From Institute of Environmental Medicine
Karolinska Institutet, Stockholm, Sweden

BIOMEDICAL FACTORS IN THE RISK AND PROGNOSIS OF AMYOTROPHIC LATERAL SCLEROSIS

Jiangwei Sun

孙江伟

Karolinska Institutet

Stockholm 2022

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2022
© Jiangwei Sun, 2022
ISBN 978-91-8016-490-0
Cover illustration by Jiangwei Sun.

Biomedical factors in the risk and prognosis of amyotrophic lateral sclerosis
THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Jiangwei Sun

The thesis will be defended in public at Lecture hall Samuelssonsalen, Tomtebodavägen 6, Karolinska Institutet, Solna, Sweden, Friday 29 April 2022, 13:00 pm

Principal Supervisor:
Prof. Fang Fang
Karolinska Institutet
Institute of Environmental Medicine
Unit of Integrative Epidemiology

Co-supervisor(s):
Prof. Yudi Pawitan
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Dr. Ulrika Zagai
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Dr. Caroline Ingre
Karolinska Institutet
Department of Clinical Neuroscience

Opponent:
Prof. Alberto Ascherio
Harvard University
Harvard T.H. Chan School of Public Health
Brigham and Women's Hospital and Harvard Medical School

Examination Board:
Prof. John Öhrvik
Uppsala University
Centre for Clinical Research

Associate professor Sofia Carlsson
Karolinska Institutet
Institute of Environmental Medicine

Associate professor Einar Eriksson
Karolinska Institutet
Department of Clinical Neuroscience

To my beloved family

致我亲爱的家人

竹杖芒鞋轻胜马，谁怕？一蓑烟雨任平生。

回首向来萧瑟处，归去，也无风雨也无晴。

------北宋 苏轼

ABSTRACT

Amyotrophic lateral sclerosis (ALS) is a relatively rare but incurable and relentlessly progressive neurodegenerative disease, characterized by motor neuron loss in the brain and spinal cord. Majority of patients die within 3-5 years after symptom onset, commonly due to respiratory failure. Until now, except for older age, male sex, family history, and specific genetic mutations, other risk factors of ALS remain largely unknown. Alterations in energy metabolism and neuroinflammation are known features of ALS, which can be directly or indirectly modulated by human gut microbiome and microbial metabolites. Gut microbiome and microbial metabolites have also been suggested to have a role in neurodegenerative disease through modulating the process of protein misfolding and aggregations and subsequently exacerbate disease phenotype. Exploring the association of biomedical factors related to gut microbiome alteration with the risk and prognosis of ALS might therefore help to elucidate the etiology of ALS. Further, as the survival of patients with ALS varies greatly, ranging from several months to more than 10 years, identification of prognostic predictors that are routinely measured in clinical practice might help to supervise treatment and improve patient care.

The first three studies focused on the etiology of ALS.

Study I examined the association between previous gastrointestinal (GI) biopsy of normal mucosa and non-specific inflammation and risk of ALS in a matched cohort study based on ESPRESSO (Epidemiology Strengthened by histoPathology Reports in Sweden). After excluding the first two years of follow-up after the biopsy from analysis, we found that individuals with a GI biopsy result of normal mucosa had an increased risk of ALS, compared with their matched reference individuals randomly selected from the general population. However, no risk alteration was observed for a GI biopsy result of non-specific inflammation. Besides, a GI biopsy result of normal mucosa or non-specific inflammation was not related to mortality risk in patients with ALS.

Study II analysed the potential role of antibiotic use on risk of ALS in a nested case-control study conducted using several Swedish national healthcare registers. After excluding all prescriptions within one year before diagnosis, patients with ALS were more likely to have antibiotic prescriptions before diagnosis, compared with controls, and there was a dose-response relationship between numbers of antibiotic prescriptions and ALS risk.

Study III explored the association between hospital-treated infection and risk of three neurodegenerative diseases, including Alzheimer's disease (AD), Parkinson's disease (PD), and ALS, in a nested case-control study conducted in the Swedish national healthcare registers. A hospital-treated infection was associated with an increased risk of AD and PD, regardless of infection type (bacterial, viral, or other infection) and site (central nervous system, gastrointestinal, or genitourinary infection). The associations were primarily noted among individuals younger than age 60. Moreover, individuals with multiple events of hospital-treated infections before age 40 appeared to have the greatest risk of developing AD and PD. However, no association was observed for ALS.

The fourth study focused on the use of routinely measured blood markers in the prognosis of ALS.

Study IV assessed the predictive role of eight commonly measured blood markers on ALS prognosis in a population-based cohort study within the SCREAM (Stockholm CREAtinine Measurement) project. At the time of diagnosis, lower serum levels of creatinine and albumin, as well as higher serum levels of C-reactive protein (CRP) and glucose, were associated with an increased risk of mortality among ALS patients. After ALS diagnosis, decreasing serum levels of creatinine and albumin or increasing serum levels of CRP and glucose were indicative of an increased mortality risk.

