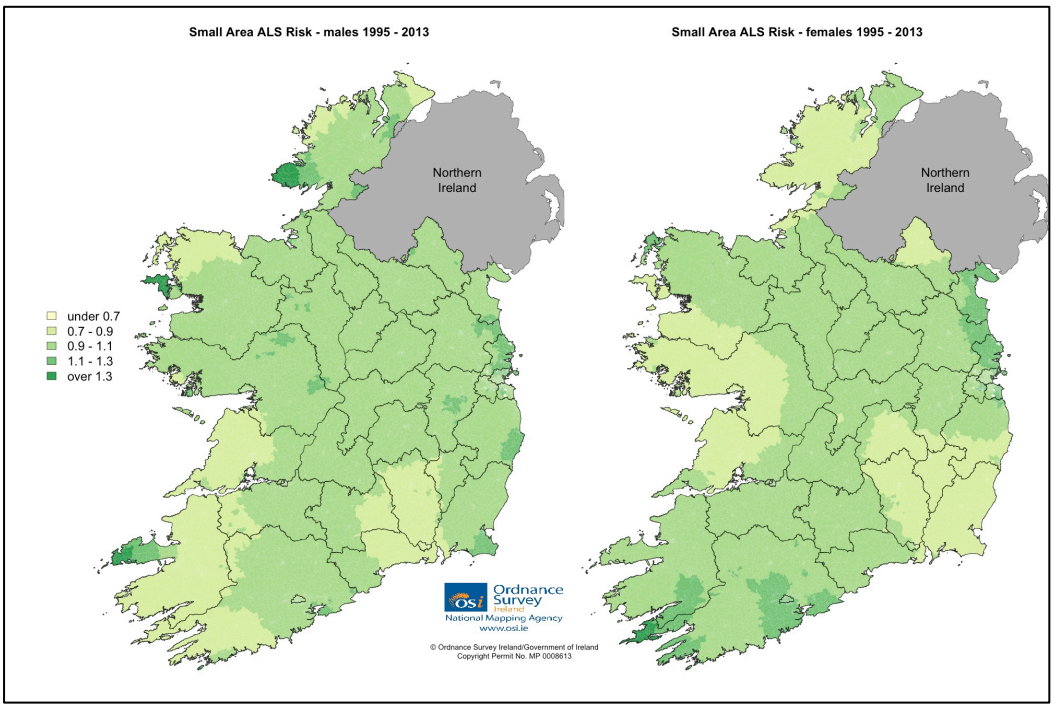


Figure 3-13 Bayesian smoothed SIR per Census small area in Ireland stratified by site of onset

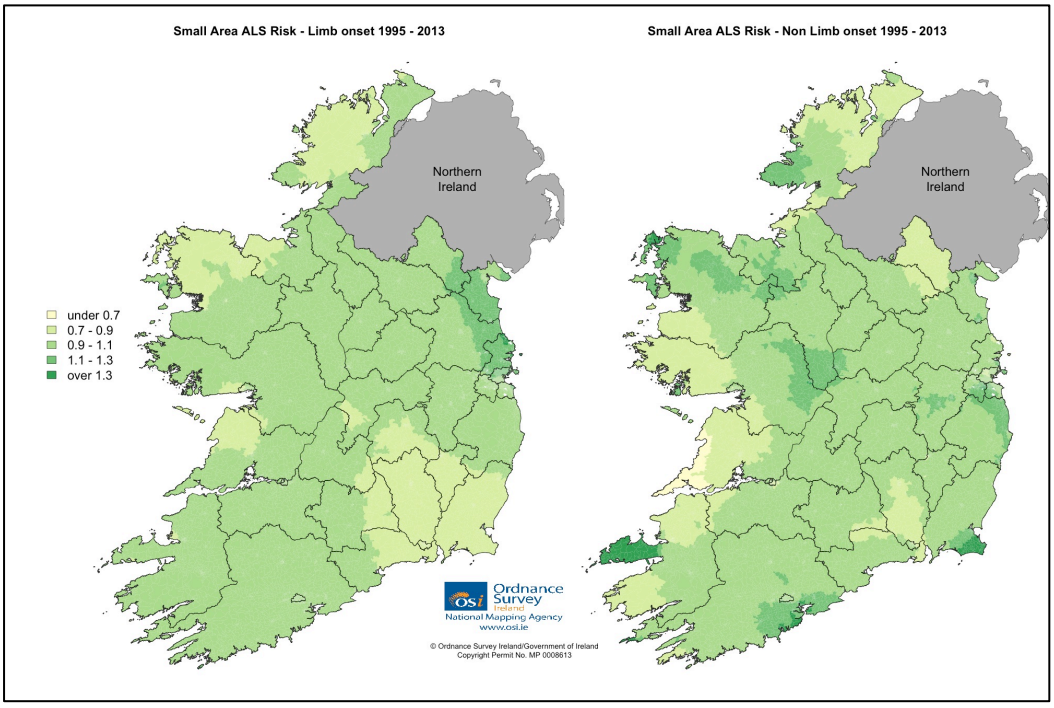


Table 3-5 shows regression coefficients and model deviance information criterion (DIC) for the 'all cases' model and additionally for models including population density and social deprivation. The 95% credible intervals of the regression coefficients for both population density and social deprivation overlapped zero indicating these variables did not explain the spatial variation of ALS risk.

Table 3-5 Comparison of Bayesian models with and without covariates

Model	**DIC**	**Covariate coefficient (95% Credible Interval)**
All cases	11232	n/a
All cases + population density	11233	-3.1 (-18.8, 11.6)
All cases + social deprivation	11233	-0.002 (-0.008, 0.003)

Post Smoothing Analysis

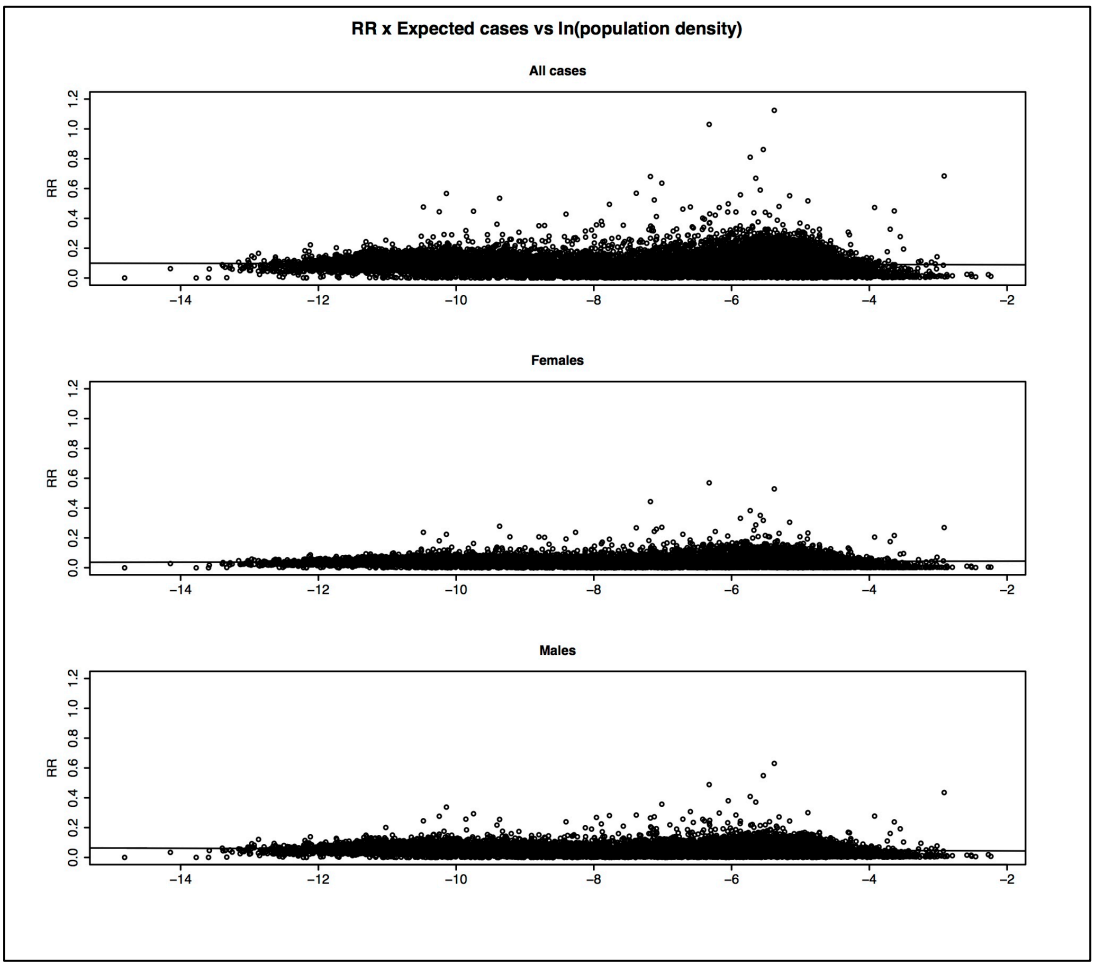
Weighted regression of post Bayesian smoothed RR's vs log of population density failed to show the same relationship as was seen at the ED level – although statistically significant unweighted coefficients were found the values were extremely small (Table 3-6) and weighted regressions for males and all cases had negative coefficients in contrast to our previous findings (Table 3-6). Graphical examination demonstrated poor correlation with no clear relationship (Figure 3-14). Similar regression of post Bayesian smoothed RR's vs log of social deprivation scores also failed to reveal any meaningful relationship.

Table 3-6 Results of linear regression of weighted and un-weighted post Bayesian smoothing RR for all cases, males and females versus ln(population density)

	Unweighted regression	**Weighted regression**
Small Area Geometry	**Coefficient (95% CI); R^2**	**Coefficient (95% CI); R^2**
All cases	0.014 (0.013, 0.0145); R^2 = 0.09	-0.0008 (-0.0011, -0.0004); R^2 = 0.05
Males only	0.011 (0.010, 0.011); R^2 = 0.09	-0.0015 (-0.0017, -0.0013); R^2 = 0.03
Females only	0.015 (0.014, 0.016); R^2 = 0.11	0.0006 (0.0004, 0.0007); R^2 = 0.03
Electoral Division Geometry	**Coefficient (95% CI); R^2**	**Coefficient (95% CI); R^2**
All cases	0.015 (0.13, 0.16); R^2 = 0.09	0.23 (0.21, 0.24); R^2 = 0.31
Males only	0.009 (0.008 – 0.011); R^2 = 0.06	0.12 (0.11, 0.13); R^2 = 0.31
Females only	0.17 (0.015, 0.019); R^2 = 0.10	0.11 (0.10 – 0.11) ; R^2 = 0.33

Table 3-6 legend: P << 0.00001 in all models, however R^2 values and graphical examination (figure 5) show that the relationship is not meaningful.

Figure 3-14 Linear Regression of post Bayesian smoothing RR x expected cases for all cases, males and females versus ln(population density)



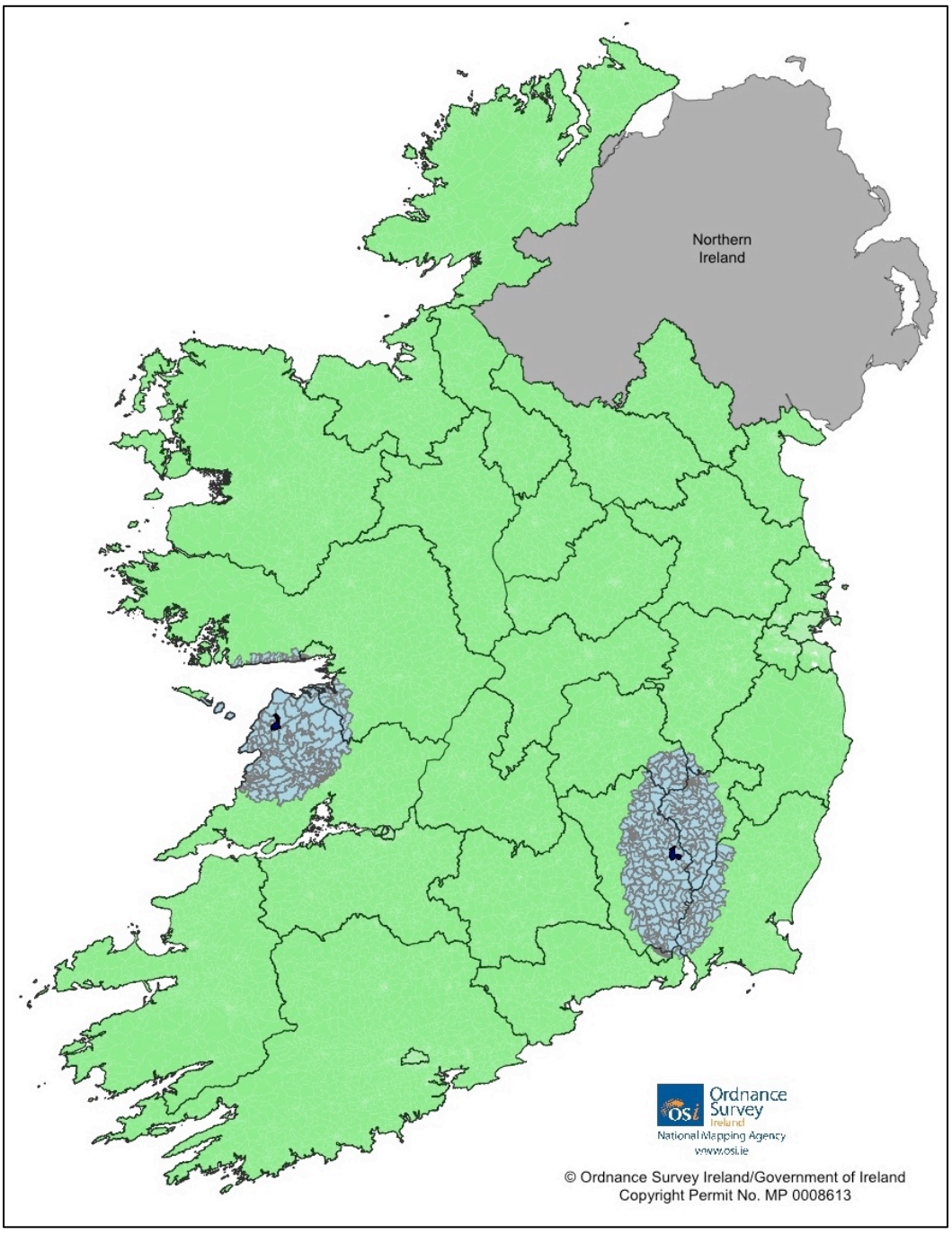
Cluster Analysis

SaTScan failed to identify any high-risk clusters, but did identify two areas of significantly low risk (Figure 3-15). An elliptically shaped cluster consisting of 656 SAs (population: 152,508) was located centred on the Kilkenny / Carlow border region and included 25 cases when 61.9 were expected (RR = 0.390; P = 0.048), while a second cluster consisting of 271 SAs (population: 56,638) was located on the coastline of County Clare where 2 cases were observed instead of the 24 expected (RR = 0.082; P = 0.028). Mean values of population density and social deprivation for the identified clusters are summarized in Table 3-7. These results showed that both clusters had population density lower than that of the national average, whilst the Kilkenny cluster had a social deprivation index lower than the national average, but the Clare cluster had a higher social deprivation index.

Table 3-7 Comparison of properties of low risk area to national values

Geography	**Population Density – persons/km^2 (95% CI)**	**Social deprivation Index (95% CI)**
National	2984 (2913 , 3055)	0.03 (-0.12 , 0.18)
Kilkenny low risk area	1985 (1772 , 2197)	-1.50 (-2.21 , -0.80)
P Kilkenny vs National	<0.001	<0.001
Clare low risk area	1494 (1171 , 1817)	1.67 (0.62, 2.72)
P Clare vs National	<0.001	0.003

Figure 3-15 Map of significant low-risk areas for ALS incidence in Ireland at small area geometry



3.3.4. Discussion

In general, the results showed good agreement between ALS risk estimated at SA level when compared to our previous ED level analysis (Chapter 3.1). The higher spatial resolution and more uniform population distribution of the units used in this analysis allowed us to more accurately define the boundary of the low risk clusters identified previously. The distribution of spatial risk was similar for all cases, males and females at both geographic levels. Risk maps newly constructed at SA level for limb vs non-limb ALS onset revealed a more heterogeneous distribution of risk for non-limb onset patients, although this could be in part due to the lower-case count of non-limb cases.

Weighted population density was not associated with ALS risk at the SA level, in contrast to our findings of a weak association at the ED level. This is most likely due to the more symmetrical distribution of population count across areas at the SA level when compared to the ED level – the later determined by the shapes of the polygons. As such our ED level findings are likely an example of the Modifiable Areal Unit Problem - a form of statistical bias long recognized in spatial statistics whereby changing the locations of area boundaries can affect statistical findings [196]. Our findings at SA level also conflict with findings from the South East of England which included 368 cases distributed across 25 districts and found that population density was associated with areal ALS risk and explained 25% of the variation in ALS rates [193]. However a recent Bayesian smoothing study from Italy including 1,216 patients found that ALS rates were higher in rural areas, although the analysis did not include population density as a covariate [111]. It is possible that the relationship between population density in these studies is confounded by other factors (e.g. higher air pollution in London, greater/different pesticide usage or genetic isolation in Italian rural and/or mountainous areas). However, the current results are based on the largest cohort using the smallest areal units, and comparison with our previous studies has revealed subtle methodological effects that may operate in the UK and Italian studies. Therefore, we conclude that population density does not directly explain spatial patterns of ALS risk.

Social deprivation has not been previously examined as a risk factor for ALS, however it is frequently included as a cofactor in cancer epidemiology [197,198], and may act as a proxy for other local factors that could potentially influence small area case ascertainment – such as local health service variation [197]. Our results indicate that social deprivation does not correlate with small area ALS risk in Ireland. It should be noted however, that social deprivation is modelled only as an areal parameter, and that individual patient social factors may differ significantly from that of their neighbourhood.

Cluster analysis at SA level identified two areas of significantly low ALS risk that closely matched the low risk areas identified at ED level (Chapter 3.2). At SA level however these areas cover a smaller geographic area that is likely a consequence of the higher spatial resolution of the SAs. Our analysis indicates that population density and social deprivation do not explain these low risk areas and further work is underway to explore whether the distribution of genetic and environmental factors may offer insight into their aetiology.

An important difference between this study and our previous ED level study is that we have only used population data from one census at SA level versus data averaged from 4 censuses at ED level. Thus, comparisons between the spatial patterns of risk and cluster analysis at both geometries need to be interpreted with some caution. However, the closely matching results from both Bayesian and cluster analyses indicate that any effect is likely very small. A weakness of our analysis is that it is based on residency at time of diagnosis only. Given the age profile of ALS patients most are likely resident in a family home for some years prior to diagnosis. Therefore exposure to risk factors in the previous 5-10 years are likely adequately captured, however, the analysis is not useful for assessing early life exposure to risk factors.

In summary, we have replicated our previous results for ED level data at a much higher spatial resolution including 18,222 small areas. Bayesian risk mapping shows a similar overall spatial pattern of risk at both geometries for all cases, males and females. In addition, we have visualized spatial risk for limb vs non-limb

cases. Cluster analysis via SaTScan identified two statistically significant low risk areas centering in counties Clare and Kilkenny as previously, however the SA geometry allowed us to define these clusters more precisely. Population density was not significantly associated with SA ALS risk on Bayesian or post-Bayesian analysis in contrast to our previous findings. Finally, small area social deprivation was not associated with SA ALS risk.

3.4. No association between soil constituents and ALS relative risk in Ireland[9]

3.4.1. Introduction

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative condition with a median survival in Ireland of 2.39 years from symptom onset [11]. Despite this bleak prognosis, progress in identifying the causes of ALS has been slow. Current theories postulate that multiple genes and environmental factors may in combination result in clinical disease [199], with recent epidemiological analysis suggesting a 6 step process leading to clinical disease [9]. Several disease causing mutations have been identified (most notably C9orf72) to date, and familial disease is present in up to 16% of cases [14]. However, progress toward identifying environmental factors has been slow, with some preliminary evidence for higher risk in those occupationally exposed to pesticides [65,66], and conflicting results regarding physical exercise [59,200].

Exposure to heavy metals has been proposed as a risk factor for ALS, particularly lead and mercury, and elemental nutrients for example selenium [64], with the possibility that metal-gene and metal-epigenome interactions could also play a role [68,201,202].

We have recently mapped ALS spatial risk in Ireland using Bayesian and cluster analysis methods at electoral division (ED) and small area (SA) levels (Chapter 3.3). Here we extend this analysis to include soil mineral levels from the Irish National Soils Database as covariates in Bayesian conditional auto-regression models to estimate associations with ALS small area risk.

---

[9] Chapter 3.4 has been published as:

Rooney J, Vajda A, Heverin M, Crampsie A, Tobin K, McLaughlin R, et al. No association between soil constituents and amyotrophic lateral sclerosis relative risk in Ireland. Environ Res. Elsevier; 2016;147:102–7

3.4.2. Methods

ALS case ascertainment was through the Irish ALS Register including patients from 1995 to 2013. Population data were obtained for Small Areas (SAs) from the 2006 census of Ireland via Trutz Haase [195] and Small Area shapefiles were obtained from Ordnance Survey Ireland [194]. Calculation of observed and expected ALS cases per small area and methods for Bayesian smoothing with covariates have been described previously (Chapter 3.3). Data on soil concentrations were obtained under license from the Irish EPA [203]. The National Soils Database contains data on 45 different soil parameters, mainly element and nutrient levels, from 1,310 selected sample sites around Ireland [204]. The sampling strategy for the National Soils Database was designed to ensure 2 samples per 100km^2 [204]. Using the *automap* package [205] of R ver 3.1.1 [190], we interpolated average values of each soil constituent for each SA polygon using ordinary kriging (a geostatistical method to interpolate an unknown parameter at a given location(s) modeled upon known values at known locations and allowing for spatial structure).

Exploratory approach

Next, Bayesian spatial smoothing was implemented using the Besag-York-Mollié model [156,180] via the R-INLA package [188,189]. Separate models were built as baseline reference models for all patients, males, females, age of onset (under or over 55), site of onset (limb or non-limb onset). Subsequent models were then built including each parameter from the national soils database as a covariate in the BYM model. Models were compared using the deviance information criterion (DIC).

Hypothesis testing approach

In addition to the above exploratory approach, a number of specific hypotheses designed to explore whether combined exposures might be masking significant associations were tested based on previous interest in the literature. Lead,

mercury and selenium have received particular interest as potential risk factors for ALS [68,201,202,206–208]. Of these elements, mercury and lead are known neurotoxins, and selenium is neurotoxic when doses exceed recommended daily intakes [209]. Additionally, both selenium and zinc are also known to affect the distribution and toxicity of mercury in vivo [210]. Little is known about combined exposure to toxic metals, however both lead and mercury are known to share some polymorphisms that predispose to toxicity (e.g. GSTP1). Therefore, we have specifically built and tested models including combinations of these variables as Bayesian covariates.

3.4.3. Results

1,701 ALS patients incident between 1995 & 2013 were included. Of these 959 (56%) were male, 938 (55%) had limb onset ALS while 6.5% had familial disease. Table 3-8 summarizes the parameters included from the National Soils Database. Base models for all cases, males, females, under 55's, over 55's, limb and non-limb onset cases have been previously described (Chapter 3.3). These were used for baseline comparison with models including covariates via DIC. 315 separate models were built in total. Table 3-9 displays results from successive models for all cases including all models for each covariate, ordered by DIC. In each case the covariate included had coefficients with credible intervals which crossed zero, indicating lack of association between the included covariates and ALS risk. Of the 315 models built, only one resulted in a coefficient that did not cross zero. For limb onset cases, total magnesium % of soil had a mean coefficient of 0.319 (credible interval 0.033 – 0.607). Figure 3-16 displays the distribution of total magnesium % of soil and the spatial risk for limb onset disease without covariate for comparison. Figure 3-17 displays the relationship between the basic Bayesian model of limb onset cases (i.e. with no soil covariate) versus total magnesium for all areas. Visual examination of both figures shows that there is no obvious overall relationship between the map of total magnesium and small area relative risk (RR) for limb onset disease.

Model results for tests of specific hypotheses are shown in Table 3-10. None of the more complex models resulted in lower DIC values, and all of the covariate coefficients were overlapping zero. Therefore, none of these models held explanatory power for the spatial variation in ALS risk.

Table 3-8 Summary of soil parameters from the National Soils Database

Parameter	Units	Mean Value	Std Dev.	Min	Max
Soil pH	n/a	5.33	0.98	3.20	7.70
Soil organic carbon	Percent	13.35	14.01	1.40	55.80
Available Phosphorous	mg/l	10.61	13.77	0.56	316.41
Available Potassium	mg/l	145.42	94.09	4.66	949.23
Available Magnesium	mg/l	218.82	132.16	13.49	1001.97
Aluminium	Percent	3.53	1.93	0.06	9.74
Arsenic	mg/kg	9.05	9.19	0.35	122.70
Barium	mg/kg	233.41	137.54	6.60	1296.90
Calcium	Percent	0.76	1.56	0.03	26.63
Cadmium	mg/kg	0.56	0.70	0.03	15.15
Cerium	mg/kg	34.30	18.55	0.60	136.40
Cobalt	mg/kg	6.86	5.26	0.20	58.70
Chromium	mg/kg	44.49	25.09	2.10	221.70
Copper	mg/kg	19.48	16.76	1.10	272.40
Iron	Percent	1.95	1.33	0.05	19.43
Gallium	mg/kg	9.04	5.05	0.14	25.16
Germanium	mg/kg	1.25	0.50	0.10	2.58
Mercury	mg/kg	0.11	0.14	0.02	3.45
Total Potassium	Percent	0.97	0.54	0.02	2.64
Lanthanum	mg/kg	19.16	9.71	0.50	75.20
Lithium	mg/kg	24.61	17.66	2.00	165.70
Total Magnesium	Percent	0.35	0.23	0.04	2.10
Manganese	mg/kg	703.05	1184.99	7.00	21077.00
Molybdenum	mg/kg	1.25	1.43	0.07	21.14

Sodium	Percent	0.42	0.32	0.02	2.25
Noibium	mg/kg	6.76	4.16	0.06	38.88
Nickel	mg/kg	20.56	15.50	0.80	176.00
Total Phosphorous	Percent	0.11	0.05	0.01	0.49
Lead	mg/kg	31.40	75.61	1.10	2634.70
Parameter	**Units**	**Mean Value**	**Std Dev.**	**Min**	**Max**
Rubidium	mg/kg	55.34	35.39	0.60	222.00
Sulphur	Percent	0.11	0.10	0.01	0.70
Antimony	mg/kg	0.67	0.51	0.05	5.29
Scandium	mg/kg	6.00	3.61	0.12	17.11
Selenium	mg/kg	1.09	1.31	0.08	17.44
Tin	mg/kg	2.03	1.69	0.22	17.84
Strontium	mg/kg	57.99	51.17	9.20	1252.50
Tantalum	mg/kg	0.50	0.27	0.05	2.71
Thorium	mg/kg	4.58	2.42	0.10	11.15
Titanium	mg/kg	2093.44	1169.71	39.00	8704.00
Thallium	mg/kg	0.46	0.25	0.02	2.66
Uranium	mg/kg	2.31	2.85	0.10	64.19
Vanadium	mg/kg	54.22	31.44	2.10	240.30
Tungsten	mg/kg	0.71	0.51	0.10	7.72
Yttrium	mg/kg	11.31	8.04	0.22	111.78
Zinc	mg/kg	70.07	57.60	3.60	1384.40

Table 3-9 Summary of BYM models for all cases including individual soil paramaters as a covariates

Covariate	DIC	Mean	2.5%	Median	97.5%
Strontium (mg/kg)	11229.1	-0.23	-0.703	-0.214	0.194
Rubidium (mg/kg)	11229.3	-0.14	-0.334	-0.136	0.050
Thorium (mg/kg)	11229.5	-0.16	-0.423	-0.160	0.082
Vanadium (mg/kg)	11230.3	-0.17	-0.684	-0.158	0.300
Thallium (mg/kg)	11230.7	-0.11	-0.323	-0.110	0.095
Aluminium (%)	11230.7	-0.12	-0.363	-0.123	0.105
Molybdenum (mg/kg)	11230.8	-0.10	-0.339	-0.097	0.126
Lanthanum (mg/kg)	11230.8	-0.20	-0.540	-0.202	0.126
Tungsten (mg/kg)	11231.0	-0.08	-0.285	-0.079	0.114
Tantalum (mg/kg)	11231.1	-0.09	-0.301	-0.093	0.107
Cerium (mg/kg)	11231.1	-0.12	-0.405	-0.123	0.145
Chromium (mg/kg)	11231.2	-0.11	-0.405	-0.107	0.175
Tin (mg/kg)	11231.3	-0.06	-0.266	-0.053	0.144
Germanium (mg/kg)	11231.4	-0.06	-0.337	-0.060	0.194
Titanium (mg/kg)	11231.6	-0.08	-0.305	-0.075	0.145
Niobium (mg/kg)	11231.6	-0.08	-0.303	-0.080	0.135
Potassium (%)	11231.6	-0.07	-0.316	-0.068	0.167
Cadmium (mg/kg)	11231.6	-0.08	-0.242	-0.080	0.079
Scandium (mg/kg)	11231.7	-0.08	-0.536	-0.076	0.351
Basic model (no covariates)	11231.8	NA	NA	NA	NA
Barium (mg/kg)	11231.8	-0.10	-0.361	-0.100	0.155
Zinc (mg/kg)	11231.9	-0.06	-0.324	-0.063	0.186
Sodium (%)	11232.0	-0.03	-0.213	-0.033	0.139
Available magnesium (mg/l)	11232.2	-0.32	-0.809	-0.313	0.177
Soil pH	11232.2	-0.01	-1.090	-0.017	1.074
Cobalt (mg/kg)	11232.2	-0.02	-0.283	-0.021	0.235
Lithium (mg/kg)	11232.2	-0.07	-0.297	-0.074	0.147
Gallium (mg/kg)	11232.4	-0.12	-0.419	-0.115	0.185
Iron (%)	11232.5	-0.06	-0.326	-0.056	0.209

Available potassium (mg/l)	11232.6	-0.09	-0.659	-0.092	0.467
Soil Organic Carbon (%)	11232.6	-0.12	-0.324	-0.123	0.079
Sulphur (%)	11232.6	-0.13	-0.395	-0.127	0.146
Calcium (%)	11232.6	-0.05	-0.190	-0.054	0.082
Magnesium (%)	11232.6	0.05	-0.222	0.048	0.309
Lead (mg/kg)	11232.7	0.01	-0.193	0.016	0.214
Nickel (mg/kg)	11232.7	0.01	-0.186	0.007	0.196
Uranium (mg/kg)	11232.8	-0.08	-0.392	-0.083	0.228
Yttrium (mg/kg)	11232.8	-0.06	-0.263	-0.062	0.139
Arsenic (mg/kg)	11232.8	-0.14	-0.500	-0.144	0.216
Antimony (mg/kg)	11232.8	0.07	-0.102	0.071	0.240
Mercury (mg/kg)	11232.9	-0.03	-0.260	-0.030	0.200
Manganese (mg/kg)	11232.9	0.10	-0.280	0.097	0.458
Available phosphorous (mg/l)	11233.1	-0.03	-0.341	-0.024	0.283
Selenium (mg/kg)	11233.2	-0.01	-0.167	-0.007	0.153
Copper (mg/kg)	11233.3	-0.01	-0.146	-0.008	0.128

Table 3-9 legend: Models are ordered by decreasing model fit as assessed by DIC – i.e. the best fitting models are listed first.

Figure 3-16 Maps of relative risk of limb onset ALS and kriged percent of total magnesium % of mineral soil per small area

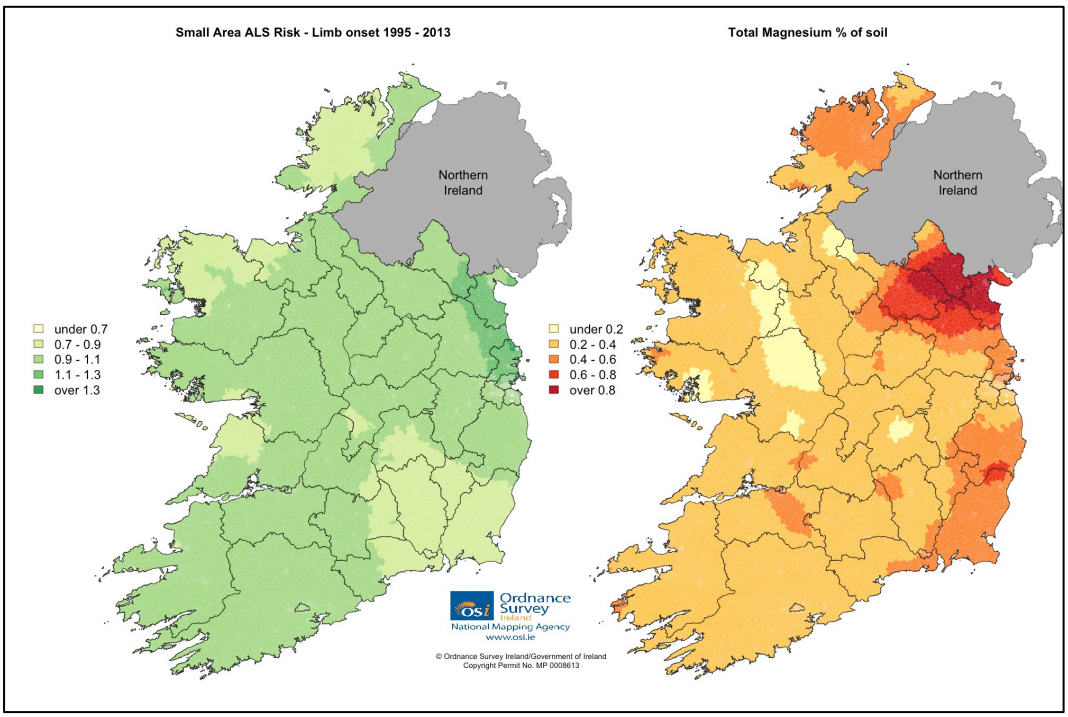


Figure 3-16 Legend: As can be seen above the overall spatial pattern of ALS and total magnesium % of soil are distinct. There is an area of correlated high total magnesium and small area relative risk in the North-East, however there are also several discordant areas in the South-east, North and North-West - therefore considering the overall pattern of both maps we do not recognize an overall similarity of patterns.

Figure 3-17 Plot of relative risk of limb onset ALS versus percent total magnesium % of soil by small area

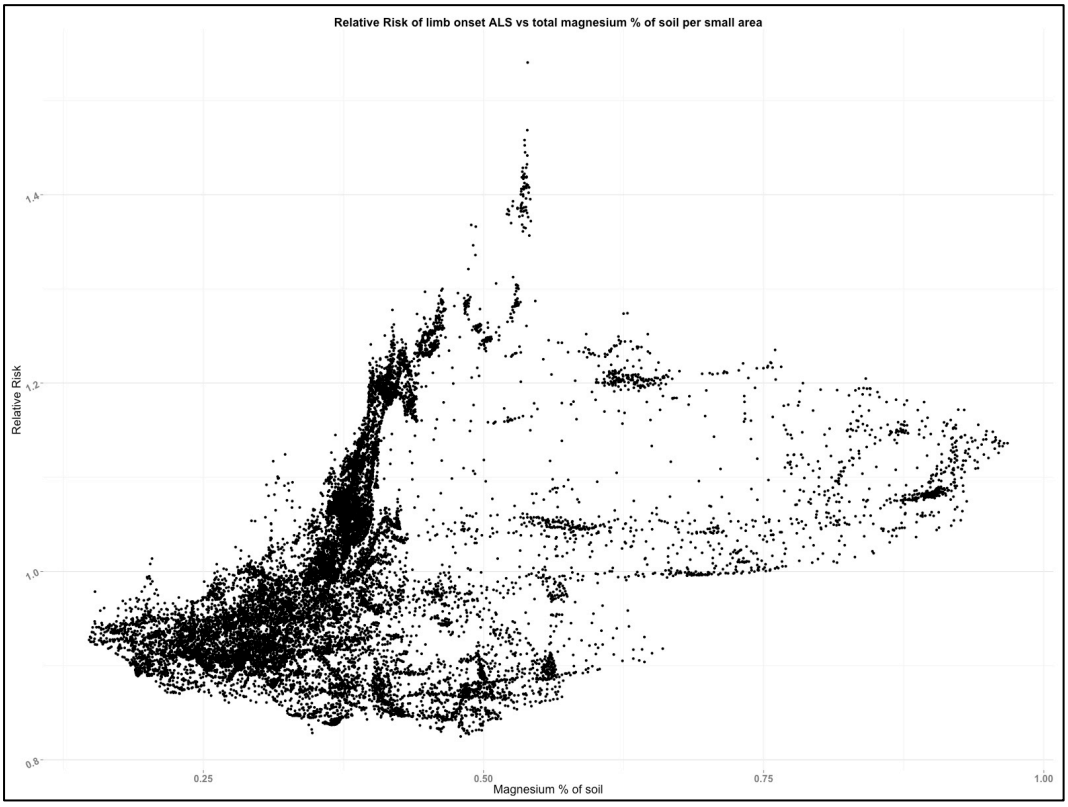


Figure 3-17 Legend: There is no clear relationship between Bayesian smoothed relative risk of limb onset ALS (without soil parameter covariate) and total magnesium % of soil.

Table 3-10 Summary of hypothesis testing models including all cases

All cases – no covariates			DIC = 11231.8		
Mercury + Selenium			DIC = 11234.1		
Covariate	Mean	SD	2.5%	Median	97.5%
log(Hg mg/kg)	-0.03	0.128	-0.284	-0.031	0.219
log(Se mg/kg)	0.00	0.089	-0.172	0.002	0.178
Mercury + Lead			DIC = 11234.1		
Covariate	Mean	SD	2.5%	Median	97.5%
log(Hg mg/kg)	-0.04	0.126	-0.288	-0.043	0.207
log(Pb mg/kg)	0.03	0.111	-0.196	0.029	0.242
Mercury + Zinc			DIC = 11233.2		
Covariate	Mean	SD	2.5%	Median	97.5%
log(Hg mg/kg)	-0.02	0.123	-0.25624	-0.01573	0.22676
log(Zn mg/kg)	-0.06	0.135	-0.33135	-0.05922	0.20178
Lead + Zinc			DIC = 11233.2		
Covariate	Mean	SD	2.5%	Median	97.5%
log(Pb mg/kg)	0.04	0.113	-0.185	0.042	0.258
log(Zn mg/kg)	-0.08	0.140	-0.362	-0.084	0.189
Lead + Selenium			DIC = 11234.0		
Covariate	Mean	SD	2.5%	Median	97.5%
log(Pb mg/kg)	0.01	0.104	-0.19466	0.01499	0.21405
log(Se mg/kg)	-0.01	0.082	-0.16656	-0.00568	0.15581
Mercury + Lead + Selenium			DIC = 11235.3		
Covariate	Mean	SD	2.5%	Median	97.5%
log(Hg mg/kg)	-0.05	0.140	-0.321	-0.047	0.229
log(Pb mg/kg)	0.03	0.113	-0.199	0.029	0.248
log(Se mg/kg)	0.01	0.091	-0.171	0.008	0.187

Mercury + Lead + Zinc			DIC = 11234.6		
Covariate	Mean	SD	2.5%	Median	97.5%
log(Hg mg/kg)	-0.03	0.128	-0.281	-0.031	0.223
log(Pb mg/kg)	0.05	0.118	-0.189	0.049	0.275
log(Zn mg/kg)	-0.08	0.143	-0.362	-0.079	0.199

3.4.4. Discussion

Our analysis, the first using a prospective population based dataset collected over 18 years, failed to find any significant associations between small area spatial risk and any of the 45 soil parameters tested. This likely indicates that none of the soil parameters tested play a role in ALS, or alternatively that levels of the various soil parameters are not informative as to the levels of those parameters in food or other sources of personal exposure. This is congruent with the findings in Guam, which showed that despite high levels of soil aluminium and low levels of calcium in the soil, there was no difference in exposure in food when compared to other locations with lower rates of ALS [211].

However, there are a number of alternate possible explanations for our findings. Primary amongst these is the possibility of the 'ecological fallacy'. An implicit assumption in areal spatial models of this type is that the areal covariates apply to all individuals within a given area, and so associations are determined between the area level covariate and the area level population risk. Therefore, the soil parameters of a given area are not measures of individual exposure, and the exposure of individuals in an area, ALS patient or non-patient, may vary around the mean areal level. We also must consider that soil parameters may vary differentially across areas and thus the areal mean level may or may not reflect soil values in the vicinity of a given individuals residence depending on the local pattern of variability. This may in turn be complicated by other factors such as urban / rural residence, occupation (e.g. farming vs office work), consumption of local vs imported food and others [211]. Other methods to assess individual exposures to given minerals or toxins, such as personal exposure questionnaires, occupational histories, blood borne/biochemical (e.g. cerebro-spinal fluid[212])/radiological markers or epigenetic markers are needed (and may be best used in combination).

A further possible explanation may relate to the recent observation that ALS likely has a 6 step disease process on average [9]. These steps may include genetic and environmental factors, however given that multiple steps could be environmental, spatial auto-regression methods examining single potential risk factors are

unlikely to find positive associations unless a given environmental variable is both strongly representative of personal exposures in the local area and also an important causative agent along the pathogenic trajectory. Bayesian maps of posterior relative risks have also proven to be conservative on sensitivity testing [213] - which may also explain why we only found one non-zero coefficient out of 315 models. An added source of uncertainty with our method was the use of a kriging step to produce small area estimates for each soil parameter, from a sparse set of original measurements, before Bayesian conditional auto-regression. Standard deviation of the kriged estimates increases with increased distance from measurement points, and kriged estimates were averaged over SAs, which themselves vary in size – therefore we cannot rule out the introduction of errors which may have obscured any weak associations.

In summary, this study did not find any evidence to support the relative concentrations of soil constituents to carry explanatory power for the pattern of small area ALS spatial risk in Ireland. Hypothesis testing of association with soil mercury, lead, zinc or selenium did not find evidence of any credible associations either for single elements or in combination. As soil parameters are an ecological assessment of exposure in a given area, individual level measures of exposure to elements of interest are required.

Chapter 4: Progression and Survival in ALS

Chapter Outline:

4. Progression and Survival in ALS

4.1. A multidisciplinary clinic approach improves survival in ALS: A comparative study of ALS in Ireland and Northern Ireland[10]

4.1.1. Introduction

Amyotrophic lateral sclerosis (ALS) is a progressive debilitating neurodegenerative disease with a life expectancy of 3-5 years from first symptom, and death from respiratory failure [214]. Riluzole is the only medication that slows disease progression [215]. Non-invasive ventilation (NIV) confers a survival and quality of life benefit in ALS patients [77,216]. It remains unclear whether gastrostomy improves survival, as there have been no randomized controlled trials [217].

There is compelling evidence that those who attend a multidisciplinary clinic experience improved survival (Ireland [11,82], Holland [83], Italy [84], Sheffield [85]). Processes by which survival advantage is conferred remain unclear. While those attending specialist clinics are younger and more likely to have familial ALS, inclusion of these factors in multivariate analysis preserves the beneficial effect of specialist clinics [11,82]. Moreover, a study in Puglia (Italy), which provides services through a network of clinics, did not demonstrate a survival benefit for ALS patients attending a centralized clinic, suggesting that decentralization with good access to care may be of equal benefit to patients [84]. Whether decentralized care with a trained coordinator with expertise in ALS is equivalent to the provision of centralized multidisciplinary clinics has not been assessed.

[10] Chapter 4.1 has been published as:

Rooney J, Byrne S, Heverin M, Tobin K, Dick A, Donaghy C, et al. A multidisciplinary clinic approach improves survival in ALS: a comparative study of ALS in Ireland and Northern Ireland. J Neurol Neurosurg Psychiatry. 2015;86(5):496–501.

The island of Ireland is divided into two countries, Republic of Ireland (RoI) and Northern Ireland (NI), each with an independent healthcare system. Differences between the two systems of care for ALS patients have offered the unique opportunity to carry out a population based study of survival between the two simultaneously ascertained incident cohorts of ALS patients of similar ancestral origin in 2 contiguous geographic regions. A centralized multidisciplinary ALS clinic has been in place in the RoI since 1995, providing care for up to 80% of the ALS population. In 2004 an ALS Care Network was put in place in NI. An ALS Care Network Coordinator was employed to support patients throughout their disease as well as provide education to non-specialist allied health professionals to facilitate timely coordinated care.

This study compared demographic and survival information for ALS patients using data from the two independent ALS Registers over a six-year period, 2005-2010. The purpose was to explore the outcome of patients in the Republic and Northern Ireland, and to determine whether centralization of services confers any extra advantage to community based care supported by a specialist care worker.

4.1.2. Methods

Register

The Irish ALS Register was set-up in 1993, and the Northern Irish Register was set-up in 2004. The Republic of Ireland covers 70,283 square kilometres, with a population of 4.6 million people, compared to Northern Ireland, which covers 14,148 square kilometers, and has a population of 1.8 million people. The Irish ALS Register records information on all people diagnosed with ALS over the age of fifteen years. ALS cases are captured by contacting all neurologists and gerontologists in the Republic of Ireland, reviewing Hospital Inpatient Enquiry Scheme (HIPE) computerized hospital discharge records, referrals from community based services and general practitioners, through the Irish Motor Neuron Disease Association (a patient-support organization), and via patient self-referral to the National ALS clinic. The Central Statistics Office provides

information on all people who have MND or ALS recorded as primary, secondary or tertiary causes of death on their death certificates.

For inclusion on the Register, extensive confirmatory measures such as clinical examination by a specialist, direct chart review, and assessment by neurophysiologist, is required. Clinical progression is tracked by regular telephone contact between the register coordinator, health care professionals, patients and carers, and by home visits by members of the ALS Research Group. Extensive methods of ascertainment within Northern Ireland are similar to the Republic of Ireland Register, on which the design has been modelled [218,219].

Demographic data on incident ALS cases from RoI and NI, diagnosed between 1st January 2005 and 31st December 2010 were extracted from each Register and entered onto a PASW database (version 18). All cases had definite, probable or possible ALS using the revised El Escorial research criteria.

Variables recorded included sex, date of birth, date of first symptom onset, site of onset (bulbar/limb), date of diagnosis, attendance at a specialist ALS clinic, use of riluzole, use of non-invasive ventilation, use of gastrostomy, and date of death or censor.

Multidisciplinary ALS clinic in the Republic of Ireland

On each visit the patient and their carer are seen by a Neurologist with specialist expertise in ALS, a specialist ALS nurse and a neuromuscular multidisciplinary team including physiotherapist, occupational therapist, speech and swallow therapist, and dietician and direct next day access of Respiratory Medicine where indicated. Each visit takes between two and three hours.

Clinical assessment of respiratory function is made at each clinic visit, using same-day pulmonary function assessment and overnight oximetry, and if needed non-invasive ventilation is initiated at home by a specialist nurse. The physiotherapy service has expertise in respiratory impairment in addition to neurological disability, and provides on-site training in secretion management using techniques

including breath stacking and the use of cough assist machines. The speech and swallow therapist addresses the management of swallow to prevent aspiration pneumonia. Particular attention is paid to pharmacological management of secretions either with anti-cholinergic medications (amitriptyline, glycopyrronium), hyoscine patches, or in refractory cases with botulinum toxin injection or radiation to the salivary glands. Adequate nutrition is addressed at every clinic and patients are weighted, reviewed by the dietician and encouraged to take dietary supplementation. Patients with marked weight loss (greater than 10%), increasing dysphagia, early respiratory impairment or evolving bulbar symptoms are discussed by the multidisciplinary team during the clinic. Decisions regarding intervention including gastrostomy insertion, introduction of non-invasive ventilation, training in secretion management and introduction of cough assist are made collectively.

Three specialist ALS nurses (one hospital based, two community based and funded by the Irish Motor Neurone Disease Association) are available to make home visits to patients and their families to answer questions and to provide advice on management of ALS. The clinic works closely with community based clinical professional services (physiotherapy, speech and language therapy and occupational therapy) and palliative care services. On-site training is provided to community based clinical professionals. A rapid access service is provided for patients with increasing symptoms within the community.

Irish patients not attending the specialized multidisciplinary clinic attend general neurology clinics and do not access integrated specialist multidisciplinary care. Care is provided by community based clinical professionals with links to services within regional hospitals.

ALS/MND Care Network in Northern Ireland

In 2004 an ALS Care Network was established, funded initially by the Motor Neurone Disease Association of England, Wales and Northern Ireland and later by the Belfast health and social care trust. The ALS Care Network Coordinator, who has a nursing background, was appointed to coordinate the care of ALS patients

in NI from diagnosis to death. The Coordinator makes home visits, attends clinic appointments and is a source of support and education to patients and their local allied health professionals to ensure optimal timeliness and quality of care. A multidisciplinary ALS clinic, comparable to that in the RoI was not set up in NI until the end of the study period (2010).

Decision to introduce NIV and Gastrostomy

Criteria for introduction of NIV in the RoI are aligned to the EFNS Guidelines [220]. In the RoI, NIV is initiated either in hospital or at home by a specialist ALS nurse with training in management of respiratory impairment in ALS. In NI, patients are referred to the local respiratory physician, who initiates NIV according to local clinical practice and in accordance with NICE guidelines.

Radiological gastrostomy tube insertion is offered to patients in the RoI and NI to those experiencing a weight loss of greater than 10% of baseline or increasing dysphagia, and preservation of respiratory function (SNIP >40 and FVC >50% predicted). Gastrostomy is also offered to those with declining respiratory function following successful initiation of NIV in both jurisdictions. Gastrostomy is discussed with family members of those with severe cognitive and behavioural impairment and decisions are made on a case-by-case basis.

Statistical analysis:

Incidence rates reported for geographical area and age categories were manually corrected using the direct method to reflect the underlying population structure in RoI (www.cso.ie) and NI (http://www.nisra.gov.uk/).

Demographic and clinical characteristics of the participants are reported as percentages for categorical variables and mean or median values for continuous variable. Comparisons were made using Chi-square or Fisher Exact test or 2-sample t-test as appropriate.

To better model the survival effect associated with type of care post diagnosis (i.e. this effect does not exist between onset and diagnosis) [11], survival time was defined as time from diagnosis to death, co-varied for time from onset to diagnosis. Time from symptom onset to diagnosis was compared between groups to ensure lack of lead-time bias. Patients were followed up from their time of diagnosis until death or censor date (31st December 2010). Patients who were alive at the time of analysis were censored. The date for point prevalence was set at 31st December 2010.

In all survival analysis univariate assessment of the survival effect of categorical variables was carried out using Kaplan-Meier (KM) survival analysis and equality of outcome was assessed using the log-rank test. Multivariate survival analysis was performed using the Cox Proportional Hazards model. Multivariate survival analysis adjusted for time from onset to diagnosis, age at diagnosis, site of onset, sex, attendance at multidisciplinary ALS clinic, use of riluzole, use of gastrostomy, use of non-invasive ventilation (NIV), and use of both gastrostomy and NIV, and Irish or Northern Irish ALS patient. Separate multivariate models were built for NI vs RoI comparisons and for general/community care versus multidisciplinary care – this was because collinearity between these variables prevented them from being included in the same model.

Finally, to determine if distance from the specialist centre was a factor in survival for the RoI cohort, the distance from the centroid of each county to the specialist centre was calculated for each patient based on their county of residence at the time of diagnosis. The distance was then included in a multivariate Cox model using the following categories: < 50km, 50 – 100km, 100 – 200km, > 200km.

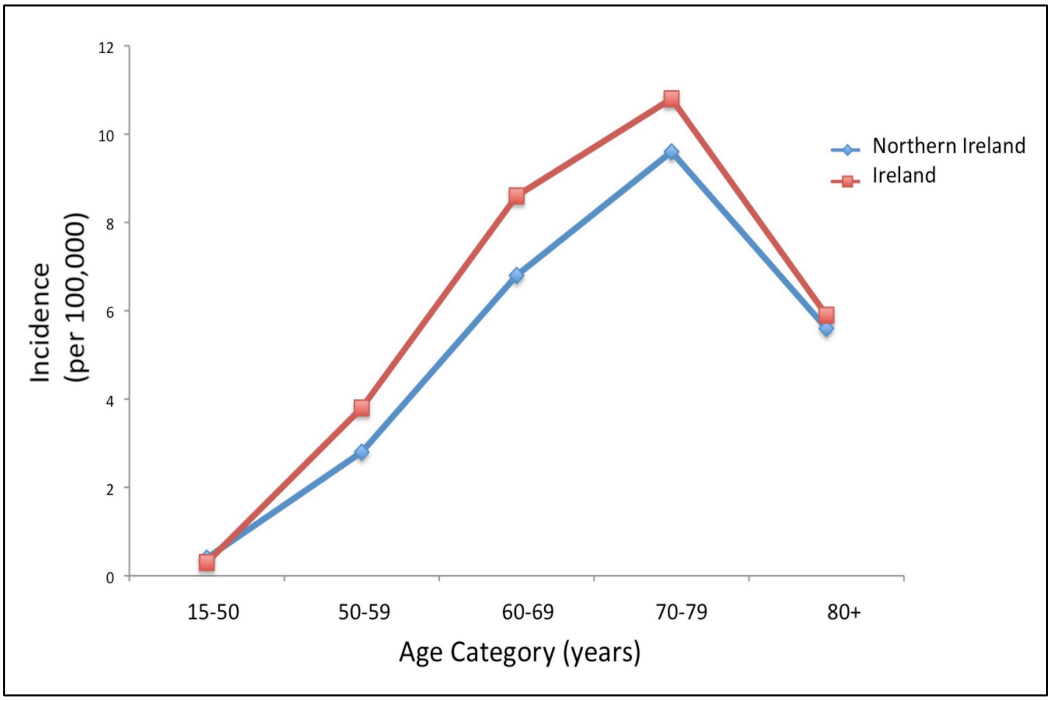
All tests were 2-tailed and statistical significance was set at $p < 0.05$. Statistical analysis was carried out using Stata version 11 [221]. Proportional hazards testing was performed using the *estat phtest* command in Stata.

4.1.1. Results

Five hundred and eleven incident patients from RoI and 208 people from NI were diagnosed with definite, probable, or possible ALS by El Escorial criteria in the six-year period from January 1st 2005 to December 31st 2010.