In conclusion, findings from this thesis work support that specific biomedical factors such as previous GI dysfunction, antibiotic use, and hospital-treated infections are associated with the later risk of ALS development, as risk factors, triggers, or prodromal symptoms. Moreover, commonly measured biomedical markers can be of predictive value in ALS prognosis.

LIST OF SCIENTIFIC PAPERS

I. **Jiangwei Sun**, Jonas F. Ludvigsson, Bjorn Roelstraete, Yudi Pawitan & Fang Fang. Gastrointestinal biopsies and amyotrophic lateral sclerosis - results from a cohort study of 1.1 million individuals. Amyotroph Lateral Scler Frontotemporal Degener, 2021, 22(5-6): 410-418.

II. **Jiangwei Sun**, Yiqiang Zhan, Daniela Mariosa, Henrik Larsson, Catarina Almqvist, Caroline Ingre, Ulrika Zagai, Yudi Pawitan, Fang Fang. Antibiotics use and risk of amyotrophic lateral sclerosis in Sweden. Eur J Neurol, 2019, 26(11): 1355-1361.

III. **Jiangwei Sun**, Jonas F. Ludvigsson, Caroline Ingre, Fredrik Piehl, Karin Wirdefeldt, Ulrika Zagai, Weimin Ye, Fang Fang. Hospital-treated infections in early and mid-life increase the risk of Alzheimer's disease and Parkinson's disease: nationwide nested case-control study in Sweden. *(Submitted)*

IV. **Jiangwei Sun**, Juan Jesus Carrero, Ulrika Zagai, Marie Evans, Caroline Ingre, Yudi Pawitan, Fang Fang. Blood biomarkers and prognosis of amyotrophic lateral sclerosis. Eur J Neurol, 2020, 27(11): 2125-2133.

LIST OF ADDITIONAL PUBLICATIONS

[*] Equal contribution, [#] Shared corresponding author.

I. **Jiangwei Sun**, Tingting Huang, Justine W. Debelius, Fang Fang. Gut microbiome and amyotrophic lateral sclerosis: A systematic review of current evidence. J Intern Med, 2021, 290(4): 758-788.

II. Can Cui[*], **Jiangwei Sun**[*#], Yudi Pawitan, Fredrik Piehl, Honglei Chen, Caroline Ingre, Karin Wirdefeldt, Marie Evans, John Andersson, Juan-Jesus Carrero, Fang Fang[#]. Creatinine and C-reactive protein in amyotrophic lateral sclerosis, multiple sclerosis and Parkinson's disease. Brain Communications, 2020, 2(2): fcaa152.

III. Hakan Cetin, **Jiangwei Sun**, Catarina Almqvist, Berthold Reichardt, Matthias Tomschik, Fritz Zimprich, Fang Fang, Caroline Ingre. No association between proton pump inhibitor use and ALS risk: a nationwide nested case-control study. Sci Rep, 2020, 10(1): 13371.

CONTENTS

LIST OF ABBREVIATIONS

AD	Alzheimer's disease
ALS	Amyotrophic lateral sclerosis
ALSFRS-R	ALS Functional Rating Score-revised
ANS	Autonomic nervous system
ATC	Anatomical Therapeutic Chemical
BMI	Body mass index
C9orf72	Chromosome 9 open reading frame 72
CDR	Cause of Death Register
CI	Confidence interval
CNS	Central nervous system
CRP	C-reactive protein
ENS	Enteric nervous system
ESPRESSO	Epidemiology Strengthened by histoPathology Reports in Sweden
FUS	Fused in sarcoma
GBA	Gut-brain axis
GI	Gastrointestinal
HR	Hazard ratio
IBD	Inflammatory bowel disease
ICD	the International Classification of Disease
LISA	Longitudinal Database for Health Insurance and Labor Market Studies
NBHW	National Board of Health and Welfare (Socialstyrelsen)
NF-L	Neurofilament light chain
NM	Normal mucosa
NPR	National Patient Register
NSI	Non-specific inflammation
OR	Odds ratio

PD	Parkinson's disease
PDR	Prescribed Drug Register
pNFH	Phosphorylated form of neurofilament heavy chain
RBP4	Retinol-binding protein 4
SCFA	Short-chain fatty acid
SCREAM	the Stockholm CREAtinine Measurement project
SD	Standard deviation
SNOMED	Swedish version of the Systematized Nomenclature of Medicine
SOD1	Superoxide Dismutase 1
TARDBP	Trans-activation response DNA-binding protein
TBK1	TANK binding kinase 1
TDP-43	Trans-activation response DNA-binding protein 43
TPR	Total Population Register
Treg	Regulatory T cell