The corrected incidence of ALS for population aged fifteen years and above (Figure 4-1) was the same in both regions; RoI – 2.5 per 100,000 (95% CI 2.0-3.0); NI – 2.5 per 100,000 (95% CI 1.7-3.8). The corrected prevalence for population aged fifteen years and above on 31st December 2010 was 6.8 per 100,000 (95% CI 5.9-7.6) for RoI, and 6.4 per 100,000 (95% CI 5.1-7.7) for NI.

Figure 4-1 Corrected incidence per age category for Ireland compared[11]



[11] Figure 4-1 prepared by Dr Susan Byrne, Trinity College Dublin

Fifty five percent (283/511) of the RoI patients were male, compared to 51% (106) of the NI patients (p=0.281). There was no difference in the age at diagnosis for RoI men compared to NI men. However, RoI women were diagnosed with ALS at a significantly younger age than NI women (Table 4-1). There was no difference in the site of onset between the two geographic groups, nor was there any difference between the number of months between first symptom onset and a diagnosis of ALS.

Table 4-1 Demographic information from ALS patients in Ireland and Northern Ireland

Variable	Ireland General care (N = 169)	Ireland: MDT (N= 340)	P value	Ireland: all cases (n = 511)	Northern Ireland (n = 208)	P value: RoI general vs NI	P value: RoI MDT vs NI	P value: all RoI vs NI
Age at diagnosis in years (SD)								
All patients	68.6 (10.8)	62.8 (11.1)	<0.001	64.7 (11.3)	65.9 (11.6)	0.021	0.002	0.219
Male	66.3 (12.0)	61.8 (10.7)	0.002	63.4 (11.2)	62.6 (11.6)	0.027	0.571	0.544
Female	71.6 (8.1)	64.0 (11.5)	<0.001	66.4 (11.1)	69.3 +/- 10.5	0.126	0.000	0.027
Site of Onset								
Limb	50% (82)	67% (222)		63% (306)	66.5% (135)			
Bulbar	48% (78)	32% (104)		37% (182)	33.5% (68)			
Both	2% (3)	1% (2)	<0.001¶	1% (5)	0% (0)	0.001¶	0.625¶	0.278¶
Mean Time of symptom onset to diagnosis (months)								
All	12.7	13.4	0.6113	13.3	14.9	0.1486	0.2136	0.164
Male	14.5	13.2	0.0440	13.7	14.5	0.9997	0.4594	0.667
Female	10.5	13.6	0.5403	12.7	15.3	0.0089	0.3159	0.081

¶ Fischers exact test

Interventions

80% of patients from the RoI (398/496) and 90% of patients from NI (187/207) received riluzole (p = 0.001). There was no significant difference in the use of gastrostomy between the two regions (26.7% in RoI (136/510) and 35% in NI (52/208) p=0.65). Neither did the time from diagnosis to gastrostomy tube insertion differ between the two cohorts (8.6 months in RoI patients compared to 8.5 months in NI patients, p=0.93). Pro-rata twice as many RoI patients were initiated on non-invasive ventilation (30.9%, 158/511), compared to NI patients (14.4%, 30/208, p<0.001). There was no difference in the time between diagnosis and initiation of NIV between RoI and NI patients (9.5 months compared to 10.1 months, p=0.81). A larger proportion of RoI patients received both gastrostomy tube and NIV (14.3%, 73/511) compared to those receiving care in NI (2.4%, 5/208, p<0.001).

In the RoI, 340 ALS patients (66.5%) attended a multidisciplinary ALS clinic, while the remainder (169) attended local neurology services. The NI cohort also attended local neurology services but within the setting of an established and integrated care network supported by a trained care worker. Of those who attended the specialist multidisciplinary clinic in the RoI, 22.9%, (78/340) received gastrostomies compared to 29.2%, (110/376) of those attending general clinics – a finding of borderline significance (P = 0.055).

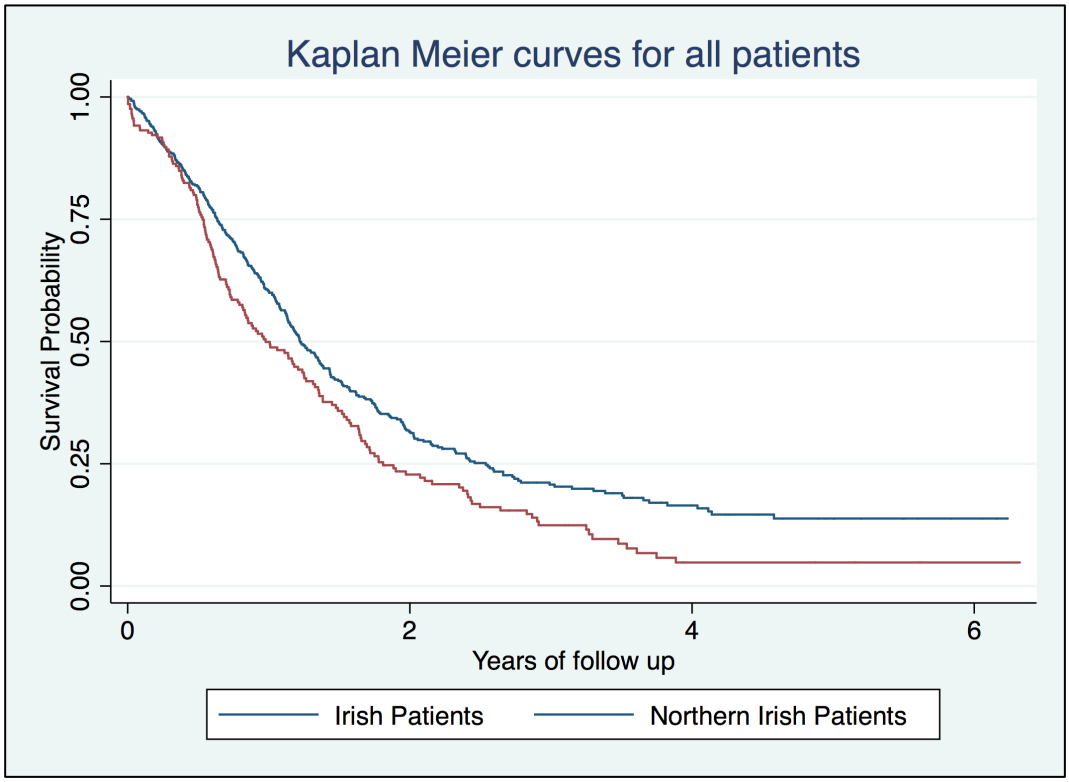
Patients attending the RoI multidisciplinary clinic were more likely to use NIV (129/340 (38%) compared to those attending general clinics in RoI or availing of integrated community care in NI (58/376 (15.4%). (P <0.001).

Additionally, significantly more patients attending a multidisciplinary ALS clinic received both gastrostomy tube and NIV (15.9%, 54/340) compared to patients attending a general (RoI) or regionalized (NI) service (6.4%, 24/376, p <0.001).

Survival in Republic of Ireland compared to Northern Ireland

Univariate analysis comparing survival time for all patients within RoI and NI demonstrated a survival advantage in patients from RoI with a median survival from diagnosis to death of 1.22 years compared to 0.98 years in patients from NI (RR = 1.34, p=0.0019) (Figure 4-2). NI bulbar onset patients (median survival 0.88 yrs 95% CI 0.64 - 1.44) showed a modest improvement in median survival compared to RoI bulbar onset patients attending the general service (median survival 0.82yrs 95% CI 0.64 - 1.04).

Figure 4-2 Survival difference between Ireland and Northern Ireland



Multivariate analysis, adjusting for country (RoI, NI), time from onset to diagnosis, age at diagnosis, site of onset, sex, use of riluzole, use of gastrostomy, use of non-invasive ventilation (NIV) (Table 4-2), demonstrated that RoI patients continued to experience a significant survival advantage (HR 0.72, 95% CI 0.59 - 0.87, P < 0.001), where a hazard rate of less than 1.0 confers a survival benefit.

Table 4-2 Hazard ratios from Cox multivariate models

		NI vs RoI		MDT vs general care	
Variable	**Strata**	**Hazard Ratio (95% CI)**	**P value**	**Hazard Ratio (95% CI)**	**P value**
Age at diagnosis	Per 10 years	1.40 (1.32 - 1.58)	<0.001	1.40 (1.28 - 1.53)	<0.001
Diagnostic Delay (weeks)	< 31	1	-	1	-
	31 – 55	0.96 (0.78 - 1.20)	0.745	0.96 (0.77 – 1.19)	0.695
	> 55	0.66 (0.52 - 0.83)	<0.001	0.67 (0.53 – 0.84)	0.001
Country	NI	1	-	-	-
	RoI	0.72 (0.59 - 0.87)	0.001	-	-
MDT	No	-	-	1	-
	Yes	-	-	0.59 (0.49 - 0.71)	<0.001
Sex	Female	1	-	1	-
	Male	1.07(0.89 - 1.39)	0.460	1.02 (0.84 – 1.24)	0.829
Site of onset	Limb	1	-	1	-
	Bulbar	1.06(0.85 - 1.31)	0.611	1.04 (0.84 – 1.29)	0.701
	Both	1.73 (0.64 - 4.71)	0.283	1.80 (0.66 – 4.89)	0.249
Riluzole use	No	1	-	1	-
	Yes	0.61 (0.48 - 0.78)	<0.001	0.68 (0.54 – 0.87)	0.002
Gastrostomy	No	1	-	1	-
	Yes	0.98 (0.78 - 1.24)	0.893	0.90 (0.71 – 1.13)	0.355
NIV	No	1	-	1	-
	Yes	1.23 (1.00 - 1.51)	0.048	1.38 (1.12 – 1.69)	0.003

Survival models of RoI patients only including distance from county centroid showed that the distance from the specialist centre was not a factor in survival (Table 4-3).

Table 4-3 Hazard ratios for patient county centroid distance to specialist centre

Distance from county centroid to specialist clinic	Hazard Ratio (95% CI)*	P value
< 50km	1	-
50 – 100km	1.03 (0.66 – 1.61)	0.882
100 – 200km	0.98 (0.74 – 1.29)	0.873
> 200km	0.99 (0.74 – 1.32)	0.940

*** Adjustment for age at diagnosis, diagnostic delay, sex, site of onset, riluzole use, gastrostomy, NIV and attendance at the MDT. These results were not significantly altered by stratifying the model by MDT, or by including patients attending the MDT only.**

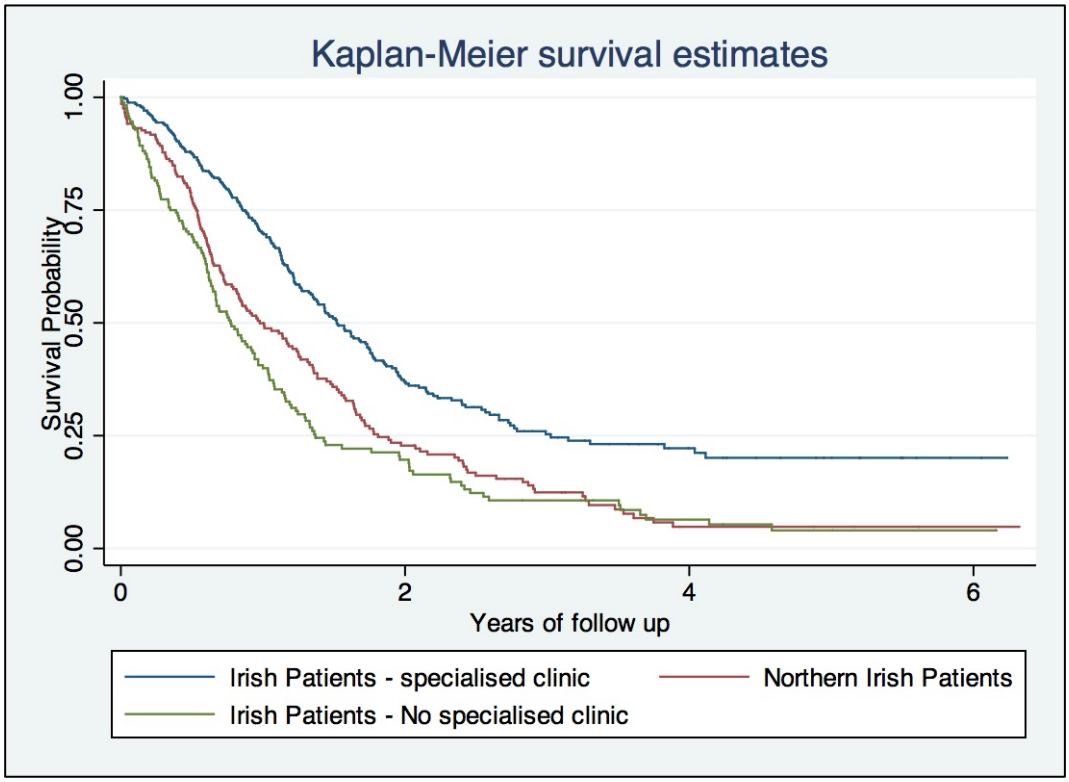
Survival based on attendance at a specialized ALS multidisciplinary clinic

RoI patients who attended the multidisciplinary ALS clinic had a younger age of disease onset compared to those who did not attend (62.7 years compared to 68.7 years, p<0.0001). Multivariable analysis adjusting for time from first symptom onset to diagnosis, age at diagnosis, site of onset, sex, use of riluzole, use of gastrostomy, use of non-invasive ventilation (NIV), was used to compare survival in RoI patients who attended a multidisciplinary ALS clinic with RoI and NI patients who did not attend a centralized multidisciplinary ALS clinic (Table 4-2). There was a demonstrable survival benefit for patients who attended the multidisciplinary ALS clinic compared to RoI and NI patients who did not attend the clinic (HR 0.59, 95% CI 0.49-0.71, p <0.001). Although RoI patients attending the multidisciplinary clinic were more likely to use NIV, this survival benefit could not be accounted for by NIV use alone, gastrostomy, or combined use of NIV and gastrostomy, as previously demonstrated (HR 0.72, 95% CI 0.61–0.85) [11].

Uni-variate analysis showed no significant difference between ROI patients who attended general neurology ALS clinics and NI patients under the care of the ALS Care Network Coordinator (RR = 1.21 95% CI: 0.97 – 1.51 p = 0.09). However, Kaplan Meier

survival analyses (figure 3) showed a trend towards modest survival advantage of NI patients over general care RoI patients for up to 3 years of follow up.

Figure 4-3 Survival difference between all three care modalities



4.1.2. Discussion

The island of Ireland is separated by a political divide, which means that the hospital systems and ALS Registers in the RoI and NI are entirely independent of one another. Despite this, population demographics, rates of capture and epidemiological features of ALS patients are remarkably similar, reflecting the common ancestral origins of patients in both jurisdictions. This unique separation has permitted the design of population based survival study comparing two cohorts, differences between which are primarily a function of the types health services to which they have access. Our findings suggest that the provision of a centralized multidisciplinary clinic confers a distinct survival advantage over the provision of an integrated community based care system for ALS patients.

We used a process of multivariable analysis to co-vary for factors that are likely to differentiate patients who attend a multidisciplinary specialist clinic from those who do not, including time from symptom onset to diagnosis, and age at diagnosis. The survival advantage of 8 months experienced by patients attending a multidisciplinary clinic was not explained in full by the use of interventions or medication. We postulate that the survival benefits accrued from the multidisciplinary clinic experience relate in part to the complex decision making processes that take place both within the clinic, and in communication with those in a community setting. We hypothesise that these multiple decision making processes involve the incorporation of different perspectives on the patient journey, and accordingly lead to an enriched set of clinical encounters for the patient and carer that are then incorporated into the care plan.

Timing of gastrostomy insertion and NIV initiation did not differ between specialist clinic and non-specialised services, however, a larger number of patients from the specialised service used both simultaneously. 38% of RoI patients attending the specialist clinic had access to NIV compared to 15.4% in a regionalised service. However, multivariable analysis did not associate the use of NIV and gastrostomy with increased survival. Whilst this contrasts with clinical trials on NIV [77], the finding is in keeping with recent analysis

of the RoI cohort from 1995 – 2010 [11] and may be explained in part by statistical power, and in part by unmeasured confounders including compliance with NIV use and variance in settings.

Although we could not demonstrate that patients in NI experienced an increased survival compared to those in the RoI seen only in the general clinic, the confidence limits on these estimates were wide and may be a function of reduced power. This is supported by the trend towards survival advantage of NI patients over general care RoI patients for up to 3 years of follow up, and suggests that the role of the ALS Network Coordinator in supporting and educating patients and health professionals in ALS management is important.

It was not possible to determine whether the integrated care service in Northern Ireland provided further advantages with respect to quality of life or carer burden. A previous study in the Netherlands suggested that within the context of multidisciplinary ALS care teams, case managers conferred no extra benefit for patients with ALS or their caregivers [222]. Although intuitively the case, it remains to be established as to whether the availability of a care worker in the absence of a centralized clinic also provides quantifiable benefits such as improved quality of life and reduced carer burden.

Differences in social deprivation and smoking are unlikely to explain the findings as mapping of all Ireland social deprivation revealed less extremes of social deprivation in NI in comparison to RoI [223], and smoking prevalence amongst adults are in fact lower in NI (18.7%) [224] compared to RoI (21.5%) [225]. It is possible that cultural and religious differences between NI and RoI may have an influence on survival. However, such an effect size is likely to be too small to determine with a study of this power. We demonstrated that distance to the specialist centre was not a factor in survival, although this analysis was not performed with a high spatial resolution. Further work at higher spatial resolution to confirm this and findings by others of a coastal/urban effect on survival [102] is planned.

This study has limitations. It was not possible to quantify the impact of individual disciplines in the RoI multidisciplinary clinic to determine whether a specific intervention

contributed to the survival advantage. It is possible that those not attending the centralized clinic exhibited other negative prognostic indicators that were not taken into account. We have shown previously that executive impairment is predictive of shorter survival [1], and it was not possible to include cognitive status as a variable in this analysis as cognitive data were not routinely collected in all patients during the study period. However, as the demographic features of the two populations were similar in all other regards, we believe this to be an unlikely explanation for the differences.

Our data supports our previous findings and those of others, that attendance at a specialist multidisciplinary ALS clinic is beneficial to patients, and should be considered a standard of care in the management of ALS. The precise advantage conferred by the multidisciplinary clinic remains to be established, but is likely related to the complex decision making processes from differing vantage points that take place among experienced professionals within the clinic. Further work to deconstruct the functionality of the multidisciplinary clinic is required to test this hypothesis. In the absence of a centralized specialist clinic, the availability of a specialist care network coordinator may be advantageous.

4.2. Survival analysis of geospatial factors in the Irish ALS cohort[12]

4.2.1. Introduction:

The clinical phenotype of ALS is heterogeneous. There is evolving evidence of a multi-step pathogenic process in which both genetic and environmental factors are likely to play a role [9], although the nature of any gene-environment interactions in pathogenesis and disease course remain to be determined. From an environmental perspective, exposures that increase the incidence risk of ALS, might also increase the rate of progression of the disease and hence impact negatively on survival. However, few studies have investigated the effect of spatially structured environmental factors on ALS survival.

In this study, we sought to explore this area by examining the influence of several geospatial factors on the survival of Irish ALS patients. Geospatial factors examined were population density (previously associated with ALS survival) [102], social deprivation (known to be an important factor in survival in breast cancer and other diseases [226,227]), distance of address of usual residence to the coast (coastal residency was previously reported to be associated with improved survival)[102] and finally distance to the multidisciplinary ALS care team (MDT) (patients attending the MDT have improved survival) (Chapter 4.1).

[12] Chapter 4.2 has been published as:

Rooney J, Heverin M, Vajda A, Burke T, Galvin M, Tobin K, et al. Survival analysis of geospatial factors in the Irish ALS cohort. Amyotroph Lateral Scler Frontotemporal Degener. 2016;17(7–8):555–60.

4.2.2. Methods

Data sources

The survival characteristics of the Irish ALS cohort have been well described in Chapter 4.1 and previous publications [11]. For the current analysis an updated survival analysis data-set was generated with follow-up on all patients to at least July 2014. Due to the previously described effects of missing values in the early years of data collection [11] and the need to allow sufficient follow-up time for analysis of survival, we included only those patients diagnosed between 1st January 1998 and 31st December 2012.

To examine the effect of geospatial factors on survival, we used the location of the address of the usual residence at the time of diagnosis to determine the values of various geospatial variables for each patient. The geospatial variables were: population density of the local small area; distance from the coast; distance from the location of the multidisciplinary clinic and small area (SA) social deprivation (small areas are areas of maximum 200 households each for which census data is available). Distances were determined 'as the crow flies', whilst population density and social deprivation data were obtained from the Pobal HP social deprivation index for 2006 at the SA geometry [195]. The Pobal HP social deprivation index uses Census data to quantify social deprivation using mulitple parameters falling under 3 core dimensions: Demographic Profile, Social Class Composition and Labour Market Situation, quantified for each small area [195]. These core dimensions are in turn drawn from a larger number of census variables designed to capture information regarding demographics, educational achievement, professional achievement, living circumstances, unemployment rates and others [195].

Date of disease onset was used to mark the time of entry to the survival analysis as any effects on survival by geospatial variables may have existed prior to diagnosis, and survival was followed up to December 2014.

Statistical Analysis

We have previously shown that being in the 'definite' El Escorial category, the presence of executive impairment, bulbar onset disease, diagnostic delay, age at onset, riluzole prescription, and attendance of the multidisciplinary clinic are important factors in the survival in the Irish ALS cohort (Chapter 4.1 & [1,11]). Therefore, we constructed a base multivariate survival model including all the aforementioned factors, with the exception of executive dysfunction. Diagnostic delay was included as a categorical variable as a marker of disease progression rate (diagnostic delay data was categorized as three equally sized groups).

Categorisation of the geo-spatial variables: population density, distance from the National ALS Clinic and distance from the coast was performed after consideration of the distribution of the variables and considering the geographic distribution of the patients. For areal social deprivation (Pobal HP index), strata were obtained from the distribution of all small areas based on the authors definitions [195]. The strata were as follows: -10 or lower considered deprived, -10 to +10 considered average, +10 or over considered affluent.

Survival analysis was performed using Royston-Parmar flexible parametric regression, which has several advantages over the more usual Cox regression, both in modelling the underlying hazard, dealing with breaches of the proportional hazards assumption and for displaying regression results [228]. It has also previously proven useful in modelling ALS survival [11]. Therefore, we constructed models using this method. Proportional hazards were used as the model scale and comparisons of AIC and BIC were used to choose the number of degrees of freedom as 3 for the cubic spline knots [228].

Next successive models were built including each spatial variable in turn. Models were compared to the base model using likelihood ratio testing and by examining the effects on hazard ratios. Stata: Release 14 [229] was used for all statistical analysis.

4.2.3. Results

After exclusions, 1,232 patients met the inclusion criteria, with 693 (56%) male, 724 (59%) spinal cases and 7.5% were reported as being familial, (with an apparent Mendelian pattern of inheritance for ALS, and affecting at least 2 family members). Table 4-4 shows a summary of their baseline characteristics and the crude survival data including unadjusted hazard ratios. For all of the geospatial variables analyzed, the unadjusted hazard ratios had 95% CI overlapping 1 (no deficit or surplus hazard), and so did not offer any explanatory power. Table 4-5 shows the hazard ratios for the covariates included in the base model, along with the hazard ratios for each of the geospatial variables added to the base model in turn. None of those variables were found to significantly improve the model fit by likelihood ratio testing – whether as categorical or continuous variables.

As shown in Figure 4-4 which displays the Kaplan Meier survival curve, confidence intervals and the mean survival from the final selected model, they are in close agreement, while Figure 4-5 displays hazard ratio by age of onset for the model taking 60 years as a reference age.

Our findings were in close agreement with our previous analyses of the Irish ALS cohort (Chapter 4.1 & Ref. [11]).

Table 4-4 Baseline demographic and survival characteristics by geographical factor

Variable	Strata	N= (1,232)	Crude median survival time in years (95% CI)	Crude Hazard ratio (95% CI)
Sex	Female	539 (44%)	2.23 (2.13 – 2.42)	1
	Male	693 (56%)	2.37 (2.23 – 2.51)	0.84 (0.75 -0.95)
Site of onset	Spinal	724 (59%)	2.68 (2.51 – 2.88)	1
	Bulbar	432 (36%)	1.96 (1.79 – 2.11)	1.59 (1.4 - 1.8)
	Other	61 (5%)	2.44 (1.60 – 2.91)	1.21 (0.92 – 1.60)
Familial Disease	Sporadic	1140 (92.5%)	2.29 (2.21 – 2.40)	1
	Familial	92 (7.5%)	2.60 (1.99 – 2.95)	1.17 (0.93 – 1.47)
El Escorical category	Suspected	16 (1%)	5.33 (1.69 – NA)	0.7 (0.37 – 1.32)
	Possible*	161 (14%)	2.92 (2.54 – 3.34)	1
	Probable	325 (28%)	2.45 (2.21 – 2.68)	1.20 (0.98 – 1.47)
	Definite	671 (57%)	2.16 (2.04 – 2.25)	1.57 (1.30 – 1.89)
Attended MDT ?	No	410 (33%)	1.83 (1.70 – 2.00)	1
	Yes	822 (67%)	2.55 (2.43 – 2.68)	0.66 (0.58 – 0.75)
Diagnostic Delay	< 31 weeks	423 (34%)	1.52 (1.33 – 1.62)	1
	31 – 55 weeks	412 (34%)	2.17 (2.03 – 2.26)	0.69 (0.60 -0.79)
	> 55 weeks	390 (32%)	3.50 (3.22 – 3.71)	0.43 (0.37 – 0.50)
Prescribed riluzole ?	No	250 (20%)	1.84 (1.64 – 2.01)	1
	Yes	900 (73%)	2.48 (2.38 – 2.61)	0.71 (0.62 – 0.82)
	Unknown	82 (7%)	2.19 (1.56 – 2.81)	0.75 (0.57 – 0.98)
Distance from MDT	< 100 km	600 (49%)	2.26 (2.16 – 2.43)	1
	100 – 200 km	360 (29%)	2.35 (2.16 – 2.55)	0.97 (0.84 – 1.11)
	> 200km	272 (22%)	2.38 (2.09 – 2.64)	0.99 (0.86 – 1.16)
Small area social deprivation	Disadvantaged	231 (19%)	2.26 (2.03 - 2.61)	1.02 (0.88 – 1.19)
	Average	868 (70%)	2.29 (2.19 - 2.42)	1
	Affluent	133 (11%)	2.50 (2.23 – 2.80)	0.93 (0.77 – 1.13)
Population density	< 2000 /km^2	694 (56%)	2.39 (2.23 – 2.55)	1
	2000 – 7500 / km^2	417 (34%)	2.26 (2.15 – 2.43)	1.10 (0.97 – 1.25)
	> 7500 /km^2	121 (10%)	2.19 (1.93 – 2.61)	1.07 (0.87 – 1.31)
Distance from coast	< 10km	680 (55%)	2.26 (2.15 – 2.45)	1
	10 – 30km	258 (21%)	2.39 (2.22 – 2.66)	0.89 (0.77 – 1.04)
	> 30km	294 (24%)	2.32 (2.13 – 2.52)	0.93 (0.81 – 1.08)

Table 4-5 Hazard ratios for multivariate Royston Parmer base model on hazard scale with 3 degrees of freedom

Variable	Strata	Hazard Ratio	P value (Wald test)
Age at onset (per 10 yrs)		1.26 (1.18 – 1.32)	**<0.001**
Diagnostic delay	< 31 weeks	1	
	31 – 55 weeks	0.64 (0.55 – 0.74)	**<0.001**
	> 55 weeks	0.33 (0.29 – 0.39)	**<0.001**
Site of onset	Spinal	1	
	Bulbar	1.50 (1.32 – 1.72)	**<0.001**
	Other	0.92 (0.68 – 1.24)	0.581
El Escorial	Suspected	0.64 (0.34 – 1.22)	0.175
	Possible	1	
	Probable	1.21 (0.98 – 1.49)	0.071
	Definite	1.46 (1.21 – 1.77)	**<0.001**
Attended MDT?	No	1	
	Yes	0.69 (0.60 – 0.79)	**<0.001**
Prescribed riluzole	No	1	
	Yes	0.77 (0.65 – 0.90)	**0.002**
	Unknown	0.81 (0.58 – 1.13)	0.206
Hazard ratios for geospatial variables individual added to base model in turn			
Distance from MDT	< 100 km	1	
P =0.281	100 – 200 km	1.14 (0.98 – 1.31)	0.081
	> 200km	1.03 (0.88 -1.21)	0.682
Small area social deprivation	Disadvantaged	1.05 (0.90 – 1.24)	0.498
P = 0.136	Average	1	
	Affluent	0.84 (0.69 – 1.03)	0.086
Population density	< 2000 /km^2	1	
P = 0.947	2000 – 7500 / km^2	1.02 (0.89 – 1.16)	0.816
	> 7500 /km^2	1.03 (0.83 – 1.27)	0.778
Distance from coast	< 10km	1	
P = 0.824	10 – 30km	0.95 (0.81 – 1.12)	0.549
	> 30km	0.98 (0.84 – 1.14)	0.803

Figure 4-4 Comparison of Kaplan-Meier and Royston-Parmar survival curves for the Irish ALS cohort 1998 – 2012

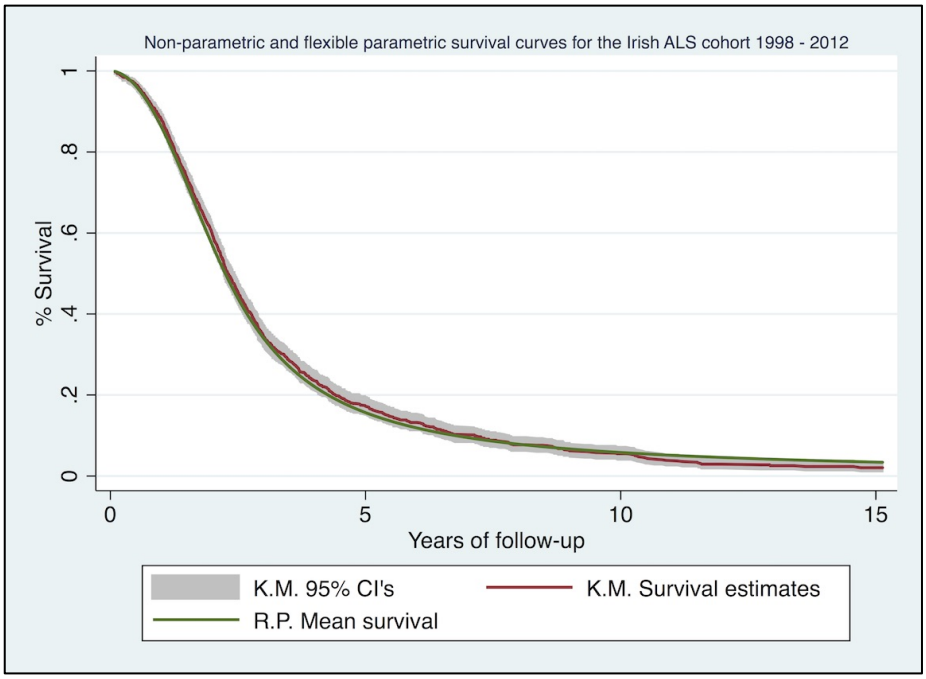
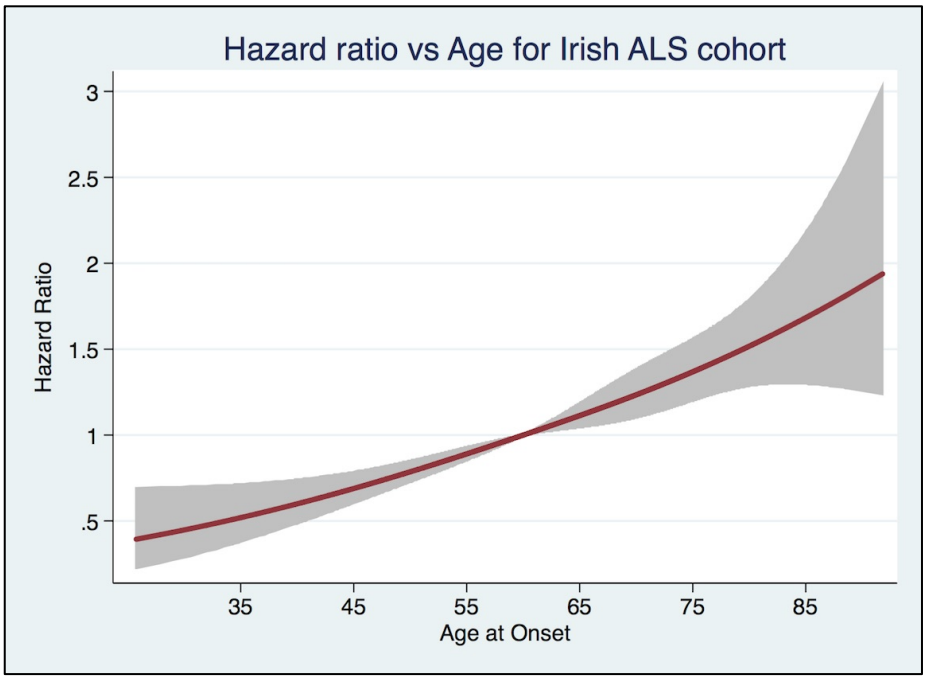


Figure 4-5 Hazard Ratio vs age from final multivariate model



Age = 60yrs is taken as the reference age

4.2.4. Discussion

Our findings imply that none of the geospatial variables were associated with survival in the Irish cohort which contrasts with results from the UK [102]. Survival analysis of 933 patients from the SEALS cohort in the South East of England found that living by the coast was associated with improved survival, whilst living in London was associated with a poor prognosis [102]. Factors that might confer a poorer prognosis for those dwelling in London have not been established, but could be related to social deprivation, known to be higher in London than along the South-East coast [230], or other environmental factors such as pollution [102]. Our study has several advantages over the UK study including a fully prospective cohort, a larger cohort size and higher spatial resolution. While the UK study is based on relatively large postcode areas, our geospatial variables were calculated from over 18,000 small areas with 50 - 200 households each [195]. We have also confirmed our own recently published finding that distance from the MDT was not associated with ALS survival – this time with higher spatial accuracy (Chapter 4.1).
Ireland has a socialized health care system where riluzole prescription and MDT attendance are both free of charge to the patients. This fact may, at least partially, explain our observation that small area social deprivation was not associated with overall survival in the Irish population. This finding is also congruent with our recent finding that social deprivation did not influence ALS incidence (Chapter 3.3). A recent study investigating an ALS cohort in New Jersey reported that socioeconomic status was associated with ALS incidence [130]. This contrasts with our findings and suggests that the relationship between social deprivation and ALS incidence may differ between countries with different health-care models.

Despite our high spatial resolution, it is important to be aware of the ecological fallacy and to remember that small area/neighbourhood social deprivation and individual level social position could potentially be distinct prognostic factors associated with disease progression and survival. Strengths of this study include a large prospective population based cohort, high spatial resolution for spatial variables and the use of Royston-Parmar flexible parametric regression. The generalizability of the results is limited however, since effects from environmental factors such as population density, distance from the

MDT or social deprivation may vary by geography and by healthcare system. Furthermore, distances have been calculated 'as the crow flies' – actual travel times may be affected by other factors such as terrain and the quality of travel services.

In summary, none of the four geospatial factors were significantly associated with ALS survival. This confirms our previous finding that distance from the MDT clinic was not associated with survival.

4.3. *C9orf72* expansion differentially affects males with spinal onset Amyotrophic Lateral Sclerosis[13]

4.3.1. Introduction

Amyotrophic lateral sclerosis (ALS) is a debilitating disease with a poor prognosis. Progress towards developing new treatments has been limited both by disease heterogeneity, and by the likely interaction between genetic and environmental factors in disease pathogenesis [199]. A pathological expansion of a hexanucleotide repeat in the *C9orf72* gene [3,4] accounts for up to 10% of those with ALS in populations of European extraction, and is associated with a distinctive clinical phenotype that includes fronto-temporal dementia in some instances [88–94]. Although the *C9orf72* repeat expansion has been shown to be an important negative prognostic factor in survival analyses [88–94], to date no study has been sufficiently large to permit robust analysis of interactions between the variant and demographic features including age, gender and site of onset. Here we used our combined clinical datasets to determine whether the presence of the expanded variant differentially modulates survival based on gender and site of onset.

4.3.2. Methods

Data sources /Case capture

Clinical data from ALS cases incident from January 2000 to April 2015 were collected from Belgium, Ireland, Italy, The Netherlands and the United Kingdom. All patients fulfilled the diagnostic criteria for ALS, and core data elements as defined by the ENCALS consortium were harmonised across data-sets for consistency based on existing

[13] Chapter 4.3 has been published as:

Rooney J, Fogh I, Westeneng H, 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 23;88(4):281

consortia agreements [231]. Due to the lack of an agreed international definition of 'familial ALS', and given a previous population based familial aggregation analysis from Ireland demonstrated a much higher familial ALS occurrence (16%) than usually recognised [14], we did not exclude cases based on family history. The Belgian and UK cases were collected from clinical research centre cohorts, while the Dutch, Irish and Italian cases were sourced from the prospective population based national registers [12,104,111,133,200,232,233].

In accordance with existing Consortia agreements, data were collated using the following variables: age of onset, date of onset, date of diagnosis, date of death / last known follow-up date, site of onset, revised El-Escorial diagnostic category (except Belgium), and *C9orf72* status (normal or expanded). For all study participants, *C9orf72* status was determined by repeat-primed PCR as described previously (with individual laboratory-based validation and quality control by Southern blot analyses) [4].

Survival analysis strategy

Initially exploratory models were constructed using Cox proportional hazards regression to explore the effect of different time of entry to the studies. Cox models were generated including known important survival covariates including age of onset, site of onset, diagnostic delay and *C9orf72*. Cox models were compared using a likelihood ratio test, and by testing the validity of the proportional hazards assumption of each covariate at each timescale.

A base model using Royston-Parmar flexible parametric regression [228] was built on the preferred timescale, with a proportional hazards scale and a number of degrees of freedom selected by comparison of the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) from models with increasing degrees of freedom, and the variance covariance matrix clustered by country. Survival follow-up was limited to 5 years from entry. Models were then built to explore the effect of *C9orf72* status in sex and site of onset subgroups. The *stpm2* [234] and *ipdmetan* [235] commands from Stata MP 14.0 were used to perform the survival analysis and produce the meta-analysis

graphs, whilst the *ggplot2* [236] package in R 3.1.1 was used to generate selected final graphs.

4.3.3. Results

Descriptive statistics and basic survival model

In total, 5,106 ALS cases met the inclusion criteria, of which 457 (8.95%) carried the *C9orf72* repeat expansion. Breakdown of the demographics of the overall cohort by country is shown in Table 4-6. Missing values were minimal, affecting 181 cases (3.5%). Appendix B summarizes the basic survival model. An individual patient data (IPD) meta-analysis of *C9orf72* status in the base model estimated a hazard ratio of 1.36 (1.18 – 1.57) for those carrying the expansion vs those not.

Table 4-6 Baseline demographics of C9orf72-tested ALS cases by country

Variable	Belgium n=482	Ireland n=645	Italy n=897	The Netherlands n=2153	UK n=929	p Value($\chi^{2)}$	Combined N=5106
Included after missing values removed	477 (99.0%)	640 (99.2%)	867 (96.7%)	2037 (94.6%)	904 (97.3%)		4925 (96.5%)
Sex							
Female	180 (37.3)	266 (41.6%)	411 (47.4%)	844 (41.4%)	352 (38.9%)		2053 (41.7%)
Male	297 (62.3)	374 (58.4%)	456 (52.6%)	1193 (58.6%)	552 (61.1%)	<0.001	2872 (58.3%)
Age of onset							
Median	61.4	63.0	66.8	63.0	61.3		63.1
Diagnostic delay							
Short	161 (33.8%)	193 (30.2%)	285 (32.9%)	702 (43.5%)	300 (33.2%)		1641 (33.3%)
Medium	163 (34.2%)	211 (33.0%)	293 (33.8%)	683 (33.5%)	298 (33.0%)		1648 (33.5%)
Long	153 (32.1%)	236 (36.9%)	289 (33.3%)	652 (32%)	306 (33.8%)	0.559	1636 (33.2%)
Site of onset							
Spinal	326 (68.3%)	437 (68.3%)	583 (67.2%)	1333 (65.4%)	610 (67.5%)		3289 (66.8%)
Bulbar	151 (31.7%)	203 (31.7%)	284 (32.8%)	704 (34.6%)	294 (32.5%)	0.540	1636 (33.2%)
C9orf72							
Normal	392 (82.2%)	578 (90.3%)	805 (92.8%)	1861 (91.4%)	841 (93.0%)		4477 (90.9%)
Expanded	85 (17.8%)	62 (9.7%)	62 (7.2%)	176 (8.6%)	63 (7%)	<0.001	448 (9.1%)

Diagnostic delay is defined by three tertiles per country labelled 'short', 'medium' and 'long' diagnostic delay to allow for variation in diagnostic delay between countries.
ALS, amyotrophic lateral sclerosis.

C9orf72, gender and site of onset subgroup analysis

Survival curves were generated to evaluate the effect of *C9orf72* status on gender and site of onset (Figure 4-6) which suggested three-way interaction. Therefore, gender, site of onset and *C9orf72* status were categorised into one variable with eight levels as demonstrated in Table 4-6. Through comparison of survival curves, redundant sub-

groups were combined (Appendix B) leaving 3 groups: spinal onset males with the *C9orf72* expansion, other spinal onset patients, and all bulbar onset patients.

Figure 4-6 Predicted survival by sex and C9orf72 status after multivariate Royston-Parmar regression including interaction terms

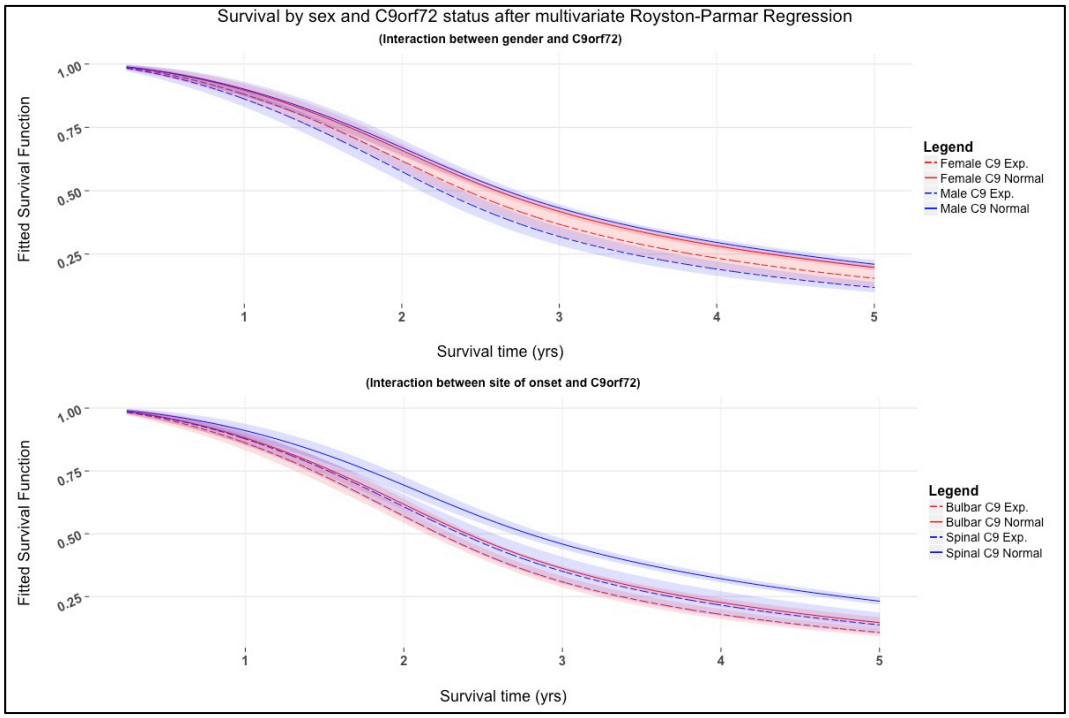


Figure 4-6 Legend: C9 Normal = group not carrying the C9orf72 expansion; C9 Exp. = group carrying the C9orf72 expansion present. Shaded areas represent 95% confidence intervals. Predicted survival curves for interaction models between gender and *C9orf72* status (upper), and site of onset and C9orf72 status (lower) after multivariate regression using a Royston Parmar model on the hazard scale (3 d.f.), correcting for age of onset (time varying), diagnostic delay group (time varying), site of onset, country and using a variance-covariance matrix clustered by country. The upper graph shows a wider spread by *C9orf72* status in males compared to females, while the lower shows a wider spread in spinal onset cases vs bulbar onset cases – however hazard ratios for these interactions were not significant.

Survival curves for these groups showed that male spinal onset patients with the *C9orf72* repeat expansion had a prognosis distinct from other spinal onset patients and similar to bulbar onset patients (Figure 4-7). Meta-analysis calculated a survival hazard ratio of 1.56 (95% CI: 1.25 – 1.96) for male spinal onset patients with the *C9orf72* repeat expansion (Figure 4-8). The finding was in the same direction in each country, although only the pooled estimate was statistically significant (Figure 4-8).

Figure 4-7 Predicted survival function for male spinal onset C9orf72 expanded patients vs other groups

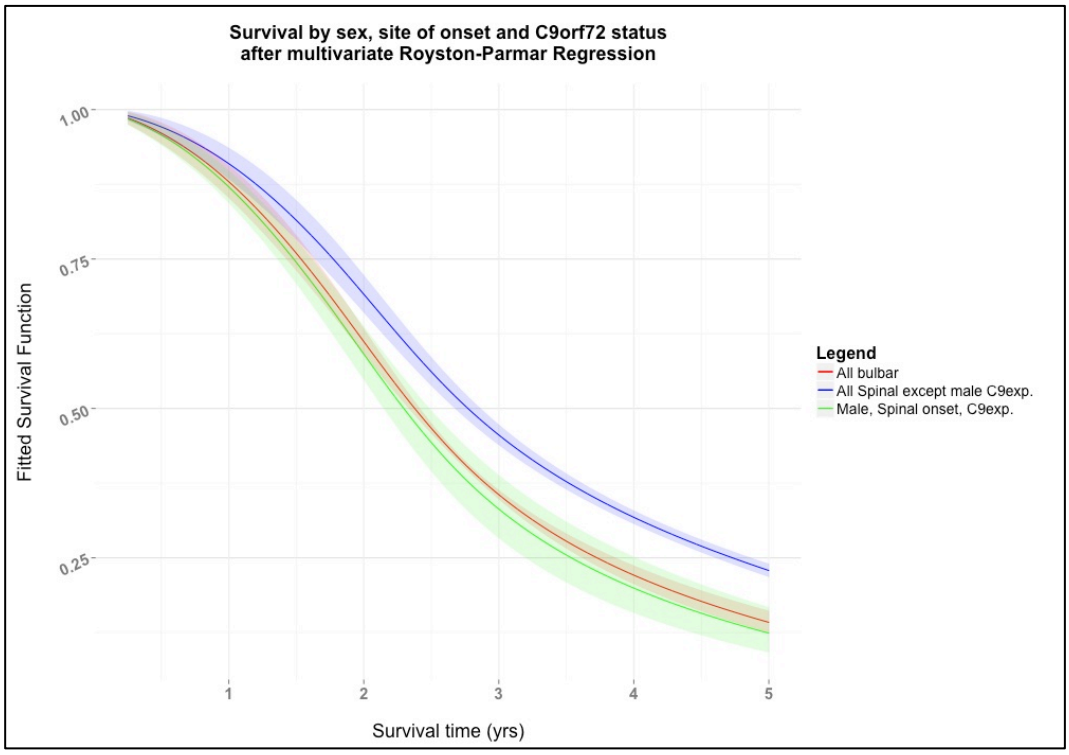


Figure 4-7 Legend: C9 exp=group carrying the C9orf72 expansion. Shaded areas represent 95% CIs. Survival curves show three subgroups of patients categorised by sex, site of onset and C9orf72 status. Survival curves for male spinal onset patients with the C9orf72 repeat expansion are markedly worse than other spinal onset patients and are in fact inseparable from bulbar onset patients. Median predicted survival in the three groups were: spinal excluding C9orf72 expanded males=2.77 years (95% CI 2.67 to 2.87); bulbar onset=2.38 years (95% CI 2.33 to 2.42); spinal onset males with C9orf72 repeat expansion=2.29 years (95% CI 2.15 to 2.49).

Figure 4-8 Individual patient data meta-analysis of the HR of male spinal onset patients with amyotrophic lateral sclerosis (ALS) carrying the C9orf72 repeat expansion

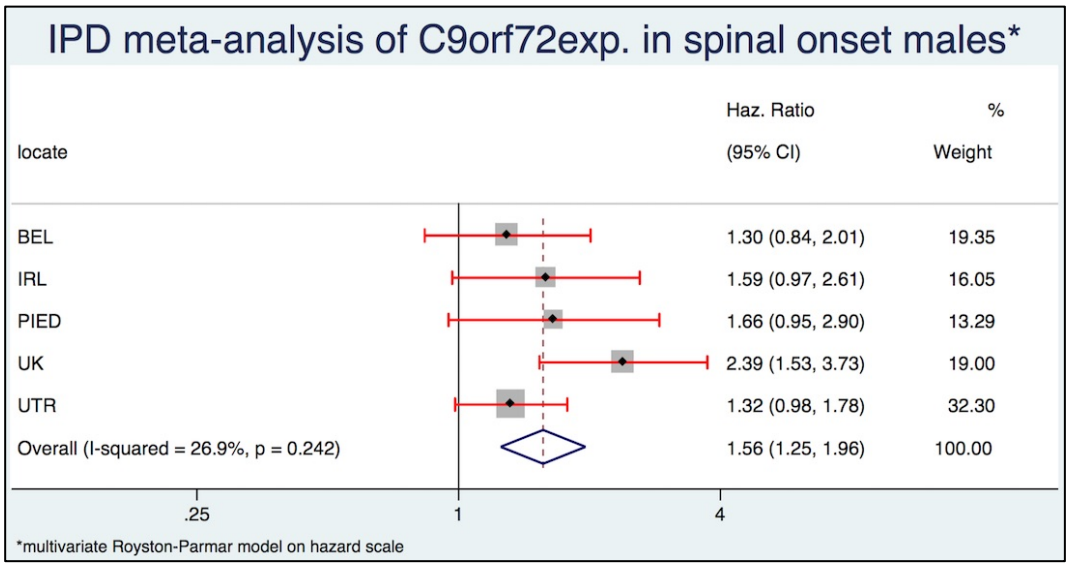


Figure 4-8 Legend: C9orf72exp, group carrying the C9orf72 expansion. IPD meta-analysis of C9orf72 repeat expansion in male spinal onset patients with ALS versus spinal onset C9orf72 normal patients pooled by country and analysed using a Royston-Parmar flexible parametric model with three degrees of freedom on the hazard scale correcting for age at onset (time varying), site of onset and diagnostic delay, based on the three level categorical breakdown of sex, site and C9orf72 status described in Appendix B. IPD = individual patient data.

The median ages and distribution of diagnostic delay across the final sub-groups is shown in Table 4-7. While age of onset was oldest in the bulbar onset group and youngest in the male spinal onset *C9orf72* expanded group, the male spinal onset *C9orf72* expanded group also had the highest proportion in the "short" diagnostic delay category, consistent with the finding that the *C9orf72* expansion differentially affects disease course in a gender specific manner. Adjustment for El Escorial category (Table 4-8) did not substantially alter the hazard ratio (HR = 1.57 95% CI:1.26 – 1.97).

Table 4-7 Age of onset and diagnostic delay for hybrid sex/ site of onset/ C9orf72 variable

Level	Age at onset		Diagnostic delay		
	Median (IQR)		Short	Medium	Long
Spinal onset excluding C9orf72 expanded males	62.3 (54.0 – 69.7)		965 (30.8%)	1014 (32.4%)	1151 (36.8%)
All bulbar onset	65.0 (58.4 – 71.9)		611 (37.4%)	590 (36.1%)	435 (26.6%)
Male spinal onset C9orf72 only	59.3 (52.3 – 64.7)		65 (40.8%)	44 (27.7%)	50 (31.5%)
P value	0.0001¶		<0.001†		

¶ = Kwallis test ; † = Chi² test

Table 4-8 Hazard ratios for El Escorial criteria after inclusion in final model

El Escorial Category	HR	95% CI	Wald test
Suspected	0.86	0.68 – 1.08	0.199
Possible	1	-	-
Probable – Lab supported	1.38	1.31 – 1.44	<0.001
Probable	1.49	1.39 – 1.60	<0.001
Definite	2.09	1.99 – 2.21	<0.001
Sex/Site of onset/C9orf72			
Spinal onset excluding male C9orf72 expanded cases	1	-	-
All bulbar onset cases	1.33	1.21 – 1.45	<0.001
Male C9orf72 expanded cases only	1.57	1.26 – 1.96	<0.001

Table 4-8 **Legend: Hazard ratios for each of the El Escorial criteria from a Royston Parmar model on the hazard scale with 3 degrees of freedom and including age of onset (time varying), diagnostic delay, site of onset, country and sex-site-C9 as a hybridized variable with variance-covariance matrix allowed to pool by country. Note that Belgium was omitted as the El-Escorial criteria was not available, the UK data included only Probable and Definite category ALS, whilst Ireland does not use the "Probable lab supported" category as shown in Table 4-6. Nevertheless, hazard ratios are in line with expectations showing a gradually increasing hazard with increasing category severity.**

4.3.4. Discussion

Previously, studies have shown that people with ALS carrying a *C9orf72* repeat expansion in blood present at a younger age and have reduced survival when compared to patients without the expanded variant (Table 4-9). However, studies to date have not been sufficiently powered to determine whether the expanded variant differentially affects outcome in subgroups based on gender and site of onset. Our findings demonstrate an intriguing and previously unrecognized interaction between the expanded variant and male patients with spinal onset disease, which appears to drive the overall survival effect. Within this cohort, the median age of onset was 59.3 and the median survival was 2.29 years. This compared to a median age of onset of 62.3 and median survival of 2.77 yrs in all other spinal onset disease, and a median age of onset of 65 and median survival of 2.38 years in all bulbar onset disease. Moreover, and contrary to the usual pattern in young onset disease, male spinal onset C9orf72 expanded cases were also more likely to have experienced a shorter diagnostic delay, suggesting rapidly progressing disease.

Female gender has previously been reported as an independent predictor of faster functional decline in ALS [237]. However, our observation of an interaction between site of onset, gender and *C9orf72* has not been previously noted possibly due to limitations in the power of previous studies due to lower numbers (Table 4-9). Taken together, our findings and those of previous studies imply that distinct processes may operate in differing subgroups of ALS even when a known genetic factor is present as the underlying cause, and demonstrate that male gender is likely to be an important interacting factor in the biology of C9orf72 related disease.