1 INTRODUCTION

Amyotrophic lateral sclerosis (ALS), with a key feature of degeneration of motor neuron in the brain and spinal cord, is a relatively rare but incurable neurodegenerative disease. The incidence of ALS is estimated as 1.75 per 100,000 person-years whereas the prevalence is estimated as 4.1-4.8 cases per 100,000 individuals. ALS commonly occurs after middle age. Although rare, the disease is restlessly progressive. Death due to respiratory paralysis typically occurs within 3 to 5 years after symptom onset, and no medicine offers a substantial clinical benefit for ALS patients as of today. Around 10%-15% of ALS cases are familial where genetic factors are the main underlying contributors, the etiology is however elusive for sporadic ALS. Although many risk factors have been suggested in either observational studies or Mendelian randomization studies (summarized in Table 1 of the **Appendix** [1]), until now, the established risk factors of ALS only include older age, male sex, family history, and genetic mutations.

Beyond motor neuron degeneration, patients with ALS tend to have defects in energy metabolism (e.g., hypermetabolism, weight loss, and diabetes) and neuroinflammation (a well-established pathological feature of ALS, characterized by activation of microglia and astroglia in the central nervous system (CNS)). Due to the rapidly growing research on gut microbiome, evidence has accumulated to support its involvement in energy metabolism, immune responses, and development of neurodegenerative disease. We therefore, in this thesis, investigated the roles of several biomedical factors which might be related to the composition and functionality of gut microbiota on ALS risk, including gastrointestinal (GI) dysfunction (**Study I**), antibiotics use (**Study II**), and hospital-treated infections (**Study III**). As survival for patients with ALS varies greatly, ranging from several months to more than 10 years, we in **Study IV** assessed the prognostic values of several blood markers that are commonly measured in clinical practice in ALS.

2 BACKGROUND

2.1 AMYOTROPHIC LATERAL SCLEROSIS

ALS, also known as Lou Gehrig's disease in the United States and motor neuron disease in the UK, is a relatively rare but incurable and relentlessly progressive neurodegenerative disease, characterized by motor neuron loss in the brain and spinal cord. [2] The neuropathological features of ALS from gross, microscopic to molecular levels are summarized in the Table 1 of the **Appendix**. [1]

ALS begins insidiously with local symptoms, but rapidly spreads to majority of muscles, which eventually leads to death commonly due to respiratory failure. [3] Although around 10% of ALS patients survive 10 years or longer, vast majority of patients die within 3-5 years after symptom onset.

Motor neurons are grouped into corticospinal motor neurons in the motor cortex (upper motor neurons) and bulbar or spinal motor neurons (lower motor neurons). In a healthy person, the upper motor neurons make direct or indirect connections with the lower motor neurons, which subsequently innervate skeletal muscles and control movement. In ALS patients, however, the communication between brain and muscles is interrupted by the deficit of either the upper motor neurons (presenting as stiffness and spasticity) or lower motor neurons (presenting as fasciculation and amyotrophy), or both.

2.2 CLINICAL SYMPTOMS AND DIAGNOSIS

As any volunteer muscle can be affected, clinical presentations of patients with ALS are heterogeneous. Depending on the involved motor neurons, the clinical symptoms can include weakness and atrophy of the extremities (progressive muscular atrophy, affecting mainly lower motor neurons), hyperreflexia and spasticity with few findings of lower motor neuron dysfunction (primary lateral sclerosis, affecting mainly the corticospinal motor neurons), prominent dysarthria (difficulty in articulation) and dysphagia (difficulty in swallowing) with tongue atrophy (bulbar ALS, affecting mainly the brainstem motor neurons), and emotional lability often accompanied by facial spasticity (pseudobulbar palsy, affecting mainly affect frontopontine motor neurons). [2-5] Evidence has further shown that about half of ALS patients have cognitive and behavioural impairment, and about 13% of ALS patients might have frontotemporal dementia. [4]

The heterogeneous clinical presentation and varying prognostic profile make the diagnosis of ALS challenging. Until today, no definitive diagnostic test exists for ALS, and its diagnosis is still based on clinical findings and exclusion of mimics. [6] The current guideline for ALS diagnosis is the EI Escorial criteria, [5] which was developed by the World Federation of Neurology Research Group on Motor Neuron Disease in 1994 and revised in 2000 (known as Airlie House criteria) [7] and 2008 (known as Awaji-Shima criteria). [8] Diagnosis of ALS requires a history of progressive weakness spreading within a region or to other regions (bulbar, cervical, thoracic, or lumbar), with evidence of involvement of lower motor neuron and upper

motor neuron, and that no alternative disease can explain these presentations. [4-6] According to the clinical test result, clinicians can classify the patients into three groups: definite, probable or possible ALS. [5, 9] Therefore, diagnostic delay (i.e., time from the first symptom to diagnosis) exists, which is approximately one year on average and might make the patients miss the best therapeutic window. [3]

2.3 TREATMENT

2.3.1 Disease modifying therapies

No medicine offers a substantial clinical benefit for ALS patients as of today. Riluzole, acting by supressing excessive motor neuron firing through reducing glutamatergic neurotransmission on presynaptic neurons, [6] is the only widely available medicine approved by the US Food and Drug Administration for the treatment of ALS. It can prolong the median survival of ALS patients by an average of 2-3 months. [4] However, the efficacy of riluzole appears to be better amongst bulbar-onset and older patients, and in the early stage of the disease. [10] Adverse effects of riluzole include asthenia, vomiting, dizziness, and liver damage.