Table 4-9 Summary of previous analyses of survival by C9orf72 status

Study	Population	C9orf72 Normal	C9orf72 Expanded	Median survival delta¶	Median age at onset delta¶	HR (CI)
Byrne et al[4]	Ireland	170	21	-6 months	-3.2 yrs	1.9 (1.1 – 3.7)
Van Rheenen et al[5]	The Netherlands	1422	78	-2.5 months	-2.6 yrs	1.46 (1.17 – 1.83)
Sabatelli et al[6]	Italy & Sardinia	1688	69	-12 months	-3.8 yrs	1.79 (1.26 – 2.98)
Borghero et al[7]	Sardinia	375	51	-18 months ¶¶	-0.9yrs	na
Debray et al[8]	Belgium	513	77	fALS -38.3 months sALS -5.8 months	fALS -5.9 yrs sALS -0.3 yrs	fALS 2.5 (1.5 – 4.3) sALS 1.1 (0.8 – 1.5)
Garc´ıa-Redondo et al[9]	Spanish	936	67	-12 months	-2.6 yrs ¶¶¶	na
Irwin et al[10] ¶¶¶¶	United States (Pennsylvania)	69	64	-6 months¶¶¶	-3.0 yrs ¶¶¶	na

¶ Negative figures imply C9orf72 expanded survive for shorter time, or are younger at onset than C9orf72 normal cases

¶¶ Calculate as median survival in C9orf72 expanded group – median survival in overall cohort median

¶¶¶ Calculated from mean data instead of median

¶¶¶¶ Mixed ALS & FTD cases

fALS = familial ALS ; sALS = sporadic ALS

A number of pathogenic mechanisms have been proposed to explain the role of the *C9orf72* repeat expansion in ALS. These include haplo-insufficiency, toxic RNA interfering with the function of RNA-binding proteins or other cellular factors, and the presence of toxic dipeptide repeat proteins through RAN translation [44,238]. Recent work has pointed also towards *C9orf72* induced pathology of nucleocytoplasmic transport processes [44–47]. However, the pathobiology of the observed interaction between the C9orf72 variant and gender remains unclear, but it is congruent with observations in the SOD1 mouse model, in which transgenic mutant males have shorter

survival compared to their transgenic female littermates with similar copy numbers [239]. The mechanism for this gender effect in animals, although well recognized has not been characterized, but can be attenuated when mice are bred on a different genetic background [239].

A potential weakness of our study is that it did not include clinical scores for the presence of cognitive change, which is a known prognostic indicator in ALS. We and others have shown that those with C9orf72 repeat expansions are more likely to experience cognitive and behavioural change. However, to our knowledge to date, no gender mediated effect has been demonstrated in C9orf72 related cognitive profiling. Moreover, as C9orf72 is part of the causal pathway for some forms of FTD, inclusion of cognitive status as a variable would have introduced a selection bias based on 'conditioning on a common effect' [240]. A further limitation to this study is that our analysis does not include C9orf72 repeat expansion analysis by Southern blot, although Individual definition of pathological expansion performed by each centre using repeat primed PCR was validated by Southern blot. While the length of expansion varies from tissue to tissue [49,241,242], diagnostic testing within a clinical setting uses blood samples from which all previous prognostic and clinical correlative studies have been performed. Finally, we did not account for riluzole use in the analysis, as this is routinely prescribed in all patients from the participating countries, and it was not possible to determine the level of compliance using available data.

In conclusion, we have performed an analysis of the effect of the C9orf72 expansion on survival in almost 5000 European ALS patients. We have shown for the first time that *C9orf72* repeat expansion is a significant negative prognostic indicator in males with spinal onset disease only. These findings suggest a hitherto unrecognized interaction between the *C9orf*72 repeat expansion, site of onset and gender. This has important implications in understanding both the pathobiology of *C9orf72*-mediated disease, and in the development of future disease-related prognostic models.

4.4. What does the ALSFRSR Really Measure? A longitudinal and survival analysis of functional dimension sub-scores in Amyotrophic Lateral Sclerosis[14]

4.4.1. Introduction:

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disorder with a typically poor prognosis, although progression is recognized to be highly variable [243]. In the absence of an established biomarker, the rate of decline on the ALS functional rating scale (revised) (ALSFRS-R) has become the most commonly used measure of progression by clinicians, in research, and in clinical trials. The ALSFRS was designed as a 10-item functional score across fine and gross motor, bulbar and respiratory functional domains [244,245]. This was later extended to the ALSFRS-R by including extra respiratory scores to better capture the respiratory functional domain [246]. The longitudinal analysis of change in ALSFRS-R, which is generally considered to be linear for the majority of the illness, is complicated both by heterogeneity [247], and by longitudinally informative censoring due to drop-out of more ill patients, and/or the mortality of patients over time [248]. This problem has led to the development of alternative measures, such as the Combined Assessment of Function and Survival (CAFS), a non-parametric ranked score combining functional and survival outcomes for use in clinical trials in ALS [249].

The rate of change of the ALSFRS-R, or 'ALSFRS-R slope', has gained widespread acceptance as a prognostic indicator [5,250,251]. However, recent Rasch analysis of the ALSFRS-R has identified multi-dimensionality within the score and suggests that the use of sub-scores corresponding to bulbar, motor, and respiratory domains may be superior

[14] Chapter 4.4 has been published as:

Rooney J, Burke T, Vajda A, Heverin M, Hardiman O. 2017. What does the ALSFRS-R really measure? A longitudinal and survival analysis of functional dimension subscores in amyotrophic lateral sclerosis. J. Neurol. Neurosurg. Psychiatry 88:381–385; doi:10.1136/jnnp-2016-314661.

to the use of a single combined score[109]. To date, the prognostic value of these sub-scores has not been evaluated.

Using data from the Irish ALS register, our aims were to explore the prognostic and longitudinal characteristics of ALSFRS-R functional sub-scores, defined by the Rasch analysis of Franchignoni *et al.* [109,252].

4.4.2. Methods

Case ascertainment

We identified all incident cases from the Irish population based register [11,218] for whom at least one ALSFRS-R measurement was available. A small number (n = 8) of respiratory onset cases were excluded, as this group was too small for analysis. We then collated demographic, clinical, survival information, and longitudinal ALSFRS-R measurements. Information included: gender, age and site of onset, diagnostic delay, revised El-Escorial category, attendance at the multidisciplinary clinic (MDT) and Riluzole use. ALSFRS-R scores were generated by physiotherapy and medical assessors, all of whom had undergone standardised training using online tools developed by the NEALS and ENCALS groups for clinical trial purposes.

Definition of ALSFRS-R sub-scores

The Rasch analysis carried out by Franchignoni *et al.* determined that the ALSFRS-R scale captured information corresponding to 3 latent domains corresponding to bulbar, motor and respiratory functions [109]. The domains were defined as follows: Bulbar score = the sum of ALSFRS-R questions 1-3 (maximum score of 12), motor score = the sum of ALSFRS-R questions 4-9 (maximum score of 24) and finally respiratory score = the sum of ALSFRS-R questions 10-12 (maximum score of 12). We shall refer to these domain scores as ALSFRS-R_bulb, ALSFRS-R_motor and ALSFRS-R_resp respectively through the remainder of this manuscript. (Franchignoni *et al.* further suggested reducing ALSFRS-R responses from 5 levels (score 0 – 4) to 3 levels (score 0 -2). Appendix C contains this analysis duplicated using these collapsed scores).

Calculation of ALSFRS-R slopes

Slopes were calculated using the common convenience method for calculating ALSFRS-R slope = (48 – ALSFRS-R / time from onset) based on the first ALSFRS-R for each person, as this methodology has been reported in prognostic models [5,250], and can be calculated from a single score. This method was extended to produce convenience estimates for ALSFRS-R_bulb, ALSFRS-R_motor and ALSFRS-R_resp slopes. These were calculated as ALSFRS-R_bulb slope = (12 – Bulbar-score / time from onset), ALSFRS-R_motor slope = (24 – motor-score / time from onset), and ALSFRS-R_resp slope = (12 – respiratory -score / time from onset) – based on the first ALSFRS-R for each person.

Survival model

For the survival analysis, time of entry to follow-up was date of onset. Using Cox regression, we constructed a survival model including the recognized prognostic variables of age at onset, site of onset, diagnostic delay and the revised El-Escorial criteria. Using this survival model as a comparison point, we first added the convenience estimate for ALSFRS-R slope to this model (i.e. based only on the first ALSFRS-R of each case and assuming a total score of 48 prior to illness onset). Next we replaced ALSFRS-R slope by the convenience estimate of slope for each ALSFRS-R sub-score for each sub-score individually. Finally, sub-scores were added to the model in combination. Model fit was assessed using the Akaike Information Criterion (AIC).

Longitudinal model

A linear mixed effects multi-level model was used to account for repeated measures of ALSFRS-R within individuals. Models were estimated using maximum likelihood and compared using ANOVA. The final model for total ALSFRS-R as outcome was then altered to consider ALSFRS-R sub-scores as the outcome variable; that is a separate model was fit for the motor, bulbar, and respiratory sub-scores. Graphs of model fit were plotted with separate curves for spinal and bulbar onset cases.

Software

All data formatting and statistical analysis were carried out using *R* version 3.2.3 [190], with additional packages (*tableone*, *ggplot2, nlme, survival, rms, caret, classInt, gridExtra*) [236,253–259].

4.4.3. Results

407 incident cases were identified with at least one complete ALSFRS-R. Two hundred and thirty three (57%) were male, while 125 (30.7%) had bulbar onset disease. The data included 1550 ALSFRS-R scores, with a median of 2 per patient (IQR: 1 – 5). Baseline demographics are shown in Table 4-10.

Table 4-10 Baseline demographics of patients included in ALSFRS-R analysis

	Spinal Onset	**Bulbar Onset**	**P value**
n (%)	282 (69%)	125 (31%)	
Male (%)	176 (62%)	57 (46%)	**0.002**
Age of onset (*M*±SD)	62.0 (11.4)	63.4 (10.5)	0.274
Diagnostic delay in months (*M*±SD)	15.7 (12.1)	11.9 (9.0)	**0.002**
El Escorial Category			
Suspected	4 (1.5%)	1 (0.8%)	
Possible	50 (18.7%)	18 (15.0%)	
Probable	77 (28.9%)	25 (20.8%)	
Definite	136 (50.9%)	76 (63.3%)	0.153
Attends MDT	271 (96.1%)	113 (90.4%)	**0.044**
ALSFRS-R per patient (median)	3 (IQR: 1-6)	2(IQR: 1-4)	

Survival Analysis

94% of the patients attended the multi-disciplinary team (MDT) and 86% were prescribed riluzole. 20 cases with missing values were omitted from the survival analysis. Median follow-up of survival was 3.0 years (CI: 2.8 – 3.3 years) with 294 (73%) dying during the follow-up period. Hazard ratios from the basic multivariate survival model, models for ALSFRS-R slope and sub-score slopes (all estimated via the convenience method) added to the basic model in isolation, and in combination are shown in Appendix C Table C1. The best-fit model, model 6, included both ALSFRS-R_motor (HR 4.7 CI: 3.7 – 6.0) and ALSFRS-R_bulb (HR 9.9, CI: 5.6 – 17.6) ALSFRS-R slopes, but not the ALSFRS-R_resp slope. (Using collapsed scores (Appendix C Table C2), the best fit model included the collapsed ALSFRS-R_resp slope).

Notably in model 6 site of onset lost its importance (bulbar onset HR 1.0; CI: 0.74 – 1.45), indicating that the sub-score slopes likely capture the same variation in survival as captured by the site of onset. To confirm this, we tested the ability of ALSFRS-R sub-score slopes to distinguish between spinal and bulbar onset cases. The cases were split into a training set of 204 and a test set of 203 cases, and a logistic model was generated with bulbar onset as the dependent variable. Convenience estimates of ALSFRS-R_bulb slope and ALSFRS-R_motor slope were entered as the independent variables. This model was then applied to the test set to estimate probabilities of bulbar onset disease. Probabilities over 50% were categorized as bulbar, below 50% as spinal. The model correctly identified 47 out of 61 bulbar cases in the test set with 133 out of 142 correctly identified as spinal, leaving 9 spinal cases incorrectly labelled as bulbar, and 14 bulbar cases incorrectly labelled as spinal. Thus, the model has sensitivity of 90% and specificity of 84% for identifying bulbar onset cases.

Longitudinal Analysis

All 407 cases and 1550 ALSFRS-R measurements were included in the longitudinal analysis. Figure 4-9 displays the longitudinal linear trend in ALSFRS-R for spinal and bulbar onset patients modelled by the linear mixed effects model.

Figure 4-9 Linear trend in total ALSFRS-R over time as modelled via linear mixed effects model

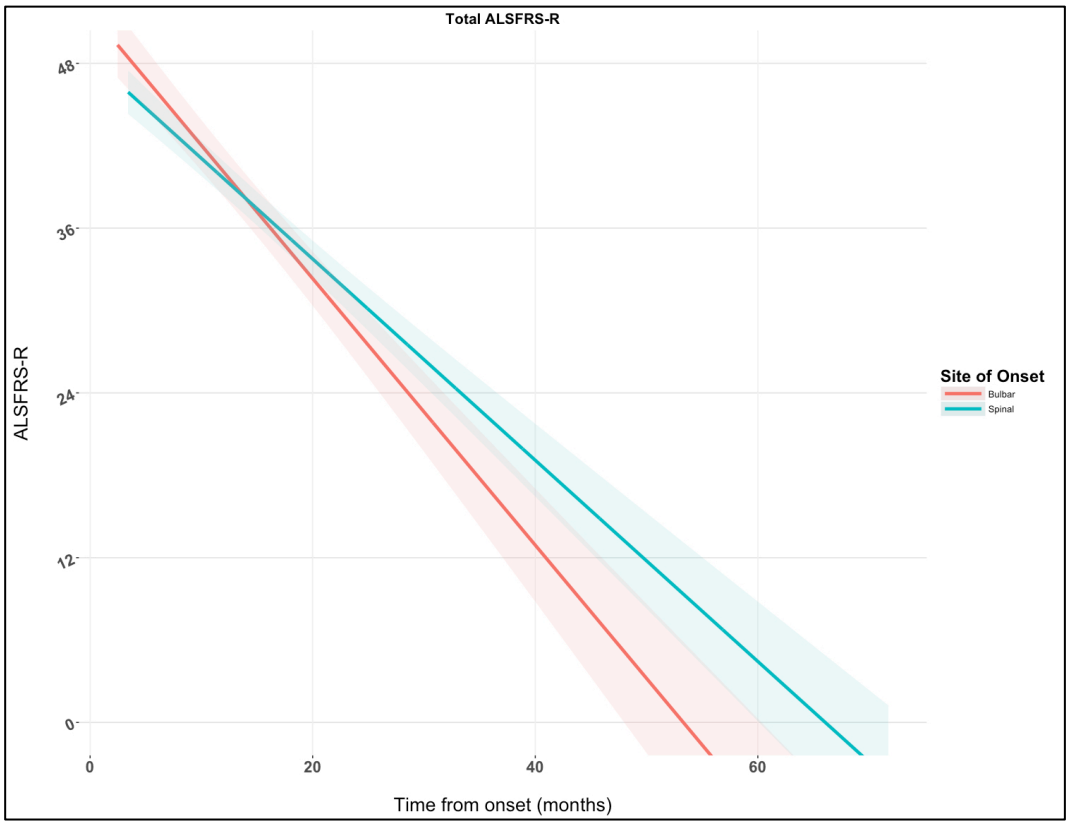


Figure 4.9 Legend: Although the fitted lines are in close agreement throughout with overlapping CIs, it is noteworthy that the patients with bulbar onset show a steeper slope of decline in overall ALSFRS score.

After inclusion of spline terms, further model comparisons indicated the inclusion of site of onset as a fixed effect modifier of total ALSFRS-R over time ($p < 0.0001$). Using this final model, graphs of total fitted models of ALSFRS-R and ALSFRS-R sub-scores over time were then produced (Figure 4-10). Total ALSFRS-R shows crossover of spinal and bulbar scores. However, ALSFRS-R_motor, and in particular ALSFRS-R_bulb, show greater deviance over time. For ALSFRS-R_motor, spinal and bulbar onset patients symptomatically decline at a similar rate in the first 4 years before converging. However, Figure 4-10 indicates that this symptomatic decline may have occurred prior to the reported disease onset time in spinal onset patients.

Figure 4-10 Graph of longitudinal total ALSFRS-R and ALSFRS-R subscores

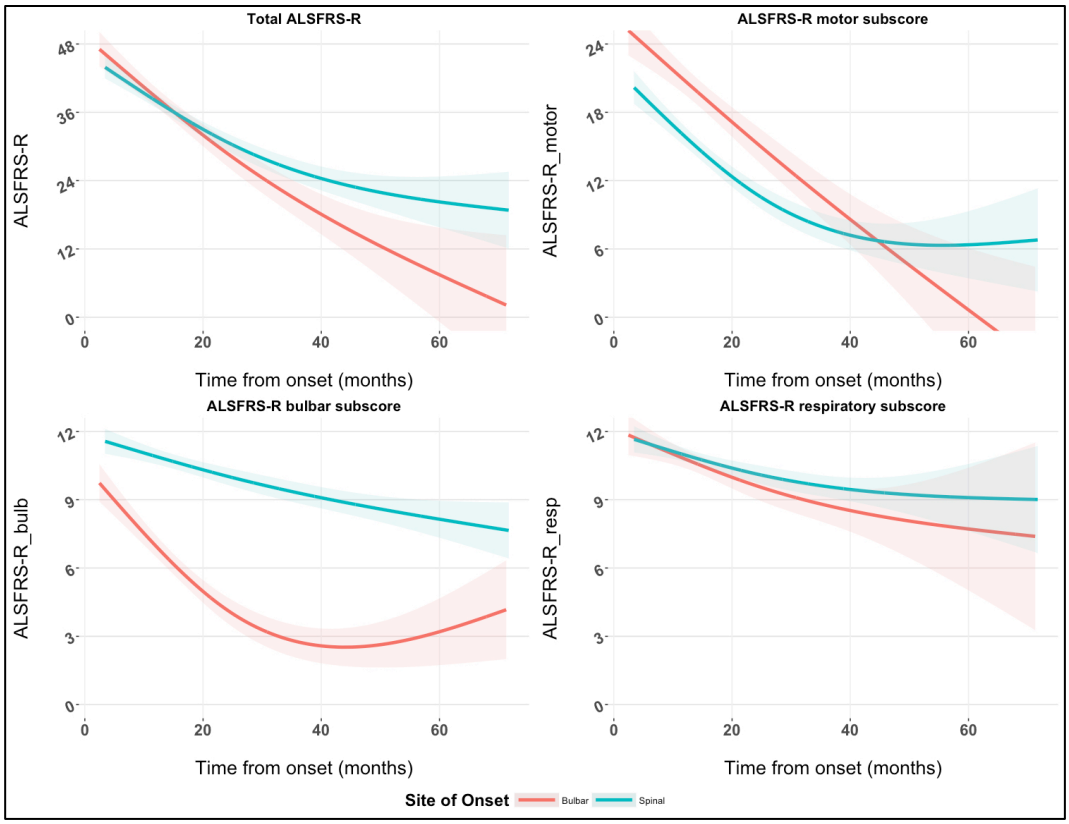


Figure 4.10 Legend: Linear mixed models for total ALSFRS-R and ALSFRS-R sub-scores in the Irish cohort when fitted with a restricted cubic spline.

Bulbar onset patients deteriorate at a markedly increased rate, with spinal onset patients rarely declining below 8 / 12 on the ALSFRS-R_bulb score. This also suggests that decline of the ALSFRS-R_bulb score may begin before reported onset in bulbar onset patients. Conversely, ALSFRS-R_resp was relatively non-informative, showing only a mild separation between bulbar and spinal onset patients over time, and perhaps the most revealing feature is that respiratory sub-scores rarely drop lower than 8 / 12. (Longitudinal models of collapsed scores (Appendix C Figure C1), did not exhibit different longitudinal trends from the longitudinal models of full scores.)

4.4.4. Discussion

The ALSFRS-R slope is well established as a prognostic indicator in ALS [5,250,251]. However, by compressing multi-dimensional domains into one combined score, information may be lost. ALSFRS-R sub-scores have previously been used to quantify regionality of disease, which was seen to be a prognostic factor [260]. Furthermore, it has been suggested that cognitive impairment, particularly executive dysfunction at onset, predicts bulbar sub-score decline [125]. Our analysis complements and expands upon these previous ALSFRS-R sub-score analyses.

Our results indicate that the use of ALSFRS-R_bulb and ALSFRS-R_motor slopes may offer improved prognostication. We have shown that ALSFRS-R_bulb slope and ALSFRS-R_motor slope can be used to confirm site of onset with very good sensitivity and specificity despite a very simple model and a moderate sample size, and observed that site of onset became unimportant in prognostic models when sub-score slopes were included. We have further shown that longitudinal analysis of ALSFRS-R_bulb & ALSFRS-R_motor revealed distinct characteristics of spinal vs bulbar onset cases. Taken together these results indicate that ALSFRS-R_bulb & ALSFRS-R_motor capture the divergent clinical and prognostic characteristics of spinal and bulbar onset patients. The observed longitudinal trend in sub-scores may also explain observations that increased regionality (estimated from ALSFRS-R sub-scores) of disease was prognostic [260], since typically the bulbar score in bulbar patients declines very rapidly (Figure 4-10).

In assessing ALSFRS-R_resp we must recognize that the cohort did not, by design, include any respiratory onset patients and that the use of Non-Invasive Ventilation (NIV) was not taken into account. Although the leading cause of death in ALS patients is respiratory failure [261], our finding that ALSFRS-R_resp slope was not prognostic in the final survival model, and the longitudinal characteristics of ALSFRS-R_resp shown in Figure 4-10, likely indicate that the ALSFRS-R_resp questions are not sensitive to the respiratory burden of the majority of patients, further reinforcing the findings of Franchignoni *et al.* [109]. The respiratory questions of the ALSFRS-R focus on chronic progressive symptoms (dyspnea, orthopnea, and ventilation), and it may be that the

absence of a question covering acute exacerbations or frequency of chest infections and questions relating to compliance in utilization of NIV account for the absence of sensitivity. However, in our analysis of collapsed ALSFRS-R sub-scores the slopes based on the collapsed respiratory scores suggested by Franhignoni *et al*. did show prognostic value (Appendix C).

Our observation that the decline in ALSFRS-R_bulb in bulbar onset patients, and decline in ALSFRS-R_motor in spinal onset patients may begin before reported symptom onset times contrasts with recent analysis by Proudfoot *et al*. [248], who reported a median lag of 5.8 months between reported symptom onset and the beginning of decline in ALSFRS from the maximum score of 48 in the PROACT dataset. Proudfoot *et al*. considered that a lack of sensitivity of the ALSFRS-R in early stages of the diseases may offer an explanation. However, our results here would suggest the opposite. Since PROACT constitutes data only from clinical trials with very specific entry criteria for patients, selection bias in favour of prevalent cohorts of patients may explain the discrepancies, as our cohort was based on mixed incident cases of MDT attenders & non-attenders. Alternatively, the different mathematical models used could be responsible. Nevertheless, our results imply that unrecognised minor symptomology amounting to a small ALSFRS-R decrease may be present before the self-reported time of onset of symptoms. We speculate that this may reflect a tendency of patients to dismiss or ignore minor initial symptoms that would otherwise have represented the true onset of disease. This is also consistent with an insidious onset of disease that would be characteristic of a long pre-clinical prodromal phase of the disease [262], and we note that long prodromal phases have been observed in SOD1 mice [262] and human SOD1 case reports [263].

Although our study is limited by relatively small numbers for the complexity of the models used, our findings shed new light on the evolution of ALSFRS-R sub-scores over time. We have shown than ALSFRS-R_bulb slope and ALSFRS-R_motor slope together provided a better fit in survival modelling than ALSFRS-R slope as a single variable, and should be considered separately in future analyses of clinical decline. Longitudinal analysis showed that ALSFRS-R_bulb and ALSFRS-R_motor show distinct trajectories for

bulbar and spinal onset patients. Convenience estimates of ALSFRS-R_bulb slope and ALSFRS-R_motor slope were not only important prognostic variables, but can confirm site of onset with high sensitivity and specificity. Our results offer improved prognostication and may allow for more detailed analysis of progression in both epidemiological studies and clinical trials, and provide strong evidence that ALSFRS-R should not be reported as a single score, but rather analysed as domain specific sub-scores.

4.5. Progression and Survival in ALS: Implications and Future Directions

The work carried out for this chapter has led to several refinements in the understanding of prognostic factors in ALS, particularly with regard to the importance of the multidisciplinary clinic in the treatment of ALS, deepening our understanding of *C9orf72*'*s* impact on survival and in highlighting that a more sophisticated treatment of ALSFRS-R data can lead to new understanding of progression. However, much remains to be understood.

Chapter 4.1 demonstrated that attending the MDT is of similar survival benefit to that of riluzole prescription. It remains unclear why exactly this is the case. While we consider that the most likely explanation relates to shared decision making amongst the multidisciplinary experts that make up the team, other factors may also be at play, for example patients who attend the clinic may have better compliance with riluzole treatment or night-time NIV use. Such complexities are difficult to untangle given that we lack data on compliance with treatments for patients who do not attend the MDT. This may be further complicated as NIV appears to be less effective in cases of more severe bulbar dysfunction [264,265]. The findings of Chapter 4.1 also have implications for healthcare delivery, particularly when combined with the recent findings from my colleagues regarding the patient paths to multidisciplinary care [266]. They found that patients frequently experienced delays, unnecessary referrals and subsequent investigations in their path to the specialist clinic at mean costs of €3,486 per patient. Earlier assessment by neurologist was associated with shorter time to the specialist multidisciplinary clinic and significant savings [266]. Taking my findings from Chapter 4.1 together with those of Galvin *et al.*[266], the case is strong for the prioritisation of multidisciplinary clinic attendance for ALS patients (or indeed those in need of expert diagnosis) in healthcare delivery.

ALSFRS-R sub-scores hold promise for refining clinical trial entry criteria and measurement of progression between trial arms. Recent methodological advancements in the joint modelling of longitudinal and time to event outcomes and new software packages provide new tools to understanding the effects of longitudinal variables on

prognosis. Indeed, a recent study combining trial data from PRO-ACT and simulation studies demonstrated that joint modelling of total ALSFRS-R and survival was the most efficient study design out of 6 designs including designs used in recent trials [267]. This approach could potentially be further enhanced using ALSFRS-R sub-scores as multiple longitudinal variables. ALSFRS-R sub-scores may also help to clarify whether particular treatments or genes affect function in a domain specific manner.

Chapter 5: Environmental Exposures in ALS: the Euro-MOTOR study

Chapter Outline:

5. Environmental Exposures in ALS: The Euro-MOTOR study

5.1. Critical issues in ALS case-control studies[15,16]

5.1.1. Introduction

Amyotrophic lateral sclerosis (ALS) is a rare neurodegenerative disease of unknown etiology for which no definite risk factors have been identified with certainty, with the notable exception of genetic factors, which account for up to 10-20% of cases [14,268]. Identification of environmental risk factors of ALS can be achieved using various observational designs, such as cohort studies and case-control studies. Although cohort studies may be preferred, they are limited in power due to the low incidence of ALS and the long and unknown interval between exposure and disease onset due to misclassification of the disease. For this reason, case-control studies remain the most informative study design in ALS research.

The failure to identify potentially modifiable risk factors can be explained by several factors: firstly, the heterogeneity of the disease; secondly, the relative rarity of ALS, making it difficult to collect a large enough number of patients to reach adequate study power; and thirdly, the methodological limitations of many published reports, including poor representativeness of the source population, small sample size, incorrect and/or varying definitions of exposures and inadequate control of confounders [9].

Studies of non-representative samples of patients are likely to confound results and to prevent comparisons across different populations. Small samples are unlikely to provide precise estimates of the risks and increase heterogeneity in published results. The use

---

[15] Chapter 5.1 has been published as:

D'Ovidio F **, Rooney JPK,** Visser AE, Vermeulen RCH, Veldink JH, Van Den Berg LH, et al. 2017. Critical issues in ALS case-control studies: the case of the Euro-MOTOR study. Amyotroph. Lateral Scler. Frontotemporal Degener. Aug;18(5-6):411-418.

[16] Chapter 5.1 and publication originally drafted by Fabrizio D'Ovidio.

D'Ovidio F**, Rooney JPK,** and Visser AE contributed equally to this work.

of subjective (often unmeasurable) definitions of exposures prevent correct estimates of the risks and comparisons across studies.

Accordingly, there is a clear need for large studies using a standardized protocol. An international multicenter study offers a unique opportunity to examine several risk factors at the same time and with the same methodology in different geographic areas, while aiming to obtain a truly representative sample of ALS patients in a population-based design. This approach also explores the possibility that the same risk factor may not have the same effect in different populations. Another possibility is that different populations may exhibit subtle differences in phenotypic characteristics [105]. All of these issues have been addressed in a case-control study by the Euro-MOTOR consortium, which has collected clinical and biological variables in a large sample of patients with ALS and representative controls across Europe, using a population-based design to investigate the etiology and mechanisms of the disease. Our previous experience with the investigation of the role of traumatic events [269] and physical exercise [200] was helpful in identifying the weaknesses of a case control study and served to improve the study design with particular reference to sample size and criteria used for data collection. This provides sufficient power to investigate research hypotheses regarding previous disputed environmental risk factors for ALS.

The purpose of Chapter 5.1 is to describe the study design and methods, highlighting the strengths and limitations of the basic structure of the project.

5.1.2. Aims of Euro-MOTOR project

The main objective of Euro-MOTOR is to discover new causative and disease-modifying pathways to pave the way for novel therapies. The epidemiological component of Euro-MOTOR is a pooled case-control study, which consists of a paper and pencil survey through face-to-face interviews undertaken in three European countries (The Netherlands, Ireland and Italy), where incident ALS patients and matched controls were recruited in a population-based design based on age, gender and area of residency. Core demographic and ALS-related clinical data were collected by trained investigators. The information was obtained from the patient/control or, for cognitively impaired individuals, from an informed caregiver. The data were entered into a centralized online database, along with biological samples collected to investigate genetic and biochemical profiles of cases and controls. The recruitment of ALS cases and controls spanned from February 2011 to January 2014.

5.1.3. Representativeness of the study population

In the participating population-based registers, several sources are periodically interrogated to ensure full case ascertainment. These include neurologists, neurophysiologists, and pulmonologists with interest in ALS, riluzole pharmacy records, lay association archives, general practitioners' records, and administrative sources (hospital discharge records, disability lists, etc.). The validity of these sources has been documented in published reports [12,16,17,270,271]. In addition, when comparing the study sample with the original ALS population from each country, no relevant differences in the main demographic and clinical indicators were detected, with the exception of patients with bulbar site disease from the Irish cohort (23.7% vs. 35.3% in the registry) (Table 5-1).

Table 5-1 Comparison of annual incidence, gender, age and bulbar site of onset between local registries and the Euro-MOTOR study

	Apulia	Lombardy	Piedmont & d'Aosta Valley	Ireland	Netherlands
Registry ALS population					
Incidence*	1.6	2.1	2.9	2.0	2.8
Male:female ratio	1.6	1.3	1.2	1.4	1.5
Age at onset	64.6 (median)	63.6 (mean)	64.8 (mean)	66.0 (median)	63.0 (median)
Bulbar site onset (%)	21	35.2	37.5	35.3	30
Euro-MOTOR study ALS population					
Male: female ratio	1.6	1.2	1.2	1.5	1.5
Age at onset	62.6 (median)	63.6 (mean)	64.2 (mean)	64.4 (median)	64.1 (median)
Bulbar site onset (%)	26.5	29.6	37.0	23.7	35.4

Data from published studies [12,16,17,270,272]**.**

*** per 100,000 person-years**

The underlying populations have intrinsic differences in terms of education and clinical characteristics of the disease at enrolment. Since the classification of education in Italy did not take into account the attendance at primary and secondary school separately, and a national reform occurred during the 1960s, and to make education comparable across countries, the six categories were aggregated into two: low level (none, grade school, high school and technical or trade school) and high level (university and graduate school). Comparing education across all participants (Table 5-2), the majority of subjects reported low educational levels (International Standard Classification of Education

(ISCED) 0-4) (77.1%). The educational levels were highest in The Netherlands and lowest in Italy.

Table 5-2 Frequency distributions of level of educational attainment by country in the case-control study and in the official EUROSTAT statistics on education, 2014 (%)

	Italy		EUROSTAT reference figures
Education level	Cases	Controls	Italy
Low	526 (92.4%)	552 (80.5%)	39,271* (86.1%)
High	43 (7.6%)	134 (19.5%)	6,355* (13.9%)
	Ireland		
Education level	Cases	Controls	Ireland
Low	148 (83.6%)	273 (78.2%)	2,146* (66.1%)
High	29 (16.4%)	76 (21.8%)	1,102* (33.9%)
	The Netherlands		
Education level	Cases	Controls	The Netherlands
Low	583 (74.5%)	1,348 (71.7%)	8,943* (71.5%)
High	205 (26.0%)	532 (28.3%)	3,563* (28.5%)
	Overall		
Education level	Cases	Controls	Italy, Ireland, Netherlands combined
Low	1,257 (81.9%)	2,173 (74.5%)	50,360* (82.1)
High	277 (18.1%)	742 (25.5%)	11,020* (17.9%)

*** Figures in thousands**

Education was classified into two levels, following the International Standard classification of Education (ISCED): low level (ISCED 0-4) and high level (ISCED 5-8); EUROSTAT statistics on education of population 15–75 years old resident in the three countries were obtained from the following source: http://ec.europa.eu/eurostat/web/education-and-training/data/main-tables

Descriptive statistics stratified by country revealed that significant differences between cases and controls in the two educational levels were detected only in Italy ($p<0.001$). The better educational level of controls is generally consistent across countries and is to be expected, as educated healthy individuals are more likely to participate in epidemiologic studies.

5.1.4. Disease phenotype

The site of ALS onset of patients was mainly spinal (64%), followed by bulbar (33%) and thoracic-respiratory (3%). Bulbar onset ALS is usually considered to account for up to one third of cases. However, the proportion varies across populations [9]. In the present study, significant differences across countries were detected ($p<0.01$), especially regarding Ireland where bulbar onset patients (24%) were fewer than in Italy (32%) and The Netherlands (35%) (Table 5-3). These differences are partly explained by the interval between onset of symptoms and diagnostic assessment (medians were 9.2 (IQR=9) months in The Netherlands, 9.8 (IQR=9) in Italy and 11.4 (IQR=9.5) in Ireland). In this regard, patients with bulbar onset – who have a worse prognosis than patients with spinal onset – are more likely to have declined participation in the Irish cohort due to the severity of their illness.

With reference to the El-Escorial classification, over half of cases were classified as probable ALS at diagnosis (54%), of whom 18% were laboratory supported, but frequency distributions of El-Escorial categories were significantly different across countries ($p<0.001$). In Ireland, the term laboratory supported probable ALS was not used whereas 50% of cases fulfilled the clinical criteria for definite ALS at the time of categorization. This compared to 18% in The Netherlands and 26% in Italy respectively. These differences cannot be explained by discordance in the diagnosis of ALS among local investigators because inter-rater agreement was satisfactory ($K=0.83$; $p<0.0001$), as measured by the Kendall (K) Coefficient of Concordance, a non-parametric statistic assessing agreement among raters and ranging from 0 (no agreement) to 1 (complete agreement). The agreement was tested among nine raters (from the Euro-MOTOR

consortium) on 52 patients with ALS or other clinical conditions. ALS patients were classified according to the revised El-Escorial categories [6]. These differences most likely reflect differences in methods of case ascertainment, subtle differences in the timing and method of assignment of diagnostic category (e.g. assignment at a specialist clinic, or by chart review or at the time of clinical evaluation): differences in practice with respect to when patients receive a definitive diagnosis of ALS (in Ireland this frequently takes place at the time of evaluation at the Specialist Clinic, rather than at the time of first encounter with a neurologist).

Table 5-3 Frequency distributions of site of onset and El Escorial categories by country

	Italy	Ireland	The Netherlands	Overall
Site of onset				
Bulbar	188 (31.9%)	42 (23.7%)	275 (34.8%)	505 (32.4%)
Spinal	385 (65.4%)	132 (74.6%)	476 (60.2%)	993 (63.8%)
Thoracic/respiratory	11 (1.9%)	3 (1.7%)	27 (4.6%)	41 (2.6%)
Missing	5 (0.8%)	0 (0%)	13 (1.6%)	18 (1.2%)
Total	589 (100%)	177 (100%)	791 (100%)	1,557 (100%)
El Escorial Category (Revised)				
Definite ALS	155 (26.3%)	88 (49.7%)	144 (18.2%)	387 (24.9%)
Probable ALS	202 (34.3%)	48 (27.1%)	316 (39.9%)	566 (36.4%)
Probable (lab. supported	80 (13.6%)	0 (0%)	192 (24.2%)	272 (17.5%)
Possible ALS	148 (25.1%)	32 (18.1%)	116 (14.7%)	296 (19.0%)
Missing	4 (0.7%)	9 (5.1%)	23 (2.9%)	36 (2.3%)
Total	589 (100%)	177 (100%)	791 (100%)	1,557 (100%)

5.1.5. Matching

In a case-control study, matching is done on one or more variables. Matching may be performed on an individual basis (individual matching) or for groups of subjects (frequency matching). Individual matching implies the identification of one or more controls matched to each case for the values of the matching factor. Frequency matching involves the selection of an entire stratum of reference subjects with value(s) of the matching factor(s) equal to the value(s) of the stratum of the index case. Matching can be justified by the need to adjust for variables known to be potential confounders and provide more efficient stratified analyses. However, matching on factors that do not affect the risk of disease but with effects on exposure may bias the results of the study. In the Euro-MOTOR study, both individual and frequency matching were used in different countries. Matching variables included age, sex and area of residency. Italy and Ireland adopted individual matching while The Netherlands adopted a frequency matching approach. Irish cases were matched *a priori* by age (±5 years), sex and location (general practitioner practice when possible, otherwise local area), with a 1:2 case-control ratio. The same criterion was followed by the three Italian centers, but location was here represented by local area (i.e., distance in kilometers and urban/rural setting), and the case-control ratio was slightly greater than 1:1. Cases in The Netherlands were frequency matched using general practitioners' registries, in order to select a representative population list by age (±5 years), sex and location (based on general practitioners' area), to achieve a 1:2 case-control ratio.

5.1.6. Environmental exposures

Environmental epidemiology attempts to characterize the health effects of known environmental exposures. The quality of exposure measurement will determine the validity of the results. A crude binary variable is necessary, but is not sufficient to determine a dose-response relationship and therefore ordinal categories are needed to increase the sensitivity of the study. It is important to distinguish between exposure setting, a complex mixture, and a single agent. Exposure to environmental agents can be assessed using different instruments, including interviews, structured diaries, measurements in external media, measurements of concentrations in human tissues,

and markers of physiologic effects. Several methods for assessment of exposure are available. These include, among others, Job Exposure Matrices (JEMs), a system linking occupations to profiles of environmental exposures by providing (semi-)quantitative assessments of exogenous exposures for each occupation. JEMs are dependent on quality of job history and assignment at job level and may result in exposure misclassification for an individual subject. In addition, occupational studies using job classifications may have large within-category variability. Moreover, there may also be some intra-individual variability. Other exposure measures may have even greater limitations: concentrations in tissues may reflect primarily or exclusively recent exposure; use of concentrations in external media may result in misclassification for individuals; markers of physiological effects may be non-specific.

Research on environmental exposures for ALS has produced a large number of studies, although findings are conflicting. A significant excess of deaths for ALS has been reported in soccer players [56,273]. A history of repeated (head) injuries has been reported more frequently by ALS patients than by individuals with other clinical conditions [269]. An association between exercise and ALS has also been suggested [59,200] and dietary supplements, including branched-chain amino acids, have been also implicated [271]. Recent reviews have also concluded that smoking should be considered a risk factor for sporadic ALS [28,199]. Industrial and agricultural chemicals, such as pesticides, solvents and heavy metals, have also been reported as risk factors for ALS, but inconsistency of findings and the poor methodological quality of most studies in this field, in particular regarding exposure assessment, has limited the potential to establish a causal relationship with ALS [64].

Based on existing preliminary data, a number of environmental exposures have been considered in the Euro-MOTOR study. We did our best to minimize recall bias by using a very detailed questionnaire that included focused questions on the history of selected exposures (i.e. occupations, sports, hobbies, traumatic events, places of residence, familial diseases, drugs intake, smoking and alcohol habits).

5.1.7. Definition and classification of exposures

The English version of the study protocol is reported in Appendix D. Each exposure was defined using conventional definitions or, where not available, using pragmatic definitions that were accepted by consensus among the principal investigators. Where possible, a given exposure was graded to capture a gradient. Definitions and classifications are available as supplementary data.

Two neuro-psychological tests were also used to identify cognitive impairment: the Frontal Assessment Battery (FAB) and the Verbal Fluency Index (VFI), administered to both cases and controls.

5.1.8. Genetic classification

A hexa-nucleotide (G_4C_2) repeat expansion in the 5' non-coding region, C9orf72, is the most common genetic cause of amyotrophic lateral sclerosis and frontotemporal lobar degeneration (FTLD) [267,268]. A total of 1,257 (77%) patients underwent a genetic analysis to identify the presence of the GGGGCC repeat expansion in the first intron of the gene C9orf72, according to the methodology described elsewhere [3,4]. Recently, a survival study of the C9orf72 expansion that included Euro-MOTOR patients from Ireland, The Netherlands and Italy (Piedmont and Val d'Aosta regions) was published (Chapter 4.3).

5.1.9. Population at risk

The total population at risk in the Euro-MOTOR project was 34.1 million, distributed over three countries and five cohorts: The Netherlands, 16 million; Ireland, 4.5 million; Italy, 13.6 million (Piedmont and Valle d'Aosta, 4.5 million; Lombardy, 5 million; Apulia, 4.1 million) (Table 5-4).

Table 5-4 Distributions of cases and controls by study site

Study Site	Source population (millions)	Cases	Controls	Total	Ratio
Apulia	4.1	141 (8.6%)	213 (7.3%)	354 (7.8%)	1 : 1.5
Lombardy	5.0	186 (11.4%)	190 (6.5%)	376 (8.3%)	1 : 1.0
Piedmont & d'Aosta Valley	4.5	262 (16.1%)	290 (9.9%)	552 (12.1%)	1 : 1.1
Ireland	4.5	177 (10.9%)	349 (11.9%)	526 (11.6%)	1 : 2.0
The Netherlands	16.0	791 (48.5%)	1,880 (64.3%)	2,671 (58.7%)	1 : 2.4
Total	34.1	1,557 (100%)	2,922 (100%)	4,479 (100%)	1 : 1.9

On this basis, and considering country-specific annual incidence rates, as shown in Table 5-1, a conservative estimate of the number of patients with ALS to be recruited was 838, but only 626 per year (or 1.84 per 100,000 per year) should be considered assuming that approximately 75% of patients would agree to participate. The Euro-MOTOR study enrolled 519 cases per year (83% of total cases estimated) and 1.9 controls per case, achieving a power of 76% powered to detect an Odds Ratio (OR) of 2, with a 1% level of exposure in the controls and a 2% level of exposure in the cases. Our preliminary results indicated that, with these numbers, the exposures we investigated can be evaluated with sufficient statistical power.

The population considered in the analyses (4,479) was represented by 1,557 cases and 2,922 controls. The distribution of definite and probable ALS cases and controls included from the three countries and five cohorts is given in Table 4.

5.1.10. Statistical analysis

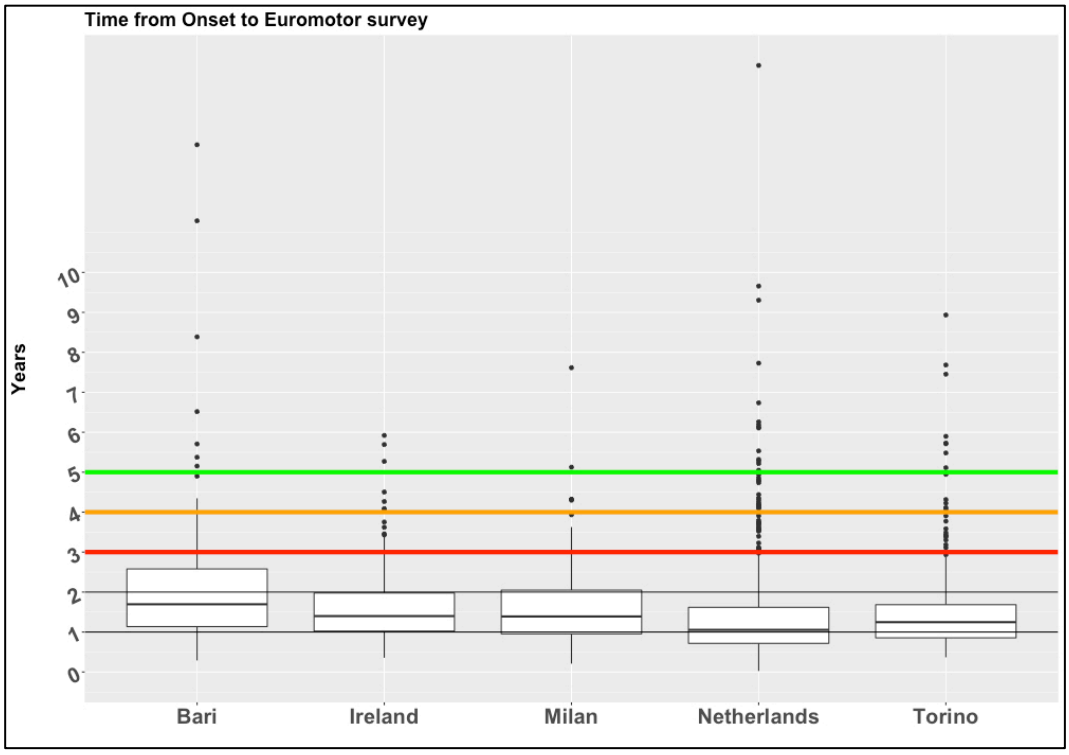
The statistical analysis of data collected from case-control studies is performed under the conventional assumptions of independence and identical exposure in patients with and without the disease of interest. Parametric and non-parametric tests are used after verification of the distribution of each variable's values. In Euro-MOTOR study, the Student's t test (or the Wilcoxon test when a non-parametric equivalent is required) was used to test height, weight, BMI, MET, duration of occupation, duration of physical exercise, number of traumatic events, duration of drug and toxin exposure. The chi-square test was used to test all categorical variables, as center or demographics. Generalized logistic regressions models were used with additional use of splines to investigate continuous exposures, and stratification/mixed-models/forest plots to investigate country-wise variations. To account for the matching conditions, models were at a minimum corrected for age and sex before addition of further co-variates[274]. Therefore, any stratification that was finer than the original matching criteria was potentially biased and such fine stratification was avoided.

Two sensitivity models were used: the first adjusted for age and sex, the second adjusted for age, sex, education, smoking and METs. Analysis of repeated measures per patient such as BMI at different ages used generalized linear mixed models by default. In addition, sensitivity analyses were performed with and without imputation of factors of interest. Imputation was done using multiple iterations (n=100) of predictive mean matching with optional weighted probability sampling of the other variables.

The primary aim for each risk factor is to identify an overall exposure response curve, while the secondary aim is to identify a temporal relationship between the exposure time to the given risk factor and development of the disease. It is also important to compare different exposure (induction) periods. Analysis of exposures at different times

was carried out in the case-control study. All exposures were truncated at 3 years before dates of survey, a threshold below which all intervals between dates of onset and dates of survey, up to the third quartile and of all cohorts, were amounted (Figure 1). This operation was performed for both cases and controls, because failure to adjust cases and controls for timing and extent would introduce a systematic bias.

Figure 5-1 Number of years between date of onset and date of survey by cohort



5.1.11. Discussion

A case-control design is the best compromise to reconcile the low incidence of ALS in a population at risk and the time lapse between any given exposure and the onset of the disease. Our study is population-based and the ALS sample largely reflects the origin populations in each country. Our data suggest that comparison of patients from different countries has the potential to introduce bias including the time lapse between onset of symptoms and enrolment of patients, and variance in the methods used to classify patients. It is also possible that the disease presents with different phenotypes as shown when comparing published reports from different countries [138].

In the present study all exposures have been defined according to standard and/or objective definitions. The use of exposures with the same definition in a large multi-institutional cohort is advantageous, within the limits of a retrospective design, as several factors (and combinations of factors) have been assessed using the same methodology.

Some exposures (e.g., the occupations) are inherently poor surrogate markers of individual risk factors, even if models quantifying dose-response relationships are available and have been used in the project. In this regard, the duration of the induction periods and the intervals between the end of the exposures and the onset of symptoms may be used to provide quantitative measures of exposure.
In this complex scenario, exposures were compared within differing strata of cases and controls and differing regression models, in order to test the consistency of the results on all variables and relevant variable combinations.

Our study design is limited by different socio-cultural, economic and institutional systems among the three countries, which could play an important role in defining, systematizing and interpreting both questionnaires and results. Additionally, the level of education varies across countries and on average controls are better educated than the cases. For these reasons, we can conclude that the results are generalizable to populations having similar characteristics to those participating in this study. Sensitivity

analyses are, however, necessary to verify the consistency of the study findings even though some effects on the dimension of the risk can still be present.

5.1.12. Conclusions

The etiology of ALS is still unknown. The research on genetic factors, which affect up to 10-20% of total ALS patients, has moved forward but no definite environmental risk factors have been defined yet.

Identifying and studying environmental and epigenetic factors can improve our knowledge of the ALS etiology, but methodological limitations can confound analyses. The representativeness and comprehensiveness of population-based registers in The Netherlands, Ireland and Italy represent a great opportunity for studying in a large-scale exposures and incidence of ALS, stratified by country, single cohort, phenotype and genetic mutation. This cross-national study has identified methodological challenges that must be recognized and addressed to minimize bias and assess the independent role of well-defined exposures.

5.2. Hormonal factors and ALS risk in European women[17]

5.2.1. Introduction:

Amyotrophic lateral sclerosis (ALS) is a fatal disease of variable phenotype and unknown etiology. There is evidence that disease manifestation during the second half of life is due to interactions between genetic and environmental factors over time [9,199]. Variants in at least 26 major genes are known to be associated with ALS – the most notable in European populations being the pathological hexa-nucleotide repeat expansion of *C9orf72* [3,4]. However, definitive evidence of environmental factors has proven elusive. The Euro-MOTOR Consortium was established to undertake a systems biology approach towards ALS, which included generation of data on potential environmental etiological factors from ALS cases and controls across three different countries (Ireland, Italy and The Netherlands).

A higher incidence in males has been observed across population based ALS registries, and women are typically older at onset [10,12,16,17]. Furthermore, imaging studies of male and female ALS patients suggest that sexual dimorphism of the anatomical patterns of cortical and sub-cortical pathology in ALS contributes to disease heterogeneity [275]. A previous case-control study found that index–to-ring finger length ratio, a marker of in utero testosterone exposure, differs between ALS cases and controls [276]. Female ALS patients reportedly have fewer children than matched controls [277], while studies of ALS cases and matched controls have demonstrated greater lifetime endogenous estrogen exposure to be a protective factor against ALS [103,278], and to be associated with improved survival amongst cases [103]. Moreover,

---

[17] Chapter 5.2 has been accepted for publication as:

Rooney JPK, Visser AE, MD; D'Ovidio F, Vermoulen R, Beghi E, Chio A, Veldink J, Logroscino G, Van den Berg L, Hardiman O. Euro-MOTOR: A case-control study of hormonal exposures as etiological factors for ALS in women. Neurology. 2017 Sep 19;89(12):1283-1290.

Rooney JPK, Visser AE , and D'Ovidio F contributed equally to this work.

recently we demonstrated that the effect of *C9orf72* on survival in ALS differs in male spinal onset patients compared to females and bulbar onset patients (Chapter 4.3).

Taken together, these observations indicate that sex hormones may play an important role in modulating ALS risk. To further investigate the role of hormones as etiological factors in ALS, the Euro-MOTOR study recorded detailed hormonal histories from study participants.

5.2.2. Methods

Case ascertainment

The recruitment procedures of the Euro-MOTOR study were described in detail previously (Chapter 5.1). Briefly, incident cases and controls matched by age, sex and location were recruited through population-based ALS registers and general practitioner (GP) practices respectively in Ireland, The Netherlands and three regions of Italy: Apulia, Piedmont and Valle d'Aosta, and Lombardy, between 2011 and 2014. Controls were matched to cases for age at survey ± 5 years. Individual matching was used in Ireland and Italy, while frequency matching was used in The Netherlands. All patients met the diagnostic criteria for definite, probable or possible ALS (according to the revised El-Escorial criteria [6]).

Hormonal exposure assessment: Female participants were asked to answer additional questions pertaining to hormonal factors regarding: age at menarche, age at menopause, monthly cycle length and cycle regularity, history of pregnancy (miscarriages and live births were included), breastfeeding history, oral and non-oral contraceptive use, details of surgical gynaecological history (oophorectomy and hysterectomy), and history of hormone replacement therapy (HRT). The full survey questions are provided (Appendix D).

Calculation of lifetime unopposed estrogen exposure: While estrogen exerts a neuro-protective effect, progesterone may oppose such effects [103], and estrogen levels are high and unopposed by progesterone only at certain times [103,279]. Therefore,

calculation of lifetime unopposed endogenous estrogen exposure entailed several steps. Firstly, as estrogen levels are low before menarche and after menopause, the reproductive span was calculated as the number of years from menarche to menopause. Next, since the ovulatory cycle consists of two-phases, a high estrogen pre-ovulatory phase of varied length and a post-ovulatory phase of falling estrogen opposed by high progesterone more consistently two weeks in length, we adjusted the reproductive span by subtracting the last two weeks of every cycle [103,279]. Then, to calculate total endogenous lifetime exposure, nine months were subtracted for each completed pregnancy, three months for each miscarriage (pregnancy invokes high estrogen levels, however this is opposed by high progesterone levels) [103,279]. Most hormonal contraceptives contain combined estrogen and progesterone or progesterone only thus interrupting natural hormonal cycles. Therefore, the total duration of oral contraceptive (OCP) use was subtracted from the adjusted reproductive span. The above adjustments were used previously in studies on ALS and cardiovascular disease [103,279]. All exposure variables were truncated to three years before survey date for both patients and controls in the attempt to remove exposures that may have occurred after ALS onset (Figure 5-1).