Edaravone, a free-redial scavenger, is approved in Japan (2015) and the United States (2017), but not Europe, to treat ALS. It has been shown to slow disease progression amongst patients in early stage with definite or probable ALS. [11]

2.3.2 Symptomatic treatments

Currently, the mainstay of care for ALS patients is to manage symptoms with pharmacological and non-pharmacological interventions. [3] Pharmacotherapy can be indicated to treat spasticity (muscle relaxants, e.g., baclofen and tizanidine), sialorrhoea (anticholinergic drugs, e.g., atropine), pain (neuropathic pain: gabapentin, pregabalin, and tricyclic antidepressants; and nociceptive pain: non-steroidal anti-inflammatory drugs, opioids, and cannabis), muscle cramps (e.g., levetiracetam and mexiletine), and depression (e.g., selective serotonin reuptake inhibitors). [6]

Non-pharmacological interventions can be indicated to treat dysphagia (e.g., dietary modifications, oral and pharyngeal range-of-motion exercise, and placement of gastrostomy tube), dysarthria (e.g., speech therapy and communication techniques based on brain-computer interface), as well as respiratory insufficiency (e.g., non-invasive positive-pressure ventilation or tracheostomy). [6]

Overall, multidisciplinary care, including neurologist, psychologist, nutritionist, physical therapist, speech therapist and specialized nurse, can prolong survival and increase the quality of life of patients with ALS. [6]

2.4 EPIDEMIOLOGY

2.4.1 Incidence and prevalence

The incidence of ALS is estimated as 1.75 cases per 100,000 person-years (male: 2.03; female: 1.45). [12] The incidence differs according to ancestral origin. Studies in populations of European origin have shown an incidence of 1.59 in Northern America, 1.92 in Northern Europe, 2.22 in Southern Europe, and 2.35 in Western Europe, per 100,000 person-years. [12] Incidence is lower in East (~0.8 per 100,000 person-years) and South (~0.7 per 100,000 person-years) Asia [6]. The lower incidence of ALS in Asia may be partially explained by the lower prevalence of known ALS genes in Asian population. [13] In addition to the ancestral differences, the incidence of ALS appears to be increasing globally, likely due to the improved ascertainment, [14] as well as improving survival rate of competitive diseases and the overall aging of the global population. [15] Because of the rapid disease progression, the prevalence of ALS remains low, with an estimated prevalence of 4.1- 4.8 per 100,000 individuals. [13]

2.4.2 Risk factors

About 10%-15% of ALS patients have a clear family history (i.e., familial cases), whereas the remaining patients are sporadic cases. [16] Although many risk factors have been suggested in either observational studies or Mendelian randomization studies (summarized in Table 1 of the **Appendix** [1]), the established risk factors of ALS to date only include older age, male sex, family history, and genetic mutations.

Genetic risk factors

Gene mutations account for about 70% of all familial cases and approximately 15% of sporadic cases. [17] The first genetic mutation found to cause ALS was reported in 1993 and affects the gene *SOD1*. [3] Due to the evolving technologies for gene mapping and DNA analysis, more than one hundred genes have later been identified. [3] These genes can be grouped into several categories, including a) genes that perturb protein homoeostasis (e.g., *SOD1*), b) genes that alter RNA homoeostasis and trafficking (e.g., trans-activation response DNA-binding protein (*TARDBP*)), and c) genes that disturb cytoskeletal dynamics in the motor neuron axon and distal terminal. [2, 3] The most important ALS genes identified to date include *SOD1*, *TARDBP*, fused in sarcoma (*FUS*), and chromosome 9 open reading frame 72 (*C9orf72*).