Statistical Analysis

Multivariate logistic regression was used to calculate odds ratios for the risk of ALS from each hormonal factor, after adjustment for age, education and study site. These models were then further extended to adjust for smoking history, alcohol use and physical activity. To examine hormonal factors by study site, stratified models were constructed. OCPs were only made legally available (by prescription) in Ireland in 1980 and in Italy in 1978, but had been available in The Netherlands since 1962. Therefore, the age of commencement of OCPs is expected to be heterogenous across sites and may present an important confounding factor. To mitigate expected heterogeneity in this variable across study sites, we categorised the age of first use of OCPs as: 1) never used OCP 2) commenced OCP aged less than or equal to the control median age at commencement 3) commenced OCP aged greater than the control median age at commencement. This

variable was used to model whether the age of commencement of the OCP was a confounding factor.

Analysis was performed as a complete case analysis, and after imputation of missing values. Imputation used multiple iterations (n=30) of predictive mean matching for each region individually implemented using the R package mice [280]. As sensitivity analysis, those experiencing menopause due to surgery (oophorectomy or hysterectomy), those on HRT, or those using OCPs at the time of menopause were excluded from the lifetime endogenous estrogen calculation [103]. Additionally, stratification by education and by site of onset was performed. Finally, analyses were recalculated after exclusion of known *C9orf72* expanded cases. All analysis was carried out using *R* version 3.2.3[190], with additional packages [236,253,257,258,280–282].

5.2.3. Results

Descriptive statistics

Six hundred and fifty three patients and 1,217 controls were included. Two hundred and seventy patients (42%) had bulbar onset disease. Table 5-5 summarizes the demographic characteristics of patients and controls stratified by study site. Overall, cases were 1.2 years older than controls (P = 0.018). More controls were highly educated compared to patients (P = 0.001), a finding driven by the Lombardy and Piedmont and Valle d'Aosta cohorts. Dutch controls were marginally younger than patients (P = 0.045).

Table 5-5 Demographics of female Euro-MOTOR participants by study site

Study site	Case/Control	N	Age at survey		Education		
			Mean (SD)	P (t)	ISCED 0-4 N (%)	ISCED 5-6 N (%)	P (f)
Apulia	Cases	60	63.8 (11.1)		54 (93.1)	4 (6.9)	
	Controls	100	63.7 (11.4)	0.939	92 (92.0)	8 (8.0)	1.000
Lombardy	Cases	86	65.7 (10.3)		73 (88.0)	10 (12.0)	
	Controls	89	65.8 (10.8)	0.971	62 (69.7)	27 (30.3)	0.005
Piedmont & Valle d'Aosta	Cases	124	67.1 (11.0)		112 (94.1)	7 (5.9)	
	Controls	137	65.2 (12.1)	0.191	116 (84.7)	21 (15.3)	0.017
Ireland	Cases	71	65.1 (11.6)		57 (80.3)	14 (19.7)	
	Controls	138	65.4 (11.7)	0.856	110 (79.7)	28 (20.3)	1.000
The Netherlands	Cases	312	64.8 (9.5)		244 (78.7)	66 (21.3	
	Controls	753	63.6 (9.4)	0.047	564 (74.9)	189 (25.1)	0.206

ISCED = International Standard Classification of Education; t = Student's t test; f = Fishers's test

Table 5-6 and Table 5-7 summarize the endogenous and exogenous hormonal factors of patients and controls stratified by country. Generally, there were minor differences between cases and controls within study sites with a lack of consistent direction across sites. Age of menarche and reproductive span were comparable across study sites and between cases and controls (Table 5-6). Notably, the use of contraceptives was lower in cases compared to controls within all countries, and reached statistical significance in Lombardy, The Netherlands, and overall (Table 5-7). Duration of OCP use was lower in cases compared to controls across all sites except Apulia, and with $P < 0.05$ in Ireland, The Netherlands, and overall. Comparing controls across study sites, 25% of women from Apulia, 43% of women from Lombardy, 38% of women from Piedmont and Valle d'Aosta, 49% of Irish and 76% of Dutch women had a history of hormonal contraceptive use. Few participants used non-oral hormonal contraceptives, therefore the remainder of our analysis of contraceptives is focused on OCPs.

Lifetime unopposed endogenous estrogen exposure was lower in the Dutch controls (mean 11.3 years) versus the Irish (mean 14.0 years) and the Italians (Apulia: 15.3 years, Lombardy: 15.8 years, Piedmont and Valle d'Aosta: 15.2 years). Lifetime unopposed estrogen exposure was 14.2 years in all cases versus 12.9 years in all controls ($P < 0.001$).

Table 5-6 Summary of endogenous hormonal exposure factors by study site

Study Site	N	Age at menarche Mean (SD)	Reached meno-pause N (%)	Age at meno-pause Mean (SD)	History of irregular Periods N (%)	Mean reproductive span in years† (SD)	Number of pregnancies Median (IQR)	Live births median (IQR)	Has breastfed ? N (%)
Apulia									
Cases	60	12.7 (1.5)	49 (84.5)	47.2 (5.9)	11 (18.3)	33.8 (5.9)	3.0 (2.0, 4.0)	2.0 (2.0, 3.0)	45 (77.6)
Controls	100	12.7 (1.8)	89 (89.0)	48.3 (5.5)	13 (13.0)	34.9 (6.0)	2.0 (2.0, 3.0)	2.0 (1.8, 3.0)	63 (63.0)
P value		0.921 (t)	0.461 (f)	0.289 (t)	0.110 (f)	0.281 (t)	0.104 (m)	0.039 (m)	0.076 (f)
Lombardy									
Cases	86	12.7 (1.4)	69 (84.1)	50.6 (4.7)	16 (18.6)	38.1 (6.8)	2.0 (1.0, 3.0)	2.0 (1.0, 3.0)	49 (59.0)
Controls	89	12.8 (1.5)	74 (84.1)	51.4 (4.4)	19 (21.3)	37.6 (5.3)	2.0 (1.0, 3.0)	2.0 (1.0, 2.0)	56 (64.6)
P value		0.756 (t)	1.000 (f)	0.272 (t)	0.747 (f)	0.610 (t)	0.309 (m)	0.613 (m)	0.529 (f)
Piedmont & Valle d'Aosta									
Cases	124	13.2 (1.6)	110 (90.9)	50.5 (4.1)	31 (25.0)	36.2 (5.3)	2.0 (1.0, 3.0)	2.0 (1.0, 2.0)	85 (71.4)
Controls	137	12.7 (1.4)	115 (84.6)	49.4 (5.3)	52 (38.0)	35.4 (6.4)	2.0 (1.0, 2.0)	2.0 (1.0, 2.0)	88 (64.2)
P value		0.005 (t)	0.134 (f)	0.085 (t)	0.016 (f)	0.272 (t)	0.058 (m)	0.249 (m)	0.231 (f)
Ireland									
Cases	71	13.5 (1.5)	62 (87.3)	48.0 (6.4)	13 (18.3)	33.9 (7.1)	3.0 (1.5, 5.0)	3.0 (1.0, 4.0)	25 (35.2)
Controls	138	13.2 (1.5)	124 (89.9)	48.1 (6.5)	27 (19.6)	33.9 (7.1)	4.0 (2.0, 5.0)	3.0 (2.0, 4.0)	59 (42.8)
P value		0.273 (t)	0.643 (f)	0.903 (t)	0.431 (f)	0.980 (t)	0.081 (m)	0.134 (m)	0.302 (f)
The Netherlands									
Cases	312	13.4 (1.5)	272 (89.2)	49.6 (5.8)	79 (25.7)	35.6 (6.2)	2.0 (2.0, 3.0)	2.0 (1.0, 3.0)	130 (56.5)
Controls	753	13.4 (1.60)	664 (88.8)	49.4 (5.70)	153 (20.4)	35.5 (6.1)	2.0 (2.0, 3.0)	2.0 (2.0, 3.0)	329 (62.7)
P value		0.503 (t)	0.914 (f)	0.627 (t)	0.189 (f)	0.650 (t)	0.242 (m)	0.298 (m)	0.124 (f)
Overall									
Cases	653	13.2 (1.5)	562 (88.2)	49.5 (5.5)	150 (23.0)	35.7 (6.3)	2.0 (2.0, 3.0)	2.0 (1.0, 3.0)	334 (59.5)
Controls	1217	13.2 (1.6)	1066 (88.1)	49.3 (5.7)	264 (21.7)	35.4 (6.2)	2.0 (2.0, 3.0)	2.0 (2.0, 3.0)	595 (60.3)
P value		0.432 (t)	1.000 (f)	0.471 (t)	0.782 (f)	0.284 (t)	0.394 (m)	0.190 (m)	0.787 (f)

† Truncated at 3 years prior to survey date to exclude exposure post onset

P values <0.05 are highlighted in bold text

t = Student's t test; m = Mann-Whitney U test; f = Fishers's test

Table 5-7 Summary of exogenous hormonal factors by study site

Study Site	N	Hormonal contraceptive use, N (%)	OCP use N (%)	Mean years taking OCPs† (SD)	Used HRT N (%)	Hysterec-tomy N (%)	Oophorec-tomy N (%)
Apulia							
Cases	60	11 (19.0)	11 (19.0)	1.5 (5.6)	9 (15.5)	11 (19.0)	10 (18.9)
Controls	100	25 (25.0)	24 (24.0)	1.7 (4.4)	7 (7.0)	24 (24.0)	23 (24.0)
P value		0.436 (f)	0.553 (f)	0.795 (t)	0.104 (f)	0.553 (f)	0.541 (f)
Lombardy							
Cases	86	22 (26.5)	22 (26.5)	1.86 (5.56)	12 (14.5)	13 (16.0)	6 (8.1)
Controls	89	38 (42.7)	38 (42.7)	3.68 (6.30)	16 (18.6)	18 (21.2)	10 (12.0)
P value		0.037 (f)	0.037 (f)	0.051 (t)	0.538 (f)	0.431 (f)	0.443 (f)
Piedmont & Valle d'Aosta							
Cases	124	33 (27.7)	32 (26.9)	1.65 (4.36)	15 (12.6)	14 (11.8)	12 (10.4)
Controls	137	52 (38.0)	51 (37.2)	2.52 (5.11)	19 (14.0)	20 (14.6)	22 (16.3)
P value		0.086 (f)	0.083 (f)	0.148 (t)	0.384 (f)	0.581 (f)	0.199 (f)
Ireland							
Cases	71	25 (35.2)	24 (33.8)	1.37 (3.00)	25 (35.2)	16 (22.5)	7 (15.2)
Controls	138	67 (48.6)	64 (46.4)	3.24 (6.24)	45 (32.6)	33 (24.3)	15 (20.3)
P value		0.078 (f)	0.104 (f)	0.022 (t)	0.758 (f)	0.864 (f)	0.629 (f)
The Netherlands							
Cases	312	215 (68.9)	212 (67.9)	8.82 (9.59)	32 (10.5)	59 (19.0)	31 (10.2)
Controls	753	574 (76.3)	565 (75.1)	10.36 (9.96)	120 (16.0)	141 (18.8)	76 (10.3)
P value		0.014 (f)	0.019 (f)	0.023 (t)	0.020(f)	0.931 (f)	1.000 (f)
Overall							
Cases	653	306 (47.6)	301 (46.8)	5.1 (8.3)	93 (14.6)	113 (17.7)	66 (11.1)
Controls	1217	756 (62.2)	742 (61.0)	7.5 (9.3)	207 (17.1)	236 (19.5)	146 (13.0)
P value		<0.001 (f)	<0.001 (f)	<0.001 (t)	0.051 (f)	0.349 (f)	0.316 (f)

OCP = oral contraceptive pill; HRT = hormone replacement therapy. † Truncated 3 years prior to survey date. P values <0.05 are highlighted in bold text. t = Student's t test; m = Mann-Whitney U test; f = Fishers's test

Table 5-8 Odds Ratios for reproductive hormonal factors after multivariable logistic regression

	Complete Cases analysis (N = 1,467)		Complete Case analysis excluding known C9orf72 expanded cases (N = 1,366)		After imputation of missing values (N=1,870)		After imputation and adjusted for additional lifestyle factors (N=1,870)	
Reproductive Factor	OR (95% CI)†	P value	OR (95% CI)†	P value	OR (95% CI)†	P value	OR (95% CI)††	P value
Age at menarche	1.06 (0.99 – 1.14)	0.113	1.06 (0.99 – 1.14)	0.115	1.05 (0.99 – 1.12)	0.106	1.05 (0.99 – 1.12)	0.123
Reproductive span (per year)	1.00 (0.99 – 1.02)	0.702	1.00 (0.98 – 1.02)	0.941	1.00 (0.98 – 1.02)	0.849	1.00 (0.98 – 1.02)	0.978
Ever pregnant	0.96 (0.70 – 1.33)	0.813	0.92 (0.66 – 1.28)	0.602	0.92 (0.68 – 1.25)	0.589	0.92 (0.67 – 1.25)	0.582
Time spent pregnant (per year)	0.99 (0.88 – 1.10)	0.220	0.98 (0.88 – 1.09)	0.701	0.97 (0.88 – 1.07)	0.594	0.98 (0.89 – 1.08)	0.714
Ever breastfed	0.91 (0.73 – 1.14)	0.410	0.89 (0.71 – 1.12)	0.329	0.94 (0.76 – 1.14)	0.518	0.96 (0.78 – 1.17)	0.678
Time spent breastfeeding (per year)	1.08 (0.95 – 1.23)	0.220	1.07 (0.94 – 1.21)	0.330	1.11 (0.99 – 1.24)	0.072	1.12 (1.00 – 1.26)	0.049
Ever used oral contraceptives	0.65 (0.51 – 0.84)	<0.001	0.64 (0.50 – 0.84)	<0.001	0.66 (0.53 – 0.83)	<0.001	0.67 (0.53 – 0.84)	<0.001
Duration of oral (contraceptive use (per year)	0.98 (0.96 – 0.99)	0.001	0.98 (0.96 – 0.99)	0.002	0.98 (0.97 – 0.99)	0.001	0.98 (0.96 – 0.99)	0.001
Lifetime endogenous estrogen exposure (per year)	1.03 (1.01 – 1.05)	0.017	1.03 (1.00 – 1.05)	0.036	1.02 (1.00 – 1.04)	0.051	1.02 (1.00 – 1.04)	0.049

† Adjusted for age, education and cohort (n = 5)

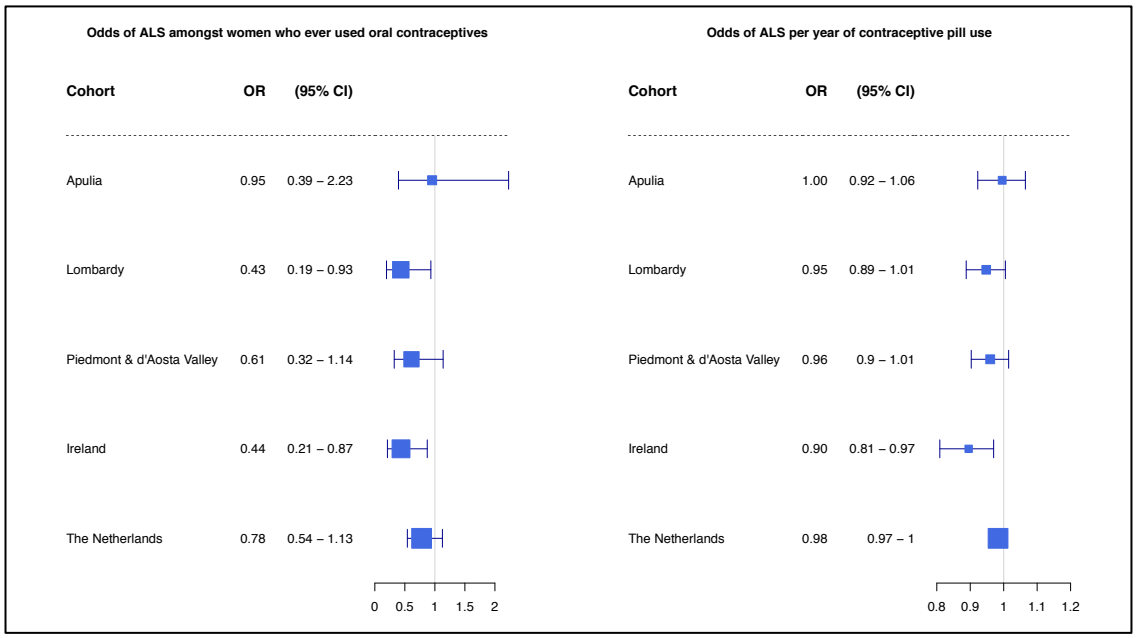
†† Adjusted for age, education, cohort (n = 5), lifetime physical activity, smoking and alcohol status

Reproductive hormonal factors

Table 5-8 shows OR's for ALS for each of the hormonal factors in turn after multivariate logistic regression. On complete case analysis, a history of ever having used OCP was significantly associated with reduced OR for ALS (OR = 0.65; 95% CI: 0.51 – 0.84). A dose response relationship was evident between duration of OCP use and reduced OR (0.98; 95% CI: 0.96 – 0.99 per year of use). These findings remained significant after multiple imputation of missing values and after correction for other suspected lifestyle etiological

factors including lifetime physical activity, smoking and alcohol consumption. Those who never used the OCP were excluded. However, this had minimal effect on the apparent dose response relationship with OR = 0.98 (95% CI: 0.96 – 1.00) per year of use after exclusions. Models stratified by education level found that history of OCP use (OR 0.61; 95% CI: 0.46 – 0.81) and duration of OCP use (OR 0.97; 95% CI: 0.95 – 0.99 per year) were significant in the lower education group, but not in the higher education group (history of OCP: OR 0.80; 95% CI: 0.45 – 1.45), duration of OCP (OR 0.99; 95%CI: 0.96 – 1.03 per year). They also remained significant in models of spinal or bulbar patients only vs controls. Similarly, exclusion of women carrying the *C9orf72* expansion did not significantly alter results (Table 5-8). OR's for "ever use" of OCP and duration of OCP stratified by cohort are shown in Figure 5-2. It is noteworthy that the OR's were in the same direction towards reduced OR in each study site (Figure 5-2). "Ever use" of OCP was clearly associated with decreased OR in Ireland (OR = 0.44; 95% CI: 0.21 – 0.87) and Lombardy (OR = 0.42 95%CI: 0.20 – 0.89), while duration of OCP use was clearly associated with decreased OR in Ireland only (OR = 0.90; 95% CI: 0.81 – 0.97). However, the mean OR's were less than one across all study sites (except Apulia) per year of OCP use (Figure 5-2).

Figure 5-2 Forest-plot showing odds ratio for history and duration of OCP use stratified by study site



A reduced OR for ALS was found in those who started the OCP at younger ages (OR 0.56 95%CI: 0.41 – 0.77) and those who started the OCP at older ages (OR 0.74 95%CI: 0.55 – 0.99) when compared to those who never used the OCP. Broken down by study site, these results showed significant heterogeneity across sites with small numbers in some subgroups (Table 5-9). Imputation of missing values and correction for lifestyle factors did not change heterogeneity across sites meaningfully (Table 5-9). However, it did reduce the difference in ORs between those starting at younger ages (OR 0.54, 95%CI: 0.41 – 0.72) and older ages (OR 0.67, 95%CI: 0.53 – 0.86) overall (Table 5-9).

Lifetime endogenous estrogen exposure

Lifetime endogenous estrogen exposure was associated with a mildly increased OR for ALS with borderline significance even after imputation and correction for other lifestyle cofactors (OR = 1.02; 95%CI: 1.00 – 1.04) (Table 5-8). After exclusion of women who experienced surgical menopause, lifetime endogenous estrogen exposure was no longer significant (OR = 0.99; 95% CI: 0.96 – 1.02). None of the remaining examined variables in Table 5-8 revealed significant associations. A model was built including each term contributing to lifetime endogenous estrogen exposure as a separate term. In this model, only duration of OCP use remained significant with OR 0.98 (95%CI: 0.96 – 0.99) per year of OCP use.

Hormone replacement therapy, hysterectomy and oophorectomy

HRT use was low amongst cases and controls across all study sites except Ireland (Table 5-7). Overall, the OR for HRT use after adjustment for age, education and study site was 0.82 (95% CI: 0.62 – 1.08). Stratified by study site, HRT use was significantly associated with ALS only in The Netherlands (OR 0.57, 95% CI: 0.37 – 0.85). Age at commencement did not differ between cases and controls overall or for each study site. The duration of HRT use was not associated with ALS either overall (OR 1.03, 95% CI: 0.99 – 1.08) or in The Netherlands only (OR 1.03, 95%CI: 0.97 – 1.08). A history of hysterectomy (OR 0.85, 95%CI: 0.66 – 1.10), or oophorectomy (OR 0.77, 95%CI: 0.56 – 1.06) was not associated with ALS.

Table 5-9 OCP commencement age stratified by cohort

Age started OCP	Cases	Controls	Complete case analysis OR† (95% CI)	After imputation and adjusted for lifestyle factors†† OR (95% CI)
Apulia				
Never used	47 (78%)	76 (76%)	1	1
< Median age	8 (13%)	12 (12%)	1.16 (0.42 – 3.12)	0.92 (0.34 – 2.52)
≥ Median age	3 (5%)	7 (7%)	0.74 (0.15 – 2.94)	0.90 (0.21 – 3.97)
Lombardy				
Never used	61 (71%)	51 (57%)	1	1
< Median age	13 (15%)	19 (21%)	0.56 (0.21 – 1.44)	0.60 (0.27 – 1.37)
≥ Median age	8 (9%)	18 (20%)	0.32 (0.11 – 0.82) *	0.34 (0.12 – 0.92) *
Piedmont & Valle d'Aosta				
Never used	8 (70%)	86 (63%)	1	1
< Median age	17 (14%)	26 (19%)	0.59 (0.27 – 1.31)	0.61 (0.27 – 1.37)
≥ Median age	15 (12%)	25 (18%)	0.62 (0.29 – 1.32)	0.64 (0.29 – 1.38)
Ireland				
Never used	47 (66%)	74 (54%)	1	1
< Median age	9 (13%)	36 (26%)	0.21 (0.07 – 0.58) **	0.24 (0.09 – 0.66) *
≥ Median age	14 (20%)	27 (20%)	0.69 (0.30 – 1.50)	0.70 (0.32 – 1.55)
The Netherlands				
Never used	100 (32%)	187 (25%)	1	1
< Median age	94 (30%)	297 (39%)	0.62 (0.39 – 0.99) *	0.67 (0.45 – 1.01)
≥ Median age	115 (37%)	266 (35%)	0.88 (0.59 – 1.32)	0.90 (0.65 – 1.26)
Overall				
Never used	342 (54%)	474 (39%)	1	1
< Median age	141 (22%)	390 (32%)	0.56 (0.41 – 0.77) **	0.54 (0.41 – 0.72) **
≥ Median age	155 (24%)	343 (28%)	0.74 (0.55 – 0.99) *	0.67 (0.53 – 0.86) **

*** P <0.05 ; ** P < 0.005; †Adjusted for age and education. ††Adjusted for age, education, lifetime physical activity, smoking and alcohol status. Note the median age is defined by the median age of starting the OCP amongst controls at each site**

Fertility

Overall there was little difference in fertility between cases and controls (Table 5-6). After adjustment for study site, education, age and contraceptive use via multivariate logistic regression, the number of pregnancies or number of live births was not significantly associated with case/control status overall. However, ever being pregnant was associated with a reduced OR in Ireland (OR 0.35; 95% CI: 0.14 – 0.84), but with increased OR in Apulia (OR 3.43; 95% CI:1.15 – 13.0).

5.2.4. Discussion

The analysis of sex hormones, pregnancy, the use of oral contraception and disease risk is complicated and prone to hidden confounders, particularly across different countries with different levels of access to contraception, and different cultural attitudes towards fertility. Notwithstanding, this large case control study across five different sites has demonstrated a negative association between ALS and hormonal contraception use in women. We found reduced odds of ALS for a history of ever using OCP and that the risk decreased further with longer duration of OCP use, even among OCP users only.

This study did not confirm previous findings from the Netherlands which showed a negative association between lifetime endogenous estrogen exposure and ALS risk [103], nor findings from Italy showing a negative association between reproductive span and ALS risk [283]. Conversely, our data suggests a weak association of increased risk with lifetime endogenous estrogen exposure, and found no effect for reproductive span (Table 5-8). By modelling the individual terms contributing to lifetime endogenous estrogen exposure we determined that this finding was driven by subtraction of the duration of OCP use in the calculation of lifetime endogenous exposure.

We found that HRT use was strongly associated with reduced risk of ALS in The Netherlands only. This discrepancy might be due to the relatively low prevalence of HRT users in others sites compared to The Netherlands, or the use of different formulations of HRT. We were unable to explore this possibility due to poor responses to questions about HRT drug names. This finding also contrasts findings from the U.S. that associated

HRT use with increased ALS risk (OR 1.9, 95% CI 0.9 – 3.8) [278]. However, our finding for HRT is compatible with our finding that OCP use is associated with lower ALS risk, since both are forms of exogenous estrogens and progestogens. Stratifying by age of OCP commencement suggests that it may be an important confounding factor, although we interpret this result with caution due to the heterogeneity between sites and the smaller difference between age groups after imputation of missing values and correction for lifestyle covariates (Table 5-9). Still, these findings are interesting in light of the 'timing hypothesis' of neuroprotection due to HRT commenced at younger ages in Alzheimer's research [284,285], although we did not observe any differences in age of commencement of HRT in ALS cases versus controls. It has been reported that the past use of OCPs and HRTs differently associate with long term reductions in circulating sex hormones [286]. Therefore, timing and formulation of exogenous hormones may both be important.

None of the other hormonal parameters examined were associated with ALS risk, although it was not possible to determine other factors of likely importance such as progesterone exposure, androgen levels, cyclical fluctuation in estrogen and progesterone levels, or follicle-stimulating hormone and luteinizing hormone levels. Notwithstanding, our findings were robust to sensitivity analysis and after exclusion of those who underwent surgical menopause.

The differences between our findings and those of others may be in part explained by the relative complexity of hormonal profiles throughout life and confounding factors. While the previous Dutch study argued that endogenous estrogens may exert neuroprotective effects [103], our current findings indicate exogenous hormones might exert such effects. While there is evidence from in-vitro studies for neuro-protective mechanisms of estrogen [287,288], and also that estrogens may slow progression in female hSOD1 G93A transgenic mice [289], it is also the case that estrogens exhibit complex interactions with androgens [290] and we note that higher prenatal testosterone determined by digit length has been linked to increased ALS risk in a case control study [276]. Moreover, testosterone has recently been associated with brain maturation in puberty [291], an age range which significantly overlaps the

commencement age range of the OCP in our study. This is consistent with observations from animal studies that a long preclinical period lasting decades may exist in ALS [262] and with the recently proposed idea that ALS results from a multi-step process [9].

Our analysis was complicated by different usage rates of contraceptives in the five study sites. The age at commencement of OCP was lower and duration of use was longer in Dutch cases and controls compared to Italian and Irish cases and controls. Contraceptive use has been found to associate with age, marital status and social class in Ireland [292] and contraceptive practices are known to vary with socio-demographic factors across Europe (and between North and South Italy) [293,294]. Therefore, the differences in contraceptive use we observed are likely due to different cultural/religious attitudes towards birth control between the five study sites, along with the late availability of hormonal contraceptives in Ireland and Italy. In addition, there are other potential sources of confounding – such as compliance with medication instructions or the prescription of OCPs for indications other than birth control (e.g. treatment of hormonal conditions such as polycystic ovary syndrome). We cannot exclude the possibility that the formulation of OCPs and HRT drugs vary across countries.

The EuroMOTOR survey of hormonal etiological factors for ALS in women is the largest study on this topic to date. It benefits from prospective recruitment of incident patients and matched controls across three countries from a source population of 34.1 million individuals and the application of standardised protocols across all sites. Although our findings are at variance with previous work, there is sufficient evidence to support a role between sex hormones and risk of ALS in women. Our study shows the novel finding of a negative association between OCP use and the risk of ALS along with an apparent dose response effect of the OCP. These findings need replication and would be reinforced by prospective studies.

5.3. Premorbid BMI

5.3.1. Introduction

Weight-loss is a ubiquitous symptom in ALS, particularly in the setting of bulbar onset disease. Decreased BMI has been associated with prognosis [75,295–297], and gastrostomy insertion, including percutaneous endoscopic gastrostomy (PEG) and radiologically inserted gastrostomy (RIG), are frequently used as therapeutic interventions in ALS despite unclear evidence for effectiveness.

There are numerous reasons why weight loss is a strong feature in ALS. The most obvious factors are the loss of swallowing ability particularly in bulbar onset patients which in turn leads to difficulty maintaining caloric intake, and the wasting of muscles due to disuse. However, loss of BMI appears to include mechanisms beyond this. A survey of 51 patients found that loss of appetite was a major feature and that this was not explained by the level of dysphagia present [298]. There is also evidence of disturbed energy metabolism and some patients may exhibit hyper-metabolism at rest [299]. Thus loss of BMI in ALS is complex and is likely the result of numerous factors.

Furthermore, it is unclear when loss of BMI begins during the course of ALS although in rare cases unintended weight loss may be the first symptom noticed by the patient. The Euro-MOTOR project survey included questions regarding longitudinal weights throughout life. In this chapter we analyse this data to investigate the age at which weight loss begins in ALS patients versus controls.

5.3.2. Methods

Data collection

The Euro-MOTOR survey included a section regarding biometrics, including height at the time of survey and weight at the time of survey and at specified ages throughout adult life (i.e. aged 20, 30, 40, 50, 60 and 70 years). This data was extracted from the survey responses, along with several other variables including study site, gender, patient or control status, educational level and in patients site of onset and *C9orf72* status where

known. Historical BMI's were calculated by assuming the height at survey date was constant throughout adult life.

Statistical Analysis

Descriptive statistics were performed and missing value rates examined.

Longitudinal BMI data were modelled using a mixed effects model to account for repeated measures per individual. A spline term was included to allow deviation from a linear fit. Models were adjusted a-priori for gender and study-site as fixed effects. Models allowing for interaction of fixed effects were compared using ANOVA tests. When a final model was selected, the model was re-specified with patients stratified by *C9orf72* status (only participants from Ireland, The Netherlands and Piedmont and Valle d'Aosta were included in the *C9orf72* analysis as many patients from Apulia and Lombardy were not tested for the *C9orf72* expansion).

5.3.3. Results

Descriptive statistics

For historical BMI's, there were a total of 20,355 data-points out of a possible 26,378. Missing values were notably higher for Ireland than for other study sites (Table 5-10). This was due to the wording of the English version of the survey which asked the question regarding historical BMI's conditional on weight loss at time of survey, whilst it was mandatory in the Dutch and Italian versions. Missing values were least in the Dutch and Piedmont and Valle D'Aosta sites. Due to the high rate of missing values, multiple imputation was not used for the BMI analyses.

Table 5-10 Response rates for historical BMI data

	Apulia	Lombardy	Piedmont & d'Aosta Valle	Ireland	The Netherlands
Patients					
Num. valid	624	873	1389	436	3793
Num. Missing (%)	201 (30%)	235 (26%)	186 (14%)	610 (65%)	862 (22%)
Controls					
Num. valid	956	836	1443	1100	8905
Num. Missing(%)	281 (30%)	298 (32%)	251 (19%)	979 (56%)	2120 (23%)

Table 5-11 displays summary statistics for height and for male and female participants stratified by study site at time of survey. As expected, male and female patients across all sites have significantly lower BMI when compared to controls. In an unexpected finding, both male and female patients were significantly shorter than controls in both Lombardy and The Netherlands, while the shorter height of male and female patients approached significance in Piedmont and D'Aosta Valley.

Table 5-11 Height and BMI of Euro-MOTOR participants at survey date

	Females			Males		
	N	Height (SD)	BMI (SD)	N	Height (SD)	BMI (SD)
Apulia						
Patients	60	1.58 (5.8)	24.2 (3.8)	81	1.69 (6.6)	24.6 (4.1)
Controls	100	1.59 (6.3)	26.5 (4.8)	113	1.71 (7.6)	27.9 (4.1)
P value		0.709	**0.003**		0.092	**<0.001**
Lombardy						
Patients	86	1.59 (6.3)	23.2 (4.8)	100	1.73 (6.7)	24.1 (3.4)
Controls	89	1.61 (6.0)	24.7 (4.0)	101	1.75 (7.1)	25.5 (3.6)
P value		**0.015**	**0.034**		**0.030**	**0.005**
Piedmont & Valle D'Aosta						
Patients	124	1.59 (6.8)	23.3 (4.4)	138	1.72 (6.2)	24.2 (3.6)
Controls	137	1.61 (6.2)	25.6 (5.9)	153	1.73 (6.8)	26.4 (3.6)
P value		0.057	**<0.001**		0.067	**<0.001**
Ireland						
Patients	71	1.62 (7.1)	24.8 (5.1)	106	1.75 (6.8)	25.5 (3.7)
Controls	138	1.62 (6.4)	26.4 (5.1)	211	1.76 (6.3)	27.1 (3.8)
P value		0.796	**0.030**		0.207	**<0.001**
The Netherlands						
Patients	313	1.66 (6.4)	24.6 (4.3)	478	1.79 (7.4)	24.8 (3.4)
Controls	753	1.67 (6.4)	26.0 (4.3)	1127	1.80 (6.7)	26.3 (3.2)
P value		**<0.001**	**<0.001**		**0.018**	**<0.001**

Bold indicates that P ≤ 0.05

Longitudinal Models of BMI

Mixed effects models of BMI vs age including 4 spline knots were chosen as optimum. ANOVA tests indicated that 5 knots improved fit (P <0.001). However, on visualisation of models there was very marginal difference and therefore the simpler model with 4 knots was preferred. Final model fits are plotted for females and males in Figure 5-3 and Figure 5-4 respectively.

Figure 5-3 Longitudinal BMI of female Euromotor cohort

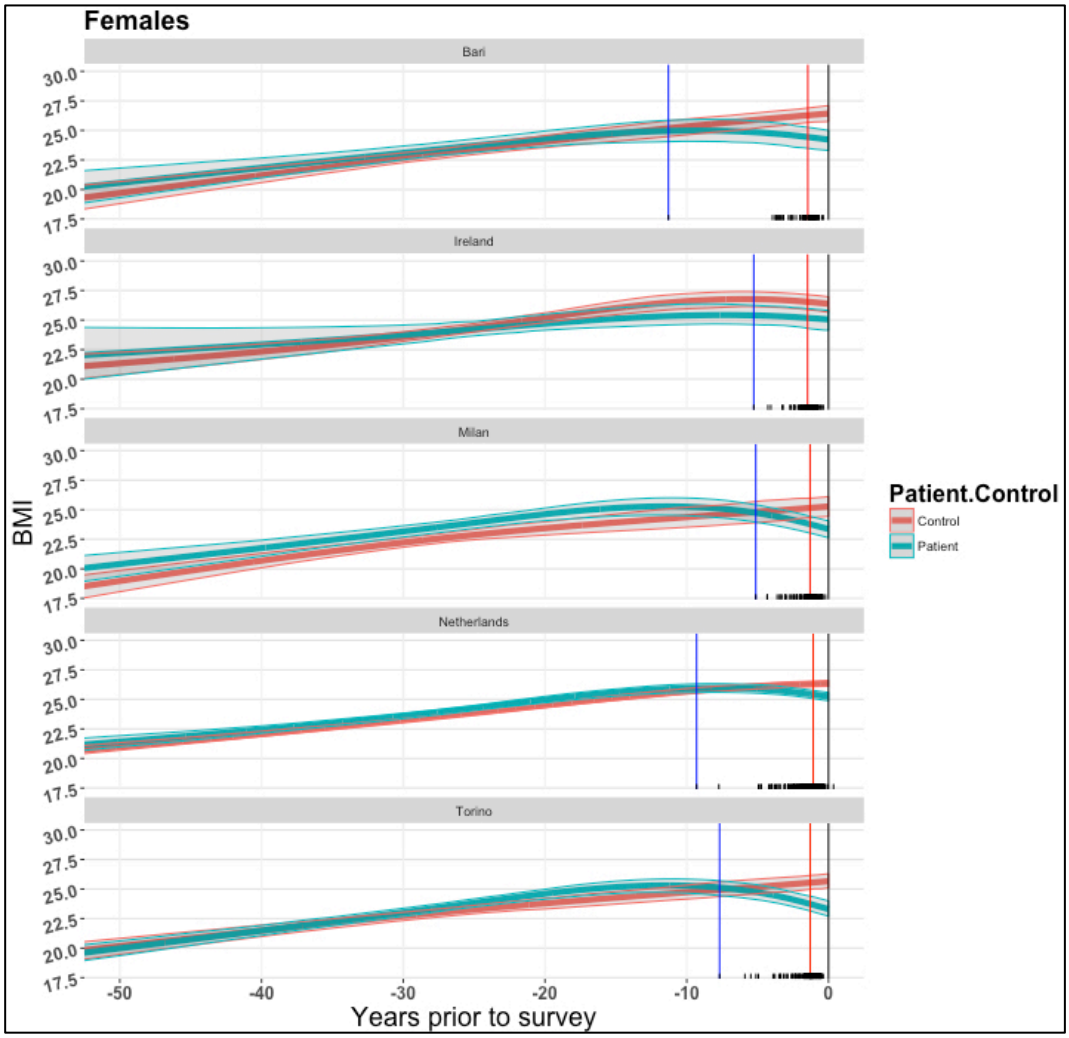


Figure 5-3 Legend: The rug plot represents individual times of onset of patients. The blue vertical lines depict the time of onset of the first patient by site, the red vertical lines predict the median onset time, and the black vertical line is the time of survey.

Figure 5-4 Longitudinal BMI of male Euromotor cohort

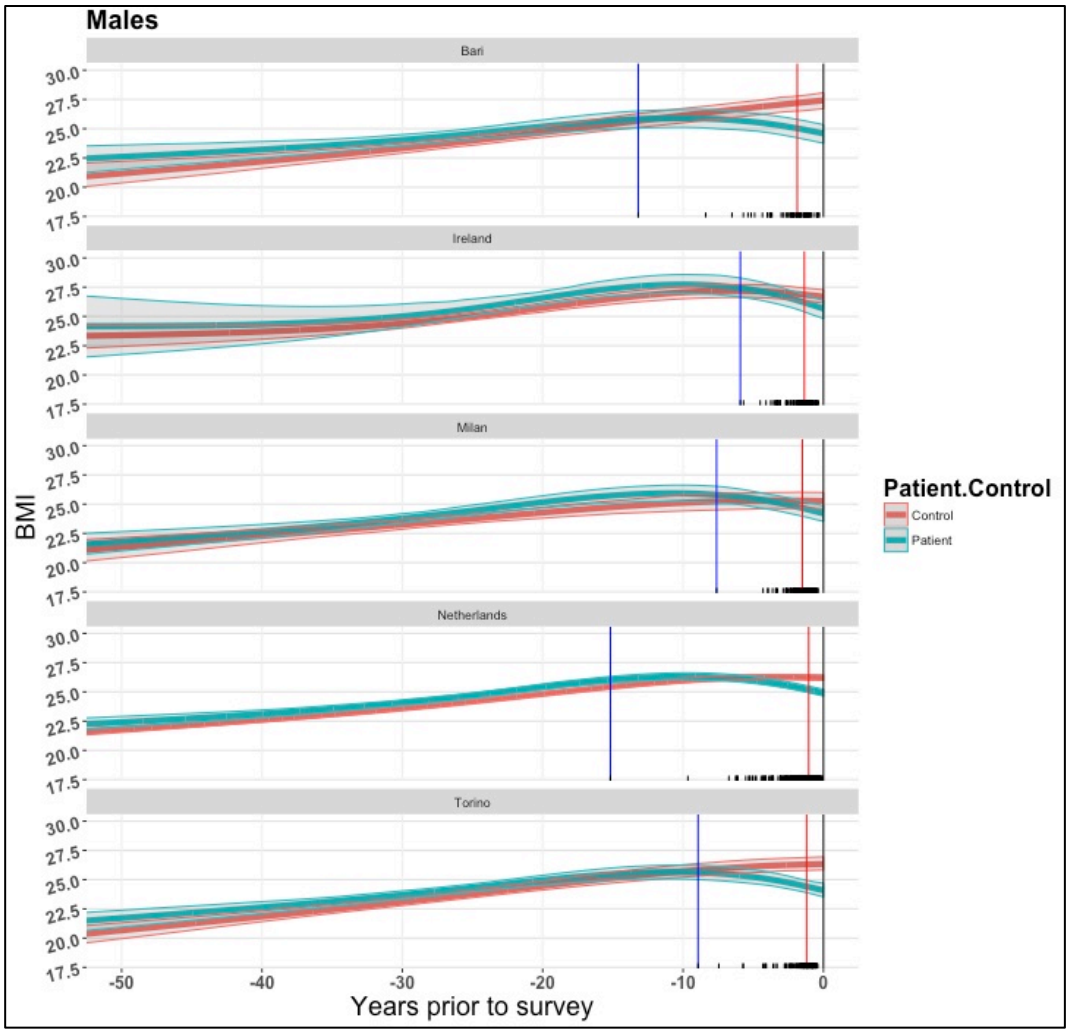


Figure 5-4 **Legend: The rug plot represents individual times of onset of patients. The blue vertical lines depict the time of onset of the first patient by site, the red vertical lines predict the median onset time, and the black vertical line is the time of survey.**

Notably, in Figure 5-3 and Figure 5-4 loss of BMI appears to begin up to a decade before survey time. However as depicted by the rug plots and blue vertical lines in the plots above, there were outliers with earlier onset and we cannot out-rule earlier onset patients are influencing the patterns seen, although BMI loss does appear to begin considerably before the median onset time (red vertical lines).

On stratifying patients by *C9orf72* expansion status (Figure 5-5 and Figure 5-6), it is apparent that loss of BMI has begun at earlier times for carriers of the expansion. In particular, for the Netherlands and Piedmont and D'Aosta Valley, which have the least missing values for historical BMI data, *C9orf72* patients appear to have reduced BMI for many decades prior to disease onset, perhaps even as a lifelong trait.

Figure 5-5 Longitudinal BMI of female Euromotor cohort stratified by C9orf72 status

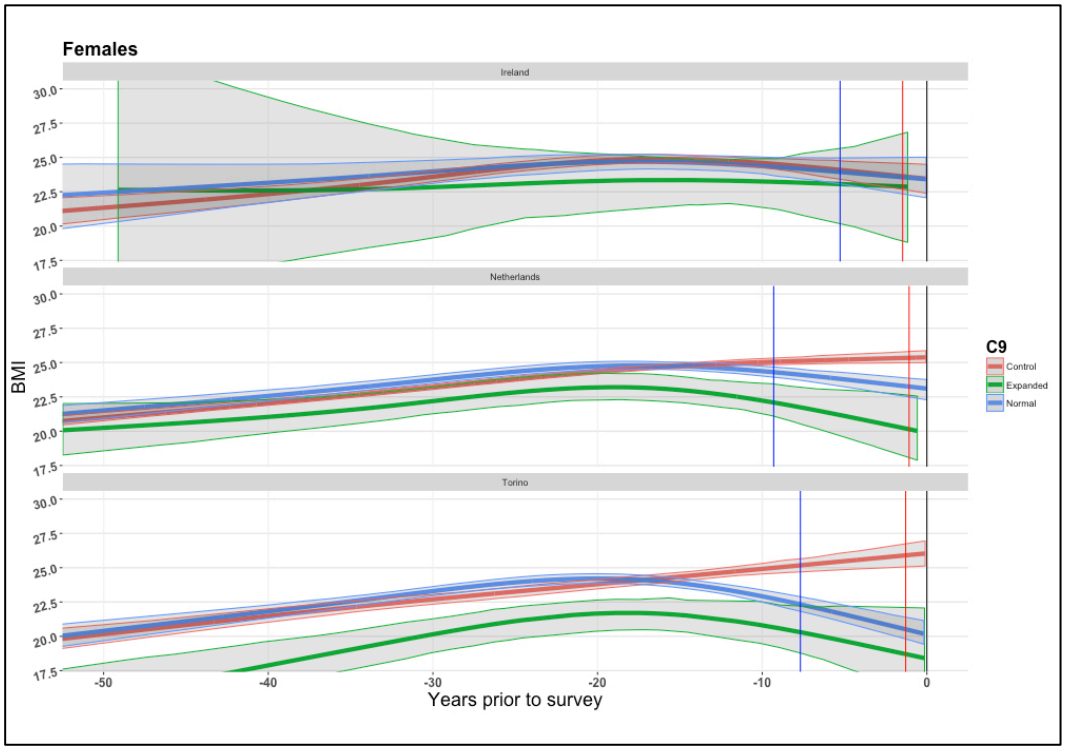


Figure 5-5 **Legend: The blue vertical lines depict the time of onset of the first patient by site, the red vertical lines predict the median onset time, and the black vertical line is the time of survey.**

Figure 5-6 Longitudinal BMI of male Euromotor cohort stratified by C9orf72 status

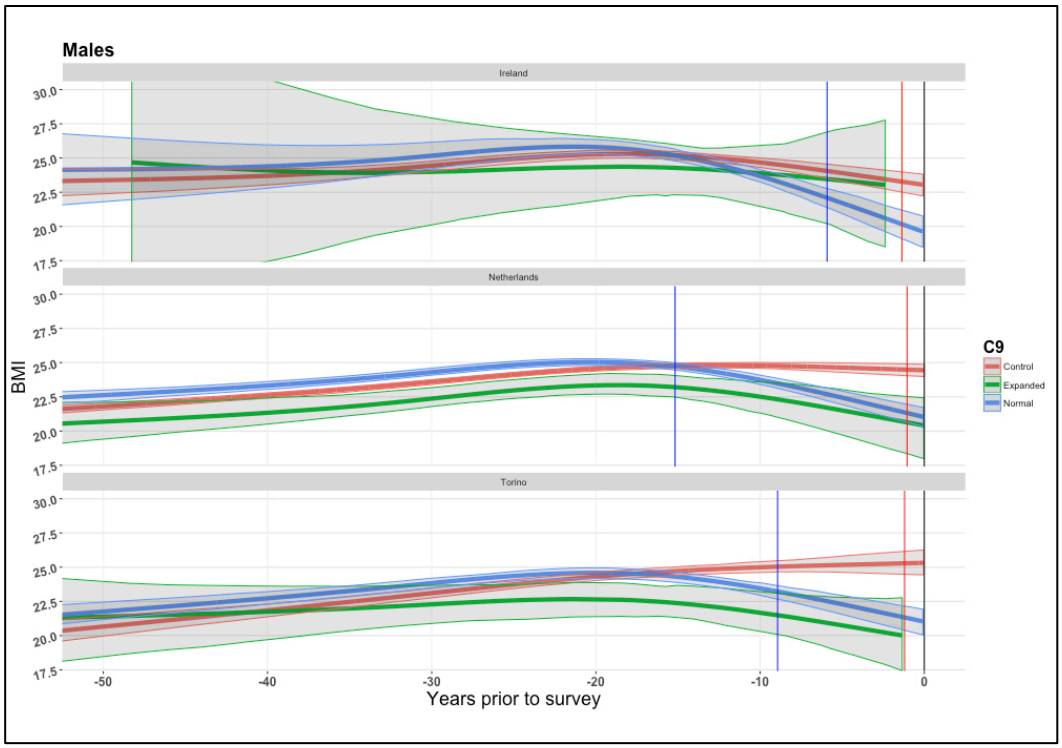


Figure 5-6 **Legend: The blue vertical lines depict the time of onset of the first patient by site, the red vertical lines predict the median onset time, and the black vertical line is the time of survey.**

5.3.4. Discussion

These results provide historical BMI data for a large cohort of ALS patients and healthy controls from the corresponding populations. Our analysis indicates that the loss of BMI in ALS patients may begin up to a decade before reported onset of symptoms, although there was some heterogeneity by study site. There was no clear difference in longitudinal plots of BMI for male vs female participants.

On stratification by *C9orf72* status, a clear pattern emerged that *C9orf72* expanded males and females had a lower BMI for several decades prior to disease onset – suggesting low BMI may be a life-long trait of *C9orf72* expansion carriers. To our knowledge, this is the first report of premorbid low BMI associated with the *C9orf72* expansion phenotype. Interestingly, mice bred with conditional knock-down of *C9orf72* showed a similar life-long reduction of body weight, although they did not develop neuro-motor symptoms [300].

We also found evidence that ALS patients are shorter in height than controls, although not across all study sites. To our knowledge this has not been previously reported and therefore should be interpreted with caution. A selection bias with regard to height may be responsible since we have noted (Chapter 5.1) that controls have higher education than cases (particularly in Italy), and correlations between height and education are described in the literature [301].

The main strength of this analysis is the large number of patients and controls included. There are several limitations – the most obvious being the high missing values for past weights for some study sites (particularly Ireland). Recall bias with regard to the recall of past weights is also a possibility – patients are well aware of the weight loss that occurs in their condition and this may have led them to wonder if weight loss had begun in the years prior to the disease. However, recall bias or selection bias are unlikely to explain the *C9orf72* expansion findings, since those patients are unaware of their genetic status and presumably therefore no more likely than *C9orf72* normal patients to ask such questions.

In conclusion, loss of BMI appears to begin prior to disease onset in ALS, possibly by up to a decade. In *C9orf72* expansion carrying patients, BMI appears to be low for several decades before disease onset and may be a lifelong trait.

5.4. Physical activity and ALS risk

5.4.1. Introduction

Physical activity has been proposed as a risk factor for ALS [302], and numerous studies have reported on the association between physical activity and ALS with conflicting results, in part due to heterogeneous study design [303]. Previous studies have not all agreed on the definition and quantification of physical activity, used varied diagnostic criteria, and have not all distinguished between leisure time physical activity and occupation related physical activity [303].

Previous studies within the populations included in the Euro-MOTOR study have themselves produced conflicting results despite similar methodology. Amongst 636 Dutch cases (and 2,166 controls), leisure time physical activity was associated with an increased risk of ALS (OR 1.08, 95% CI: 1.02 to 1.14) [59]. However, the EURALS study included cases (N = 652) and controls (N = 1,166) from England, France, Italy, Ireland and Serbia found overall physical activity (OR 0.65, 95% CI: 0.48 to 0.89) to be associated with reduced risk of ALS, as were occupational physical activity (OR 0.56, 95% CI: 0.36 to 0.87) and sports (OR 0.49, 95% CI: 0.32 to 0.75) independently [200]. However, there was significant heterogeneity across study sites [200].

The aim, therefore, was to investigate the association between physical activity and ALS, across all sites partaking in the Euro-Motor case control study using methodology validated in the previous European studies [59,200].

5.4.2. Methods

Quantification of Physical Activity

To quantify the cumulative lifetime physical activity level of participants objectively, all reported activities were scored and coded based on the Compendium of Physical Activities [304]. The Compendium provides a coding scheme that links specific activities performed in various settings with their respective metabolic equivalent of task (MET), an expression of the energy expenditure as a ratio of the standard resting metabolic rate (sitting quietly). A MET score of 1.0 is defined as 1 kcal $\times$ kg^{-1} body weight $\times$ h^{-1}. The cumulative scores per participant were calculated as follows:

$$\sum_{k=1}^{n} (\text{MET score}_k \text{ x duration in years}_k \text{ x hours per week}_k)$$

where k represents an activity from the lifetime physical activity history. Because of the magnitude of the cumulative score, it was divided by 1000. MET scores were then aggregated per person truncated to three years prior to survey to exclude physical activity undertaken after the disease had begun.

Statistical Analysis

To assess the associations of METs with ALS risk, multivariate logistic regression models were used. Complete case models were adjusted for age at survey, education, cohort and gender. Next, missing values were imputed using predictive mean matching by study site. Pooled odds ratios were then obtained from imputed data. Imputed models were then rebuilt excluding cases known to carry the *C9orf72* expansion. After exclusion of *C9orf72* cases, imputed models were further adjusted for smoking and alcohol consumption. Finally, to allow for the variation of effects by country/region, we built Bayesian mixed effects multi-level models and results were visualized using forest plots.

5.4.3. Results

The demographic and clinical characteristics of cases and controls have been described in Chapter 5.1. Table 5-12 summarises the physical activity of cases and controls by study site. Of note occupational METs in patients exceeded that in controls with $P < 0.05$ in all study sites except The Netherlands, while only in Ireland did leisure time METs in patients exceed that in controls with $P < 0.05$.

Table 5-12 Occupational and leisure time physical activity for patients and controls

	N	**Occupational METS Mean (SD)**	**Leisure time METS Mean (SD)**
Apulia			
Patients	141	4.31 (4.34)	0.46 (0.97)
Controls	213	2.98 (2.97)	0.38 (0.83)
P value		**0.002**	0.457
Lombardy			
Patients	186	3.48 (2.88)	0.81 (1.42)
Controls	190	2.32 (1.85)	0.76 (1.62)
P value		**<0.001**	0.780
Piedmont and d'Aosta Valley			
Patients	262	4.04 (2.60)	1.01 (1.20)
Controls	290	3.39 (2.37)	0.97 (1.31)
P value		**0.003**	0.752
Ireland			
Patients	177	4.66 (3.96)	1.85 (2.35)
Controls	349	3.90 (2.98)	1.33 (1.50)
P value		**0.026**	**0.016**
Netherlands			
Patients	791	3.17 (2.51)	1.32 (1.91)
Controls	1880	3.15 (2.56)	1.35 (1.45)
P value		0.889	0.676

In a complete-case multivariate regression model both occupational (OR 1.06 95%CI: 1.03 – 1.10) and leisure time METs (OR 1.05 95%CI: 0.99 – 1.11) were associated with increased ALS risk (Table 5-13). These associations were minimally affected by the imputations of missing values, the exclusion of *C9orf72* cases, or by further adjustment for smoking and alcohol status. The relationship for occupational METs had a more significant P value for all models (P <0.001) than leisure time METs which had P marginally below 0.05 after exclusion of *C9orf72* and imputation of missing values. Models fit with restricted cubic splines did not result in improved model fit when compared via likelihood ratio tests.