SOD1 gene is located on chromosome 21q22.11 and encodes SOD1 protein (CuZn-superoxide dismutase), [16] which is one of the three superoxide dismutase and has the function of eliminating free radicals and transforming free superoxide radicals into molecular oxygen and hydrogen peroxide. *SOD1* mutation accounts for 10-20% of familial ALS and 1-5% of sporadic ALS cases, [18] and can trigger neurotoxicity and result in accumulation of aggregated proteins and influence the degradation of intracellular protein. The discovery of *SOD1* mutation in ALS has greatly promoted the understanding of the disease, and many *SOD1* transgenic animal models of ALS are available to date. [19]

TARDBP gene is located on chromosome 1 and encodes TDP-43, a nuclear ribonucleoprotein implicated in exon splicing, gene transcription, regulation of mRNA stability, and mRNA biosynthesis. [20] *TARDBP* mutation accounts for ~5% of familial ALS and 1% of sporadic ALS, [18] and can lead to the aggregation of TDP-43 characterized by abnormal phosphorylation, truncation, and cytoplasmic mislocalization. The dysfunction of TDP-43 has been found in both ALS and other neurodegenerative diseases (e.g., frontotemporal dementia). [20] ***FUS*** gene is located on chromosome 16 and encodes another RNA-binding protein, similar to TDP-43, and its mutation leads to self-assembly of proteins, which is also observed in ALS patients. [21]

C9orf72 is located at locus 9p21 of chromosome 9 and is the most commonly mutated gene of familial ALS and accounts for ~10% of sporadic ALS cases. [22] It is characterized by an expansion of a noncoding hexanucleotide (GGGGCC) up to hundreds and thousands repeats, whereas ALS-free individuals have approximately 30 repeats. [19] This expansion has been shown in frontotemporal dementia as well. Several mechanisms may contribute to the neurotoxicity of the expansion, including a) sequestering RNA binding protein and disabling RNA processing machinery by forming RNA foci, [23] b) generating five potentially toxic repeat dipeptides after escaping to the cytoplasm, [24] and c) causing vulnerability of motor neurons to Ca^{2+} permeable receptor-mediated excitotoxicity. [22]

Non-genetic risk factors

Many lifestyle and environmental factors have been investigated in the past decades as potential risk factors for ALS. Exposure to heavy metal (e.g., lead), [25, 26] pesticides, [16] electromagnetic field, [27] and air pollution [28] has been shown to be associated with increased risk of ALS. Military service [29] and trauma [30] have been related to ALS risk as well.

ALS patients tend to have a higher level of physical fitness and lower body mass index (BMI) [31] than average, [18] and this impression is further supported by the evidence that athletes (i.e., professional soccer or football players) appear to have a higher risk of ALS. [32] Previous studies also explored the influence of diet on ALS risk, and showed that higher intake of antioxidants (e.g., vitamin E), fruits, vegetables, [33] and chicken [34] might be associated with a lower risk of ALS, whereas fat, [35] red and processed meat, and animal protein [36] might be related to a higher risk of ALS. However, the association between alcohol consumption and ALS is inconsistent. [35, 37, 38]

Because observational studies are prone to methodological problems including selection bias, measurement error, and confounding, drawing conclusions about causality for the abovementioned associations are challenging. Such issue could however be partially tackled by using Mendelian randomization analysis, which has the potential to investigate causal relationship between risk factors and a health outcome avoiding common methodological concerns of observational studies. [39, 40] For example, a causal relationship between smoking, [41, 42] physical activity, [42] and blood lipid levels [42, 43] and risk of ALS has been

suggested in Mendelian randomization study. We summarized risk factors suggested in observational studies or Mendelian randomization studies in the Table 1 of the **Appendix**. [1]

2.4.3 Prognostic factors

Survival for patients with ALS varies greatly, ranging from several months to more than 10 years. Predictors for better prognosis include younger age at onset of symptoms or diagnosis, spinal onset, higher score of the revised ALS Functional Rating Scale (ALSFRS-R) at diagnosis, longer diagnostic delay, higher BMI at diagnosis, weight gain after diagnosis, and interdisciplinary care. [13, 44] In contrast, presence of frontotemporal dementia and depression, respiratory or genitourinary comorbidities, worse nutritional status, and *C9orf72* mutation has been suggested to be related with worse prognosis. [13, 45]

Because diaphragm is usually affected at the end stage of ALS, tracheostomy is commonly used in ALS care. Patient survival could be greatly extended once tracheostomy is in place. However, there are considerable differences in the extension of survival among patients with tracheostomy compared with patients without this procedure, ranging from 16 months in Italy (47 months vs 31 months), [46] 33.9 months in Denmark (56.8 months vs 22.9 months), [47] 4.1 years in Norway (7.3 years vs 3.2 years), [48] to approximately 7 years in Japan (11.33 years vs 4.61 years). [49] The exact reasons of this difference need to be investigated further.

Many biomarkers have emerged for ALS prognosis in the past decade. [50] For instance, higher levels of blood creatinine, albumin, [51] and serum retinol-binding protein 4 (RBP4) [52] are related with better prognosis, whereas higher levels of C-reactive protein (CRP) in blood, [53] neurofilament light chain (NF-L) [54] and phosphorylated form of neurofilament heavy chain (pNFH) [55] in the cerebrospinal fluid are associated with worse prognosis. However, the ability of the commonly measured blood markers to serve as the predicators for disease progression in ALS remains largely unknown, we therefore explored the predictive ability of eight blood markers in **Study IV**.

2.5 MICROBIOME

2.5.1 Microbiome and health

Human being has coevolved with trillions of microbes (bacteria, archaea, fungi, and viruses) inhabiting in the body, and the main colonization sites are skin, airway, urogenital tract, and GI tract. [56] Although until now only minority of these can be cultured, advance in culture-independent technologies, such as next-generation sequencing technologies, and microbiome bioinformatic pipelines have greatly accelerated microbiome research over the past two decades. [56] Several large collaborative efforts have been conducted to explore the effect of microbiome on human health. These collaborative efforts include Human Microbiome Project, [57] the European Union's Metagenomics of the Human Intestinal Tract consortium, the

Japanese Human Metagenome consortium, the Canadian Microbiome Initiative, and the Irish Metagenomics of the Elderly programme. [58]

In human, GI tract hosts the majority of microbial inhabitants, and the function of gut microbiome is currently best studied, compared with the other colonization sites. Numerous studies have indicated that gut microbiota, microbial metabolites, and their interaction with the host, are correlated with a wide array of physiological functions, [59] such as promoting the maturation of immune system [60] and regulating intestinal endocrine function. [61] Altered microbiome composition has been related to a myriad of diseases, including a) GI diseases (e.g., irritable bowel syndrome and inflammatory bowel disease [62, 63]), b) metabolic conditions (e.g., diabetes [64, 65] and obesity [66, 67]), c) neuropsychiatric disorders (e.g., depression, [68] schizophrenia, [69] autism spectrum disorder, [70] and attention-deficit/hyperactivity disorder [71]), d) neurodegenerative disease (e.g., Alzheimer's disease (AD) and dementia, [72] multiple sclerosis, [73] Parkinson's disease (PD), [74] and ALS [75]), and e) others, such as cancer and cancer treatment, [76, 77] stroke, [78] cardiovascular disease, [79] and brain injury. [80]

2.5.2 Gut microbiome and ALS

Due to the rapidly growing research on gut microbiome, evidence has accumulated to support its involvement in energy metabolism and neuroinflammation (both of which are key features of ALS) and in neurodegeneration in general. [1]

In order to provide a summary of the existing evidence on gut microbiome and ALS, we conducted a comprehensive review of the existing studies (see **Appendix** [1]). We found relatively consistent results from animal studies, including a) a shifted microbiome composition in the pre-symptomatic stage of ALS and after disease onset, b) microbial composition was unstable with disease progression, c) microbial metabolites had a role on modifying disease progression (e.g., butyrate and nicotinamide), and d) gut microbiome could modify motor function and affect survival through immune responses. However, publication bias might be a concern.

Although we observed similar findings in human studies, the results are weaker, compared with the findings from animal studies. Multiple reasons might contribute to such conflicting results, such as random error duo to small sample size, as well as different study populations, recruitment methods of cases and controls, sample processing procedures, handling of sequencing data, bioinformatics pipelines, and statistical methods.

We also discussed future perspectives in the microbiome research in human ALS**,** focusing on optimizing study size and design, standardizing approaches, investigating the interactions between gut microbiome and other factors, and considering gut microbiome in the prevention and therapy of ALS.

In this thesis, we focused on several biomedical factors closely related to gut microbiome, including GI dysfunction (**Study I**), antibiotics use (**Study II**) and hospital-treated infections (**Study III**), and explored their associations with risk of ALS.

2.6 POTENTIAL MECHANISMS FOR GUT MICROBIOME AND ALS

Although the precise mechanisms remain unclear, dysfunction of the gut-brain axis (GBA), which comprises the CNS, the autonomic nervous system (ANS), the enteric nervous system (ENS), the immune system, and the hypothalamic–pituitary–adrenal axis, might play a role in linking gut microbiome to ALS. [81-83] The GBA is bidirectional in terms of both anatomical and biochemical aspects. The CNS communicates with the ENS, muscles, and gut mucosa via the ANS, both sympathetic and parasympathetic, to regulate the GI motility and permeability, mucus secretion, and immunity, which in turn modulate the composition and function of gut microbiota. [82, 84] Meanwhile, gut microbiome transmits neurochemical signals to the CNS through its derived metabolites, which play key roles in the maintenance and modulation of CNS function, host metabolism and immune responses. [83, 84] We discuss below a few potential mechanisms linking gut microbiome and ALS in the context of the GBA, with a focus on pathways of relevance to energy metabolism, immune responses, and pathological protein aggregates (Figure 2.1).