Table 5-13 ALS risk associated with physical activity quantified via METs

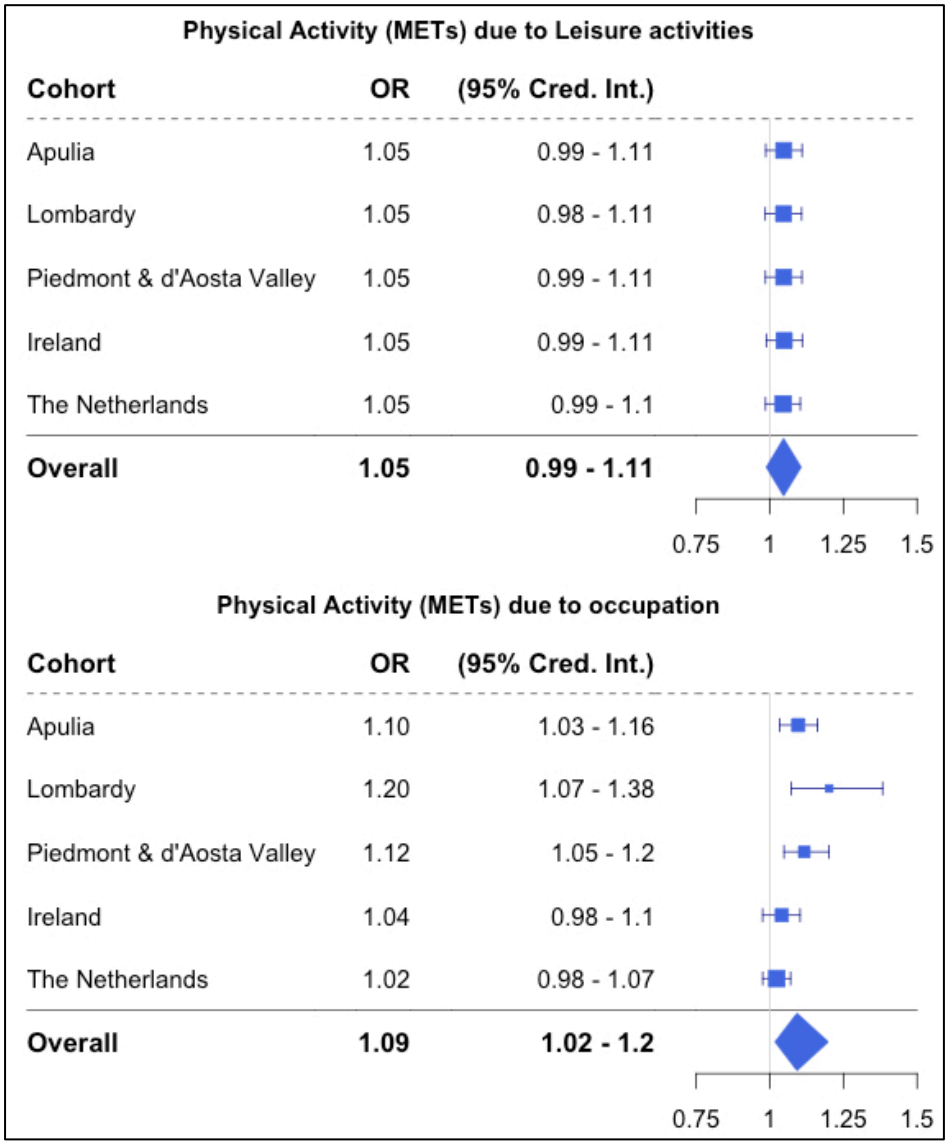
Model	Occupational METs		Leisure time METs	
	Odds ratio (95% CI)	P value	Odds ratio (95% CI)	P value
Complete cases (N = 2,816) †	1.06 (1.03, 1.10)	**<0.001**	1.05 (0.99, 1.11)	0.118
All cases after imputation of missing values (N = 4,479) †	1.06 (1.03, 1.08)	**<0.001**	1.04 (0.99, 1.09)	0.084
Imputed missing values excluding 118 *C9orf72* expanded cases (N = 4,361) †	1.06 (1.03, 1.08)	**<0.001**	1.05 (1.00, 1.10)	**0.047**
Imputed missing values excluding *C9orf72* expanded and adjusted for lifestyle risk factors (N =)††	1.05 (1.02, 1.08)	**<0.001**	1.05 (1.00, 1.10)	**0.049**
Spinal cases vs controls (N = 2,495) †	1.07 (1.03, 1.11)	**<0.001**	1.07 (1.00, 1.14)	**0.038**
Bulbar cases vs controls (N = 2,142) †	1.06 (1.01, 1.11)	**0.013**	0.98 (0.89, 1.07)	0.718

† Adjusted for age, gender, education and cohort. Occupational and leisure time METs included in same model. †† Adjusted for age, gender, education, cohort, smoking and alcohol status. Occupational and leisure time METs included in same model.

Several sensitivity analyses were performed to test the model assumptions. Models including effect modification by gender were compared to the complete case model using likelihood ratio tests and did not exhibit superior fit for occupational (P = 0.654) or leisure (P = 0.547) physical activity. In spinal cases both occupational and leisure time physical activity were associated with ALS risk, while for bulbar cases only occupational METs were significant (Table 5-13).

Mixed effects models allowing for heterogeneity by study site did not alter the odds ratio for leisure time METs (OR 1.05, 95% Cred.Int: 0.99 – 1.11), however the odds ratio for occupational METs was increased (OR 1.09, 95% Cred.Int: 1.02 – 1.20). Odds ratios by study site are shown in Figure 5.7.

Figure 5-7 Odds Ratios for physical activity by Euro-MOTOR study site



5.4.4. Discussion

The analysis of the Euro-MOTOR cohort across five study sites revealed a linear association between physical activity and ALS risk. This relationship appears to be more certain for occupational METs, although estimates were very similar for both occupational and leisure time METs (Table 5-13). The METs expended on occupational physical activity for both patients and controls were considerably higher than those expended on leisure time activity (Table 5-12) for all study sites. Therefore, this greater magnitude of exposure likely explains the clearer relationship between occupational METs and ALS risk. We further found that including spline terms in models did not improve fit (likelihood ratio test), indicating a linear relationship between physical activity and ALS risk.

Heterogeneity by study site was evident for occupational METs, but not for leisure time METs, with greater OR's due to occupational METs for the Italian sites and Lombardy in particular. This pattern is similar to patterns seen for chemical exposures (Chapter 5.5), and therefore may represent different distribution of occupations by region, residual confounding by education (for example due to differences in education systems), or possibly an unrecognised source of bias that differs by site. Nevertheless, all estimates are in the same direction with regard to the null for both occupational and leisure time METs.

Our findings show rough agreement with the previous Dutch study which used the same methodology [59]. That study of 636 cases and 2166 controls found an association between leisure time physical activity (OR 1.08, 95% CI 1.02 to 1.14) but no association with occupational activity. Although our current findings for leisure time physical activity were only significant after multiple adjustments, the odds ratios from the two studies are very similar. From our multi-level model we also found that the Dutch occupations were not associated with ALS risk (Figure 5-7). We note that the exertion of Dutch cases and controls (Table 5-12) is typically lower than that of other countries (except controls from Apulia and Lombardy), thus lower prevalence of physical labour in The Netherlands may explain this. However our results stand in contrast to the EURALS study that

included participants from Ireland, Italy, France, Serbia and the United Kingdom, and which used the same methodology [200] and reported overall physical activity to have a reduced association with ALS risk (OR 0.65, 95% CI 0.48–0.89). These conflicting results reflect contradictory results in the wider literature. Recent systematic review concluded that physical activity is not a risk factor for ALS [58]. Nevertheless some studies are still reporting increased risk of ALS in athletes. A recent cohort study of 212,246 cross-country skiers in Sweden found highly increased risk of ALS amongst the fastest skiers (HR 4.31, 95 % CI: 1.78–10.4) and those who completed the most races (HR 3.13, 95 % CI 1.37–7.17) [305].

Despite these conflicting findings, the Euro-MOTOR study benefits from large size, international multi-centre recruitment and standardised data collection and analysis. Our findings were robust to correction for other lifestyle factor, and removal of *C9orf72* expansion ALS cases. Although we found some evidence for heterogeneity, heterogeneous results were non-conflicting with regard to the null, and may partly be explained by differences in occupation profile by country. We cannot rule-out a possible influence of recall bias in the results, or unmeasured differences in recruitment of controls by study site.

In summary, the Euro-MOTOR study provides new evidence of a relationship between physical activity and ALS risk. Our results indicated a clear relationship between occupational physical activity and ALS risk, and, while the relationship between leisure time physical activity was less clear, odds ratios were in the same direction with regard to the null.

5.5. Occupational Exposures and ALS risk

5.5.1. Introduction

As summarised in Chapter 1.3, numerous exogenous exposures have been proposed as risk factors for ALS. Many of these exposures are maximal in those with occupations that place them at risk of greater exposures. Therefore, occupational exposure studies have been used to assess individual likelihood to exposure to a range of factors, from chemicals such as pesticides or solvents, to electrical fields and electric shock risk. The potential association between ALS risk and occupations have been studied in a number of studies.

5.5.2. Methods

Case ascertainment

The recruitment procedures of the Euro-MOTOR study have been described in detail previously (Chapter 5.1) [129]. In brief, ALS cases incident between 2011 and 2014, and controls matched by age, gender and location, were recruited through population based ALS registers and GP practices respectively in Ireland, The Netherlands and three regions of Italy: Apulia, Piedmont and Valle d'Aosta, and Lombardy. All patients met the diagnostic criteria for definite, probable or possible ALS (according to the revised El-Escorial criteria) [6].

Exposures assessments

Patients and controls completed an extensive survey carried out through face-to-face interviews asking detailed information about medical and lifestyle history – including lifetime occupational history. For each employment until the date of interview, participants were asked to provide job title, daily tasks, start and end years of employment and the number of hours worked per week. Jobs were then reviewed and encoded using the International Standard Classifications of Occupations 1968 & 1988 (ISCO-68 and ISCO-88 hereafter) [306].

Next, exposures were estimated by applying general population job-exposure matrices (JEMs) to the encoded jobs. Three different JEMs were used to quantify exposures in participants: DOM-JEM, ALOHA-JEM and BEN-JEM. The DOM-JEM [307] was applied to jobs via the ISCO-68 codes, the ALOHA-JEM [308] and BEN-JEM [309] utilizes ISCO-88 codes. For the DOM-JEM and ALOHA-JEM each was used to assign ordinal exposure intensity scores to each job corresponding to 'none', 'low' or 'high' exposure (coded as 0, 1 & 4 respectively), whereas for the BEN-JEM exposure was assessed as a quantitative scale for probability and intensity. Here we report on chromium and nickel exposure (DOM-JEM), airborne metals and solvents (ALOHA-JEM) and benzene exposure (BEN-JEM). All exposures in patients and controls were truncated at 3 years before date of survey to ensure all quantified exposures occurred before the onset of disease in the patients.

Statistics

For each exposure due to each job, a cumulative exposure score was calculated incorporating the relevant exposure intensity score, hours per week and years in job via the following equation:

$$Cumulative\ Exposure\ per\ job = \left(\frac{Intensity\ score\ \times\ years\ in\ job\ \times\ weekly\ hours}{40}\right)$$

Cumulative exposure scores were then aggregated for each job per individual to arrive at individualized cumulative exposure scores for each different exposure. Using individualized cumulative exposure scores, linear regression models were built for each exposure and adjusted for age, gender and education status and used to estimate odds ratios (ORs) for the association with ALS of never vs ever exposure for each exposure. As *C9orf72* status was available for 1,257 out of 1,557 patients (77%), a sensitivity analysis was performed omitting those patients with positive *C9orf72* status (N = 118). For those exposures with a significant association with ALS ($P < 0.05$), we modelled the cumulative exposure score as a continuous variable to attempt to identify an exposure-response relationship amongst those with exposure. To test multiple exposures, we counted the number of exposures to which each participant was ever exposed. The

number of exposures was entered as a variable in a regression model along with the usual covariates (age, gender, education and cohort).

Next we adjusted for covariates thought to be important environmental risk factors for ALS – e.g. smoking (never, former, current), alcohol consumption (ever vs never) and leisure time physical activity measured in Metabolic Equivalent of Tasks (METs). Multiple imputations (m = 100) via predictive mean matching was used for the covariate adjustment due to the higher rate of missing values in the METs data. As sensitivity analyses, models were constructed stratified by gender, education and site of onset. Finally, to allow for the variation of effects by country/region, we built Bayesian mixed effects multi-level models and results were visualized using forest plots. We adjusted for leisure time METS only, as the occupational METs data was generated from the same source data as the cumulative exposures and thus was strongly correlated with the exposure estimates (Figure 5-8).

Software

All analyses were performed using the *R 3.2.3* [190] statistical software and using the *tableone* [253], *mice* [280], *openxlsx* [281], *INLA* [189], *ggplot2* [236] and *gridExtra* [257] packages.

5.5.3. Results

Descriptive statistics

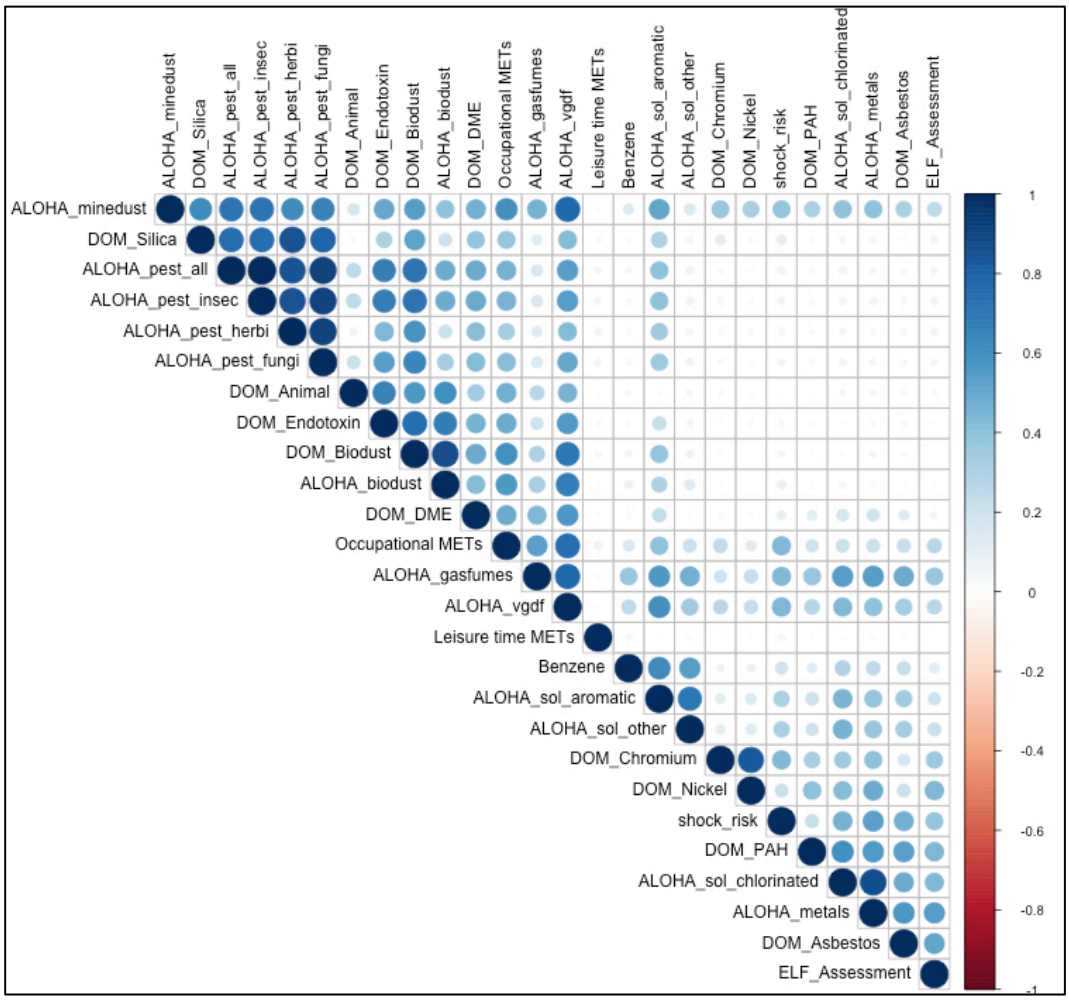
From the 5 regions, a total of 4,479 participants were recruited. Chapter 5.1 described the demographic characteristics of cases and controls by region. There were significant differences in education levels between cases and control in Lombardy, Piedmont and Valle d'Aosta and Ireland, differences in smoking, alcohol and leisure time METs in The Netherlands, and differences in alcohol between cases and controls in Apulia (Table 5-14). Missing values were <1% for key variables with the exception of leisure time METs (26%).

Appendix F displays ever exposed vs never exposed for each of the occupational exposures across each cohort with comparisons of all cases vs controls using Fisher's Exact test. We noted that for Apulia, Lombardy and to a lesser extent, Ireland, several exposure rates were <5% for controls, in some cases numbering in single digits, therefore we did not examine exposure data by tertiles. Figure 5-8 displays a correlogram of Pearson correlations for exposure variables by individual including occupational and leisure time METs. Correlations are almost exclusively positive with the lowest correlation being -0.07. Occupational METs were positively correlated with all exposure variables, whilst leisure time METS were not correlated with any (hence the decision to only adjust for leisure time METs in multivariable models). The exposure correlate in two large clusters – the first consisting of mineral dust, silica, pesticides (all types), biological dust, animal dust, endotoxin, diesel motor exhaust, gas fumes, VGDF (vapours, gasses, dusts and fumes) and occupational METs, the second consisting of benzene, solvents (all types), metals, chromium, nickel, asbestos, ELF-MF and shock risk, and PAH (poly-aromatic hydrocarbons).

Table 5-14 Descriptive statistics of lifestyle variables of Euro-MOTOR participants by study site

		Ever smoked (%)	Ever drank alcohol	Median Leisure time METs (IQR)
Apulia	Patients (n = 141)	100 (47.4)	93 (43.7)	0.0 (0.0, 0.3)
	Controls (n = 213)	106 (55.7)	84 (63.6)	0.0 (0.0, 0.3)
Lombardy	Patients (n = 186)	102 (56.2)	126 (69.2)	0.3 (0.0, 1.0)
	Controls (n = 190)	99 (52.4)	133 (70.0)	0.2 (0.0, 0.9)
Piedmont and d'Aosta valley	Patients	121 (47.6)	165 (64.7)	0.59 (0.19, 1.46)
	Controls	145 (50.2)	197 (68.2)	0.45 (0.12, 1.35)
Ireland	Patients	98 (56.0)	148 (83.6)	1.07 (0.36, 2.18)
	Controls	179 (51.4)	286 (81.9)	0.80 (0.26, 1.84)
The Netherlands	Patients	530 (67.4)	691 (87.4)	0.76 (0.23, 1.71)
	Controls	1,248 (66.4)	1,728 (92.0)	0.96 (0.38, 1.84)

Figure 5-8 Correlogram displaying the Pearson correlations between cumulative occupational exposures in the Euro-MOTOR study



DME = diesel motor exhaust. VGDF = Vapors, Gasses, Dusts and Fumes. PAH = Poly-aromatic Hydrocarbons. Minedust = mineral dust. ELF_Assessment = extremely low frequency magnetic field.

Linear regression models

Linear regression models for JEM exposures with adjustment for cohort, age, gender and education level are shown in Table 5-15 and Table 5-16. Of note, from the ALOHA-JEM all exposures are associated with increased OR for ALS with the exception of chlorinated and other solvents on complete case analysis. Benzene, itself an aromatic solvent, was also associated with increased odds of ALS. From the DOM-JEM, diesel motor exhaust, PAH, silica, biodust and endotoxin were associated with increased OR for ALS on complete case analysis. ELF-MF and shock risk were also associated with ALS. A history of ever being exposed to silica had the highest OR of any exposure (OR 1.55 95% CI: 1.26 – 1.90), while ever exposure to mineral dust, pesticides (all), fungicides, benzene, and biological dust all had very strong associations with increased OR (i.e. P < 0.005).

Table 5-15 ALS risk associated with ALOHA-JEM and BEN-JEM exposures

Exposure	Complete case analysis	Complete cases excluding *C9orf72* expanded cases†	All cases after imputation of missing values and adjusted for lifestyle factors††
ALOHA-JEM			
Biodust	1.16 (1.02 – 1.33)*	1.15 (1.00 – 1.31)*	1.13 (0.98 – 1.29)
Mineral dust	1.29 (1.12 – 1.49)**	1.33 (1.15 – 1.55)**	1.33 (1.14 – 1.54)**
Gas fumes	1.18 (1.02 – 1.36)*	1.18 (1.02 – 1.37)*	1.17 (1.01 – 1.36)*
VGDF	1.19 (1.03 – 1.38)*	1.20 (1.03 – 1.39)*	1.21 (1.04 – 1.40)*
Metals	1.23 (1.03 – 1.48)*	1.28 (1.06 – 1.54)*	1.26 (1.10 – 1.50)**
Pesticides (all)	1.34 (1.09 – 1.63)**	1.36 (1.11 – 1.67)**	1.34 (1.09 – 1.65)**
Herbicides	1.35 (1.06 – 1.71)*	1.35 (1.05 – 1.73)*	1.33 (1.04 – 1.71)*
Insecticides	1.32 (1.07 – 1.62)*	1.33 (1.07 – 1.64)*	1.32 (1.06 – 1.63)*
Fungicides	1.38 (1.11 – 1.70)**	1.41 (1.13 – 1.75)**	1.40 (1.13 – 1.74)**
Aromatic solvents	1.20 (1.03 – 1.40)*	1.22 (1.04 – 1.43)*	1.21 (1.04 – 1.42)*
Chlorinated solvents	1.13 (0.96 – 1.34)	1.15 (0.97 – 1.37)	1.16 (0.98 – 1.38)
Other solvents	1.02 (0.89 – 1.17)	1.02 (0.88 – 1.17)	1.02 (0.88 – 1.18)
BEN-JEM			
Benzene	1.26 (1.08 – 1.47)**	1.29 (1.11 – 1.51)**	1.29 (1.10 – 1.50)**

*** P ≤ 0.05 ; ** P ≤ 0.005. † Adjusted for age, gender, education and cohort †† Adjusted for age, gender, education, cohort, lifetime physical activity due to leisure, smoking and alcohol status and *C9orf72* cases were excluded**

Omission of the 118 *C9orf72* expanded cases increased the odds ratio for 16 out of 25 exposures, and the OR for asbestos and chromium exposure became significant at the P < 0.05 level. After further adjusting of models for alcohol, smoking and total leisure time METs the odds ratio for biological dust, asbestos and ELF-MF metrics were no longer significant at the P < 0.05 level. When modeled as continuous cumulative exposures including those with no history of exposure only aromatic solvents (OR 1.08, CI: 1.01 – 1.15) and benzene (OR 1.08, CI: 1.01 – 1.15) differed from the null. However, when limited to only those with an exposure history these relationships were no longer significant.

Table 5-16 ALS risk associated with DOM-JEM, ELF-JEM and SHOCK-JEM exposures

Exposure	Complete case analysis	Complete cases excluding 118 *C9orf72* expanded cases†	All cases after imputation of missing values and adjusted for lifestyle factors††
DOM-JEM			
Asbestos	1.19 (0.99 – 1.42)	1.22 (1.01 – 1.46)*	1.19 (0.99 – 1.43)
Chromium	1.24 (0.98 – 1.58)	1.31 (1.03 – 1.67)*	1.33 (1.04 – 1.70)*
Diesel Motor Exhaust	1.23 (1.04 – 1.45)*	1.28 (1.08 – 1.51)**	1.24 (1.04 – 1.46)*
Nickel	1.12 (0.84 – 1.48)	1.18 (0.89 – 1.57)	1.21 (0.91 – 1.61)
PAH	1.25 (1.03 – 1.52)*	1.28 (1.05 – 1.56)*	1.25 (1.02 – 1.52)*
Silica	1.55 (1.26 – 1.90)**	1.58 (1.28 – 1.94)**	1.55 (1.26 – 1.92)**
Animal	1.19 (0.93 – 1.51)	1.26 (0.99 – 1.61)	1.22 (0.96 – 1.56)
Biodust	1.22 (1.07 – 1.41)**	1.22 (1.06 – 1.40)*	1.19 (1.04 – 1.37)*
Endotoxin	1.24 (1.06 – 1.45)*	1.24 (1.06 – 1.45)*	1.21 (1.03 – 1.42)*
ELF-JEM			
ELF-MF	1.14 (1.00 – 1.31)*	1.15 (1.00 – 1.32)*	1.12 (0.98 – 1.29)
SHOCK-JEM			
Shock risk	1.24 (1.06 – 1.45)*	1.23 (1.05 – 1.45)*	1.17 (1.00 – 1.37)*

*** P ≤ 0.05 ; ** P ≤ 0.005**

† Adjusted for age, gender, education and cohort

†† Adjusted for age, gender, education, cohort, lifetime physical activity due to leisure, smoking and alcohol status and *C9orf72* cases were excluded

Multiple exposures

A linear regression model indicated an increased odds ratio of 1.03 (1.02 – 1.05) per exposure (considered as a binary ever vs never exposure) after adjustment for age, gender education, and cohort. Fitting this model with a restricted cubic splines did not improve the fit (P = 0.523 via likelihood ratio test).

Exposure as continuous variables

When exposures were modelled as continuous cumulative exposures and including only those with a history of exposure, only mineral dust, gas-fumes and VGDF from the ALOHA-JEM produced estimates excluding the null value Table 5-17. Inclusion of restricted cubic splines to allow for non-linear fit did not improve these models on likelihood ratio testing.

Table 5-17 Odds ratios risk of ALS for cumulative exposures as continuous values

Exposure	Odds Ratio (95% CI)		Exposure	Odds Ratio (95% CI)
ALOHA-JEM			DOM-JEM	
Biodust	1.07 (0.97 – 1.19)		Asbestos	0.96 (0.80 – 1.13)
Mineral dust	1.15 (1.03 – 1.29)*		Chromium	0.99 (0.77 – 1.26)
Gas fumes	1.11 (1.01 – 1.23)*		Diesel Motor Exhaust	1.03 (0.89 – 1.20)
VGDF	1.18 (1.07 – 1.30)**		Nickel	0.96 (0.70 – 1.28)
Metals	0.89 (0.75 – 1.06)		PAH	1.01 (0.83 – 1.26)
			Silica	1.11 (0.89 – 1.38)
Pesticides (all)	1.17 (0.94 – 1.46)		Animal	0.90 (0.70 – 1.16)
Herbicides	1.09 (0.81 – 1.46)		Biodust	1.07 (0.97 – 1.19)
Insecticides	1.23 (0.97 – 1.55)		Endotoxin	1.05 (0.90 – 1.21)
Fungicides	1.08 (0.85 – 1.38)			
			ELF-JEM	
Aromatic solvents	1.05 (0.92 – 1.19)		ELF-MF	0.96 (0.86 – 1.06)
Chlorinated solvents	0.98 (0.84 – 1.14)			
Other solvents	1.08 (0.96 – 1.21)		SHOCK-JEM	
			Shock risk	1.09 (0.96 – 1.24)
BEN-JEM				
Benzene	1.09 (0.96 – 1.24)			

*** P ≤ 0.05 ; ** P ≤ 0.005**
Adjusted for age, gender, education and cohort

Bayesian mixed effects models

Bayesian mixed effects models (excluding the 118 *C9orf72* cases) revealed considerable heterogeneity of odds ratios across study sites for a large number of exposures. Overall Bayesian estimated odds ratios for all exposures are shown in Table 5-18, with forest plots by study site included in Appendix G. By study site, strong effects were seen in the Italian sites, particularly Lombardy, for the all exposures with the exception of the three types of solvents and benzene (Appendix G). Consequently, mixed effects estimates for all estimates except solvents (including benzene) were greater (Table 5-18) than linear regression estimates (Table 5-15 and Table 5-16). The greatest estimated OR was again for silica exposure (OR 1.90 95% Cred. Int. 1.29 – 2.81), with fungicides next (OR 1.70 95% Cred.Int. 1.16 – 2.51) closely followed by other forms of pesticides, and then other estimates. Only ALOHA-JEM biological dust, VGDF, chlorinated and other solvents, and nickel did not exclude the null effect.

Table 5-18 Overall ALS risk associated with JEM exposures estimated via mixed effects Bayesian models

Exposure	Odds Ratio (95% Credible interval)		Exposure	Odds Ratio (95% Credible interval)
ALOHA-JEM			DOM-JEM	
Biodust	1.25 (0.97 – 1.65)		Asbestos	1.46 (1.03 – 2.06)
Mineral dust	1.54 (1.16 – 2.10)		Chromium	1.50 (1.01 – 2.21)
Gas fumes	1.29 (0.99 – 1.68)		Diesel Motor Exhaust	1.51 (1.10 – 2.09)
VGDF	1.28 (0.99 – 1.66)		Nickel	1.37 (0.90 – 2.09)
Metals	1.53 (1.10 – 2.16)		PAH	1.55 (1.08 – 2.28)
			Silica	1.90 (1.29 – 2.81)
Pesticides (all)	1.63 (1.12 – 2.37)		Animal	1.51 (1.03 – 2.25)
Herbicides	1.63 (1.10 – 2.42)		Biodust	1.34 (1.01 – 1.79)
Insecticides	1.59 (1.11 – 2.32)		Endotoxin	1.46 (1.05 – 2.06)
Fungicides	1.70 (1.16 – 2.51)			
Aromatic solvents	1.19 (1.01 – 1.39)		ELF	
Chlorinated solvents	1.12 (0.95 – 1.33)		Assessment	1.38 (1.03 – 1.87)
Other solvents	0.99 (0.85 – 1.15)			
			SHOCK-JEM	
BEN-JEM			Shock risk	1.46 (1.10 – 1.96)
Benzene	1.27 (1.08 – 1.49)			

All estimates adjusted for age, gender, and education as fixed effects with mixed effects for exposure by study site. Country specific estimates are shown in Appendix G.

5.5.4. Discussion

In this large international case-control study using validated JEM's we found associations between a wide range of exposures and increased ALS risk. None of the exposures tested were associated with a decreased risk of ALS. These findings were robust to adjustment for additional lifestyle factors and sensitivity analysis. We found evidence for heterogeneity across study sites all exposures excluding solvents, and in general estimates were increased by allowing for heterogeneity via mixed effects models.

Silica

The strongest association between exposure and ALS odds ratio was for silica. Silica exposure occurred primarily in jobs in the construction, farming and horticulture industries. Silica exposure is typically associated with lung disorders including lung cancer, silicosis and chronic obstructive pulmonary disorder. To the best of our knowledge silica has not been previously associated with ALS. However, the correlogram (Figure 5-8) shows a very strong overlap between silica and pesticides exposure, which have been previously suggested as risk factors for ALS [64–67]. Given this lack of prior evidence it seems likely that the association with silica is due to co-linearity between silica, pesticides and other exposures (e.g. occupational METs, mineral dust and others).

Pesticides

The next strongest association we found was for pesticides, with the strongest association within the pesticides group found for fungicides. However, given the very strong correlation between pesticide exposures it is unsurprising that each group of pesticides led to similar estimates, and therefore it is not possible to single out one group of pesticides as being the main factor driving the associations. These findings are in keeping with results from previous studies [64–67]. However, this contradicts a recent case-cohort study from The Netherlands (N, cohort = 4,344, N, cases = 136) which found reduced OR (0.6, 95% CI: 0.27 to 1.35) for ALS for individuals with ever occupational exposure to pesticides [310]. That study had very low numbers of cases with exposure however (n = 7), and therefore our study has greater statistical validity.

Diesel Motor Exhaust

We found increased risk of ALS in association with exposure to diesel motor exhaust. Previous studies assessing exposure to DME in ALS are lacking, however our finding is in keeping with the suggestion that exposure to DME may be the common element linking increased ALS risk to a number of occupations including truck and bus drivers, farmers, military personnel and others who operate heavy machinery [72]. Furthermore, DME exposure has been linked to other neurodegenerative conditions.

Solvents (including Benzene)

Exposures due to solvents were assessed by both the ALOHA-JEM, and for benzene specifically via the BEN-JEM. Increased risk of ALS was demonstrated using both JEMs. Our findings are in keeping with a meta-analysis that found increased odds of ALS in those exposed to (unclassified) solvents (OR = 1.43 95%CI: 1.10 - 1.86) [62]. Furthermore, a recent case-control study of 51 cases vs 51 matched controls in Pennsylvania found that residential exposure to aromatic solvents was associated with increased risk of ALS in 1999 (OR = 4.27, 95% CI: 1.09 - 16.79) and 2002 (OR = 5.03, 95% CI: 1.29 - 19.53).

Metals

Our finding that the ALOHA-JEM airborne metals exposure score was associated with higher odds of ALS is also in keeping with results from meta-analyses for association between lead exposure and ALS (OR = 1.72, 95% CI: 1.33 - 2.23) [62,69], and general exposure to heavy metals (OR = 1.71, 95% CI: 1.38 - 2.11) [62,69]. In addition, two recent smaller studies have shown associations between levels of metals in cerebro-spinal fluid (CSF) and ALS risk [212,311].

ELF/EMF and Electric Shock

We found both occupational exposure to ELF-MF and electric shock risk to be associated with increased risk of ALS, although unlike other exposures the magnitude of these associations were reduced after adjustment for other lifestyle factors (Table 5-16). A 2013 meta-analysis did not support a role for ELF-MF as risk factors for ALS [312], two more recent studies using the same ELF-MF and shock JEMs as our study did support such a role. The first, a cohort study of 2.2 million Swiss workers, found increased ALS mortality for those with medium or high exposure to ELF-MF (HR 1.55, 95%: 1.11 to 2.15), with results for shock (HR 1.17 , 95% CI: 0.83 to 1.65) reduced to almost unity after adjusting for ELF-MF [313]. The second, a cohort study of 120,852 Dutch workers, found increased ALS mortality for those ever exposed to high risk of shock (HR: 2.19, 95% CI: 1.02 to 4.73), with results for high risk of exposure to shock close to unity (HR: 1.04, 95% CI: 0.57 to 1.92) [310]. Conversely, a Dutch case-control study of exposure to ELF-MF from residential proximity to power lines failed to find a significant association with ALS [314]. Therefore on balance the evidence for ELF-MF and shock risk remains mixed considering our results in the context of the literature.

Poly-aromatic Hydrocarbons

We found that occupational exposure to PAHs was also associated with increased risk of ALS. Poly-aromatic hydrocarbons (PAHs) are hydrocarbon compounds containing multiple aromatic rings. PAHs are formed by incomplete combustion, and have been associated with diesel exhaust, the burning of fossil fuels such as coal, cooking and cigarette smoking. In our data the top three occupations associated with PAH exposure were: waiters/waitresses, vehicle mechanics and fitters and cooks. Previous studies of occupational PAH exposure and ALS are lacking, however toxicological studies have demonstrated neurotoxic effects at high concentrations [315].

Biological dust, Animal dust and Endotoxin

Biological dust exposure was assessed by both JEMS. It was found to be associated with ALS risk in the DOM-JEM but not the ALOHA-JEM. However, after allowing for

heterogeneity via mixed models it was associated with ALS risk for both JEMs. Biological dust exposures are generally mixtures of substances of microbial, plant or animal origin such as bacteria, fungi, allergens, endotoxins, peptidoglycans, B-glucans, pollens, and plant fibres [308,316]. We also found that endotoxin exposure was associated with ALS risk as assessed by the DOM-JEM, but animal dust was not (Table 5-16). To the best of our knowledge biological dust exposure has not been previously associated with ALS risk.

Mineral dust, Gas fumes, VGDF

Associations of small increases in ALS risk were found for occupations with exposure to mineral dust, gas fumes and VGDF (vapour, gases, dust and fumes). Mineral dust, gas fumes and VGDF are a set of overlapping exposures typically experienced by people in a wide range of occupations including: mining, armed forces, welding, painting, carpentry, domestic cleaning and a wide range of industrial jobs. As such there is a wide range of chemical substances captured by these metrics. As with biological dusts, there is a lack of previous studies examining these exposures in ALS.

Multiple exposures

We found that patients were more likely to have ever been exposed to multiple chemicals than controls, with a linear relationship between the number of exposures and OR for ALS. This may be partly explained by the collinearity between exposures (Figure 5-8). However, different exposures can also have occurred within individuals due to different jobs. Again, there is a lack of previous studies regarding combined occupational exposures and ALS risk.

Strengths and limitations

Euro-MOTOR recruited large numbers of cases and controls from each participating study site and applied well established JEM's to determine individual occupational exposure scores for categories and specific metals and solvents. In addition, missing values for occupational data were very low for all study sites.

A limitation of the study is that we did not specifically assess exposure to some exposures that are of interest in ALS, for example lead and mercury exposure, as these were not assessed by the JEMs used. Therefore, we cannot out-rule that the associations we found with general metal exposure may represent an effect of lead exposure for example. Furthermore, due to the positive correlations between the JEM exposures, it is not possible to specifically identify any one of the exposures as driving the associations. In addition, the JEM only assesses exposures through occupation, therefore exposure from other sources could not be assessed. Nevertheless, relationship between occupational exposures and ALS risk are not expected to deviate between occupational and environmental exposures.

Finally, as our study has a case-control design, we cannot rule out the possibility of bias in the recruitment of controls. Although the study benefits from a matched case control design to minimise bias, we observed controls to have higher education status in Italy compared to cases. Since education is such an important factor in occupation, this might in part explain the heterogeneity observed across sites, and higher odds ratios found in Northern Italian sites.

Conclusions

In summary, in this large multi-centre case-control study we found that numerous occupational exposures were associated with increased risk of ALS and were robust to adjustment for multiple confounders. The strongest associations were found for occupational exposure to silica and pesticides, although increased risk was also found for exposure to solvents, metals, diesel mineral exhaust, poly-aromatic hydrocarbons, ELF-MF and electric shock, dusts gases and vapours. Removal of cases carrying the *C9orf72* expansion increased these associations for most, though not all exposures. There was significant heterogeneity by study site with the strongest associations found in Northern Italy.

Chapter 6: Summary of Results

Chapter Outline:

6. Synthesis of Results

6.1. Summary of Results

6.1.1. Spatial Epidemiology

Incident spatial risk of ALS in Ireland

The initial exploratory analysis of 1,645 incident prospective cases aggregated over 17.5 years (Chapter 3.1) provided an update on the only previous attempt to analysis the spatial incidence of ALS in Ireland [27]. This much earlier analysis of 231 incident prospective cases from 1995 to 1997 had implicated a possible increased risk for ALS in the North-West county of Donegal. Our exploratory analysis, which used much higher spatial resolution, indicated that this was not correct. Whilst the earlier analysis had been carried out at the coarse level of counties (N = 26), and thus was subject to the modifiable unit area problem[131], our use of 3,355 Electoral Divisions (EDs), and greater cohort size greatly reduces such concerns. The ED level analysis revealed no overall geographic pattern to Bayesian smoothed ALS risk in Ireland, with only localised areas of moderately increased risk apparent in some areas (Figure 3-5). These findings stand in contrast to those from other countries including Sweden[182] and Spain[153], where geographic North-South patterns of risk have been described, and in the United States where a North-Western to South-Eastern distribution has been described[154]. However, our use of an incident population based cohort and Bayesian methods represent considerable methodological advantages over those studies which did not use population based cohorts.

We built upon this exploratory analysis with a further analysis of 1,701 cases at the Small Area (SA) Geometry level (N = 18,222 areas) (Chapter 3.3). The SAs have the advantage of containing by design between 50 and 200 households when compared to larger historically defined ED's which have wide ranging populations. Thus, the SAs have better statistical properties for geospatial analysis. The SA level analysis confirmed a lack of large-scale geometric patterns of high risk areas. However, at the SA geometry, large

areas of low risk for ALS were more readily apparent. There were two such areas, one including Clare & Galway and the other including Carlow and Kilkenny and some surrounding areas. Furthermore, our SA level analysis confirmed that population density was not associated with ALS risk (this had been unclear at ED geometry). Local area social deprivation, quantified via the PobalHP deprivation index, did also not associate with the spatial distribution of ALS risk.

Finally, we investigated the association between 45 soil parameters and ALS spatial risk again at SA geometry (Chapter 3.4). In addition to examining the association of 45 different parameters with ALS risk, we also tested hypotheses for specific combinations of parameters – for example soil lead and soil mercury. Of 315 models constructed only one had a soil parameter coefficient that did not overlap zero - total magnesium percentage of mineral soil per small area. However this was likely a chance finding, and the overall conclusion is that soil mineral parameters do not associate with local ALS risk.

Spatial Cluster Analysis

In contrast to many previous studies from other countries, spatial clusters, or "hotspots", of ALS were not found in Ireland (both at ED level geometry(Chapter3.2), and at SA geometry (Chapter 3.3). This is likely due to our application of robust methodology to a complete population based data-set covering the country, thus avoiding many of the pitfalls of spatial cluster analysis, such as the 'post-hoc' investigation of purported clusters [131], or the use of spatially biased data-sets such as clinic based cohorts. On the contrary, we found two statistically significant low risk for ALS areas – risk—one centered on County Kilkenny (RR =0.390; P=0.048 at SA level geometry) and a smaller area in County Clare (RR = 0.082; P = 0.028 at SA level geometry). We investigated whether a failure of case-ascertainment might have explained these regions using a novel spatio-temporal approach and found that this did not explain the observed patterns. Conversely, this method was sensitive enough to identify lower ascertainment in County Donegal which was explained by an additional 5

later presenting cases and 2 cases from Donegal who presented to Northern Ireland (Chapter 3.2.)

Nevertheless, there are some limitations to our work. For example the analysis used only patient's address at the time of diagnosis, which likely does not reflect environmental exposures accumulated over time throughout the life-course. Because the incidence rate of ALS is relatively low and Ireland is relatively sparsely populated, case numbers may not yet be significant to identify clusters of ALS cases relating to local environmental causes. Nevertheless, the finding of low incidence areas is noteworthy, although likely requires genetic profiling of all patients and geographic data on environmental variables such as pesticides to be fully explained. Periodic spatial incidence mapping is recommended every 4 or 5 years going forward in order to a) detect any emerging hotspots and b) identify any localised failures of case ascertainment.

6.1.2. Survival and Prognostic Factors

Multidisciplinary Clinic Attendance

Evidence from previous observational studies indicates that those who attend a multidisciplinary clinic experience improved survival [11,82–85]. Differences between the two systems of care in Northern Ireland (community care) versus the Republic of Ireland (multidisciplinary clinic) for patients with ALS offered the opportunity to carry out a population-based study of survival between the two systems (Chapter 4.1). Multivariate analysis of 511 patients from the Republic of Ireland and 208 from the North demonstrated that Republic of Ireland patients experienced a significant survival advantage (HR 0.72, 95% CI 0.59 to 0.87, $p<0.001$). In addition, there was a survival benefit for patients who attended the multidisciplinary ALS clinic (MDT) compared with RoI and NI patients who did not attend the clinic (HR 0.59, 95% CI 0.49 to 0.71, $p<0.001$). Our data supports previous evidence that attendance at a specialist multidisciplinary ALS clinic is beneficial to patients, and should be considered a standard of care in the management of ALS.

Spatial variables and ALS prognosis

Variations in environmental risk factors have potential to influence the incidence and progression in complex multifactorial diseases such as ALS. However, few studies have investigated this topic. In Chapter 4.2, the potential for prognostic association of four spatially structured variables was investigated for 1,232 Irish ALS patients. The four spatial variables were: 1) small area population density, 2) small area social deprivation, 3) distance to the coast (for which there was prior evidence [102]), and 4) distance to the ALS multi-disciplinary clinic. Multivariate Royston-Parmar regression models were used to adjust for important prognostic factors including age at onset, diagnostic delay, site of onset, El-Escorial category, attendance at the MDT and riluzole prescription. After addition of the individual geospatial variables in turn, none of the four variables was found to be associated with survival with a p-value > 0.05.

Prognostic impact of the C9orf72 hexa-nucleotide expansion

A number of previous studies have indicated that the *C9orf72* repeat expansion is a negative prognostic indicator in ALS [88–94], however such studies were not large enough to permit analysis stratified by gender and site of onset. In Chapter 4.3, the association of the *C9orf72* expansion with survival in gender and site of onset subgroups was examined. Amongst 4,925 patients from Belgium, Ireland, Italy, The Netherlands and The United Kingdom, 457 (8.95%) carried the *C9orf72* expansion, which was associated with a worse prognosis (HR 1.36, 95%CI: 1.18 to 1.57). Models evaluating interaction between gender and *C9orf72* repeat expansions demonstrated that the reduced survival due to *C9orf72* expansion was being driven by spinal onset males with the expansion (HR 1.56, 95% CI 1.25 to 1.96). The median predicted survival in all spinal patients excluding spinal onset *C9orf72* expanded males was 2.77 years, compared to 2.38 years in bulbar onset patients, and 2.29 years in spinal onset males who carried the *C9orf72* repeat expansion. In addition, spinal onset males carrying the expansion were younger at onset, and had shorter diagnostic delay compared to other patients.

ALSFRS-R sub-score Analysis

The revised ALS Functional Rating Scale (ALSFRS-R) is normally analysed as a single total score. However, there is evidence that retains multidimensionality corresponding to bulbar, motor and respiratory domains [109,252]. In Chapter 4.4 the longitudinal and prognostic characteristics of the bulbar, motor and respiratory ALSFRS-R sub-scores from 407 Irish ALS patients were characterised. Survival modelling revealed that the slope of ALSFRS-R sub-scores, provided better fit in prognostic models than did the slope of the total ALSFRS-R alone. Subsequent longitudinal analysis revealed that the ALSFRS-R motor sub-score deteriorated earlier in spinal-onset disease over bulbar-onset disease, while in bulbar-onset disease the ALSFRS-R bulbar sub-score deteriorated earlier and faster than in spinal-onset disease. There was also evidence that the decline in ALSFRS-R motor sub-scores in patients with spinal-onset disease, and decline in ALSFRS-R bulbar sub-scores in patients with bulbar-onset disease, may predate reported disease onset dates. Finally it was found that the slope of ALSFRS-R respiratory sub-scores were not prognostically informative after adjusting for the bulbar and motor sub-scores.

6.1.3. Exposures

The EuroMOTOR study is a pooled case-control study which recruited ALS patients from incident population based ALS cohorts in Ireland, The Netherlands and three regions of Italy (Apulia, Lombardy, and Piedmont and D'Aosta Valley) from February 2011 to January 2014. Controls matched by age, gender and area of residency were recruited over the same period of time. All participants were asked to complete a detailed structured survey that included questions on a wide range of potential exposures including but not limited to: targeted medical history and medications; drugs of abuse, smoking and alcohol; lifelong exercise habits and hobbies; full occupational history, residential history; family history for neuropsychiatric disorders. The complete survey is provided in Appendix D. The cohort included 1,557 ALS patients and 2,922 controls across the five study sites.

Hormonal exposures and ALS risk in European women

Amongst women in the Euro-MOTOR study (653 patients and 1,217 controls) oral contraceptive use was higher among controls. For ever use of OCP, OR = 0.65, 95% CI 0.51 to 0.84), and a linear dose-response relationship was apparent (OR = 0.98, 95% CI 0.96 to 0.99 per year of use). The results for OCP's were robust to multiple confounders, but there was some heterogeneity across study sites. This may be due to differences in availability and usage rates of the OCP across countries. Hormone replacement therapy (HRT) was associated with a reduced risk of ALS only in the Netherlands (OR = 0.57, 95% CI 0.37 to 0.85). In contrast to previous studies, lifetime unopposed estrogen exposure was associated with an increased risk of ALS (OR = 1.02, 95% CI 1.00 to 1.04). However, this finding appears to be driven by the subtraction of duration of OCP use during calculation of lifetime unopposed estrogen exposure.

Premorbid BMI

Longitudinal models of premorbid BMI revealed than weight loss may begin up to a decade earlier than survey date in patients compared to controls – i.e. prior to the onset of ALS symptoms. BMI differences between cases and controls were not gender specific. Stratification by *C9orf72* expansion status (Ireland, Piedmont and D'Aosta Valley and The Netherlands only), revealed lower BMI amongst *C9orf72* expansion carrying patients for decades prior to disease onset – perhaps representing a life-long trait. Selection or recall bias are unlikely to explain this finding.

Physical Activity

The analysis of the Euro-MOTOR cohort across five study sites revealed a linear association between physical activity and ALS risk for both occupational METs (OR 1.06 95%CI: 1.03 – 1.10) and leisure time METs (OR 1.05 95%CI: 0.99 – 1.11). These associations were minimally affected by the imputation of missing values, the exclusion of *C9orf72* cases, or by further adjustment for smoking and alcohol status. There was heterogeneity across study sites evident for occupational METs, but not for leisure time

METs. Our findings show rough agreement with a previous Dutch study which used the same methodology[59].

Occupational Exposures

There were positive associations between ALS risk and a history of ever being exposed to a range of hazardous occupational exposures, including: silica, pesticides, diesel motor exhaust, metals, poly-aromatic hydrocarbons, biological dusts, mineral dust, VGDF (vapors, gasses, dusts and fumes), ELF-MF and electric shock risk. In addition, there was significant positive correlation between exposure variables, forming two large clusters: the first consisting of mineral dust, silica, pesticides (all types), biological dust, animal dust, endotoxin, diesel motor exhaust, gas fumes, VGDF and occupational METs; the second consisting of benzene, solvents (all types), metals, chromium, nickel, asbestos, ELF-MF and shock risk, and PAH (polyaromatic hydrocarbons). The strongest exposure-ALS risk associations were seen for exposure to silica (OR = 1.55, 95% CI: 1.26 to 1.90) and fungicides (OR = 1.38, 95%CI 1.11 to 1.70), with more modest associations found for other exposures. For 16 out of 25 exposures assessed, the associations with ALS risk were strengthened by removal of *C9orf72* expanded cases. After adjustment for drinking alcohol, smoking and lifestyle physical activity, the associations of most exposures with ALS risk were unaffected. However, associations for ELF-MF and electric shock risk were reduced. Ever exposure to multiple chemicals was associated with further increased risk for ALS (OR = 1.03, 95% CI: 1.02 to 1.05 per exposure). Euro-MOTOR adds to the body of evidence indicating that exposures to a range of hazardous chemicals is associated with increased risk for ALS. However, due to the strong correlations between exposures it is not possible to determine which specific chemicals are responsible. We note that exposure to lead was not assessed by the available JEMs.

6.2. General Discussion

Spatial Epidemiology

The spatial epidemiology findings for ALS in Ireland differ in several ways from those of other countries. The primary difference is the lack of statistically significant high risk clusters of ALS cases in Ireland. This is likely explained by two factors: 1) the use of a true population based incident cohort and 2) the judicious use of appropriate statistical techniques. Many studies from other countries that have reported statistically significant high risk clusters have not used prospectively collected incident cases (Table 3-2), or have recruited only from restricted areas with the aim to confirm supposed clusters[24,25]. Such approaches to cluster analysis are ill-advised and prone to Type 1 error[131]. Another distinguishing characteristic of the spatial epidemiology of ALS in Ireland is the presence of statistically significant low risk areas. Low risk areas are generally not given much focus in the literature. However, they are of note in Ireland because they were found in absence of high risk clusters, over relatively large areas and with statistical significance using the SaTScan program. None of the variables examined using Bayesian risk maps, including social deprivation, population density and 45 different soil parameters explained this pattern; thus these areas remain unexplained. Population genetics may explain the pattern, however it is inadvisable to include genetics as a covariate in cohort only maps as there is likely to be spatial bias with regard to which patients provide genetics samples. The low risk area in Clare/Galway centred on the Burren is of note since this area does not have industrial agriculture and pesticides use is very low. This is particularly interesting given the high risks associated with occupational exposures to pesticides in the EuroMOTOR study. However, spatial data on the use of pesticides at the necessary resolution to construct spatial models incorporating maps of pesticide use remain unavailable. Therefore we cannot confirm this hypothesis presently.

While the results of models including soil parameters as spatial covariates did not produce significant findings, other datasets might be of more interest. Given the results of the EuroMOTOR occupational analysis have shown a number of airborne exposures to be associated with increased ALS risk (Chapter 5.5), future spatial studies should

investigate associations between air pollution concentrations over time as risk factors for ALS. Occupational exposure to pesticides was also found to be important, consequently spatial analysis including land use and pesticide use should be performed in future. The Pesticides Registration and Control Division of the Department of Agriculture (http://www.pcs.agriculture.gov.ie) maintains extensive and precise records on the use of pesticides in Ireland including details on the amounts of individual chemicals used in accordance with European regulations that would be ideally suited to this purpose. However, I was unable to obtain access to this dataset during this thesis. Beyond this, developments in statistical methodology show significant promise in improving the sensitivity of spatial disease mapping. Log Gaussian Cox processes (LGCP) have recently been demonstrated to outperform the BYM model (Chapter 3) in the identification of clusters in synthesised data [317], and a recent ALS spatial study from Denmark employed novel general additive models (GAM) to associate ALS risk with location at birth as well as diagnosis [318]. It would be desirable to model ALS risk versus address at birth and various ages in the Irish ALS cohort, however at the time of writing this data is not available for ALS register cohort.

The advent of multi-step theories of ALS aetiology present further challenges to the use of spatial epidemiology to identify risk factors. If, as evidence suggests, six steps are involved in the development of ALS in an individual [9], then there is potential for each one of those 6 steps to vary spatially across a country. This is true even for genetic factors which may vary by area. Therefore, attempting to explain geographical patterns of ALS risk in terms of individual risk factors may be a futile endeavour until one or more of the six steps are identified and can be included in geospatial models.

Survival and Progression

This thesis presented a number of findings that help improve understanding of survival in ALS. Randomised clinical trials of multi-disciplinary clinics in ALS are not feasible for a variety of reasons. Therefore, the findings of Chapter 4.1 provide important evidence that attendance of the MDT improves the survival of ALS patients. We have also shown that survival does not relate to population density or distance from the coast (Chapter

4.2), which had previously been reported in the UK [102]. These findings are not mutually exclusive however, as population density and consequential factors such as air pollution reach higher levels in the UK compared to Ireland. Studies from other countries with high population densities are needed to clarify this further.

Chapter 4.4 confirmed previous studies by others that the ALSFRS-R is a mutli-dimensional metric, and greater clinico-pathological insight may be gained by analysing if as bulbar, motor and respiratory sub-scores. Indeed, by applying such analysis longitudinally, we identified that the respiratory aspect of the score is not very sensitive to patient deterioration. Further work is being carried out within the Trinity College Dublin ALS group to develop complementary metrics to augment assessment of patient respiratory function.