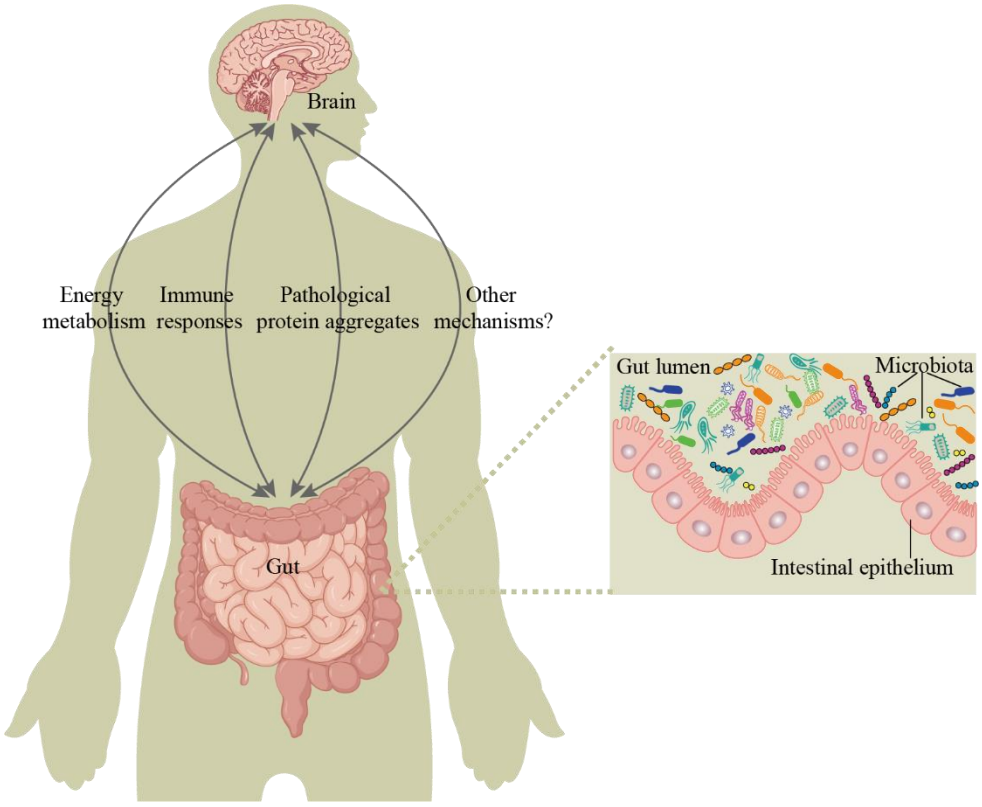

Figure 2.1 Potential mechanisms linking gut microbiome and ALS

2.6.1 Energy metabolism

Many ALS patients have been found to demonstrate disturbances in energy metabolism (e.g., hypermetabolism, weight loss, and diabetes). [85, 86] Impaired mitochondrial and glycolytic energy metabolism in the motor neurons and glia, reduced glucose uptake in the motor-sensory cortex, and insulin resistance have been found in ALS patients. [85, 87, 88] Gut microbiota can contribute to the host energy metabolism through their energy-yielding nutrients (e.g., complex carbohydrates, proteins, monosaccharides, short-chain fatty acids (SCFAs), and amino acids), [89] as well as via affecting host energy homeostasis and insulin sensitivity. [89]

Mitochondrial dysfunction and oxidative stress have been proposed as mechanisms underlying neurodegenerative disease including ALS, [90-92] which can be modulated by gut microbiota as well. [75, 93] Blancher et al. showed that *Akkermansia muciniphila* and nicotinamide treatment shared a binding site for the transcription factor nuclear respiratory factor-1, which is known to modulate mitochondrial biogenesis, electron transport chain activity, and oxidative stress. [75, 94] Another bacterial metabolite - Urolithin A - can pass through the blood-brain barrier (BBB) and induce mitophagy (a selective degradation of mitochondria by autophagy), which can subsequently prevent age-related mitochondrial dysfunction and increase muscle function in mice. [95] Two recent studies also suggested that lipopolysaccharide derived from Gram-negative bacteria might contribute to development of neurodegenerative disease through regulation of oxidative stress and inflammation. [96, 97] High nicotinamide adenine dinucleotide phosphate oxidase 2 activation has been shown among patients with neurodegenerative disease and is correlated with high levels of circulating lipopolysaccharide. [98, 99] Declined functional pathways involved in carbohydrate metabolism were also reported in the gut microbiota of ALS patients, compared with controls. [100]

2.6.2 Immune responses

Animal and human studies have shown that microbial dysbiosis can damage the integrity of gut epithelium and BBB. [101, 102] The intestinal epithelial layer forms a physical and biochemical barrier to prevent the infiltration of microbes and large molecules to the circulation. [102] Microbial dysbiosis has been shown to result in reduced expression of adherens and tight-junction proteins in gut epithelium and subsequently an increased intestinal permeability in ALS. [102-104] Probiotic *Escherichia coli Nissle* has been shown to reinforce the intestinal epithelial barrier via the redistribution of tight-junction proteins in the gut epithelium, [103] and butyrate treatment was suggested to improve the gut permeability in *SOD1*^{G93A} mice. [105] Germ-free mice were shown to have an increased permeability in the BBB during embryonic development, which might be restored by postnatal recolonization of microbiota or the administration of butyrate, [106] demonstrating therefore a causal role of gut microbiota in BBB development. [56] Similarly, gut microbial dysbiosis might increase the permeability of BBB by decreasing the expression of key tight-junction proteins in the brain microvascular endothelium. [101] As a result, it is possible that gut microbial dysbiosis leads to an increased transmission of immune cells and soluble molecules, such as hormones, neurotransmitters, proinflammatory cytokines, and metabolites - either host or microbial in origin - between the periphery and CNS through the increased permeability of the intestinal epithelium and BBB. [101]