The findings of Chapter 4.4 also carry implications for clinical trials. Contemporary clinical trials in ALS typically utilise strict enrolment criteria, including that patients should be either ”normal progressors” or “fast progressors”. Patients are usually defined as normal or fast progressors according to the slope of their ALSFRS-R. For example the recent masitinib trial patients included were “normal progressors”, defined as those with a slope < 1.1 ALSFRS-R points per month [319]. Similarly, the recent edaravone trial included a 12 weak lead in period from which only patients with an ALSFRS-R slope between 0.33 and 1.5 points per month progressed to randomisation[86]. In this context the results of Chapter 4.4 raise a number of concerns about such criteria for trials. First, the ALSFRS-R is in fact multidimensional in nature corresponding with bulbar, motor and respiratory subgroups - both with respect to longitudinal trend (Figure 4-10) and prognostic characteristics (Appendix C). Second, the ALSFRS-R sub-score longitudinal trends are markedly different with respect to site of onset categories – themselves important prognostic indicators. Third, the ALSFRS-R respiratory sub-score appears to perform badly in quantifying respiratory dysfunction – likely a highly important prognostic factor. Finally, subgroups defined by ALSFRS-R slope do not correspond to any known biological process or known causal mechanism, rendering such definitions to be largely arbitrary cut-offs. For this reason the use of total ALSFRS-R slope as cut-off criteria for clinical trials is not guaranteed to select true “normal progressors” or “fast

progressors". Furthermore, these considerations limit the generalisability of trial results that makes it unclear which presenting patients in population or clinical ALS cohorts will benefit from treatments trialled under total ALSFRS-R slope defined entry criteria.

Refined picture of C9orf72 phenotype

Several clinical features of the *C9orf72* phenotype are already well described; in particular the *C9orf72* expansion is recognised as a cause of both ALS and FTD and has helped established the two conditions as a continuum[3,4]. The negative prognostic impact of the *C9orf72* expansion has been described in numerous previous studies [88–94]. As a result *C9orf72* ALS may be considered as a distinct subtype of ALS, and novel therapies are being developed that specifically target the G_4C_2 RNA associated with the *C9orf72* expansion [320].

Several chapters of this thesis have provided new evidence to further define the distinct phenotype of the *C9orf72* expansion. Specifically, the following points were elucidated:

1. The negative prognosis of the *C9orf72* expansion is driven by males with spinal onset disease (Chapter 4.3).
2. Male spinal onset patients with the C9orf72 expansion are younger at onset and have shorter diagnostic delay than other groups (Chapter 4.3).
3. Males and females with the *C9orf72* expansion have lower BMI than controls and patients without the expansion for several decades prior to disease onset (Chapter 5.3).
4. Most occupational exposures appear to be more strongly associated with ALS risk when patients with the *C9orf72* expansion are excluded. This implies that such exposures are less important in those carrying the expansion (Chapter 5.5).

These new findings thus provide further evidence for the view of *C9orf72* ALS as a distinct sub-type of ALS. Furthermore, they hint at new biology, with the poorer prognosis, younger onset and shorter diagnostic delay of male spinal onset patients carrying the expansion pointing towards gender specific effects, and the finding of perhaps lifelong lower BMI in patients with the C9orf72 expansion pointing towards

differences in metabolism associated with the expansion. Biological research will be needed to fully explain these findings. Stratification of observational studies and trials by *C9orf72* status is recommended.

Physical activity, hormones and ALS risk

As noted in chapter 5.3, there is conflicting evidence regarding physical activity as a risk factor for ALS – both in the literature generally, and within case-control studies using identical methods in Europe (i.e. EURALS and Euro-MOTOR). This pattern of conflicting results has led to the recent hypothesis that rather than physical activity being a risk factor for ALS directly, instead cardiovascular fitness profile and ALS may share a common genetic or metabolic factor [321]. Amongst women hormonal factors may modulate ALS risk, and more specifically, there is an association between reduced risk of ALS and past history of OCP use (Chapter 5.2). The Euro-MOTOR findings may be mediated by anti-androgenic effects of oestrogens and progesterones (Chapter 5.2). Of note, a recent study of elite athletes found associations between, serum free testosterone and athletic performance in females across event types, lower testosterone and throwing events in males, and sex hormone binding globulin (SBGH) and throwing events [322]. Hormones and their interactions are very complex. However as there is evidence in the Euro-MOTOR study of an association between lifetime physical activity and ALS, and also between earlier life OCP use and reduced risk of ALS, I speculate that hormonal factors may be one factor feeding into a common metabolic mechanism underlying a favourable cardiovascular fitness profile and ALS susceptibility. Future cohort studies should capture as much data as possible regarding both lifetime physical activity and hormonal factors to better explore such hypotheses.

Occupational Exposures - limitations

The occupational analysis of EuroMOTOR found that a wide range of occupational exposures were associated with increased risk of ALS. This raises the possibility that a systematic information bias, for example over-reporting by cases, or under-reporting by controls may partly explain the results. However, since we sought to obtain a complete job listing from cases and controls and occupation was not the sole focus of the survey

this is unlikely. Alternatively, this could be consistent with the concept of volunteer bias – i.e. that controls, whom are all volunteers, may differ systematically from the general population from which cases are generated. We saw some evidence for this in that Italian controls tended to have higher education levels than Italian cases (Table 5-2). Throughout Chapter 5 education, which was not seen to be an important variable in models, was dichotomised as low (ISCED: 0 – 4) or high (ISCED: 5 - 8), due to the need to match levels across countries. Possibly, such sensitivity analyses could be improved by including more granular sub-categories of education level, although this would require interaction by country leading to a very complex model. An alternate approach would be to perform a detailed sociological analysis of the occupational data using a metric such as the European Socio-economic Classification (ESeC). This would provide a new dimension for interrogating the EuroMOTOR dataset and allow a new sensitivity analysis of the occupational data. For future studies, we have redesigned the EuroMOTOR education questions to be based on the ISCED classifications and this will be maintained across languages allowing for granular stratification by education level without the need to include interaction by country.

Exposures and implications for genetic studies

Current theories posit that ALS results from the culmination of multiple genetic and/or environmental factors, with the possibility that genes, environmental exposures and possibly gene-environment interactions form different steps along the pathway to disease [9,199]. To date, studies on ALS genetics, and studies on environmental exposures in ALS have been performed somewhat independently. The Euro-MOTOR study goes part-way to closing this divide(Chapter 5), and we have been able to show that removing *C9orf72* cases from the analysis leads to stronger associations between 16 out of 25 exposures and OR for ALS (Chapter 5.5). However this falls short of a full gene-environment interaction analysis.

Current thinking in ALS genetics is that ever-larger GWAS studies and large sequencing studies such as ProjectMine will yield results in the form of significant genes with moderate or small effect sizes that contribute to an oligogenic or polygenic causal

pathway [29]. Importantly, the Euro-MOTOR study (Chapter 5), demonstrates that occupational exposures are heterogeneous between different study populations represented in ALS studies and between the case and control groups within these studies. This carries implications for genetics studies in ALS, particularly in the context of ALS as a result of a multi-step pathway resulting from both genetic and environmental steps. Critically, our results imply that case-control studies in ALS, including genetic studies, are likely to be biased between cases and controls and across different populations with respect to occupational exposures. At best, this represents a source of noise causing greater likelihood of type 2 error (i.e. reducing statistical power). However, at worst such bias has potential to cause type 1 errors (i.e. false positives) in genetic studies. Furthermore, Gage *et al.* have recently highlighted that SNPs identified in GWAS studies, may in fact "reflect the effect of modifiable risk factors as well as direct genetic effects" [323], whilst Vanderweele *et al.* have demonstrated that uncontrolled environmental confounding may bias gene-environment analyses [324].

Therefore, both to understand current and future findings from genetics studies in ALS, and to dis-entangle the effects of correlated environmental exposures, it is a prerequisite that genetic studies include measurements of potential environmental confounders in ALS.

6.3. Final Conclusions

In conclusion, it has been shown using robust methodology that there are no clusters of incident ALS cases occurring in Ireland. The overall pattern of ALS incidence is not explained by population density, social deprivation, failure of ascertainment or soil mineral levels, nor do any of these factors associate with survival. Repeat analysis is recommended every 4 to 5 years (once per census cycle), to identify changing patterns or emerging clusters.

Survival of ALS patients is improved by attendance of the ALS multi-disciplinary clinic and therefore it is recommended that all newly diagnosed patients should attend a specialist multi-disciplinary ALS clinic. The ALSFRS-R yields greater prognostic significance when analysed as bulbar, motor or respiratory sub-scores, than as a single total score. Longitudinal characteristics of these sub-scores are distinct between spinal and bulbar onset patients and offer richer characterisation of progression in both observational studies and clinical trials. The respiratory sub-score is not sensitive to patient decline on average and enhancement with new clinical or functional metrics is recommended.

The negative prognostic effect of the *C9orf72* expansion was found to be driven by spinal onset male patients with the expansion across 5 European countries. Furthermore, in Ireland, Italy and The Netherlands patients carrying the *C9orf72* expansion were seen to have lower BMI for decades prior to ALS onset when compared to patients lacking the expansion or healthy controls. Patients without the expansion had higher odds ratios for many occupational exposures. Together these findings further delineate the *C9orf72* phenotype from sporadic ALS. Therefore, it is recommended that all ALS patients are tested for the *C9orf72* expansion, both for use in observational studies and for stratified analysis of clinical trials.

Lifetime physical activity was associated with increased odds for ALS. This relationship was stronger for occupational physical activity than leisure time physical activity. These findings are on a background of mixed previous findings in the literature. Such

conflicting findings may be a result of a common metabolic cause for both physical activity and ALS and future work is needed to evaluate this possibility. Use of the oral contraceptive pill was associated with reduced odds of ALS in women suggesting early-life oestrogens or progesterones (or their effects opposing androgens) may be a protective factor against ALS. Prospective cohort studies are needed to confirm this effect.

Occupational exposures to a range of chemicals were also associated with increased odds-ratios for ALS. The strongest associations were found for silica and pesticides exposure, with weaker associations found for solvents, metals, diesel mineral exhaust, poly-aromatic hydrocarbons, ELF-MF and electric shock, dusts gasses and vapours. Due to strong positive co-linearity between exposures and the omission of potentially important risk factors such as lead exposure, it is not possible to determine which specific exposures are driving these relationships. Future studies using exposure measurements with greater specificity are needed, and ideally future studies will be of gene-environment design.

7. References:

1. Elamin M, Phukan J, Bede P, Jordan N, Byrne S, Pender N, et al. Executive dysfunction is a negative prognostic indicator in patients with ALS without dementia. Neurology. 2011 Apr 5;76(14):1263–9.
2. Elamin M, Pinto-Grau M, Burke T, Bede P, Rooney J, O'Sullivan M, et al. Identifying behavioural changes in ALS: Validation of the Beaumont Behavioural Inventory (BBI). Amyotroph Lateral Scler Front Degener. 2017;18(1–2).
3. 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;72(2):245–56.
4. 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;72(2):257–68.
5. Elamin M, Bede P, Montuschi A, Pender N, Chio A, Hardiman O. Predicting prognosis in amyotrophic lateral sclerosis: a simple algorithm. J Neurol. 2015;262(6):1447–54.
6. 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;1(5):293–9.
7. Turner MR, Talbot K. Mimics and chameleons in motor neurone disease. Pract Neurol. 2013 Jun;13(3):153–64.
8. 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. 2017 Feb 1;46(1):57–74.
9. Al-Chalabi A, Calvo A, Chio A, Colville S, Ellis CMCM, 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.
10. Manjaly ZR, Scott KM, Abhinav K, Wijesekera L, Ganesalingam J, Goldstein LH, et al. The sex ratio in amyotrophic lateral sclerosis: A population based study. Amyotroph Lateral Scler. 2010;11(5):439–42.
11. 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;8(9).
12. Huisman MHB, Jong SW De, Doormaal PTC Van, Weinreich SS, Schelhaas HJ, Kooi AJ Van Der, et al. Population based epidemiology of amyotrophic lateral sclerosis using capture-recapture methodology. J Neurol Neurosurg Psychiatry. 2011;82(10):1165–70.
13. Byrne S, Walsh C, Lynch C, Bede P, Elamin M, Kenna K, et al. Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis. J Neurol Neurosurg Psychiatry. 2011 Jun;82(6):623–7.

14. 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;74(5):699–708.
15. Ganesalingam J, Stahl D, Wijesekera L, Galtrey C, Shaw CE, Leigh PN, et al. Latent cluster analysis of ALS phenotypes identifies prognostically differing groups. PLoS One. 2009;4(9):e7107.
16. Toole OO, 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;79(1):30–2.
17. Chiò A, Mora G, Calvo A, Mazzini L, Bottacchi E, Mutani R. Epidemiology of ALS in Italy: A 10-year prospective population-based study. Neurology. 2009;72(8):725–31.
18. Plato CC, Galasko D, Garruto RM, Plato M, Gamst A, Torres JM, et al. ALS and PDC of Guam: forty-year follow-up. Neurology. 2002 Mar 12;58(5):765–73.
19. Kihira T, Yoshida S, Kondo T, Iwai K, Wada S, Morinaga S, et al. An increase in ALS incidence on the Kii Peninsula , 1960 – 2009 : A possible link to change in drinking water source. Amyotrophic Lateral Sclerosis, 2012; 13: 347–350.
20. Kuzuhara S, Kokubo Y. Atypical parkinsonism of Japan: amyotrophic lateral sclerosis-parkinsonism-dementia complex of the Kii peninsula of Japan (Muro disease): an update. Mov Disord. 2005 Aug;20 Suppl 1:S108-13.
21. Sabel CE. Spatial Clustering of Amyotrophic Lateral Sclerosis in Finland at Place of Birth and Place of Death. Am J Epidemiol. 2003;157(10):898–905.
22. Uccelli R, Binazzi A, Altavista P, Belli S, Comba P, Mastrantonio M, et al. Geographic distribution of amyotrophic lateral sclerosis through motor neuron disease mortality data. Eur J Epidemiol. 2007;22(11):781–90.
23. Boumédiène F, Druet-Cabanac M, Marin B, Preux P-M, Allée P, Couratier P. Contribution of geolocalisation to neuroepidemiological studies: incidence of ALS and environmental factors in Limousin, France. J Neurol Sci. 2011 Oct 15;309(1–2):115–22.
24. Turabelidze G, Zhu B, Schootman M, Malone JL, Horowitz S, Weidinger J, et al. An epidemiologic investigation of amyotrophic lateral sclerosis in Jefferson County, Missouri, 1998-2002. Neurotoxicology. 2008;29(1):81–6.
25. Caller T a, Doolin JW, Haney JF, Murby AJ, West KG, Farrar HE, et al. A cluster of amyotrophic lateral sclerosis in New Hampshire: a possible role for toxic cyanobacteria blooms. Amyotroph Lateral Scler. 2009;10 Suppl 2(June):101–8.
26. Caller T a, Chipman JW, Field NC, Stommel EW. Spatial analysis of amyotrophic lateral sclerosis in Northern New England, USA, 1997-2009. Muscle Nerve. 2013;48(2):235–41.
27. Traynor BJ, Codd MB, Corr B, Forde C, Frost E, Hardiman O. Incidence and prevalence of ALS in Ireland, 1995-1997: A population-based study. Neurology. 1999;52(3):504–504.
28. Armon C. Smoking may be considered an established risk factor for sporadic ALS. Neurology.

2009 Nov 17;73(20):1693–8.

29. Al-Chalabi A, van den Berg LH, Veldink J. Gene discovery in amyotrophic lateral sclerosis: implications for clinical management. Nat Rev Neurol. 2016;13(2):96–104.
30. 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.
31. Deng HX, Hentati A, Tainer JA, Iqbal Z, Cayabyab A, Hung WY, et al. Amyotrophic lateral sclerosis and structural defects in Cu,Zn superoxide dismutase. Science. 1993 Aug 20;261(5124):1047–51.
32. Renton AE, Chiò A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. Nat Neurosci. 2014;17(1):17–23.
33. Ferrari R, Kapogiannis D, Huey ED, Momeni P. FTD and ALS: a tale of two diseases. Curr Alzheimer Res. 2011;8(3):273–94.
34. Silverman JM, Fernando SM, Grad LI, Hill AF, Turner BJ, Yerbury JJ, et al. Disease Mechanisms in ALS: Misfolded SOD1 Transferred Through Exosome-Dependent and Exosome-Independent Pathways. Cell Mol Neurobiol. 2016;36(3):377–81.
35. Kabashi E, Valdmanis PN, Dion P, Spiegelman D, McConkey BJ, Vande Velde C, et al. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat Genet. 2008;40(5):572–4.
36. Sreedharan J, Blair IP, Tripathi VB, Hu X, Vance C, Rogelj B, et al. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science. 2008;319(5870):1668–72.
37. Scotter EL, Chen HJ, Shaw CE. TDP-43 Proteinopathy and ALS: Insights into Disease Mechanisms and Therapeutic Targets. Neurotherapeutics. 2015;12(2):352–63.
38. Dormann D, Haass C. Fused in sarcoma (FUS): An oncogene goes awry in neurodegeneration. Mol Cell Neurosci. 2013;56:475–86.
39. 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;323(5918):1208–11.
40. Kabashi E, Bercier V, Lissouba A, Liao M, Brustein E, Rouleau GA, et al. Fus and tardbp but not sod1 interact in genetic models of amyotrophic lateral sclerosis. PLoS Genet. 2011;7(8):17–28.
41. Majounie E, Renton AE, Mok K, Dopper EGP, 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;11(4):323–30.
42. Mori K, Weng S-M, Arzberger T, May S, Rentzsch K, Kremmer E, et al. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science. 2013;339(6125):1335–8.
43. May S, Hornburg D, Schludi MH, Arzberger T, Rentzsch K, Schwenk BM, et al. C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. Acta Neuropathol. 2014;485–503.

44. Xiao S, MacNair L, McGoldrick P, McKeever PM, McLean JR, Zhang M, et al. Isoform-specific antibodies reveal distinct subcellular localizations of C9orf72 in amyotrophic lateral sclerosis. Ann Neurol. 2015 Oct;78(4):568–83.
45. Jovicic A, Mertens J, Boeynaems S, Bogaert E, Chai N, Yamada SB, et al. Modifiers of C9orf72 dipeptide repeat toxicity implicate nucleocytoplasmic transport impairments to FTD/ ALS. Nat Neurosci. 2015;18(9).
46. Freibaum BD, Lu Y, Lopez-Gonzalez R, Kim NC, Almeida S, Lee K, et al. GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. Nature. 2015 Aug 26;525(7567):129–33.
47. Zhang K, Donnelly CJ, Haeusler AR, Grima JC, Machamer JB, Steinwald P, et al. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. Nature. 2015; 525(7567):56-61.
48. Devenney EM, Landin-Romero R, Irish M, Hornberger M, Mioshi E, Halliday GM, et al. The neural correlates and clinical characteristics of psychosis in the frontotemporal dementia continuum and the C9orf72 expansion. NeuroImage Clin. 2017;13:439–45.
49. Cooper-Knock J, Shaw PJ, Kirby J. The widening spectrum of C9ORF72-related disease; Genotype/phenotype correlations and potential modifiers of clinical phenotype. Acta Neuropathol. 2014;127(3):333–45.
50. 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;82(4):292–9.
51. Kostić VS, Dobričić V, Stanković I, Ralić V, Stefanova E. C9orf72 expansion as a possible genetic cause of huntington disease phenocopy syndrome. J Neurol. 2014;261(1):1917–21.
52. Helgason E, Phung QT, Dueber EC. Recent insights into the complexity of Tank-binding kinase 1 signaling networks: The emerging role of cellular localization in the activation and substrate specificity of TBK1. FEBS Lett. 2013;587(8):1230–7.
53. Oakes JA, Davies MC, Collins MO. TBK1: a new player in ALS linking autophagy and neuroinflammation. Molecular Brain; 2017;10(1).
54. Alonso A, Logroscino G, Hernán MA. Smoking and the risk of amyotrophic lateral sclerosis: a systematic review and meta-analysis. J Neurol Neurosurg Psychiatry. 2010;81(11):1249–52.
55. 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;87(11):1229-1233.
56. Chiò A, Benzi G, Dossena M, Mutani R, Mora G. Severely increased risk of amyotrophic lateral sclerosis among Italian professional football players. Brain. 2005;128(Pt 3):472–6.
57. Armon C. Sports and trauma in amyotrophic lateral sclerosis revisited. J Neurol Sci. 2007;262(1–2):45–53.
58. Hamidou B, Couratier P, Besançon C, Nicol M, Preux PM, Marin B. Epidemiological evidence that physical activity is not a risk factor for ALS. Eur J Epidemiol. 2014;29(7):459–75.
59. 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;84(9):976–81.

60. 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.

61. Armon C, Nelson LM. Is head trauma a risk factor for amyotrophic lateral sclerosis? An evidence based review. Amyotroph Lateral Scler. 2012;13(4):351–6.

62. Wang M-D, Little J, Gomes J, Cashman NR, Krewski D. Identification of risk factors associated with onset and progression of amyotrophic lateral sclerosis using systematic review and meta-analysis. Neurotoxicology. 2017;61:101-130.

63. 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;46(2):96–105.

64. Sutedja N a., Veldink JH, Fischer K, Kromhout H, Heederik D, Huisman MHBB, et al. Exposure to chemicals and metals and risk of amyotrophic lateral sclerosis: a systematic review. Amyotroph Lateral Scler. 2009;10(5–6):302–9.

65. Kamel F, Umbach DM, Bedlack RS, Richards M, Watson M, Alavanja MCRR, et al. Pesticide exposure and amyotrophic lateral sclerosis. Neurotoxicology. 2012;33(3):457–62.

66. 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: pesticide exposure as a risk factor for ALS. Environ Res. 2012;117:112–9.

67. Kang H, Cha ES, Choi GJ, Lee WJ. Amyotrophic Lateral Sclerosis and Agricultural Environments : A Systematic Review. J Korean Med Sci 2014; 29: 1610-1617.

68. Callaghan B, Feldman D, Gruis K, Feldman E. The association of exposure to lead, mercury, and selenium and the development of amyotrophic lateral sclerosis and the epigenetic implications. Neurodegener Dis. 2011;8(1–2):1–8.

69. Wang M-D, Gomes J, Cashman NR, Little J, Krewski D. A meta-analysis of observational studies of the association between chronic occupational exposure to lead and amyotrophic lateral sclerosis. J Occup Environ Med. 2014;56(12):1235–42.

70. Hersi M, Quach P, Wang M-D, Gomes J, Gaskin J, Krewski D. Systematic reviews of factors associated with the onset and progression of neurological conditions in humans: a methodological overview. Neurotoxicology. 2017;61:12-18.

71. E M, Yu S, Dou J, Jin W, Cai X, Mao Y, et al. Association between alcohol consumption and amyotrophic lateral sclerosis: a meta-analysis of five observational studies. Neurol Sci. 2016;37(8):1203–8.

72. Pamphlett R, Rikard-Bell A. Different occupations associated with amyotrophic lateral sclerosis: is diesel exhaust the link? PLoS One. 2013;8(11):e80993.

73. Abhinav K, Al-Chalabi A, Hortobagyi T, Leigh PN. Electrical injury and amyotrophic lateral sclerosis: a systematic review of the literature. J Neurol Neurosurg Psychiatry. 2006;78(5):450–3.

74. Lekoubou A, Matsha TE, Sobngwi E, Kengne AP. Effects of diabetes mellitus on amyotrophic

lateral sclerosis: a systematic review. BMC Res Notes. 2014;7(1):171.

75. Jawaid A, Murthy SB, Wilson AM, Qureshi SU, Amro MJ, Wheaton M, et al. A decrease in body mass index is associated with faster progression of motor symptoms and shorter survival in ALS. Amyotroph Lateral Scler. 2010;11(6):542–8.

76. Miller RG, Mitchell JD, Moore DH. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). Cochrane Database Syst Rev. 2012;3(3):CD001447.

77. 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;5(2):140–7.

78. Chiò A, Calvo A, Moglia C, Gamna F, Mattei A, Mazzini L, et al. Non-invasive ventilation in amyotrophic lateral sclerosis: a 10 year population based study. J Neurol Neurosurg Psychiatry. 2012;83(4):377–81.

79. Stavroulakis T, Walsh T, Shaw PJ, McDermott CJ. Gastrostomy use in motor neurone disease (MND): A review, meta-analysis and survey of current practice. Amyotroph Lateral Scler Front Degener. 2013;14(2):96–104.

80. Russ KB, Phillips MC, Wilcox CM, Peter S. Percutaneous endoscopic gastrostomy in amyotrophic lateral sclerosis. Am J Med Sci. 2015;350(2):95–7.

81. Allen J a, Chen R, Ajroud-Driss S, Sufit RL, Heller S, Siddique T, et al. Gastrostomy tube placement by endoscopy versus radiologic methods in patients with ALS: a retrospective study of complications and outcome. Amyotroph Lateral Scler Frontotemporal Degener. 2013;14(4):308–14.

82. Traynor BJ, Alexander M, Corr B, Frost E, Hardiman O. Effect of a multidisciplinary amyotrophic lateral sclerosis (ALS) clinic on ALS survival: a population based study, 1996-2000. J Neurol Neurosurg Psychiatry. 2003 Sep;74(9):1258–61.

83. Van den Berg JP, Kalmijn S, Lindeman E, Veldink JH, de Visser M, Van der Graaff MM, et al. Multidisciplinary ALS care improves quality of life in patients with ALS. Neurology. 2005;65(8):1264–7.

84. Zoccolella S, Beghi E, Palagano G, Fraddosio A, Guerra V, Lepore V, et al. ALS multidisciplinary clinic and survival. Results from a population-based study in Southern Italy. J Neurol. 2007;254(8):1107–12.

85. Aridegbe T, Kandler R, Walters SJ, Walsh T, Shaw PJ, McDermott CJ. The natural history of motor neuron disease: assessing the impact of specialist care. Amyotroph Lateral Scler Frontotemporal Degener. 2013 Jan;14(1):13–9.

86. Abe K, Aoki 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;4422(17).

87. Al-Chalabi A, Andersen PM, Chandran S, Chio A, Corcia P, Couratier P, et al. July 2017 ENCALS statement on edaravone. Amyotroph Lateral Scler Frontotemporal Degener. 2017;18(7–8):471–

474.

88. 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;11(3):232–40.
89. Van Rheenen W, Van Blitterswijk M, Huisman MHB, Vlam L, Van Doormaal PTC, Seelen M, et al. Hexanucleotide repeat expansions in C9ORF72 in the spectrum of motor neuron diseases. Neurology. 2012;79(9):878–82.
90. Sabatelli M, Conforti FL, Zollino M, Mora G, Monsurrò MR, Volanti P, et al. C9ORF72 hexanucleotide repeat expansions in the Italian sporadic ALS population. Neurobiol Aging. 2012;33(8):1–6.
91. Borghero G, Pugliatti M, Marrosu F, Marrosu MG, Murru MR, Floris G, et al. Genetic architecture of ALS in Sardinia. Neurobiol Aging. 2014 Dec;35(12):2882.
92. Debray S, Race V, Crabbé V, Herdewyn S, Matthijs G, Goris A, et al. Frequency of C9orf72 repeat expansions in amyotrophic lateral sclerosis: a Belgian cohort study. Neurobiol Aging. 2013;34(12):2890.e7-2890.e12.
93. García-Redondo A, Dols-Icardo O, Rojas-García R, Esteban-Pérez J, Cordero-Vázquez P, Muñoz-Blanco JL, et al. Analysis of the C9orf72 Gene in Patients with Amyotrophic Lateral Sclerosis in Spain and Different Populations Worldwide. Hum Mutat. 2013;34(1):79–82.
94. Irwin DJ, McMillan CT, Brettschneider J, Libon DJ, Powers J, Rascovsky K, et al. Cognitive decline and reduced survival in C9orf72 expansion frontotemporal degeneration and amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 2013;84(2):163–9.
95. Chiò A, Mora G, Restagno G, Brunetti M, Ossola I, Barberis M, et al. UNC13A influences survival in Italian amyotrophic lateral sclerosis patients: A population-based study. Neurobiol Aging. 2013;34(1):357.e1-357.e5.
96. Diekstra FP, van Vught PWJ, van Rheenen W, Koppers M, Pasterkamp RJ, van Es MA, et al. UNC13A is a modifier of survival in amyotrophic lateral sclerosis. Neurobiol Aging. 2012;33(3):630.e3-630.e8.
97. Gaastra B, Shatunov A, Pulit S, Jones AR, Sproviero W, Gillett A, et al. Rare genetic variation in UNC13A may modify survival in amyotrophic lateral sclerosis. Amyotroph Lateral Scler Front Degener. 2016;8421(September):1–7.
98. Fogh I, Lin K, Tiloca C, Rooney J, Gellera C, Diekstra FP, et al. Association of a locus in the CAMTA1 gene with survival in patients with sporadic amyotrophic lateral sclerosis. JAMA Neurol. 2016;73(7).
99. Landers JE, Melki J, Meininger V, Glass JD, van den Berg LH, van Es M a, et al. Reduced expression of the Kinesin-Associated Protein 3 (KIFAP3) gene increases survival in sporadic amyotrophic lateral sclerosis. Proc Natl Acad Sci U S A. 2009;106(22):9004–9.
100. Van Hoecke A, Schoonaert L, Lemmens R, Timmers M, Staats K a, Laird AS, et al. EPHA4 is a disease modifier of amyotrophic lateral sclerosis in animal models and in humans. Nat Med.

2012;18(9):1418-22 .

101. Traynor BJ, Nalls M, Lai S-L, Gibbs RJ, Schymick JC, Arepalli S, et al. Kinesin-associated protein 3 (KIFAP3) has no effect on survival in a population-based cohort of ALS patients. Proc Natl Acad Sci U S A . 2010;107(27):12335–8.

102. Keren NOA, Scott KM, Tsuda M, Barnwell J, Knibb JA, Ellis CM, et al. Evidence of an environmental effect on survival in ALS. Amyotroph Lateral Scler Frontotemporal Degener. 2014;15(7-8):528-33.

103. De Jong S, Huisman M, Sutedja N, Van Der Kooi A, De Visser M, Schelhaas J, et al. Endogenous female reproductive hormones and the risk of amyotrophic lateral sclerosis. J Neurol. 2013;260(2):507–12.

104. 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. Biochim Biophys Acta. 2006;1762(11–12):1150–7.

105. Logroscino G, Traynor BJ, Hardiman O, Chiò A, Mitchell D, Swingler RJ, et al. Incidence of amyotrophic lateral sclerosis in Europe. J Neurol Neurosurg Psychiatry. 2010;81(4):385–90.

106. Kenna KP, C van Doormaal PT, Dekker AM, Ticozzi N, Kenna BJ, Diekstra FP, et al. NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. Nat Genet. 2016;48(9).

107. Galvin M, Madden C, Maguire S, Heverin M, Vajda A, Staines A, et al. Patient journey to a specialist amyotrophic lateral sclerosis multidisciplinary clinic: an exploratory study. BMC Health Serv Res. 2015;15(1):571.

108. de Jong SW, Huisman MHB, Sutedja NA, van der Kooi AJ, de Visser M, Schelhaas HJ, et al. Smoking, alcohol consumption, and the risk of amyotrophic lateral sclerosis: a population-based study. Am J Epidemiol. 2012;176(3):233–9.

109. Franchignoni F, Mora G, Giordano A, Volanti P, Chiò A. Evidence of multidimensionality in the ALSFRS-R Scale: a critical appraisal on its measurement properties using Rasch analysis. J Neurol Neurosurg Psychiatry. 2013;84(12):1340–5.

110. Scott KM, Abhinav K, Stanton BR, Johnston C, Turner MR, Ampong M-A, et al. Geographical clustering of amyotrophic lateral sclerosis in South-East England: a population study. Neuroepidemiology. 2009;32(2):81–8.

111. Migliaretti G, Berchialla P, Dalmasso P, Cavallo F, Ò ACHI, Chiò A. Amyotrophic lateral sclerosis in Piedmont (Italy): a Bayesian spatial analysis of the incident cases. Amyotroph Lateral Scler Frontotemporal Degener. 2013;14(1):58–65.

112. 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.

113. Chiò a., Logroscino G, Traynor BJ, Collins J, Simeone JC, Goldstein L a., et al. Global Epidemiology of Amyotrophic Lateral Sclerosis: A Systematic Review of the Published Literature. Neuroepidemiology. 2013;41(2):118–30.

114. Yeo L, Lynch C, Hardiman O. Validating population-based registers for ALS: how accurate is

death certification? J Neurol. 2010;257(8):1235–9.

115. 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;12(6):451–7.

116. Murphy M, Quinn S, Young J, Parkin P, Taylor B. Increasing incidence of ALS in Canterbury, New Zealand. Neurology. 2008;71(23):1889–95.

117. Govoni V, Granieri E, Capone J, Manconi M, Casetta I. Incidence of amyotrophic lateral sclerosis in the local health district of Ferrara, Italy, 1964-1998. Neuroepidemiology. 2003;22(4):229–34.

118. Zach N, Ennist DL, Taylor AA, Alon H, Sherman A, Kueffner R, et al. Being PRO-ACTive: What can a Clinical Trial Database Reveal About ALS? Neurotherapeutics. 2015;12(2):417–23.

119. Chiò a, Canosa a, Gallo S, Cammarosano S, Moglia C, Fuda G, et al. ALS clinical trials: do enrolled patients accurately represent the ALS population? Neurology. 2011;77(15):1432–7.

120. UKMND-LiCALS Study Group, Morrison KE, Dhariwal S, Hornabrook R, Savage L, Burn DJ, et al. Lithium in patients with amyotrophic lateral sclerosis (LiCALS): a phase 3 multicentre, randomised, double-blind, placebo-controlled trial. Lancet Neurol. 2013 Apr;12(4):339–45.

121. Beghi E, Balzarini C, Bogliun G, Logroscino G, Manfredi L, Mazzini L, et al. Reliability of the El escorial diagnostic criteria for amyotrophic lateral sclerosis. Neuroepidemiology. 2002;21(6):265–70.

122. 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;263(10).

123. 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;264(1):54-63.

124. Phukan J, Elamin M, Bede P, Jordan N, Gallagher L, Byrne S, et al. The syndrome of cognitive impairment in amyotrophic lateral sclerosis: A population-based study. J Neurol Neurosurg Psychiatry. 2012;83(1):102–8.

125. Elamin M, Bede P, Byrne S, Jordan N, Gallagher L, Wynne B, et al. Cognitive changes predict functional decline in ALS. Neurology. 2013;80(17):1590–7.

126. Vlam L, Piepers S, Sutedja NA, Jacobs BC, Tio-Gillen AP, Stam M, et al. Association of IgM monoclonal gammopathy with progressive muscular atrophy and multifocal motor neuropathy: a case-control study. J Neurol. 2015;262(3):666–73.

127. Huisman MHB, Seelen M, van Doormaal PTC, de Jong SW, de Vries JHM, van der Kooi AJ, et al. Effect of Presymptomatic Body Mass Index and Consumption of Fat and Alcohol on Amyotrophic Lateral Sclerosis. JAMA Neurol. 2015;72(10):1155-62.

128. Sutedja N a, Fischer K, Veldink JH, van der Heijden GJMG, Kromhout H, Heederik D, et al. What we truly know about occupation as a risk factor for ALS: a critical and systematic review. Amyotroph Lateral Scler. 2009;10(5–6):295–301.

129. D'Ovidio F, Rooney JPK, Visser AE, Vermeulen RCH, Veldink JH, Van Den Berg LH, et al. Critical issues in ALS case-control studies: the case of the Euro-MOTOR study. Amyotroph Lateral Scler

Front Degener. 2017;18(5-6):411-418.

130. Henry K a., Fagliano J, Jordan HM, Rechtman L, Kaye WE. Geographic Variation of Amyotrophic Lateral Sclerosis Incidence in New Jersey, 2009-2011. Am J Epidemiol. 2015;182(6):512-9.

131. Olsen SF, Martuzzi M, Elliott P. Cluster analysis and disease mapping--why, when, and how? A step by step guide. BMJ. 1996;313(7061):863–6.

132. Tesauro M, Consonni M, Filippini T, Mazzini L, Pisano F, Chiò A, et al. Incidence of amyotrophic lateral sclerosis in the province of Novara, Italy, and possible role of environmental pollution. Amyotroph Lateral Scler Frontotemporal Degener. 2017;18(3-4):284-290.

133. Rooney J, Vajda A, Heverin M, Elamin M, Crampsie A, McLaughlin R, et al. Spatial cluster analysis of population amyotrophic lateral sclerosis risk in Ireland. Neurology. 2015;84(15):1537–44.

134. Pugliatti M, Parish LD, Cossu P, Leoni S, Pirina P, Borghero G, et al. Amyotrophic lateral sclerosis in Sardinia , insular Italy , 1995 – 2009. J Neurol. 2013;260(2):572-9.

135. 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;50(11):776-83.

136. 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;79(1):6–11.

137. Zaldivar T, Gutierrez J, Lara G, Carbonara M, Logroscino G, Hardiman O. Reduced frequency of ALS in an ethnically mixed population: a population-based mortality study. Neurology. 200912;72(19):1640–5.

138. 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. Eur J Epidemiol. 2016;31(3):229-45.

139. Bucheli M, Andino A, Montalvo M, Cruz J, Atassi N, Berry J, et al. Amyotrophic lateral sclerosis: analysis of ALS cases in a predominantly admixed population of Ecuador. Amyotroph Lateral Scler Frontotemporal Degener. 2014;15(1–2):106–13.

140. Cruz-Flores S, Rabinstein A, Biller J, Elkind MS V, Griffith P, Gorelick PB, et al. Racial-ethnic disparities in stroke care: The American experience: A statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2011;42(7):2091–116.

141. Hall WJ, Chapman M V., Lee KM, Merino YM, Thomas TW, Payne BK, et al. Implicit Racial/Ethnic Bias Among Health Care Professionals and Its Influence on Health Care Outcomes: A Systematic Review. Am J Public Health. 2015;105(12):e1–17.

142. Rooney J, Byrne S, Heverin M, Tobin K, Dick A, Donaghy C, et al. A multidisciplinary clinic approach improves survival in ALS:A comparative study of ALS in Ireland and Northern Ireland. J Neurol Neurosurg Psychiatry. 2015;86(5).

143. Chiò a, Bottacchi E, Buffa C, Mutani R, Mora G. Positive effects of tertiary centres for

amyotrophic lateral sclerosis on outcome and use of hospital facilities. J Neurol Neurosurg Psychiatry. 2006;77(8):948–50.

144. van der Steen I, van den Berg J-P, Buskens E, Lindeman E, van den Berg LH. The costs of amyotrophic lateral sclerosis, according to type of care. Amyotroph Lateral Scler. 2009;10(1):27–34.

145. Berger ML, Dreyer N, Anderson F, Towse A, Sedrakyan A, Normand SL. Prospective observational studies to assess comparative effectiveness: The ISPOR good research practices task force report. Value Heal. 2012;15(2):217–30.

146. Traynor BJ, Alexander M, Corr B, Frost E, Hardiman O. An outcome study of riluzole in amyotrophic lateral sclerosis--a population-based study in Ireland, 1996-2000. J Neurol. 2003;250:473–9.

147. HRB. GDPR guidance for researchers [Internet]. 2018 [cited 2018 Sep 21]. Available from: http://www.hrb.ie/funding/gdpr-guidance-for-researchers/gdpr-overview/

148. Timmers M, Van Veen E-B, Maas AIR, Kompanje EJO. Will the Eu Data Protection Regulation 2016/679 Inhibit Critical Care Research? Med Law Rev. 2018 May 17. doi: 10.1093/medlaw/fwy023. [Epub ahead of print].

149. McCall B. What does the GDPR mean for the medical community? Lancet. 2018;391(10127):1249–50.

150. Leek JT, Peng RD. P values are just the tip of the iceberg. Nature. 2015;520(7549):612.

151. Kenna KP, Mclaughlin RL, Hardiman O, Bradley DG. Using reference databases of genetic variation to evaluate the potential pathogenicity of candidate disease variants. Hum Mutat. 2013;34(6):836–41.

152. Carpenter TE. The spatial epidemiologic (r)evolution: a look back in time and forward to the future. Spat Spatiotemporal Epidemiol. 2011;2(3):119–24.

153. Alonso V, Villaverde-Hueso A, Hens MJ, Morales-Piga A, Abaitua I, de la Paz MP. Increase in motor neuron disease mortality in Spain: temporal and geographical analysis (1990-2005). Amyotroph Lateral Scler. 2011;12(3):192–8.

154. Noonan CW, White MC, Thurman D, Wong L-Y. Temporal and geographic variation in United States motor neuron disease mortality, 1969-1998. Neurology. 2005 Apr 12;64(7):1215–21.

155. Doi Y, Yokoyama T, Tango T, Takahashi K, Fujimoto K, Nakano I. Temporal trends and geographic clusters of mortality from amyotrophic lateral sclerosis in Japan, 1995–2004. J Neurol Sci. Elsevier B.V.; 2010;298(1–2):78–84.

156. Bivand RS, Pebesma EJ, Gomez-Rubio V. Applied Spatial Data Analysis with R. 1st ed. New York: Springer; 2008. 378 p.

157. OSI. Ordnance Survey of Ireland. Electoral Divison Boundary Data. 2013. Available from: http://www.osi.ie/Home.aspx

158. Office CS. Census 2002 Volume 1 - Population Classified by Area. Dublin; 2003.

159. Office CS. Census 2006 Volume 1 - Population Classified by Area. Dublin; 2007.

160. Office CS. Census 2011. Census 2011 This is Ireland (Part 1). Dublin; 2012.
161. Office CS. Census 1996, Vol 1. Population Classified by Area. Dublin; 1997.
162. OSI. Ordnance Survey Ireland Website [Internet]. [cited 2014 Jan 1]. Available from: https://www.osi.ie
163. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012. Available from: http://www.r-project.org/
164. Nicholas J. Lewin-Koh, Roger Bivand, contributions by Edzer J. Pebesma, Eric Archer AB, Hans-Jörg Bibiko, *et al*. maptools: Tools for reading and handling spatial objects. R package version 0.8-22; 2013.
165. Bivand RS, Kleitt T, Rowlingson B. rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 0.8-5; 2013.
166. Pebesma EJ, Bivand RS. Classes and methods for spatial data in R. R News 5 (2); 2005.
167. Bivand, Roger, Leisch F, Maechler M. pixmap: Bitmap Images (``Pixel Maps'') [Internet]. 2011. Available from: http://cran.r-project.org/package=pixmap
168. Tierney L, Rossini AJ, Li N, Sevcikova H. snow: Simple Network of Workstations. 2012.
169. Roger Bivand, with contributions by Micah Altman, Luc Anselin, Renato Assunção, *et al*. spdep: Spatial dependence: weighting schemes, statistics and models. R package version 0.5-56; 2013.
170. Brunsdon C, Chen H. GISTools: Some further GIS capabilities for R. R package version 0.7-1; 2012.
171. Bivand RS. classInt: Choose univariate class intervals. R package version 0.1-19; 2012.
172. Sturtz S, Ligges U, Gelman A. R2WinBUGS: A Package for Running WinBUGS from R. J Stat Software2005. 12(3):1–16.
173. Plummer M, Best N, Cowles K, Vines K. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News. 2006;6:7–11.
174. Wickham H. Reshaping Data with the reshape Package. J Stat Softw. 2007;21(12).
175. Neuwirth E. RColorBrewer: ColorBrewer palettes. R package version 1.0-5; 2011.
176. Urbanek S. proj4: A simple interface to the PROJ.4 cartographic projections library. R package version 1.0-8; 2012.
177. Bivand RS, Rundel C. rgeos: Interface to Geometry Engine - Open Source (GEOS). R package version 0.2-13; 2013.
178. Registry NC, Cancer NI. All Ireland Cancer Atlas 1995-2007. 2007;
179. Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: Evolution, critique, and future directions. 2009;
180. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math. 1991;43(1):1–20.
181. Lunn D, Jackson C, Thomas A, Best N, Spiegelhalter D. The BUGS Book: A Practical Introduction to Bayesian Analysis. 1st ed. CRC Press, Taylor & Francis Group; 2012. 399 p.
182. Fang F, Valdimarsdóttir U, Bellocco R, Ronnevi L-O, Sparén P, Fall K, et al. Amyotrophic lateral

sclerosis in Sweden, 1991-2005. Arch Neurol. 2009 Apr;66(4):515–9.

183. Kulldorff M, Information Management Services I. SaTScan v8.0: software for the spatial and space-time scan statistics. 2009.
184. Kulldorff BM. SaTScan User Guide. 2013;
185. Tango T, Takahashi K. A flexible spatial scan statistic with a restricted likelihood ratio for detecting disease clusters. Stat Med. 2012;31(30):4207–18.
186. Takahashi K, Yokoyama T, Tango T. FleXScan v3.1: Software for the Flexible Scan Statistic. Japan: National Institute of Public Health; 2010.
187. Takahashi K, Yokoyama T, Tango T. FleXScan User Guide. 2010;(October).
188. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Ser B. 2009;71(2):319–92.
189. Martins TG, Simpson D, Lindgren F, Rue H. Bayesian computing with INLA: New features. Comput Stat Data Anal. 2013;67:68–83.
190. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. Available from: http://www.r-project.org/
191. Masseret E, Banack S, Boumédiène F, Abadie E, Brient L, Pernet F, et al. Dietary BMAA exposure in an amyotrophic lateral sclerosis cluster from southern France. PLoS One. 2013;8(12):e83406.
192. Smyth W. Map-making, Landscapes and Memory: A Geography of Colonial and Early Modern Ireland. Cork: Cork University Press; 2006.
193. Scott KM, Abhinav K, Wijesekera L, Ganesalingam J, Goldstein LH, Janssen A, et al. The association between ALS and population density: A population based study. Amyotroph Lateral Scler. 2010;11(5):435–8.
194. OSI. Ordnance Survey Ireland. Small area boundary data. 2013. Available from: http://www.osi.ie/Products/Professional-Mapping/Boundary-Data.aspx
195. Haase T, Pratschke J. The 2011 Pobal HP Deprivation Index for Small Areas (SA) Introduction and Reference Tables. 2012. Available from: http://trutzhaase.eu/deprivation-index/the-2011-pobal-hp-deprivation-index-for-small-areas/
196. Openshaw S. The modifiable areal unit problem. Norwich, united Kingdom: Geo Books; 1984.
197. Sharp L, Donnelly D, Hegarty A, Carsin A, Deady S, McCluskey N, et al. Risk of Several Cancers is Higher in Urban Areas after Adjusting for Socioeconomic Status. Results from a Two-Country Population-Based Study of 18 Common Cancers. J Urban Heal. 2014;91(3):510–25.
198. Northern Ireland Cancer Registry B. Cancer Incidence and Survival in Northern Ireland. Belfast; 2012. Available from: http://www.qub.ac.uk/research-centres/nicr/FileStore/PDF/Filetoupload,297948,en.pdf
199. Al-Chalabi A, Hardiman O. The epidemiology of ALS: a conspiracy of genes, environment and time. Nat Rev Neurol. 2013 Nov;9(11):617-28.
200. Pupillo E, Messina P, Giussani G, Logroscino G, Zoccolella S, Chiò A, et al. Physical activity and amyotrophic lateral sclerosis: a European population-based case-control study. Ann Neurol.

2014;75(5):708–16.

201. Eum K-D, Seals RM, Taylor KM, Grespin M, Umbach DM, Hu H, et al. Modification of the association between lead exposure and amyotrophic lateral sclerosis by iron and oxidative stress related gene polymorphisms. Amyotroph Lateral Scler Frontotemporal Degener. 2015;16(1-2):72-9

202. Rooney J. Further thoughts on mercury, epigenetics, genetics and amyotrophic lateral sclerosis. Neurodegener Dis. 2011;8(6):523-4.

203. Fay, D. Zhang C. Secure Archive For Environmental Research Data. Towards a National Soil Database. 2007. Available from: http://erc.epa.ie/safer/resource?id=c265bb3f-2cec-102a-b1da-b128b41032cc

204. Fay D, Mcgrath D, Zhang C, Carrigg C, Flaherty VO, Kramers G, et al. NATIONAL SOILS DATABASE END OF PROJECT REPORT RMIS 5192 Authors Johnstown Castle Research Centre Wexford July 2007 [Internet]. 2007. Available from: http://www.teagasc.ie/research/reports/environment/5192/eopr-5192.pdf

205. Hiemstra PH, Pebesma EJ, Twenhöfel CJW, Heuvelink GBM. Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. Comput Geosci. 2009;35(8):1711–21.

206. Vinceti M, Solovyev N, Mandrioli J, Crespi CM, Bonvicini F, Arcolin E, et al. Cerebrospinal fluid of newly diagnosed amyotrophic lateral sclerosis patients exhibits abnormal levels of selenium species including elevated selenite. Neurotoxicology. 2013;38:25–32.

207. Pamphlett R, Jew SK. Uptake of inorganic mercury by human locus ceruleus and corticomotor neurons : implications for amyotrophic lateral sclerosis. Acta Neuropathol Commun. 2013 May 9;1:13.

208. Johnson FO, Atchison WD. The role of environmental mercury, lead and pesticide exposure in development of amyotrophic lateral sclerosis. Neurotoxicology. 2009;30(5):761–5.

209. Flomenbaum N, Goldfrank L, Hoffman R, Howland MA, Lewin N, Nelson L. Goldfrank's Toxicologic Emergencies, Eighth Edition. McGraw-Hill Companies,Incorporated; 2006.

210. Rooney JPKJPK. The role of thiols, dithiols, nutritional factors and interacting ligands in the toxicology of mercury. Toxicology. 2007;234(3):145–56.

211. Crapper McLachlarf DR, McLachlan CD, Krishnan B, Krishnan SS, Dalton AJ, Steele JC. Aluminium and calcium in soil and food from Guam, Palau and Jamaica: Implications for amyotrophic lateral sclerosis and parkinsonism-dementia syndromes of Guam. Environ Geochem Health. 1989;11(2):45–53.

212. Roos PM, Vesterberg O, Syversen T, Flaten TP, Nordberg M. Metal concentrations in cerebrospinal fluid and blood plasma from patients with amyotrophic lateral sclerosis. Biol Trace Elem Res. 2013;151(2):159–70.

213. Richardson S, Thomson A, Best N, Elliott P. Interpreting Posterior Relative Risk Estimates in Disease-Mapping Studies. Environ Health Perspect. 2004;112(9):1016–25.

214. Kiernan MC, Vucic S, Cheah BC, Turner MR, Eisen A, Hardiman O, et al. Amyotrophic lateral sclerosis. Lancet. 2011;377(9769):942–55.
215. Bensimon G, Lacomblez L, Meininger V. A controlled trial of riluzole in amyotrophic lateral sclerosis. ALS/Riluzole Study Group. N Engl J Med. 1994;330(9):585–91.
216. Newsom-Davis IC, Lyall R a, Leigh PN, Moxham J, Goldstein LH. The effect of non-invasive positive pressure ventilation (NIPPV) on cognitive function in amyotrophic lateral sclerosis (ALS): a prospective study. J Neurol Neurosurg Psychiatry. 2001;71(4):482–7.
217. 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.
218. Donaghy C, O'Toole O, Sheehan C, Kee F, Hardiman O, Patterson V, et al. An all-Ireland epidemiological study of MND, 2004-2005. Eur J Neurol. 2009;16(1):148–53.
219. 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;11(4):374–8.
220. Andersen PM, Borasio GD, Dengler R, Hardiman O, Kollewe K, Leigh PN, et al. EFNS task force on management of amyotrophic lateral sclerosis: guidelines for diagnosing and clinical care of patients and relatives. Eur J Neurol. 2005 Dec;12(12):921–38.
221. StataCorp. Stata: Release 11. College Station, Texas. USA: StataCorp LP; 2009.
222. Creemers H, Veldink JH, Grupstra H, Nollet F, Beelen A, van den Berg LH. Cluster RCT of case management on patients' quality of life and caregiver strain in ALS. Neurology. 2014;82(1):23–31.
223. Haase T, Pratschke J, Gleeson J. All-Island Deprivation Index: Towards the Development of Consistent Deprivation for the Island of Ireland. 2012. p. 18. Available from: http://trutzhaase.eu/wp/wp-content/uploads/R-2012-All-Island-Deprivation-Index.pdf
224. Cancer Research UK. Tobacco Statistics. 2014. Available from: http://www.cancerresearchuk.org/cancer-info/cancerstats/causes/tobacco-statistics/
225. HSE. Cigarette Smoking Prevalence in Ireland. 2013. Available from: http://www.hse.ie/eng/about/Who/TobaccoControl/Research/
226. Walsh PM, Byrne J, Kelly M, McDevitt J, Comber H. Socioeconomic Disparity in Survival after Breast Cancer in Ireland: Observational Study. PLoS One. 2014;9(11):e111729.
227. Mari-Dell'Olmo M, Gotsens M, Palencia L, Burstrom B, Corman D, Costa G, et al. Socioeconomic inequalities in cause-specific mortality in 15 European cities. J Epidemiol Community Health. 2015;69(5):432-41.
228. Royston P, Lambert PC. Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model. First Edit. Stata Press; 2011.
229. StataCorp. Stata Statistical Software: Release 14. College Station, Texas. USA: StataCorp LP; 2015.
230. Hansell, AL and Beale, LA and Ghosh, RE and Fortunato, L and Fecht, D and Järup L. The

Environment and Health Atlas for England and Wales. Oxford: OUP; 2014.

231. ENCALS. ALS Core Clinical Dataset. 2014. Available from: http://www.encals.eu/public/media/ac8bacd302965623601ddfe4229c4aadcc370922/6xg7/pdf/17a00a103f602f2837d968fd038df3ca17ae0c76.pdf

232. van Rheenen W, Diekstra FP, van Doormaal PTC, Seelen M, Kenna K, McLaughlin R, et al. H63D polymorphism in HFE is not associated with amyotrophic lateral sclerosis. Neurobiol Aging. Elsevier; 2013;34(5):1517.e5-1517.e7.

233. van Es M a, Veldink JH, Saris CGJ, Blauw HM, van Vught PWJ, Birve A, et al. Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. Nat Genet. 2009;41(10):1083–7.

234. Lambert P. STPM2: Stata module to estimate flexible parametric survival models. BIOSENSORS & BIOELECTRONICS. 2012. p. 11–8.

235. Fisher DJ. Two-stage individual participant data meta-analysis and generalized forest plots. Stata J. College Station, TX: Stata Press; 2015;15(2):369–96.

236. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer; 2009.

237. Watanabe H, Atsuta N, Nakamura R, Hirakawa A, Watanabe H, Ito M, et al. Factors affecting longitudinal functional decline and survival in amyotrophic lateral sclerosis patients. Amyotroph Lateral Scler Front Degener. 2015;16(3–4):230–6.

238. Fox BW, Tibbetts RS. Neurodegeneration: Problems at the nuclear pore. Nature. 2015;525(7567):36–7.

239. Heiman-Patterson TD, Deitch JS, Blankenhorn EP, Erwin KL, Perreault MJ, Alexander BK, et al. Background and gender effects on survival in the TgN(SOD1-G93A)1Gur mouse model of ALS. J Neurol Sci. 2005;236(1–2):1–7.

240. Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, et al. Illustrating bias due to conditioning on a collider. Int J Epidemiol. 2010;39(2):417–20.

241. van Blitterswijk M, DeJesus-Hernandez M, Niemantsverdriet E, Murray ME, Heckman MG, Diehl NN, et al. Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): A cross-sectional cohort study. Lancet Neurol. 2013;12(10):978–88.

242. Gendron TF, van Blitterswijk M, Bieniek KF, Daughrity LM, Jiang J, Rush BK, et al. Cerebellar c9RAN proteins associate with clinical and neuropathological characteristics of C9ORF72 repeat expansion carriers. Acta Neuropathol. Springer Berlin Heidelberg; 2015;130(4):559–73.

243. Hardiman O, van den Berg LH, Kiernan MC. Clinical diagnosis and management of amyotrophic lateral sclerosis. Nat Rev Neurol. 2011;7(11):639–49.

244. The ALS CNTF Treatment Study (ACTS) Phase I-II Study Group. The Amyotrophic Lateral Sclerosis Functional Rating Scale. Arch Neurol. 1996 Feb 1;53(2):141.

245. Cedarbaum JM, Stambler N. Performance of the amyotrophic lateral sclerosis functional rating scale (ALSFRS) in multicenter clinical trials. J Neurol Sci. 1997;152(SUPPL. 1):1–9.

246. 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. J Neurol Sci. 1999;169(1–2):13–21.
247. Mandrioli J, Biguzzi S, Guidi C, Sette E, Terlizzi E, Ravasio A, et al. Heterogeneity in ALSFRS-R decline and survival: a population-based study in Italy. Neurol Sci. 2015;36(12):2243-52.
248. 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;8421:1–12.
249. Berry JD, Miller R, Moore DH, Cudkowicz ME, van den Berg LH, Kerr D a, et al. The Combined Assessment of Function and Survival (CAFS): a new endpoint for ALS clinical trials. Amyotroph Lateral Scler Frontotemporal Degener. 2013;14(3):162–8.
250. 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;66(2):265–7.
251. Labra J, Menon P, Byth K, Morrison S, Vucic S. Rate of disease progression: a prognostic biomarker in ALS. J Neurol Neurosurg Psychiatry. 2016 Jun;87(6):628-32.
252. Franchignoni F, Mandrioli J, Giordano A, Ferro S. A further Rasch study confirms that ALSFRS-R does not conform to fundamental measurement requirements. Amyotroph Lateral Scler Front Degener. 2015;8421:1–7.
253. Yoshida K, Bohn J. tableone: Create “Table 1” to Describe Baseline Characteristics [Internet]. 2015. Available from: https://cran.r-project.org/package=tableone
254. Pinheiro J, D B, S D, D S, (2016). RCT. nlme: Linear and Nonlinear Mixed Effects Models. 2016. Available from: http://cran.r-project.org/package=nlme
255. Therneau TM. A Package for Survival Analysis in S. 2015. Available from: http://cran.r-project.org/package=survival
256. Max Kuhn. Contributions from Jed Wing, Steve Weston, Andre Williams, *et al*. caret: Classification and Regression Training. 2016. Available from: https://cran.r-project.org/package=caret
257. Auguie B. gridExtra: Miscellaneous Functions for “Grid” Graphics [Internet]. 2015. Available from: https://cran.r-project.org/package=gridExtra
258. Frank E Harrell Jr. rms: Regression Modeling Strategies. R package version 4.4-1. 2015. Available from: https://cran.r-project.org/package=rms
259. Bivand R. classInt: Choose Univariate Class Intervals. R package version 0.1-23. 2015.
260. van der Kleij L a., Jones AR, Steen IN, Young C a., Shaw PJ, Shaw CE, et al. Regionality of disease progression predicts prognosis in amyotrophic lateral sclerosis. Amyotroph Lateral Scler Frontotemporal Degener. 2015;8421:1–6.
261. Kurian KM, Forbes RB, Colville S, Swingler RJ. Cause of death and clinical grading criteria in a cohort of amyotrophic lateral sclerosis cases undergoing autopsy from the Scottish Motor Neurone Disease Register. J Neurol Neurosurg Psychiatry. 2009;80(1):84–7.

262. Eisen A, Kiernan M, Mitsumoto H, Swash M. Amyotrophic lateral sclerosis: a long preclinical period? J Neurol Neurosurg Psychiatry. 2014;85(11):1232-8.
263. Ioannides ZA, Henderson RD, Robertson T, Davis M, McCombe PA. When does ALS start? A novel SOD-1 p.Gly142Arg mutation causing motor neurone disease with prominent premorbid cramps and spasms. J Neurol Neurosurg Psychiatry. 2016;87(9):1031–2.
264. Radunovic A, Annane D, Mk R, Brassington R, Mustfa N. Mechanical ventilation for amyotrophic lateral sclerosis / motor neuron disease. Cochrane Database Syst Rev. 2017 Oct 6;10:CD004427.
265. Sancho J, Martínez D, Bures E, Díaz JL, Ponz A, Servera E. Bulbar impairment score and survival of stable amyotrophic lateral sclerosis patients after noninvasive ventilation initiation. ERJ Open Res. 2018;4:159–2017.
266. Galvin M, Ryan P, Maguire S, Heverin M, Madden C, Vajda A, et al. The path to specialist multidisciplinary care in amyotrophic lateral sclerosis: A population-based study of consultations, interventions and costs. PLoS One. 2017;12(6):1–12.
267. van Eijk RPA, Nikolakopoulos S, Ferguson TA, Liu D, Eijkemans MJC, van den Berg LH. Increasing the efficiency of clinical trials in neurodegenerative disorders using group sequential trial designs. J Clin Epidemiol. 2018 Jun;98:80-88.
268. Marangi G, Traynor BJ. Genetic causes of amyotrophic lateral sclerosis: new genetic analysis methodologies entailing new opportunities and challenges. Brain Res. Elsevier; 2015 May;1607:75–93.
269. Pupillo E, Messina P, Logroscino G, Zoccolella S, Chiò A, Calvo A, et al. Trauma and amyotrophic lateral sclerosis: a case-control study from a population-based registry. Eur J Neurol. 2012 Dec;19(12):1509–17.
270. Logroscino G, Beghi E, Zoccolella S, Palagano R, Fraddosio a, Simone IL, et al. Incidence of amyotrophic lateral sclerosis in southern Italy: a population based study. J Neurol Neurosurg Psychiatry. 2005;76(8):1094–8.
271. Beghi E. Are professional soccer players at higher risk for ALS? Amyotroph Lateral Scler Frontotemporal Degener. 2013;14(7–8):501–6.
272. Beghi E, Millul A, Micheli A, Vitelli E, Logroscino G. Incidence of ALS in Lombardy, Italy. Neurology. 2007;68(2):141–5.
273. Beretta S, Carrì MT, Beghi E, Chiò A, Ferrarese C. The sinister side of Italian soccer. Lancet Neurol. 2003;2(11):656–7.
274. Pearce N. Analysis of matched case-control studies. BMJ. 2016 Feb 25;352:i969.
275. Bede P, Elamin M, Byrne S, Hardiman O. Sexual dimorphism in ALS: exploring gender-specific neuroimaging signatures. Amyotroph Lateral Scler Frontotemporal Degener. 2014;15(3–4):235–43.
276. Vivekananda U, Manjalay Z-R, Ganesalingam J, Simms J, Shaw CE, Leigh PN, et al. Low index-to-ring finger length ratio in sporadic ALS supports prenatally defined motor neuronal vulnerability. J Neurol Neurosurg Psychiatry. 2011;82(6):635–7.

277. Byrne S, Heverin M, Bede P, Elamin M, Hardiman O. Fecundity in ALS. Amyotroph Lateral Scler Frontotemporal Degener. 2014;15(3–4):204–6.
278. Popat RA, Van Den Eeden SK, Tanner CM, Bernstein AL, Bloch DA, Leimpeter A, et al. Effect of reproductive factors and postmenopausal hormone use on the risk of amyotrophic lateral sclerosis. Neuroepidemiology. 2006;27(3):117–21.
279. Kleijn MJJ De, Schouw YT Van Der, Verbeek ALM, Peeters PHM. Endogenous Estrogen Exposure and Cardiovascular Mortality Risk in Postmenopausal Women. 2002;155(4):339–45.
280. Van Buuren S, Groothuis-Oudshoorn K. Multivariate Imputation by Chained Equations. J Stat Softw. 2011;45(3):1–67.
281. Walker A. openxlsx: Read, Write and Edit XLSX Files. 2015.
282. Gohel D. ReporteRs: Microsoft Word, Microsoft PowerPoint and HTML Documents Generation. 2016.
283. Chió A, Meineri P, Tribolo A, Schiffer D. Risk factors in motor neuron disease: a case-control study. Neuroepidemiology. 1991;10(4):174–84.
284. Shao H, Breitner JCS, Whitmer RA, Wang J, Hayden K, Wengreen H, et al. Hormone therapy and Alzheimer disease dementia: New findings from the Cache County Study. Neurology. 2012;79(18):1846–52.
285. Rocca WA, Grossardt BR, Shuster LT. Oophorectomy, estrogen, and dementia: a 2014 update. Mol Cell Endocrinol. 2014 May 25;389(1–2):7–12.
286. Chan M-F, Dowsett M, Folkerd E, Wareham N, Luben R, Welch A, et al. Past oral contraceptive and hormone therapy use and endogenous hormone concentrations in postmenopausal women. Menopause. 2008;15(2):332–9.
287. Nakamizo T, Urushitani M, Inoue R, Shinohara a, Sawada H, Honda K, et al. Protection of cultured spinal motor neurons by estradiol. Neuroreport. 2000;11(16):3493–7.
288. Wise PM, Dubal DB, Wilson ME, Rau SW, Böttner M, Rosewell KL. Estradiol is a protective factor in the adult and aging brain: Understanding of mechanisms derived from in vivo and in vitro studies. Brain Res Rev. 2001;37(1–3):313–9.
289. Choi C Il, Lee YD, Gwag BJ, Cho SI, Kim SS, Suh-Kim H. Effects of estrogen on lifespan and motor functions in female hSOD1 G93A transgenic mice. J Neurol Sci. 2008;268(1–2):40–7.
290. Panet-Raymond V, Gottlieb B, Beitel LK, Pinsky L, Trifiro MA. Interactions between androgen and estrogen receptors and the effects on their transactivational properties. Mol Cell Endocrinol. 2000;167(1–2):139–50.
291. Koolschijn PCMP, Peper JS, Crone EA. The influence of sex steroids on structural brain maturation in adolescence. PLoS One. 2014;9(1):e83929.
292. McBride O, Morgan K, McGee H. Irish Contraception and Crisis Pregnancy Study 2010 (ICCP-2010). A Survey of the General Population. Crisis Pregnancy Programme Report No. 24. Dublin: Health Service Executive Crisis Pregnancy Programme (CPP), 2012. Available from: http://epubs.rcsi.ie/psycholrep/50/

293. Bezemer ID, Verhamme KMC, Gini R, Mosseveld M, Rijnbeek PR, Trifirò G, et al. Use of oral contraceptives in three European countries: A population-based multi-database study. Eur J Contracept Reprod Heal Care. 2016;21(1):81–7.
294. Spinelli A, Talamanca IF, Lauria L. Patterns of contraceptive use in 5 European countries. European Study Group on Infertility and Subfecundity. Am J Public Health. 2000;90(9):1403–8.
295. Desport JC, Preux PM, Truong TC, Vallat JM, Sautereau D, Couratier P. Nutritional status is a prognostic factor for survival in ALS patients. Neurology. 1999;53:1059–63.
296. Marin B, Desport JC, Kajeu P, Jesus P, Nicolaud B, Nicol M, et al. Alteration of nutritional status at diagnosis is a prognostic factor for survival of amyotrophic lateral sclerosis patients. J Neurol Neurosurg Psychiatry. 2011;82(6):628–34.
297. Paganoni S, Deng J, Jaffa M, Cudkowicz ME, Wills A-M. Body mass index, not dyslipidemia, is an independent predictor of survival in amyotrophic lateral sclerosis. Muscle Nerve. 2011 Jul;44(1):20–4.
298. Holm T, Maier A, Wicks P, Lang D, Linke P, Münch C, et al. Severe loss of appetite in amyotrophic lateral sclerosis patients: Online self-assessment study. J Med Internet Res. 2013;15(4):1–10.
299. Tefera TW, Borges K. Metabolic dysfunctions in amyotrophic lateral sclerosis pathogenesis and potential metabolic treatments. Front Neurosci. 2017;10:611.
300. Koppers M, Blokhuis AM, Westeneng HJ, Terpstra ML, Zundel CAC, Vieira De Sá R, et al. C9orf72 ablation in mice does not cause motor neuron degeneration or motor deficits. Ann Neurol. 2015;78(3):426–38.
301. Huang Y, Van Poppel F, Lumey LH. Differences in height by education among 371,105 Dutch military conscripts. Econ Hum Biol. 2015;17:202–7.
302. Longstreth WT, Nelson LM, Koepsell TD, van Belle G. Hypotheses to explain the association between vigorous physical activity and amyotrophic lateral sclerosis. Med Hypotheses. 1991;34(2):144–8.
303. Lacorte E, Ferrigno L, Leoncini E, Corbo M, Boccia S, Vanacore N. Physical activity, and physical activity related to sports, leisure and occupational activity as risk factors for ALS: A systematic review. Neurosci Biobehav Rev. 2016;66:61–79.
304. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Tudor-Locke C, et al. 2011 compendium of physical activities: A second update of codes and MET values. Med Sci Sports Exerc. 2011;43(8):1575–81.
305. Fang F, Hållmarker U, James S, Ingre C, Michaëlsson K, Ahlbom A, et al. Amyotrophic lateral sclerosis among cross-country skiers in Sweden. Eur J Epidemiol. 2016;31(3):247-53.
306. ILO. International Standard Classification of Occupations. 2010. Available from: http://www.ilo.org/public/english/bureau/stat/isco/index.htm
307. Peters S, Vermeulen R, Cassidy A, Mannetje A 'T, van Tongeren M, Boffetta P, et al. Comparison of exposure assessment methods for occupational carcinogens in a multi-centre lung cancer case-control study. Occup Environ Med. 2011;68:148–53.

308. Matheson MC, Benke G, Raven J, Sim MR, Kromhout H, Vermeulen R, et al. Biological dust exposure in the workplace is a risk factor for chronic obstructive pulmonary disease. Thorax. 2005;60(8):645–51.
309. Daniel B, Eva J, Huss A, Rischewski J, BD S. Parental Occupational Exposure to Benzene and Risk of Childhood Cancer: a Census-Based Cohort Study. Environ Int. 2017;108(July):84–91.
310. Koeman T, Slottje P, Schouten LJ, Peters S, Huss A, Veldink JH, et al. Occupational exposure and amyotrophic lateral sclerosis in a prospective cohort. Occup Environ Med. 2017;74(8):578-585.
311. Vinceti M, Filippini T, Mandrioli J, Violi F, Bargellini A, Weuve J, et al. Lead, cadmium and mercury in cerebrospinal fluid and risk of amyotrophic lateral sclerosis: A case-control study. J Trace Elem Med Biol. 2017;43:121-125.
312. Vergara X, Kheifets L, Greenland S, Oksuzyan S, Cho Y-S, Mezei G. Occupational exposure to extremely low-frequency magnetic fields and neurodegenerative disease: a meta-analysis. J Occup Environ Med. 2013;55(2):135–46.
313. Huss A, Spoerri A, Egger M, Kromhout H, Vermeulen R. Occupational exposure to magnetic fields and electric shocks and risk of ALS: The Swiss National Cohort. Amyotroph Lateral Scler Front Degener. 2015;16(1–2):80–5.
314. Seelen M, Vermeulen RCH, van Dillen LS, van der Kooi AJ, Huss A, de Visser M, et al. Residential exposure to extremely low frequency electromagnetic fields and the risk of ALS. Neurology. 2014 Nov 4;83(19):1767–9.
315. Tang Y, Donnelly KC, Tiffany-Castiglioni E, Mumtaz MM. Neurotoxicity of polycyclic aromatic hydrocarbons and simple chemical mixtures. J Toxicol Environ Health A. 2003;66(10):919–40.
316. Douwes J, Thorne P, Pearce N, Heederik D. Bioaerosol health effects and exposure assessment: Progress and prospects. Ann Occup Hyg. 2003;47(3):187–200.
317. Konstantinoudis G, Schuhmacher D, Rue H, Spycher B. Discrete versus continuous domain models for disease mapping. ArXiv (Preprint). 2018 Aug 14. Available from: http://arxiv.org/abs/1808.04765
318. Vieira VM, Hansen J, Gredal O, Weisskopf MG. Spatial analyses of ALS incidence in Denmark over three decades. Amyotroph Lateral Scler Front Degener. 2018;19(3–4):275–84.
319. Mora JS, Barbeito L, Hermine O. Masitinib as an add-on therapy to riluzole is beneficial in the treatment of amyotrophic lateral sclerosis (ALS) with acceptable tolerability: results from a randomized controlled phase 3 trial. J Neurol Sci. 2017;381(2017):183.
320. Gendron TF, Chew J, Stankowski JN, Hayes LR, Zhang Y-J, Prudencio M, et al. Poly(GP) proteins are a useful pharmacodynamic marker for C9ORF72-associated amyotrophic lateral sclerosis. Sci Transl Med. 2017;29;9(383)
321. Visser AE, Seelen M, Hulsbergen A, de Graaf J, van der Kooi AJ, Raaphorst J, et al. Exploring the fitness hypothesis in ALS: a population-based case-control study of parental cause of death and lifespan. J Neurol Neurosurg Psychiatry. 2017;88(7):550-556.
322. Bermon S, Garnier P-Y. Serum androgen levels and their relation to performance in track and

field: mass spectrometry results from 2127 observations in male and female elite athletes. Br J Sports Med. 2017 Sep;51(17):1309-1314.

323. Gage SH, Davey Smith G, Ware JJ, Flint J, Munafò MR. G = E: What GWAS Can Tell Us about the Environment. PLoS Genet. 2016;12(2):1–13.

324. Vander Weele TJ, Ko YA, Mukherjee B. Environmental confounding in gene-environment interaction studies. Am J Epidemiol. 2013;178(1):144–52.

Appendix A - Supplementary results tables for Chapter 3.2

Table A1: Poisson test results (P values) of observed vs expected by county for all cases, males only and females only

	All cases			Males only			Females Only		
County	**Observed**	**Expected**	**P value**	**Observed**	**Expected**	**P value**	**Observed**	**Expected**	**P value**
Carlow County	10	19.06	0.038	6	11.02	0.171	c	c	0.212
Dublin City	212	214.07	0.918	120	113.24	0.511	92	100.84	0.425
South Dublin	99	76.45	0.014	49	42.71	0.320	50	33.74	0.007
Fingal	88	64.37	0.005	48	36.41	0.067	40	27.96	0.029
Dun Laoghaire-Rathdown	92	88.02	0.670	46	46.56	1.000	46	41.46	0.484
Kildare County	58	54.77	0.636	39	31.67	0.183	19	23.10	0.466
Kilkenny County	18	36.52	*0.001	9	21.06	0.006	9	15.46	0.124
Laois County	19	25.90	0.201	13	15.15	0.699	6	10.76	0.169
Longford County	13	15.46	0.612	7	8.94	0.736	6	6.52	1.000
Louth County	57	42.00	0.025	28	23.44	0.351	29	18.56	0.020
Meath County	48	51.97	0.628	27	30.06	0.648	21	21.91	1.000
Offaly County	19	28.15	0.089	10	16.23	0.136	9	11.92	0.470
Westmeath County	33	30.38	0.586	13	17.30	0.398	20	13.08	0.070
Wexford County	52	54.12	0.838	36	30.92	0.367	16	23.20	0.146
Wicklow County	53	47.45	0.424	30	26.80	0.499	23	20.65	0.581
Clare County	31	46.76	0.019	17	27.11	0.054	14	19.65	0.257
Cork City	48	55.62	0.347	24	29.86	0.314	24	25.75	0.843
Cork County	168	140.31	0.023	86	80.21	0.503	82	60.10	0.007
Kerry County	58	66.48	0.326	30	38.25	0.196	28	28.23	1.000
Limerick City	25	24.81	0.920	14	13.45	0.785	11	11.36	1.000
Limerick County	47	49.19	0.831	29	28.35	0.851	18	20.84	0.660
North Tipperary	30	29.45	0.854	18	16.84	0.714	12	12.61	1.000
South Tipperary	35	37.14	0.805	20	21.34	0.914	15	15.80	1.000
Waterford City	17	18.63	0.817	7	10.22	0.431	10	8.40	0.492
Waterford County	21	27.17	0.290	14	15.72	0.800	7	11.45	0.235
Galway City	22	22.12	1.000	15	12.01	0.383	7	10.11	0.429
Galway County	71	68.62	0.763	50	40.50	0.135	21	28.13	0.219
Leitrim County	8	14.55	0.088	c	c	0.126	c	c	0.677
Mayo County	59	60.51	0.898	32	35.03	0.673	27	25.48	0.766
Roscommon County	37	29.23	0.164	24	17.09	0.113	13	12.14	0.773
Sligo County	32	28.22	0.451	19	16.10	0.453	13	12.13	0.773
Cavan County	24	27.92	0.569	15	16.51	0.806	9	11.40	0.655
Donegal County	58	64.74	0.455	35	37.59	0.744	23	27.15	0.501
Monaghan County	22	23.83	0.837	16	13.68	0.497	6	10.15	0.268

Table legend:

Statistically significant results (P < 0.0015) are designated by *

c – For observed case counts <5, both observed and expected cases are censored

Table A2: Poisson test results (P values) of observed vs expected by county for stratified time periods

	All cases 1995 - 1999			All cases 2000 - 2004			All cases 2005 - 2009			All cases 2010 - 2013		
County	**Observed**	**Expected**	**P value**	**Observed**	**Expected**	**P value**	**Observed**	**Expected**	**P value**	**Observed**	**Expected**	**P value**
Carlow County	c	c	0.057	c	c	0.641	5	5.10	1.000	c	c	0.073
Dublin City	43	48.32	0.517	61	54.00	0.340	39	54.69	0.036	59	49.65	0.178
South Dublin	14	13.68	0.892	27	18.50	0.061	29	20.54	0.076	26	21.37	0.328
Fingal	17	10.80	0.066	22	14.79	0.067	24	17.37	0.117	20	18.93	0.730
Dun Laoghaire-Rathdown	21	18.20	0.481	24	21.89	0.593	23	22.98	0.917	23	22.21	0.832
Kildare County	15	9.51	0.100	10	12.75	0.573	20	14.63	0.151	12	15.66	0.447
Kilkenny County	c	c	0.136	c	c	0.016	7	9.49	0.517	5	9.43	0.190
Laois County	c	c	0.825	c	c	0.543	7	6.72	0.846	c	c	0.437
Longford County	c	c	0.777	c	c	0.061	c	c	1.000	6	3.94	0.302
Louth County	12	8.34	0.220	12	10.08	0.526	12	11.01	0.761	20	11.00	0.014
Meath County	5	9.31	0.189	10	12.11	0.666	14	13.87	0.893	19	14.50	0.235
Offaly County	6	5.65	0.831	c	c	0.178	c	c	0.014	7	7.25	1.000
Westmeath County	8	5.97	0.406	7	7.35	1.000	6	7.97	0.597	12	7.91	0.151
Wexford County	11	9.99	0.750	20	13.00	0.069	7	14.45	0.048	10	14.67	0.293
Wicklow County	14	9.02	0.096	11	11.46	1.000	15	12.59	0.478	10	12.62	0.572
Clare County	c	c	0.012	6	11.59	0.106	11	12.31	0.886	11	11.92	1.000
Cork City	10	12.18	0.666	9	14.12	0.228	17	14.35	0.429	9	13.19	0.332
Cork County	33	27.47	0.292	43	33.88	0.121	49	36.84	0.048	41	36.88	0.459
Kerry County	10	13.51	0.414	15	16.55	0.806	16	17.64	0.812	16	16.48	1.000
Limerick City	7	5.55	0.518	5	6.25	0.840	8	6.40	0.431	c	c	0.675
Limerick County	11	9.56	0.625	10	11.95	0.771	14	12.90	0.676	12	13.00	1.000
North Tipperary	5	5.99	1.000	9	7.28	0.457	7	7.68	1.000	8	7.43	0.714
South Tipperary	c	c	0.271	10	9.15	0.739	11	9.72	0.629	7	9.33	0.620
Waterford City	5	3.76	0.435	c	c	0.641	c	c	1.000	c	c	1.000
Waterford County	5	5.35	1.000	6	6.63	1.000	6	7.20	0.851	c	c	0.342
Galway City	c	c	0.807	6	5.30	0.662	c	c	0.676	7	5.84	0.534
Galway County	16	13.90	0.503	20	16.70	0.391	19	17.80	0.722	16	17.64	0.812
Leitrim County	c	c	1.000	c	c	0.057	c	c	0.792	c	c	0.283
Mayo County	10	12.55	0.572	13	14.90	0.794	12	15.78	0.448	23	15.10	0.052
Roscommon County	5	6.17	0.840	c	c	0.344	12	7.58	0.140	11	7.26	0.187
Sligo County	7	5.91	0.539	6	6.87	1.000	9	7.35	0.461	8	7.08	0.704
Cavan County	8	5.75	0.297	c	c	0.178	c	c	0.346	7	7.08	1.000
Donegal County	16	13.09	0.404	18	15.74	0.527	17	16.83	0.903	c	c	*0.001
Monaghan County	c	c	0.503	9	5.82	0.205	c	c	0.543	6	5.96	0.838

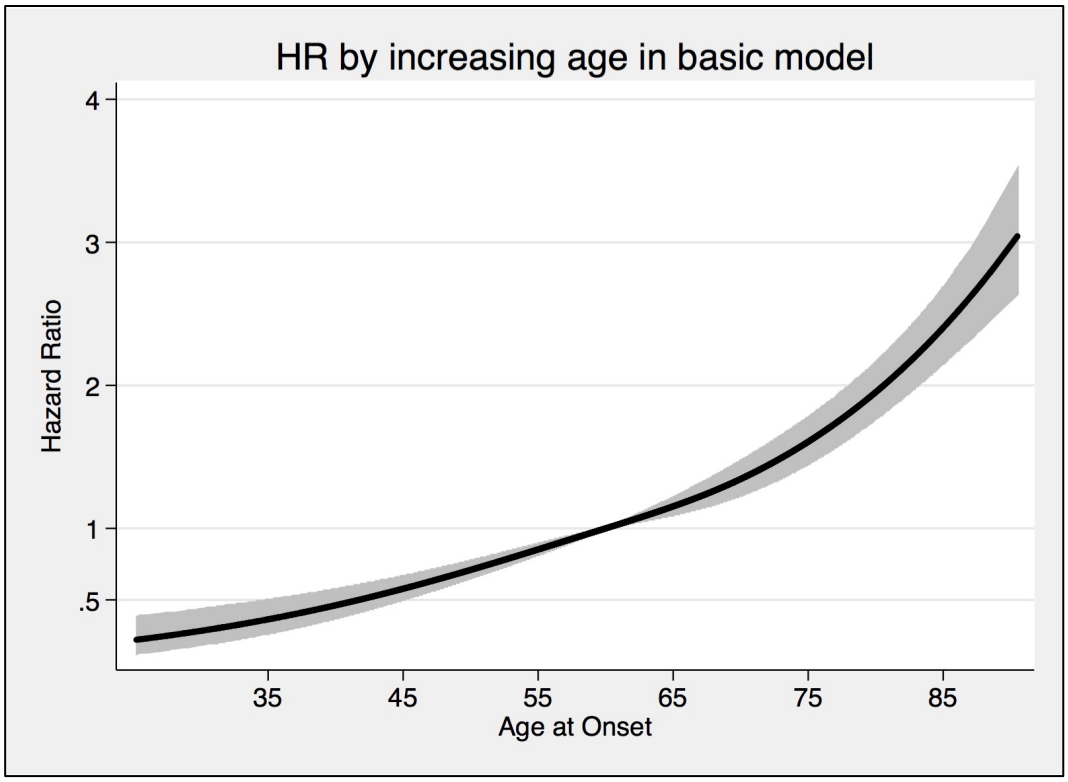
Table legend:

Statistically significant results (P < 0.0015) are designated by *

c – For observed case counts <5, both observed and expected cases are censored

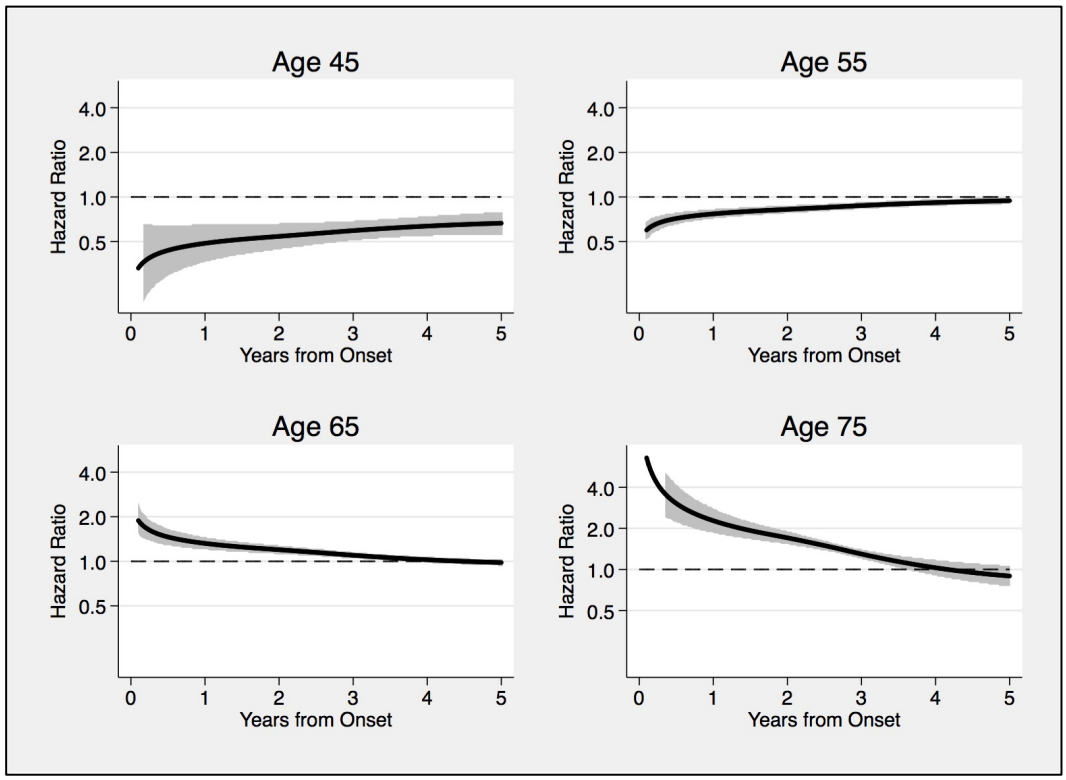
Appendix B - Base Royston-Parmar flexible parametric survival model used in Chapter 4.3

Figure B1: Hazard ratio for age as a continuous variable before addition of time varying component



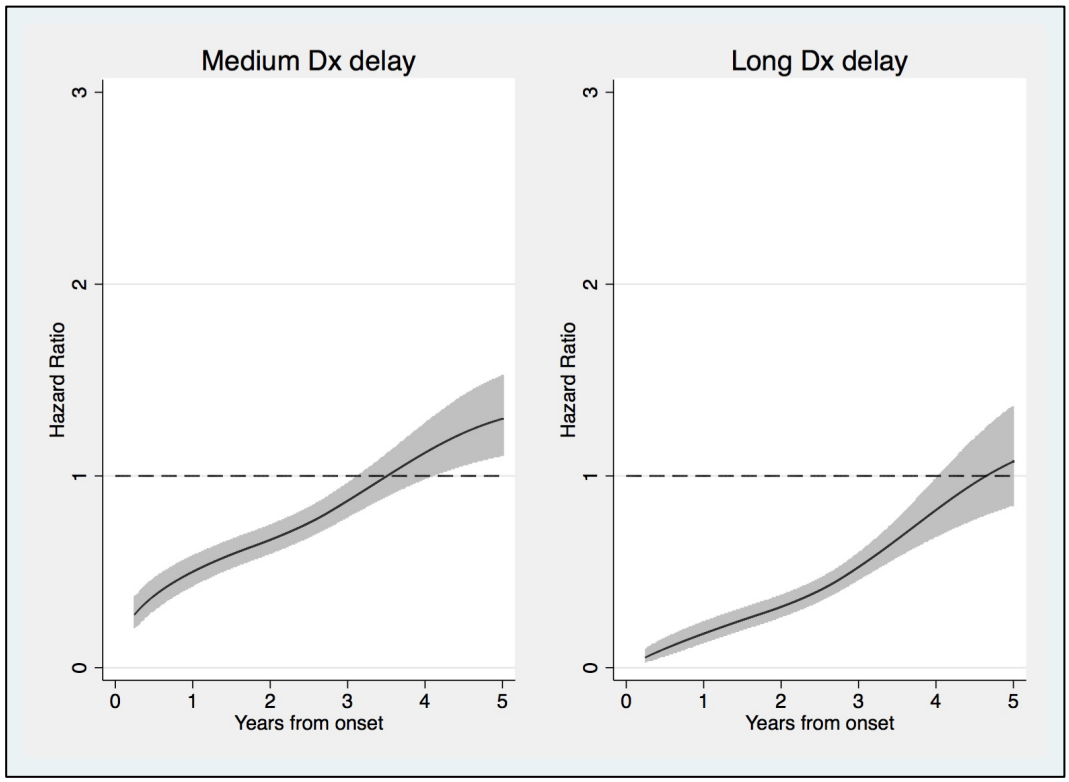
Legend: Hazard ratio for age as a continuous variable from a Royston-Parmar model on the hazard scale with 3.d.f. before the addition of time varying components. Model covariates included: site of onset, C9orf72 status, grouped diagnostic delay, country and robust variance-covariance matrix clustered by country was used.

Figure B2: Hazard ratio for age as a continuous variable after addition of a 1 d.f. time varying component to the age variable



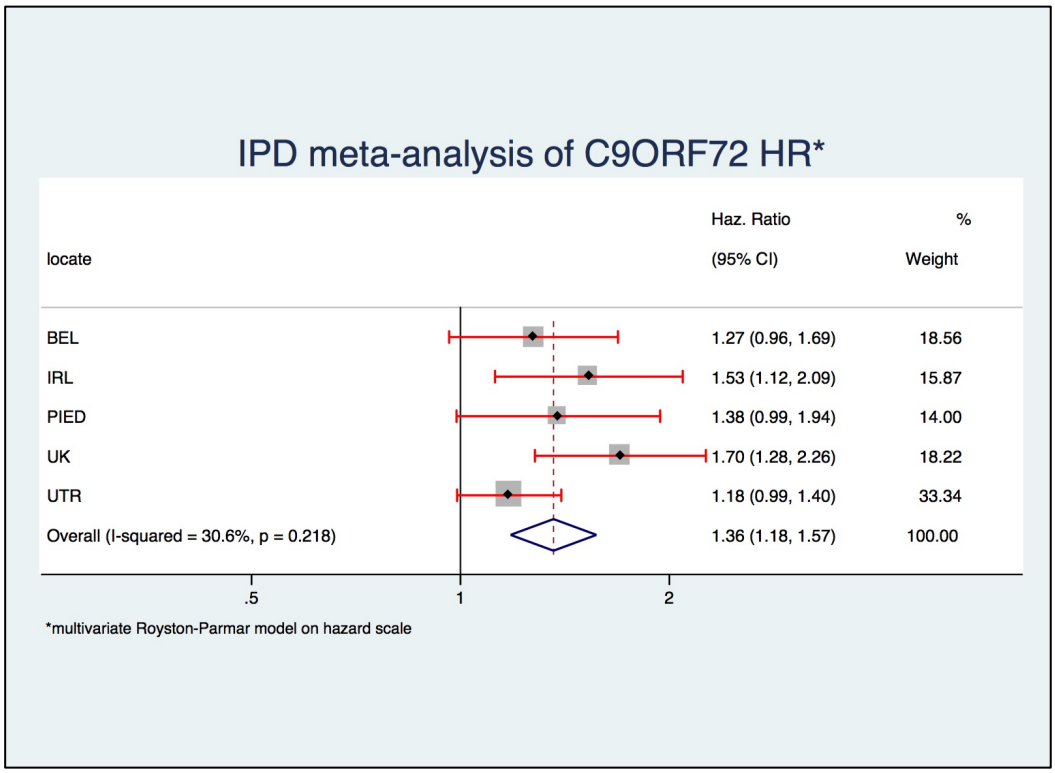
Legend: The graph shows time dependent hazard ratios for different ages of onset in a multivariate Royston Parmar flexible parametric model on the hazard scale with 3 degrees of freedom and variance-covariance matrix clustered by country. These graphs imply that the relative survival advantage or disadvantage for a given age dissipate over time – with only the 45 year old graph still showing a survival advantage after 5 years. Model covariates included site of onset, grouped diagnostic delay, country and C9orf72 status, and robust variance-covariance matrix clustered by country was used.

Figure B3: Hazard ratio for grouped diagnostic delay after addition of a 1 d.f. time varying component to the diagnostic delay variable



Legend: In addition to time varying age, the basic model included time varying diagnostic delay categories shown above. Three categories of diagnostic delay were defined: short, medium and long diagnostic delay. For each country the cohort was split into three groups defined by the 33rd and 67th centiles of diagnostic delay in that country, thus allowing for inter-country differences in diagnostic delay. Short diagnostic delay is the reference group to which the above graphs are compared, therefore they indicate that for medium delay patients the survival advantage they experience has dissipated after approx. 3.5 years, whilst for long delay patients their survival advantage has dissipated after approx. 4.5 years.

Figure B4: IPD meta-analysis of the hazard ratio for the C9orf72 expansion in the basic model



Legend: Individual patient data meta-analysis of hazard ratio for C9orf72 expanded compared to C9orf72 normal patients after multivariate Royston-Parmar regression. The model included, age and grouped diagnostic delay as time-varying covariates, site of onset, country and C9 status. The Stata *metan* command implements IPD meta-analysis as a two step process and we used a random effects model to pool the estimate across countries.

Table B1: Hazard ratios for time independent covariates included in the basic model.

Variable	Hazard Ratio (95% CI)	Wald Test
Spinal onset	1	-
Bulbar onset	1.35 (1.25 – 1.45)	<0.001
Country		
Belgium	1	-
Ireland	1.01 (0.99 – 1.03)	0.58
Italy	0.66 (0.64 – 0.69)	<0.001
The Netherlands	0.95 (0.94 – 0.96)	<0.001
United Kingdom	0.70 (0.68 – 0.71)	<0.001
C9orf72 Normal	1	-
C9orf72 Expanded	1.32 (1.14 – 1.54)	<0.001

The above hazard ratios are for the covariates included in the completed basic model alongside time dependent age at onset and diagnostic delay variables (shown in Fig.B2 & Fig.B3). Again robust errors were used. Note that the hazard ratio for C9orf72 generated using robust errors, 1.32 (95% CI: 1.14 – 1.54), is very similar to the estimate obtained via the alternate method IPD meta-analysis: 1.36 (95% CI: 1.18 – 1.57) (Fig.B4).

Appendix C - Supplementary analysis of collapsed ALSFRS-R scores for Chapter 4.4

Table C1 – Hazard ratios for the basic multivariate Cox PH model

Analysis of unadjusted ALSFRS score and sub-scores				
Basic multivariate model				
Model	Variable	Hazard ratio (95% CI)	P value (Wald test)	Model AIC
Base	Age at onset (per 10 years)	1.24 (1.11 – 1.39)	**<0.001**	
	Site of onset			
	Spinal	1		
	Bulbar	1.25 (0.97 – 1.62)	0.087	
	Diagnostic delay			
	< 8 months	1		
	8 – 15 months	0.38 (0.27 – 0.52)	**<0.001**	
	> 15 months	0.21 (0.10 – 0.41)	**<0.001**	
	El Escorial		0.559	
	Suspected	0.49 (0.12 – 2.04)		
	Possible	1		
	Probable	1.28 (0.87 – 1.90)	0.216	
	Definite	1.50 (1.06 – 2.11)	**0.021**	2812.2
Single variable added to the basic multivariate model				
Model	Variable	Hazard ratio (95% CI)	P value (Wald test)	Model AIC
1	Total ALSFRS-R slope	4.2 (3.5 – 5.0)	**<0.001**	2636.2
2	ALSFRS-R bulbar slope	11.7 (7.2 – 19.3)	**<0.001**	2747.8

3	ALSFRS-R motor slope	5.2 (4.1 – 6.5)	**<0.001**	2673.8
4	ALSFRS-R respiratory slope	15.2 (7.4 – 31.2)	**<0.001**	2774.5
Multiple variables added to the basic multivariate model				
Model	Variable	Hazard ratio (95% CI)	P value (Wald test)	Model AIC
5	ALSFRS-R bulbar slope	9.3 (5.2 – 16.8)	**<0.001**	
	ALSFRS-R motor slope	4.5 (3.5 – 5.8)	**<0.001**	
	ALSFRS-R respiratory slope	1.6 (0.7 – 3.6)	0.250	2812.21
6	ALSFRS-R bulbar slope	9.9 (5.6 – 17.6)	**<0.001**	
	ALSFRS-R motor slope	4.7 (3.7 – 6.0)	**<0.001**	**2628.8**

Definition of collapsed ALSFRS-R scores

The individual ALSFRS-R questions each allow a 5 level response scoring from 0 – 4. However, Rasch analysis by Franchignoni *et al.* revealed redundancy in these levels and suggested collapsing scores from five to three levels scoring from 0 - 2. The following table shows how individual ALSFRS responses were recoded to collapsed scores as per Franchignoni *et al.* Note that maximum total collapsed ALSFRS score is 24, and maximum collapsed sub-scores are: motor =12, bulbar =6, resp =6.

Q.1 – Q.11 of ALSFRS		Q.12 of ALSFRS	
Original score	Collapsed score	Original score	Collapsed score
0	0	0	0
1	0	1	1
2	1	2	1
3	1	3	1
4	2	4	2

For clarity our main analysis was completed using raw ALSFRS-R scores, however we present below the same analysis using the collapsed ALSFRS-R scores as suggested by Franchignoni *et al*.[110]

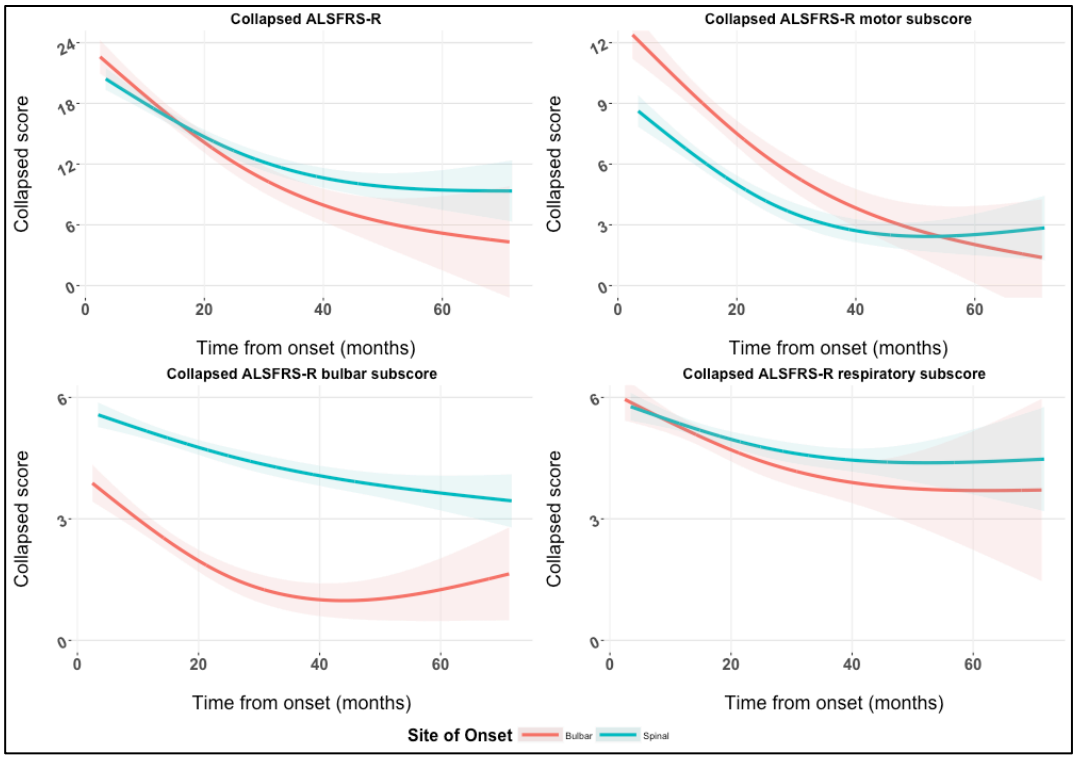
Table C2 – Hazard ratios for the basic multivariate Cox proportional hazards model using collapsed ALSFRS-R scores and sub-scores

Analysis of collapsed ALSFRS score and sub-scores				
Single variable added to the basic multivariate model				
Model	Variable	Hazard ratio (95% CI)	P value (Wald test)	Model AIC
1	Total ALSFRS-R slope	12.6 (9.1 – 17.6)	**<0.001**	2621.1
2	ALSFRS-R bulbar slope	30.1 (15.3 – 59.4)	**<0.001**	2752.1
3	ALSFRS-R motor slope	16.1 (10.9 – 23.6)	**<0.001**	2667.6
4	ALSFRS-R respiratory slope	125.7 (36.7 – 430.3)	**<0.001**	2769.8
Multiple variables added to the basic multivariate model				
Model	Variable	Hazard ratio (95% CI)	P value (Wald test)	Model AIC
5	ALSFRS-R bulbar slope	19.7 (8.3 – 46.7)	**<0.001**	
	ALSFRS-R motor slope	12.3 (7.9 – 19.0)	**<0.001**	
	ALSFRS-R respiratory slope	8.2 (2.2 – 30.3)	**0.001**	**2623.7**
6	ALSFRS-R bulbar slope	20.4 (8.9 – 46.5)	**<0.001**	
	ALSFRS-R motor slope	14.2 (9.4 – 21.5)	**<0.001**	2630.7

Note that the very large hazard ratios relative to the un-collapsed scores are driven by the reduced maximum score and hence reduced slopes of collapsed scores, therefore it is more informative to compare the AIC scores of collapsed vs un-collapsed models. In

general, collapsed scores have lower AIC than un-collapsed scores, indicating a better model fit. Furthermore, in model 5 for the collapsed scores the respiratory sub-score remains a significant prognostic indicator.

Figure C1 – Longitudinal trend in collapsed ALSFRS scores over follow-up time



The longitudinal trend of collapsed total ALSFRS-R and collapsed ALSFRS-R sub-scores does not differ from that of un-collapsed scores shown in chapter 4.4, and leads to the same conclusions (the bulbar and motor sub-scores display differences by site of onset, whilst the collapsed respiratory sub-score does not meaningfully distinguish between the two).

Appendix D - Euro-MOTOR participant questionnaire

Euro-MOTOR Questionnaire Version 1.00 May 19 2011

EURONUMBER ID

1. DEMOGRAPHICS AND EDUCATION

Person answering questions is:

Patient/control	Proxy of patient/control
☐	☐

Sex:

Male	Female
☐	☐

Date of Birth: ___ ___/___ ___/___ ___ ___ ___

M M / D D / Y Y Y Y

Date of Survey: ___ ___/___ ___/___ ___ ___ ___

M M / D D / Y Y Y Y

Highest degree attained:

☐ 1=>Grade school, (grades 1-8)

☐ 2=>High school diploma

☐ 3=>Technical or trade school diploma

☐ 4=>University degree

☐ 5=>Graduate school (PhD)

☐ 6=>None

2. BIOMETRICS

Current weight (kg) (self reported, whole numbers): _____

Current height (cm) (self reported, whole numbers): _____

Waist circumference (cm) (measured at the level of the navel, whole numbers): _____

Apart from when you were young, have you ever been more than 5 kilo's heavier or lighter than your current weight?

Yes ☐ No ☐

What did you *approximately* weigh at the following ages?

a. 20 years: _____kg

b. 30 years: _____kg

c. 40 years: _____kg

d. 50 years: _____kg

e. 60 years: _____kg

f. 70 years: _____kg

BMI (automatically calculated): _____

3. ANCESTRY

What is your country of origin: _____

What is your province/region/county of origin: _____

What is country of origin of biological father: _____

What is country of origin of biological mother: _____

What is country of origin of grandfather (paternal) : _____ & county:

_____

What is country of origin of grandmother (paternal) : _____ & county:

_____

What is country of origin of grandfather (maternal) : _____ & county:

_____

What is country of origin of grandmother (maternal) : _____ & county:

_____

4. RESIDENTIAL HISTORY

What is your current address: *will be filled in not stored and converted to geocode*

GEOCODE (provided by database): _____

Did you always live here:

Yes ☐ No ☐

If no, from _____ (age) to _____ (age)

	Home address until current	Address (*will be filled in not stored and converted to geocode*)	From (age)	To (age)
A	First			
B	Second			
C	Third			
D	Fourth			
E	Fifth			
F	Sixth			
G	Seventh			

5. SMOKING

Cigarettes /Tobacco

Have you ever smoked?:

Yes ☐ No ☐

Year started: _____

Stopped smoking? Yes ☐ No ☐

If yes, the year you stopped: _____

How many cigarettes do/did you smoke on average per day: _____

Has there ever been a period when you smoked more than you do now?

Period of smoking more:

Yes ☐ No ☐

Period 1

How many cigarettes do/did you smoke on average per day at that time?: _____

When was this: _____ year -_____ year

Period 2

How many cigarettes do/did you smoke on average per day at that time?: _____

When was this: _____ year -_____ year

Has there ever been a period when you smoked less or not at all (during the time that you were smoking cigarettes)?

Period of less smoking:

Yes ☐ No ☐

Period 1

How many cigarettes do/did you smoke on average per day at that time?: _____

When was this: _____ year -_____ year

Period 2

How many cigarettes do/did you smoke on average per day at that time?: _____

When was this: _____ year -_____ year

Do/did you (also) smoke products other than cigarettes/ tobacco?

Smoking other product:

Yes ☐ No ☐

Cigars:

Smoking cigars:

Yes ☐ No ☐

If yes, how many per day: _____

Year started: _____

If stopped, year stopped: _____

Pipe:

Smoking pipes:

Yes ☐ No ☐

If yes, how many per day: _____

Year started: _____

If stopped, year stopped: _____

6. ALCOHOL

Do you sometimes drink alcohol or have you ever done so?:

Yes ☐ No ☐

Year started: _____

Stopped:

Yes ☐ No ☐

If yes, year: _____

Has there been a period when you did not drink alcohol?

Yes ☐ No ☐

If yes, for how many years: _____

How many glasses of alcohol do you or did you drink on average per week?: _____

How many of these are glasses of red wine: _____

7. HORMONES (only women)

Menstruation/pregnancy

At what age did you have your first period: _____

How regular were your periods when you were about 25 years of age (do not include periods of using the pill, hormone-containing coils, pregnancies and breast-feeding):

☐ 1=>every 24 days or less
☐ 2=>every 25 or 26 days
☐ 3=>every 27, 28 or 29 days
☐ 4=>every 30 or 31 days
☐ 5=>every 32 or more
☐ 5=>irregular
☐ 6=>I no longer know

Has there ever been a time when your cycle was irregular?:

Yes ☐ No ☐ Cannot remember ☐

If yes, when: _____year -_____year

Has there ever been a period when you stopped menstruating for more than a year (with the exception of pregnancies)?

Yes ☐ No ☐ Cannot remember ☐

Have your periods stopped for good?:

Yes ☐ No ☐

If yes, at what age: _____ (age)

How often have you been pregnant?: _____

How many live births have you had?: _____

Breastfed your children:

Yes ☐ No ☐

How many children have you breastfed: _____

Per child, for how many months have you breastfed them:

Child 1: _____

Child 2: _____

Child 3: _____

Child 4: _____

Are you currently using hormonal contraceptives or have you ever done so?

Yes ☐ No ☐

If yes, what form:

☐ 1=>Pill

☐ 2=>Subcutaneous implant

☐ 3=>Injection

☐ 4=>Other _____

In case of pill:

Have you stopped taking the pill:

Yes ☐ No ☐

If yes, at what age: _____(age)

How old were you when you started taking the pill: _____ (age)

How long have you been using the pill?

☐ 1=>never

☐ 2=>less than 1 year

☐ 3=>1-4 years

☐ 4=>5-9 years

☐ 5=>10-14 years

☐ 6=>15-19 years

☐ 7=>20 years or more

What are the name(s) of the pill:

Are you using or have you ever used hormone replacement therapy?:

Yes ☐ No ☐

If yes, how old when started: _____(age)

For how many years: _____(number of years)

What form?

☐ 1=>Pill

☐ 2=>Estrogen plaster

☐ 3=>Subcutaneous implant

☐ 4=>Injection

☐ 5=>Cream

Name(s) of these hormones:

Have you stopped ?:

Yes ☐ No ☐

If yes, at what age: _____(age)

OPERATIONS

Have you had a hysterectomy?:

Yes ☐ No ☐

If yes, at what age: _____(age)

Were your ovaries removed? :

☐ 1=>No

☐ 2=>Yes, one site

☐ 3=>Yes, both

☐ 4=>I don't know

If yes, at what age? _____ (age)

8. MEDICAL HISTORY

Have you ever been diagnosed with diabetes?:

Yes ☐ No ☐

In which year was this diagnosis made? _____ (year)

Do you have raised cholesterol?

Yes ☐ No ☐

In which year was this first established: _____ (year)

Are you NOW using medications for raised cholesterol?:

Yes ☐ No ☐

If yes, what is the name and which year did you start?:

Name:	Year:
Name:	Year:
Name:	Year:

Have you ever used any other medication for raised cholesterol:

Yes ☐ No ☐

Name:	Year started:	Year stopped:
Name:	Year started:	Year stopped:
Name:	Year started:	Year stopped:

Have you ever had high blood pressure? (except during pregnancy):

If yes, in which year was this first found: _____ (year)

Are you CURRENTLY using medications for high blood pressure?:

Yes ☐ No ☐

If yes, what is the name and which year did you start?

Name:	Year started:	Year stopped:
Name:	Year started:	Year stopped:
Name:	Year started:	Year stopped:

Have you ever used other medications for high blood pressure?

Yes ☐ No ☐

If yes, what is the name and which year did you start?

Name:	Year started:	Year stopped:
Name:	Year started:	Year stopped:
Name:	Year started:	Year stopped:

Have you ever had heart problems

Yes ☐ No ☐

If yes, did you visit a GP/doctor or hospital with these heart problems?:

Yes ☐ No ☐

Have you ever had angina pectoris:

Yes ☐ No ☐

Have you ever had a heart attack:

Yes ☐ No ☐

If yes, when? _____(year)

Were you ever told you have narrowing of one or both carorid arteries?

Yes ☐ No ☐

Did you ever have a "TIA"

Yes ☐ No ☐

Did you ever have a stroke:

Yes ☐ No ☐

Which year did you have the (first) stroke? _____(Year)

Have you ever undergone one of the following procedures:

Heart bypass:

Yes ☐ No ☐

If yes, which year: _____(Year)

Bypass operation in the legs:

Yes ☐ No ☐

If yes, which year: _____(Year)

Balloon catheter dilatation (angioplasty) in the legs:

Yes ☐ No ☐

If yes, which year: _____(Year)

Balloon catheter dilatation (angioplasty) in the heart:

Yes ☐ No ☐

If yes, which year: _____(Year)

Have you ever had cancer?

Yes ☐ No ☐

If yes, what type? _____ (type)

If yes, which year? _____ (Year)

What kind of treatment did you receive (more that one answer possible):

☐ 1=> Radiation therapy

☐ 2=> Surgery

☐ 3=> Chemotherapy

☐ 4=> Other, namely: _____

Did you have all vaccinations as a child according to the vaccination programme of your country:

Yes ☐ No ☐ I don't know ☐

9. OCCUPATIONS

Have you done military service

Yes ☐ No ☐

If so, in which of the armed forces did you serve?

_____

Were you ever deployed?

Yes ☐ No ☐

If so, to where?

Have you been in paid or unpaid employment ?

Yes ☐ No ☐

	Occupation	Employer/ Company	City/ Village	Can you indicate a few activities related to that occupation?	Hours per week	From (Year)	To (Year)
A							
B							
C							
D							
E							
F							
G							

10. PHYSICAL ACTIVITIES

Did you ever play sport when you were young (before your 18th birthday)?

Yes ☐ No ☐

Do you/Did you play sport as an adult?

Yes ☐ No ☐

What is/was your sport (when you were young) and when were you active

	Sport	Hours per week	START (year)	STOP (year)
A				
B				
C				
D				
E				
F				
G				

In addition to the sports you may have mentioned, do/did you have any hobbies?

Yes ☐ No ☐

(If No, proceed with next question)

	Hobby	Hours per week	START (year)	STOP (year)
A				
B				
C				
D				
E				
F				
G				

Have you ever engaged in sport that required great physical effort, for instance running a marathon?

Yes ☐ No ☐

(If No, proceed with next question)

	STRENUOUS PHYSICAL EXERTION	When? (years)
1		
2		
3		
4		

11. TRAUMA / INJURY
Trauma/ Letsel

Have you ever had any injury requiring medical care?

Yes No

If yes, please fill in the table below:

	Injury type 1=>Head injury with concussion 2=>Fracture 3=>Contusion 4=>Sprain 5=>Strain 6=>Other, namely:	At what age did injury occur?	Circumstances? 1=>Work 2=>Sport 3=>Leisure (other than sport) 4=>Traffic 5=> Other, namely:	Did injury cause disability 1=>Yes 2=>No	Injury was: 1=>Temporary 2=>Permanent	Where was Injury? 1=>Head 2=>Arms 3=>Chest 4=>Abdomen 5=>Legs 6=>Spine	Severity of injury? 1=>Mild 2=>Moderate 3=>Severe
A							
B							
C							
D							
E							
F							
G							

12. USE OF DRUGS/SUBSTANCES

Gebruik van medicijnen/ middelen

DRUGS IN SPORT

Have you ever used any of the following drugs, and if so, please indicate the age when you started and stopped:

Oral

Name of the drug:	Used: Yes or No	Age started:	Age stopped:
Creatine			
Anabolic Androgenic Steroids			
Clenbuterol, tibolone, zeranol, zilpaterol			
Amphetamines			
Adrenaline			
Heroin, fentanyl hydromorphone/ Hydromorfine, methadone, morphine, oxycodone, oxymorphone/ oxymorfine, pentazocine, pethidine			

Intramuscular performance enhancing agents?

Name of the drug:	Used: Yes or No	Age started:	Age stopped:
Erythropoietin (EPO), dEPO, CERA or hematide			
Chorionic Gonadotrophin (CG)			
Luteinizing Hormone (LH)			
Growth Hormone (GH)			
Insulin-like Growth Factor-1 (IGF-1)			
Mechano Growth Factors (MGFs)			
Platelet-Derived Growth Factor (PDGF)			
Fibroblast Growth Factors (FGFs)			
Vascular-Endothelial Growth Factor (VEGF)			
Hepatocyte Growth Factor (HGF)			

ANTIDEPRESSANTS AND ANTIPSYCHOTICS

Have you ever been prescribed any of the following drugs and if so, please indicate the age when you started and stopped

Anti-Anxiety/ Anti Depressants

Name of the drug:	**Used: Yes or No**	**Age started:**	**Age stopped:**
Diazepam (Valium)			
Duloxetine (Cymbalata)			
Venlafaxine (Efexor)			
Escitalopram (Lexapro)			
Sertraline (Lustral)			
Fluoxetine (Prozac)			
Citalopram (Cipramil)			
Sodium Valproate (Epilim)			
Lamotrigine (Lamictal)			
Lofepramine (Gamanil)			
Mirtazepine (Zispin)			
Trazodone			
Paroxetine (Seroxat)			
Lithium (Priadel)			
Dothiepin (Prothiaden)			
Trimipramine (Surmontil)			
Buproprion			
OTHER			

Anti Psychotics

Name of the drug:	Used: Yes or No	Age started:	Age stopped:
Trifluoperazine (Stelazine)			
Arpiprazole (Abilify)			
Chlorpromazine (Largactil)			
Clozapine (Clozaril)			
Flupenthizol (Depixol)			
Sulpiride (Dolmatil)			
Ziprasidone (Geodon)			
Haloperidol (Haldol, Serenase)			
Fluphenazine			
Risperidone			
Quetiapine (Seroquel)			
Olanzepine (Zyprexa)			
Thioridazine (Melleril)			
OTHER			

Have you been to a GP or hospital doctor for anything else, not mentioned above?

Reason	Hospital admission	Year
	○Yes ○No	
	○Yes ○No	
	○Yes ○No	
	○Yes ○No	
	○Yes ○No	
	○Yes ○No	

Below are questions about your family history The diseases we are interested in are across the top and the family members are along the side.(*Please note, the questions only relate to your direct family and not to relations through marriage (in-laws).)*

- If a relative has (had) ALS, Parkinson's disease or dementia, you can indicate this by colouring the "YES" circle black beside the relevant family member.
- If a **female** relative has had a heart attack or a stroke before her **65**th birthday, you can indicate this by colouring the "YES" circle black beside the relevant family member.
- If a **male** relative has had a heart attack or a stroke before his 5**5**th birthday, you can indicate this by colouring the "YES" circle black beside the relevant family member.
- If a relative has not had the disease, you can indicate this by colouring the "NO" circle black beside the relevant family member.
- If you are not sure whether a relative has (had) the disease, you can indicate this by colouring the "?" circle black beside the relevant family member.

If you do not have a particular relative (they are listed in case you DO have them), leave the circles empty.

The following questions are about your father, your mother, grandfather and grandmother on father's (F) side and grandfather and grandmother on mother's side (M).

13. FAMILY HISTORY

Parents

What is your **father's** date of birth? _____

If applicable, age at death? _____ (age in years)

What was the cause of death? _____

What is your **mother's** date of birth? _____

If applicable, age at death? (age in years)

What was the cause of death? _____

Siblings

How many **brothers** do/did you have? _____

What are the dates of birth of your brothers?

If you only know the year, fill in: 01-01-year of birth.

If applicable, please also fill in age at death and cause of death.

Brother	Date of birth (dd-mm-yyyy)	Age at death (in years)	Cause of Death
1			
2			
3			
4			
5			
6			
7			
8			

How many **sisters** do/did you have? _____

What are the dates of birth of your sisters?

If you only know the year, fill in: 01-01-year of birth.

If applicable, please also fill in age at death and cause of death.

Sister	Date of birth	Age at death (in years)	Cause of Death
1			
2			
3			
4			
5			
6			
7			
8			

Twins

Do you have a twin brother or – sister?

Yes ☐ No ☐

If yes, what type of twin are you?

☐ 1=> Identical

☐ 2=> Non-identical

☐ 3=> Unknown

If unknown:

When you were children, did you and your twin brother or –sister look identical or was there only the usual family resemblance?

☐ 1=> Identical

☐ 2=> Normal family resemblance

When you were children, did your parents/brothers/sisters/ teachers have trouble telling you apart?

Yes ☐ No ☐

What is the gender of your twin?

☐ 1=> Male

☐ 2=>Female

Which disorder(s) has/have been found in your twin brother or –sister? (more than one answer possible)

☐ 1=> ALS

☐ 2=> Polyneuropathy

☐ 3=> PLS

☐ 4=> PSMA

☐ 5=> Parkinson disease

☐ 6=> Dementia

☐ 7=> Other, namely ____

☐ 8=> None of the above

Has you twin brother or –sister died?

Yes ☐ No ☐

(If applicable) When did he/she die? __/__/____(dd mm yy)

What the cause of death? _____

Uncles and Aunts

How many **brothers** does/did your **father** have? ____

In which years were your father's brothers born?

If applicable: age at death?

Brother of father	Year of birth	Age at death (in years)
1		
2		
3		
4		
5		
6		
7		
8		

How many **sisters** does/did your **father** have? ____

When were your father's sisters born?

(If applicable): age at death**?**

Sister of father	Year of birth	Age at death (in years)
1		
2		
3		
4		
5		
6		
7		
8		

How many **brothers** does/did your **mother** have? ____

When were your mother's brothers born?

If applicable: age at death?

Brother of mother	Year of birth	Age at death (in years)
1		
2		
3		
4		
5		
6		
7		
8		

How many **sisters** does/did your **mother** have? ____

When were your mother's sisters born?

If applicable: age at death?

Sister of mother	Year of birth	Age at death (in years)
1		
2		
3		
4		
5		
6		
7		
8		

Parents and grandparents

	ALS	PLS	PSMA	Poly neuro pathy	Parkinson's disease	Dementia	Stroke, brain infarction, brain haemorrhage	Depres-Sion	Alcoholism	Suicide	Heart attack
Father	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?
Mother	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?
Grand father (F)	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?
Grand mother (F)	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?
Grand father (M)	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?
Grand mother (M)	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?

Brothers

	ALS	PLS	PSMA	Poly neuro pathy	Parkinson's disease	Dementia	Stroke, brain infarction, brain haemorrhage	Depres-Sion	Alcoholism	Suicide	Heart attack
Brother 1	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?
Brother 2	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?
Brother 3	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?
Brother 4	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?
Brother 5	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?
Brother 6	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?

Sisters

	ALS	PLS	PSMA	Poly neuro pathy	Parkinson's disease	Dementia	Stroke, brain infarction, brain haemorrhage	Depres-Sion	Alcoholism	Suicide	Heart attack
Sister 1	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?
Sister 2	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?
Sister 3	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?
Sister 4	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?
Sister 5	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?
Sister 6	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?	○No ○Yes ○?

How many of these family members suffer or suffered from the following conditions:

	TOTAL NUMBER of family members you have/had:	ALS	PLS	PSMA	Polyneuopathy	Parkinson's disease	Dementia	Stroke, brain infarction, brain haemorrhage	Depression	Alcoholism	Suicide	Heart attack
Father's brothers (uncles)												
Father's sisters (aunts)												
Male cousins from father's side												
Female cousins from father's side												
Mother's brothers (uncles)												
Mother's sisters (aunts)												
Male cousins from mother's side												
Female cousins from mother's												

Appendix E – Supplementary results for Chapter 5.2

Table E1 - Clinical characteristics of female Euro-MOTOR patients by study site

Characteristic	Category	Apulia (N = 60)	Lombardy (N = 86)	Piedmont & Valle d'Aosta (N = 124)	Ireland (N = 71)	The Netherlands (N = 312)
Site of onset	Spinal (%)	35 (62.5)	49 (57.0)	58 (46.8)	52 (73.2)	172 (56.2)
	Bulbar (%)	20 (35.7)	36 (41.9)	65 (52.4)	18 (25.4)	131 (42.8)
	Thoracic (%)	1 (1.8)	1 (1.2)	1 (0.8)	1 (1.4)	3 (1.0)
El Escorial						
Category	Definite (%)	21 (36.2)	22 (25.9)	35 (28.2)	43 (65.2)	73 (24.0)
	Probable (%)	18 (31.0)	32 (37.6)	39 (31.5)	12 (18.2)	122 (40.1)
	Probable LS (%)	0 (0)	12 (14.1)	23 (18.5)	0 (0.0)	66 (21.7)
	Possible (%)	19 (32.8)	19 (22.4)	27 (21.8)	11 (16.7)	43 (14.1)
C9orf72 (%)						
	Normal (%)	0 (0)	33 (38.4)	112 (90.3)	63 (88.7)	256 (82.1)
	Expanded (%)	1 (1.7)	2 (2.3)	12 (9.7)	7 (9.9)	30 (9.6)
	Not tested (%)	59 (98.3)	51 (59.3)	0 (0)	1 (1.4)	26 (8.3)

Appendix F - Exposures of Euro-MOTOR cases and controls by study site

Apulia	Controls (n = 213)	Cases (n = 141)	
Exposure	Exposed, n (%)	Exposed, n (%)	P value (Fisher)
DOM-JEM			
Asbestos	23 (10.8)	18 (14.0)	0.396
Chromium	25 (11.8)	21 (16.2)	0.258
DME	40 (18.9)	28 (21.9)	0.576
Nickel	8 (3.8)	12 (9.2)	0.056
PAH	8 (3.8)	17 (13.1)	0.002
Silica	40 (18.9)	30 (23.6)	0.332
Animal	9 (4.2)	8 (6.1)	0.451
Biodust	57 (27.0)	44 (36.4)	0.083
Endotoxin	36 (17.1)	39 (31.5)	0.003
ALOHA-JEM			
Biodust	65 (31.0)	47 (39.2)	0.147
Minedust	76 (36.2)	55 (45.1)	0.130
Gasfumes	80 (38.1)	62 (51.2)	0.022
VGDF	104 (49.8)	70 (59.3)	0.107
Metals	19 (9.0)	23 (17.6)	0.027
Pesticides (all)	25 (11.8)	25 (19.5)	0.058
Pesticides - herbicides	25 (11.8)	20 (15.6)	0.325
Pesticides - insecticides	25 (11.8)	21 (16.4)	0.253
Pesticides - Fungicides	25 (11.8)	24 (18.8)	0.082
Solvents - aromatic	47 (22.3)	40 (32.0)	0.054
Solvents - chlorinated	20 (9.4)	22 (17.1)	0.042
Solvents - other	30 (14.2)	31 (16.4)	0.639
BEN-JEM (Benzene)	24 (11.4)	30 (22.9)	0.006
ELF-MF	48 (22.7)	43 (34.1)	0.031
Shock-JEM (shock risk)	56 (27.1)	40 (33.3)	0.257

Lombardy	Controls (n = 190)	Cases (n = 186)	
Exposure	Exposed, n (%)	Exposed, n (%)	P value (Fisher)
DOM-JEM			
Asbestos	12 (6.6)	32 (18.0)	0.001
Chromium	4 (2.2)	25 (14.0)	<0.001
DME	14 (7.8)	33 (18.9)	0.003
Nickel	2 (1.1)	14 (7.8)	0.001
PAH	6 (3.4)	26 (14.6)	<0.001
Silica	2 (1.1)	27 (15.2)	<0.001
Animal	3 (1.7)	19 (10.7)	<0.001
Biodust	42 (23.6)	57 (32.4)	0.076
Endotoxin	12 (6.6)	39 (22.0)	<0.001
ALOHA-JEM			
Biodust	53 (29.8)	66 (37.7)	0.117
Minedust	28 (15.6)	70 (40.2)	<0.001
Gasfumes	65 (37.4)	82 (47.7)	0.064
VGDF	78 (45.1)	95 (55.2)	0.067
Metals	15 (8.2)	37 (20.9)	<0.001
Pesticides (all)	3 (1.6)	18 (10.2)	<0.001
Pesticides - herbicides	1 (0.5)	14 (7.9)	<0.001
Pesticides - insecticides	3 (1.6)	16 (9.0)	0.002
Pesticides - Fungicides	1 (0.5)	17 (9.6)	<0.001
Solvents - aromatic	53 (29.4)	49 (28.0)	0.815
Solvents - chlorinated	51 (28.3)	37 (20.9)	0.112
Solvents - other	60 (33.5)	44 (24.9)	0.081
BEN-JEM (Benzene)	50 (28.6)	41 (25.8)	0.623
ELF-MF	40 (22.6)	80 (45.7)	<0.001
Shock-JEM (shock risk)	22 (13.8)	51 (35.4)	<0.001

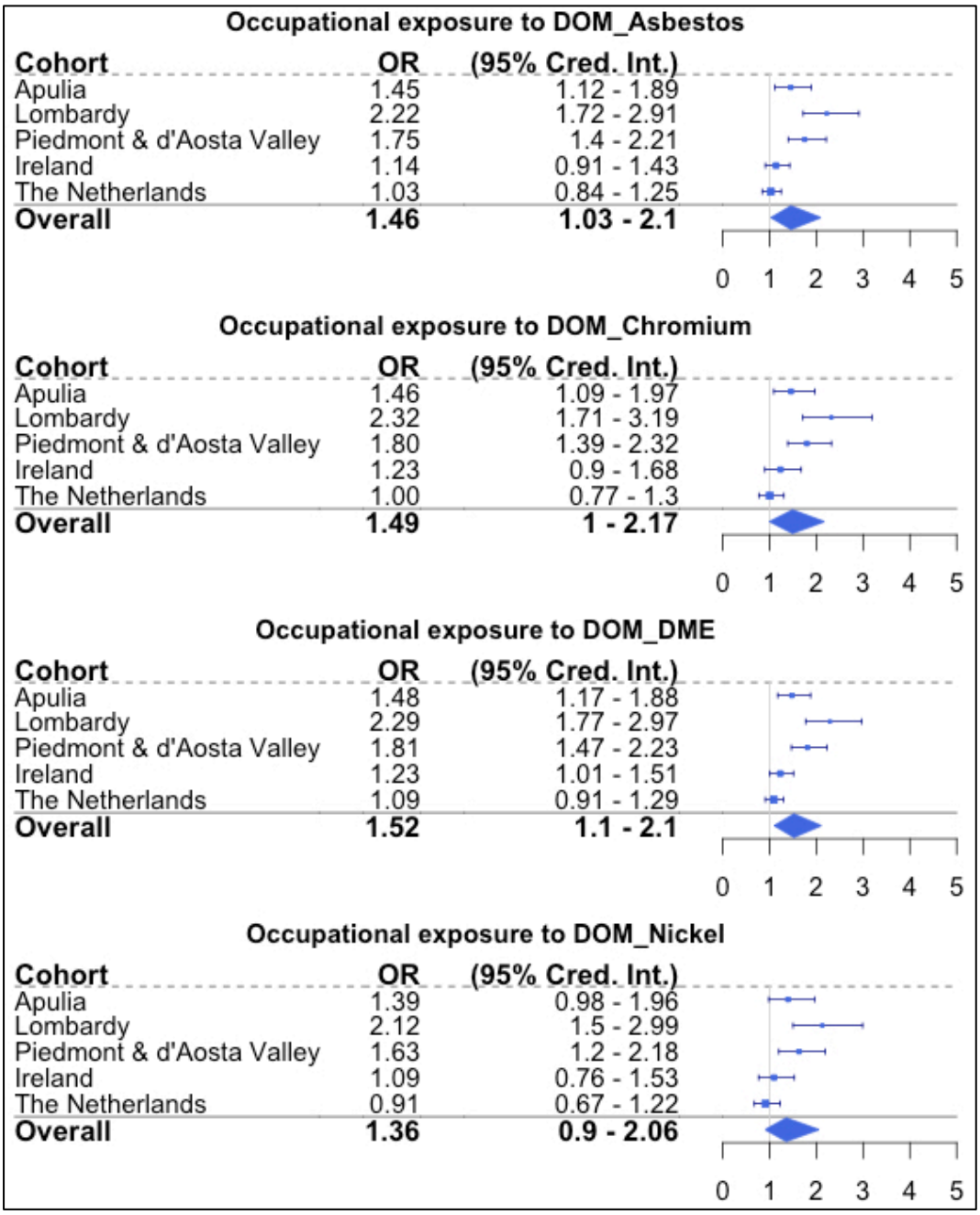
Piedmont and d'Aosta Valley	Controls (n = 290)	Cases (n = 262)	
Exposure	Exposed, n (%)	Exposed, n (%)	P value (Fisher)
DOM-JEM			
Asbestos	48 (16.6)	58 (22.8)	0.082
Chromium	38 (13.1)	49 (19.2)	0.061
DME	60 (20.8)	74 (29.0)	0.028
Nickel	32 (11.1)	31 (12.2)	0.789
PAH	58 (20.1)	61 (24.0)	0.299
Silica	22 (7.6)	41 (16.1)	0.003
Animal	19 (6.6)	22 (8.6)	0.417
Biodust	108 (37.5)	141 (55.7)	<0.001
Endotoxin	79 (27.3)	101 (39.6)	0.003
ALOHA-JEM			
Biodust	135 (47.0)	150 (59.3)	0.006
Minedust	123 (42.3)	144 (56.7)	0.001
Gasfumes	162 (56.3)	178 (70.1)	<0.001
VGDF	184 (64.1)	198 (78.6)	<0.001
Metals	61 (21.1)	77 (30.3)	0.018
Pesticides (all)	27 (93)	54 (21.2)	<0.001
Pesticides - herbicides	18 (6.2)	27 (10.6)	0.085
Pesticides - insecticides	23 (8.0)	44 (17.3)	0.001
Pesticides - Fungicides	23 (8.0)	46 (18.0)	<0.001
Solvents - aromatic	88 (30.6)	106 (41.6)	0.009
Solvents - chlorinated	70 (24.3)	90 (35.3)	0.006
Solvents - other	101 (35.2)	107 (42.0)	0.112
BEN-JEM (Benzene)	83 (29.3)	96 (39.5)	0.016
ELF-MF	147 (50.9)	157 (62.1)	0.009
Shock-JEM (shock risk)	109 (39.6)	118 (49.6)	0.026

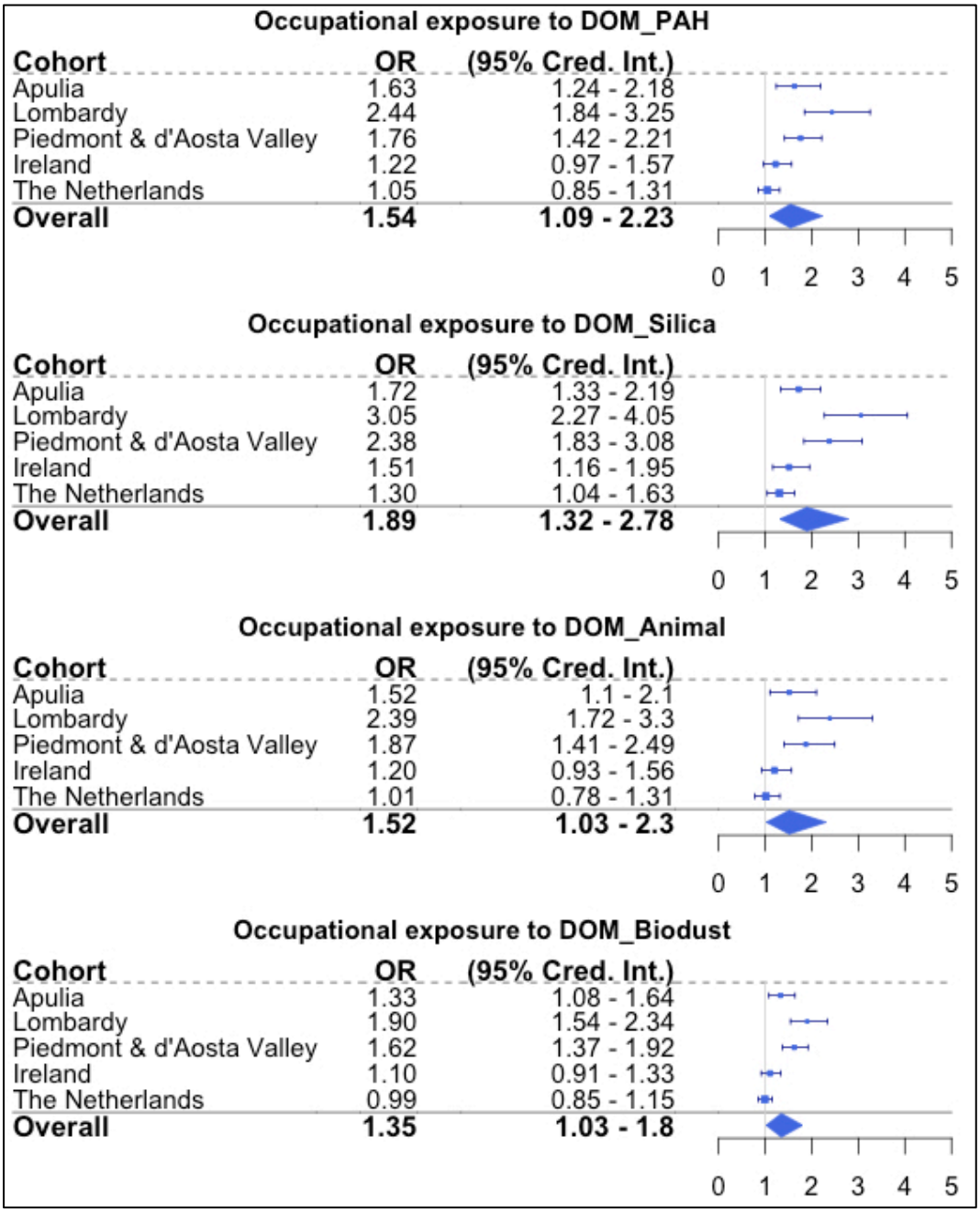
Ireland	Controls (n = 349)	Cases (n = 177)	
Exposure	Exposed, n (%)	Exposed, n (%)	P value (Fisher)
DOM-JEM			
Asbestos	77 (22.3)	37 (21.9)	1.000
Chromium	10 (2.9)	12 (6.9)	0.037
DME	85 (24.8)	48 (28.4)	0.393
Nickel	8 (2.3)	7 (4.1)	0.275
PAH	50 (14.5)	33 (19.3)	0.164
Silica	36 (10.3)	26 (15.0)	0.150
Animal	48 (13.8)	34 (19.8)	0.096
Biodust	126 (36.6)	66 (40.0)	0.497
Endotoxin	72 (20.6)	44 (25.3)	0.264
ALOHA-JEM			
Biodust	151 (44.2)	80 (48.5)	0.392
Minedust	122 (36.1)	69 (42.1)	0.204
Gasfumes	195 (58.2)	99 (63.1)	0.325
VGDF	213 (64.0)	105 (67.3)	0.479
Metals	59 (17.0)	33 (19.5)	0.540
Pesticides (all)	49 (14.0)	30 (17.3)	0.369
Pesticides - herbicides	24 (6.9)	14 (8.0)	0.721
Pesticides - insecticides	45 (12.9)	27 (15.6)	0.420
Pesticides - Fungicides	46 (13.2)	29 (16.7)	0.292
Solvents - aromatic	81 (23.4)	46 (27.4)	0.329
Solvents - chlorinated	58 (16.8)	29 (17.1)	1.000
Solvents - other	101 (29.3)	61 (36.7)	0.104
BEN-JEM (Benzene)	82 (26.0)	52 (30.8)	0.287
ELF-MF	153 (44.9)	71 (43.3)	0.775
Shock-JEM (shock risk)	118 (39.9)	63 (41.4)	0.839

The Netherlands	Controls (n = 1,880)	Cases (n = 791)	
Exposure	Exposed, n (%)	Exposed, n (%)	P value (Fisher)
DOM-JEM			
Asbestos	274 (14.8)	127 (16.6)	0.257
Chromium	115 (6.2)	40 (5.1)	0.364
DME	340 (18.2)	157 (20.5)	0.189
Nickel	90 (4.8)	32 (4.1)	0.477
PAH	194 (10.4)	84 (10.9)	0.727
Silica	155 (8.3)	82 (10.6)	0.073
Animal	119 (6.4)	43 (5.5)	0.425
Biodust	606 (33.6)	259 (68.1)	0.646
Endotoxin	394 (21.6)	162 (21.3)	0.875
ALOHA-JEM			
Biodust	790 (44.1)	329 (44.3)	0.930
Minedust	668 (41.0)	302 (43.9)	0.213
Gasfumes	955 (59.5)	395 (59.3)	0.963
VGDF	1025 (64.1)	428 (64.6)	0.847
Metals	288 (15.5)	121 (15.8)	0.859
Pesticides (all)	176 (9.5)	72 (9.2)	0.942
Pesticides - herbicides	114 (6.1)	56 (7.2)	0.298
Pesticides - insecticides	158 (8.5)	69 (8.9)	0.761
Pesticides - Fungicides	149 (8.0)	63 (8.1)	0.937
Solvents - aromatic	456 (24.7)	198 (26.0)	0.518
Solvents - chlorinated	328 (17.7)	140 (18.2)	0.780
Solvents - other	633 (34.5)	248 (32.7)	0.389
BEN-JEM (Benzene)	447 (24.7)	211 (52.8)	0.066
ELF-MF	750 (41.4)	316 (42.4)	0.659
Shock-JEM (shock risk)	548 (31.4)	241 (34.5)	0.138

Appendix G - Mixed effects estimates of odds ratios for ALS of JEM exposures by Euro-MOTOR study site

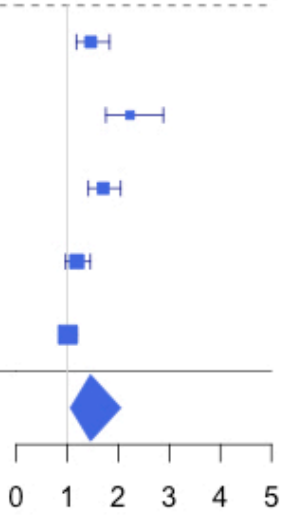
The forest plots below illustrate the odds ratios for ALS risk due to each exposures estimated by JEM as part of the Euro-MOTOR study. These estimates are generated from mixed effects models estimated using r-INLA as described in Chapter 5.5.




Occupational exposure to DOM_PAH
Cohort OR (95% Cred. Int.)
Apulia 1.63 1.24 - 2.18
Lombardy 2.44 1.84 - 3.25
Piedmont & d'Aosta Valley 1.76 1.42 - 2.21
Ireland 1.22 0.97 - 1.57
The Netherlands 1.05 0.85 - 1.31
Overall 1.54 1.09 - 2.23
0 1 2 3 4 5
Occupational exposure to DOM_Silica
Cohort OR (95% Cred. Int.)
Apulia 1.72 1.33 - 2.19
Lombardy 3.05 2.27 - 4.05
Piedmont & d'Aosta Valley 2.38 1.83 - 3.08
Ireland 1.51 1.16 - 1.95
The Netherlands 1.30 1.04 - 1.63
Overall 1.89 1.32 - 2.78
0 1 2 3 4 5
Occupational exposure to DOM_Animal
Cohort OR (95% Cred. Int.)
Apulia 1.52 1.1 - 2.1
Lombardy 2.39 1.72 - 3.3
Piedmont & d'Aosta Valley 1.87 1.41 - 2.49
Ireland 1.20 0.93 - 1.56
The Netherlands 1.01 0.78 - 1.31
Overall 1.52 1.03 - 2.3
0 1 2 3 4 5
Occupational exposure to DOM_Biodust
Cohort OR (95% Cred. Int.)
Apulia 1.33 1.08 - 1.64
Lombardy 1.90 1.54 - 2.34
Piedmont & d'Aosta Valley 1.62 1.37 - 1.92
Ireland 1.10 0.91 - 1.33
The Netherlands 0.99 0.85 - 1.15
Overall 1.35 1.03 - 1.8
0 1 2 3 4 5

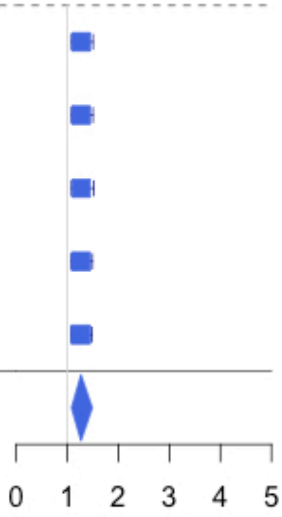
Occupational exposure to DOM_Endotoxin

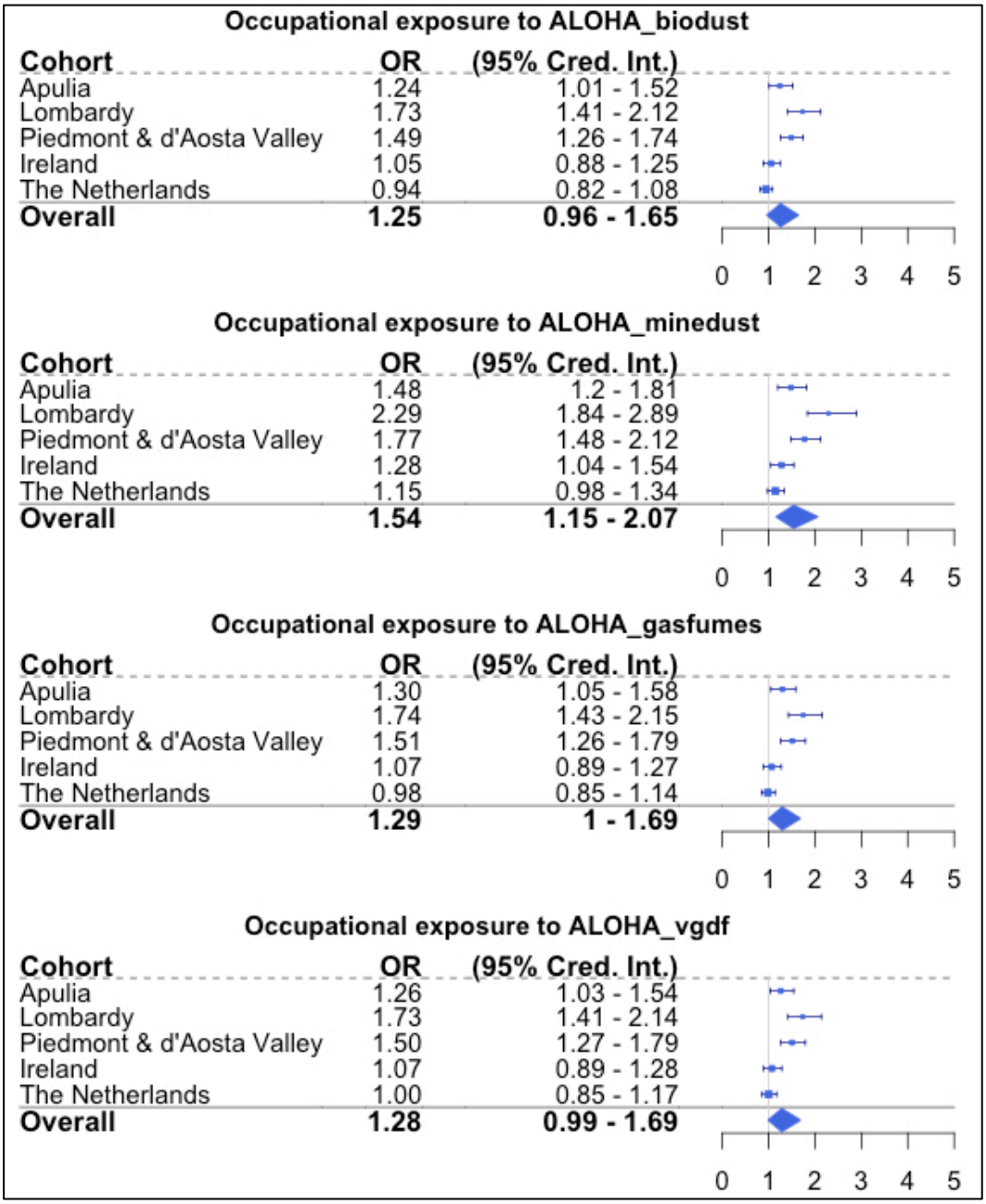
Cohort	OR	(95% Cred. Int.)
Apulia	1.46	1.18 - 1.83
Lombardy	2.23	1.76 - 2.88
Piedmont & d'Aosta Valley	1.71	1.41 - 2.05
Ireland	1.19	0.97 - 1.45
The Netherlands	1.01	0.86 - 1.19
Overall	**1.46**	**1.06 - 2.04**

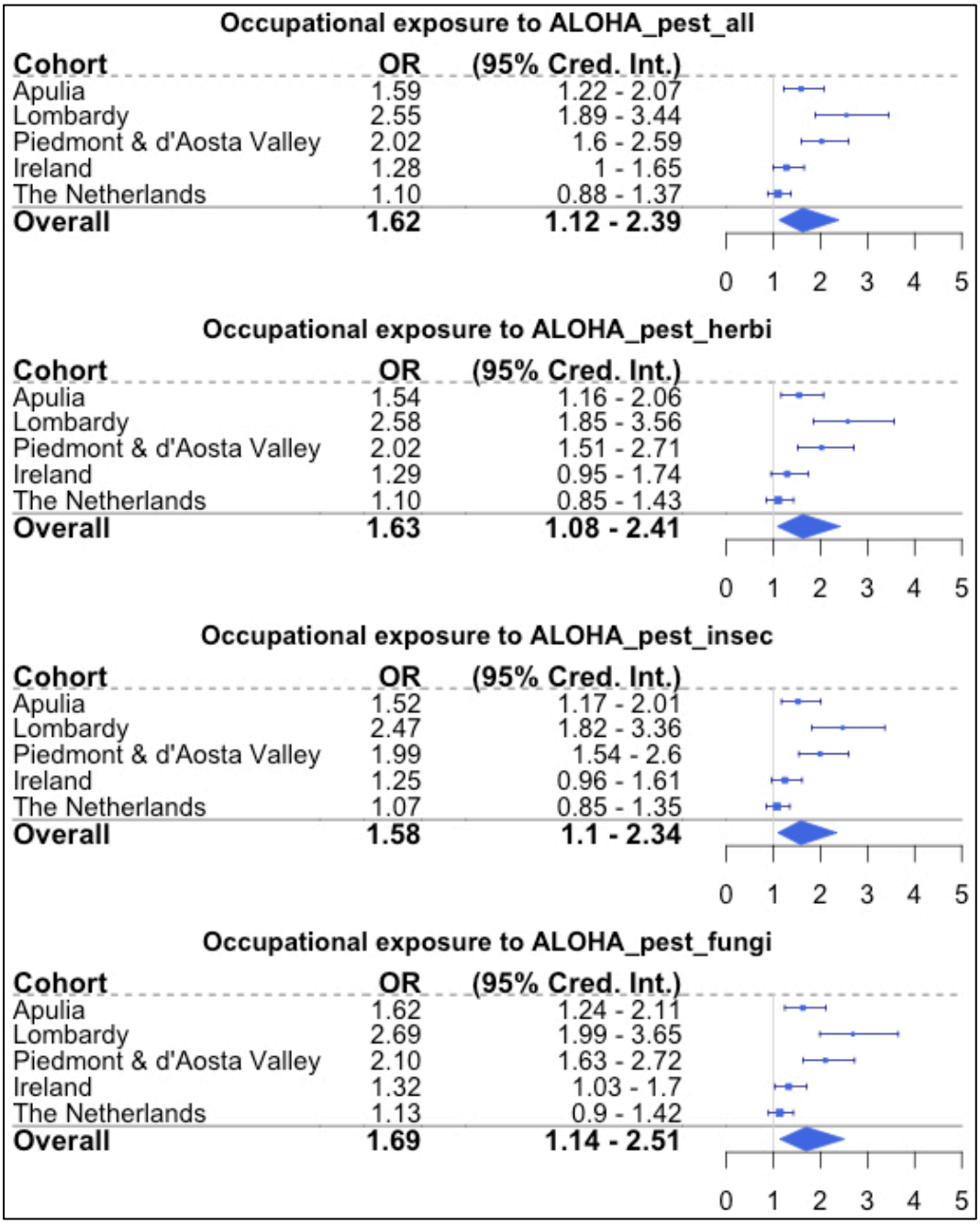


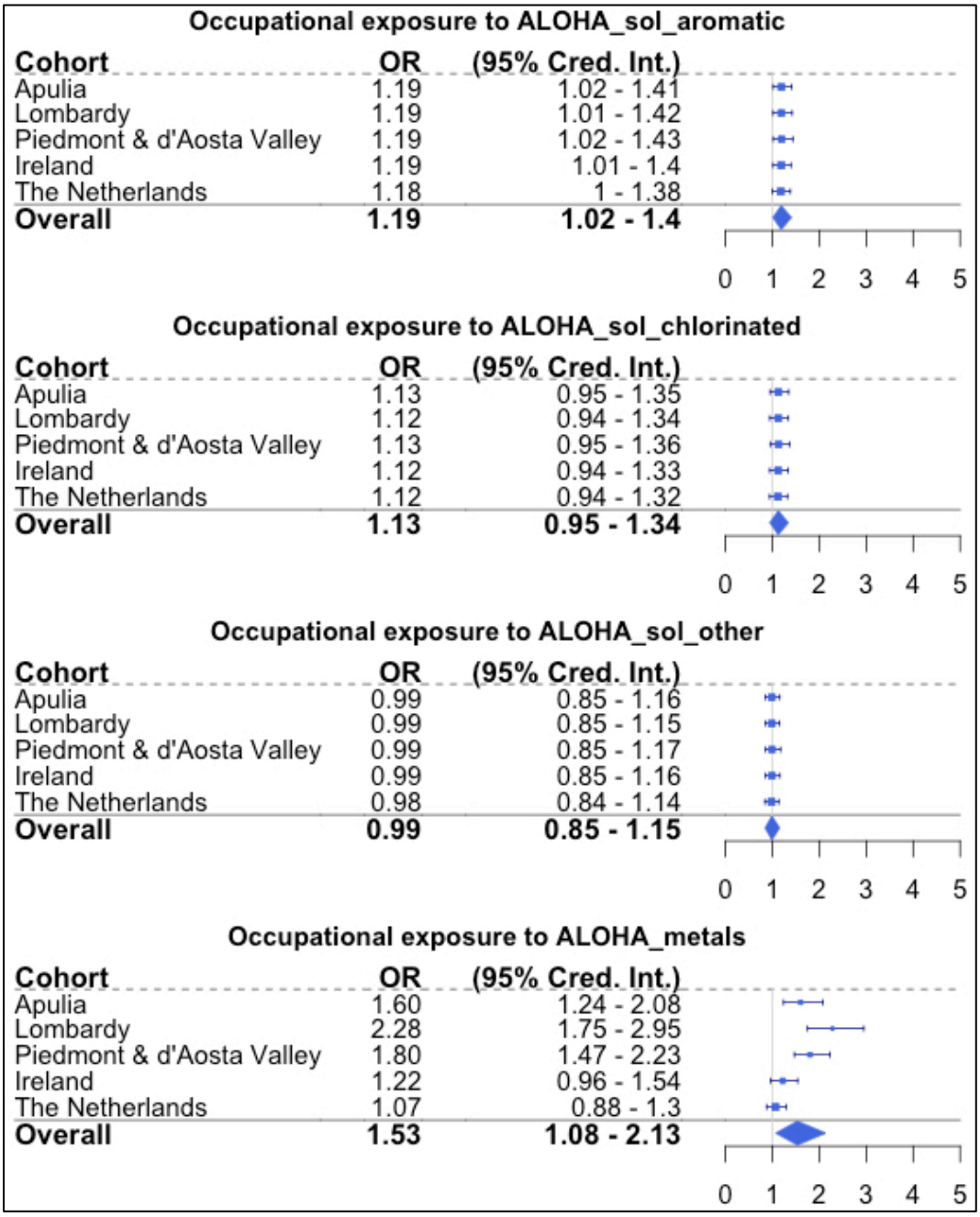
Occupational exposure to Benzene

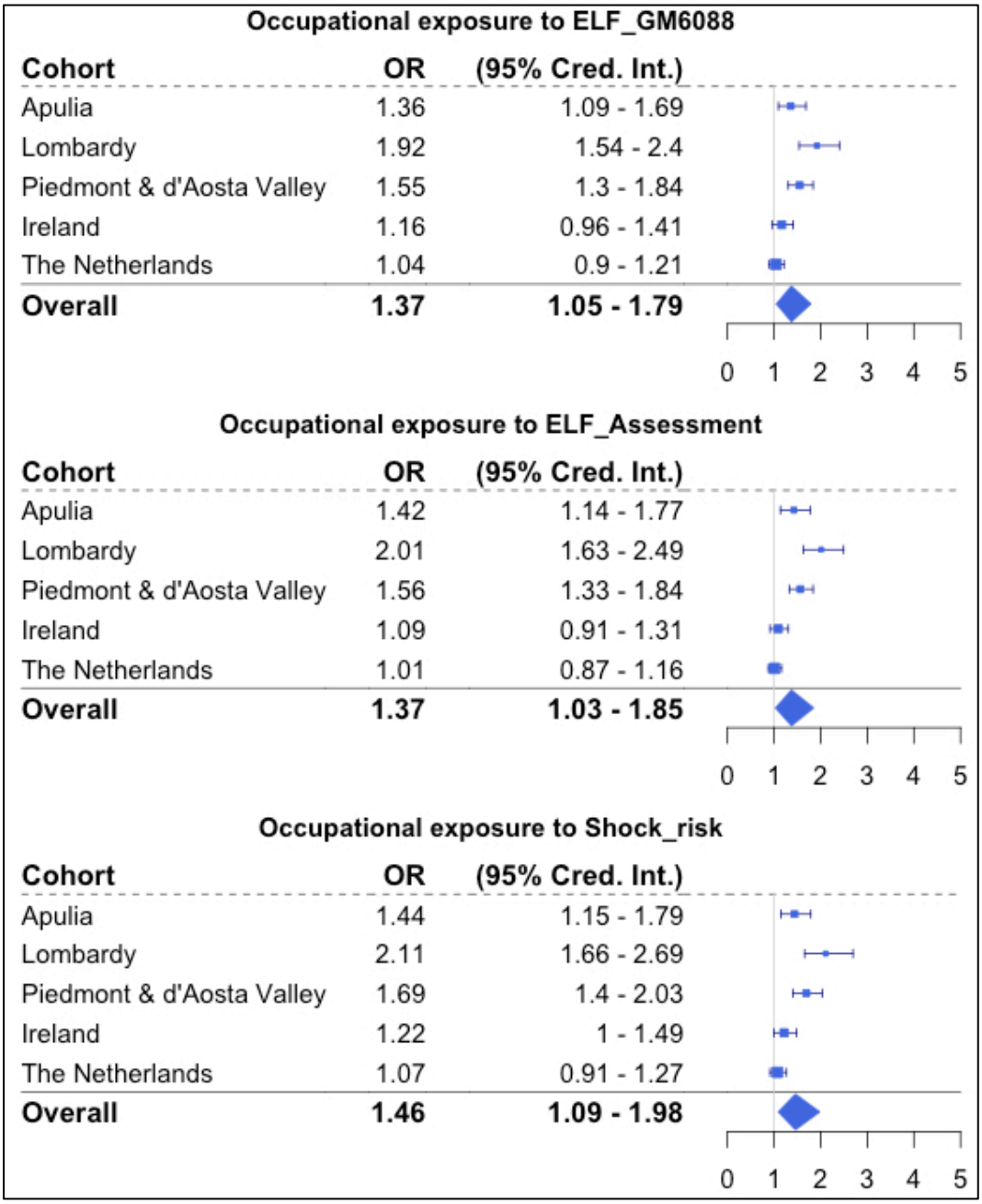
Cohort	OR	(95% Cred. Int.)
Apulia	1.27	1.09 - 1.5
Lombardy	1.27	1.08 - 1.5
Piedmont & d'Aosta Valley	1.27	1.08 - 1.51
Ireland	1.27	1.08 - 1.49
The Netherlands	1.26	1.08 - 1.48
Overall	**1.27**	**1.08 - 1.49**




Occupational exposure to ALOHA_biodust
Cohort OR (95% Cred. Int.)
Apulia 1.24 1.01 - 1.52
Lombardy 1.73 1.41 - 2.12
Piedmont & d'Aosta Valley 1.49 1.26 - 1.74
Ireland 1.05 0.88 - 1.25
The Netherlands 0.94 0.82 - 1.08
Overall 1.25 0.96 - 1.65
0 1 2 3 4 5
Occupational exposure to ALOHA_minedust
Cohort OR (95% Cred. Int.)
Apulia 1.48 1.2 - 1.81
Lombardy 2.29 1.84 - 2.89
Piedmont & d'Aosta Valley 1.77 1.48 - 2.12
Ireland 1.28 1.04 - 1.54
The Netherlands 1.15 0.98 - 1.34
Overall 1.54 1.15 - 2.07
0 1 2 3 4 5
Occupational exposure to ALOHA_gasfumes
Cohort OR (95% Cred. Int.)
Apulia 1.30 1.05 - 1.58
Lombardy 1.74 1.43 - 2.15
Piedmont & d'Aosta Valley 1.51 1.26 - 1.79
Ireland 1.07 0.89 - 1.27
The Netherlands 0.98 0.85 - 1.14
Overall 1.29 1 - 1.69
0 1 2 3 4 5
Occupational exposure to ALOHA_vgdf
Cohort OR (95% Cred. Int.)
Apulia 1.26 1.03 - 1.54
Lombardy 1.73 1.41 - 2.14
Piedmont & d'Aosta Valley 1.50 1.27 - 1.79
Ireland 1.07 0.89 - 1.28
The Netherlands 1.00 0.85 - 1.17
Overall 1.28 0.99 - 1.69
0 1 2 3 4 5


Occupational exposure to ALOHA_pest_all
Cohort OR (95% Cred. Int.)
Apulia 1.59 1.22 - 2.07
Lombardy 2.55 1.89 - 3.44
Piedmont & d'Aosta Valley 2.02 1.6 - 2.59
Ireland 1.28 1 - 1.65
The Netherlands 1.10 0.88 - 1.37
Overall 1.62 1.12 - 2.39
0 1 2 3 4 5
Occupational exposure to ALOHA_pest_herbi
Cohort OR (95% Cred. Int.)
Apulia 1.54 1.16 - 2.06
Lombardy 2.58 1.85 - 3.56
Piedmont & d'Aosta Valley 2.02 1.51 - 2.71
Ireland 1.29 0.95 - 1.74
The Netherlands 1.10 0.85 - 1.43
Overall 1.63 1.08 - 2.41
0 1 2 3 4 5
Occupational exposure to ALOHA_pest_insec
Cohort OR (95% Cred. Int.)
Apulia 1.52 1.17 - 2.01
Lombardy 2.47 1.82 - 3.36
Piedmont & d'Aosta Valley 1.99 1.54 - 2.6
Ireland 1.25 0.96 - 1.61
The Netherlands 1.07 0.85 - 1.35
Overall 1.58 1.1 - 2.34
0 1 2 3 4 5
Occupational exposure to ALOHA_pest_fungi
Cohort OR (95% Cred. Int.)
Apulia 1.62 1.24 - 2.11
Lombardy 2.69 1.99 - 3.65
Piedmont & d'Aosta Valley 2.10 1.63 - 2.72
Ireland 1.32 1.03 - 1.7
The Netherlands 1.13 0.9 - 1.42
Overall 1.69 1.14 - 2.51
0 1 2 3 4 5


Occupational exposure to ALOHA_sol_aromatic
Cohort OR (95% Cred. Int.)
Apulia 1.19 1.02 - 1.41
Lombardy 1.19 1.01 - 1.42
Piedmont & d'Aosta Valley 1.19 1.02 - 1.43
Ireland 1.19 1.01 - 1.4
The Netherlands 1.18 1 - 1.38
Overall 1.19 1.02 - 1.4
0 1 2 3 4 5
Occupational exposure to ALOHA_sol_chlorinated
Cohort OR (95% Cred. Int.)
Apulia 1.13 0.95 - 1.35
Lombardy 1.12 0.94 - 1.34
Piedmont & d'Aosta Valley 1.13 0.95 - 1.36
Ireland 1.12 0.94 - 1.33
The Netherlands 1.12 0.94 - 1.32
Overall 1.13 0.95 - 1.34
0 1 2 3 4 5
Occupational exposure to ALOHA_sol_other
Cohort OR (95% Cred. Int.)
Apulia 0.99 0.85 - 1.16
Lombardy 0.99 0.85 - 1.15
Piedmont & d'Aosta Valley 0.99 0.85 - 1.17
Ireland 0.99 0.85 - 1.16
The Netherlands 0.98 0.84 - 1.14
Overall 0.99 0.85 - 1.15
0 1 2 3 4 5
Occupational exposure to ALOHA_metals
Cohort OR (95% Cred. Int.)
Apulia 1.60 1.24 - 2.08
Lombardy 2.28 1.75 - 2.95
Piedmont & d'Aosta Valley 1.80 1.47 - 2.23
Ireland 1.22 0.96 - 1.54
The Netherlands 1.07 0.88 - 1.3
Overall 1.53 1.08 - 2.13
0 1 2 3 4 5


Occupational exposure to ELF_GM6088
Cohort OR (95% Cred. Int.)
Apulia 1.36 1.09 - 1.69
Lombardy 1.92 1.54 - 2.4
Piedmont & d'Aosta Valley 1.55 1.3 - 1.84
Ireland 1.16 0.96 - 1.41
The Netherlands 1.04 0.9 - 1.21
Overall 1.37 1.05 - 1.79
0 1 2 3 4 5
Occupational exposure to ELF_Assessment
Cohort OR (95% Cred. Int.)
Apulia 1.42 1.14 - 1.77
Lombardy 2.01 1.63 - 2.49
Piedmont & d'Aosta Valley 1.56 1.33 - 1.84
Ireland 1.09 0.91 - 1.31
The Netherlands 1.01 0.87 - 1.16
Overall 1.37 1.03 - 1.85
0 1 2 3 4 5
Occupational exposure to Shock_risk
Cohort OR (95% Cred. Int.)
Apulia 1.44 1.15 - 1.79
Lombardy 2.11 1.66 - 2.69
Piedmont & d'Aosta Valley 1.69 1.4 - 2.03
Ireland 1.22 1 - 1.49
The Netherlands 1.07 0.91 - 1.27
Overall 1.46 1.09 - 1.98
0 1 2 3 4 5