Neuroinflammation is a well-established pathological feature of ALS. It is characterized by activation of microglia and astroglia in the CNS, infiltration of lymphocytes, monocytes and macrophages to the CNS, and upregulation of proinflammatory cytokines. [107] Genetic mutations related to ALS (e.g., *SOD1* and *C9orf72*) have been shown to promote inflammation-mediated motoneuron injury and death. [108] The gut microbiome plays an important role in modulating the maturation and function of immune cells in the periphery and CNS. [109] Decreased microbial diversity may result in defective microglia, whereas microbial recolonization might lead to partially recovered microglia feature. [109] SOD^{G93A} mice with a shifted microbial composition are characterized by an increased level of proinflammatory cytokine and an abnormal amount of Paneth cells in the intestine. [104] Paneth cells are specialized intestinal epithelial cells regulating autophagy activity and releasing antimicrobial peptides in response to pathogens. [104] Administration of butyrate was shown to decrease the percentage of abnormal Paneth cells and restore the antimicrobial peptide in the $SOD1^{G93A}$ mice. [105] Butyrate serves as a primary energy source for intestinal epithelial cells [110] and has an anti-inflammatory function through promoting regulatory T (Treg) cells in the colon by inhibiting histone deacetylase activity and inhibiting inflammatory cytokine production. [111] One study suggested that the SOD^{G93A} mice exhibited peripheral leukocyte change after microbial alteration, and the infiltration of leukocytes, microglia, CD8+ T cells, and neutrophils to the spinal cord increased along with disease progression. [112] Both positive (*Porphyromonadaceae*) and negative (*Lachnospiraceae*) correlations were noted between specific bacteria and microglial activation in the CNS. [112] Based on the *C9orf72* mouse model, Burberry A et al. showed an increased infiltration of peripheral immune cells to the CNS as the *C9orf72* function declined. [113] Further, the study showed that treatment with antibiotics modulated both immune cell infiltration and microglial activation, whereas transplantation of pro-survival gut microflora significantly improved these immune phenotypes. [113]

2.6.3 Pathological protein aggregates

A pathological hallmark of ALS is the aggregation of cytoplasmic proteins, including SOD1 and TDP-43, prominently but not exclusively in the motor neurons. [3] The misfolded SOD1 and TDP-43 can propagate within and between cells in a prion-like manner, [114] similar to α-synuclein in PD and amyloid-β and tau in AD. [115, 116] Although ALS, PD, and AD all affect primarily the CNS, the origin of these pathological protein aggregates might be outside the CNS. In the case of PD for example, a transmission of α-synuclein aggregates from GI tract to the substantial nigra, via the vagus nerve, has been demonstrated. [117] Our group [118] and others [119] have shown that vagotomy is associated with a lower future risk of PD. Studies have also revealed that in mice overexpressing α-synuclein, colonization with curli-producing bacteria *Escherichia coli* accelerated α-synuclein aggregation and exacerbated α-synuclein induced behavioral deficits, while oral treatment with a gut-restricted amyloid inhibitor alleviated the acceleration of pathologic and behavioral abnormalities. [120] These results suggest a possibility that carriage of particular bacteria may be a factor triggering or exacerbating neurodegenerative disease. Although direct evidence for such gut-to-brain transmission does not exist for ALS yet, ALS patients indeed

demonstrate symptoms related to GI dysfunction (e.g., delayed gastric emptying, constipation, abdominal pain, nausea, and satiety), [121] as patients with PD. [122]

3 RESEARCH AIMS

The overarching aim of this thesis is to explore the roles of different biomedical factors in the risk and prognosis of ALS. Toward this end, four constituent studies were conducted with the following specific aims (Figure 3.1).

Study I: To explore whether individuals with GI biopsy result of normal mucosa or non-specific inflammation have an altered consequent risk of ALS, and whether a GI biopsy result of normal mucosa or non-specific inflammation is associated with mortality risk in patients with ALS.

Study II: To assess the association between previous antibiotic use and ALS risk.

Study III: To investigate the association between previous hospital-treated infection and the risk of neurodegenerative disease, including AD, PD, and ALS.

Study IV: To examine the potential ability of eight commonly measured blood markers to serve as prognostic biomarkers in ALS.

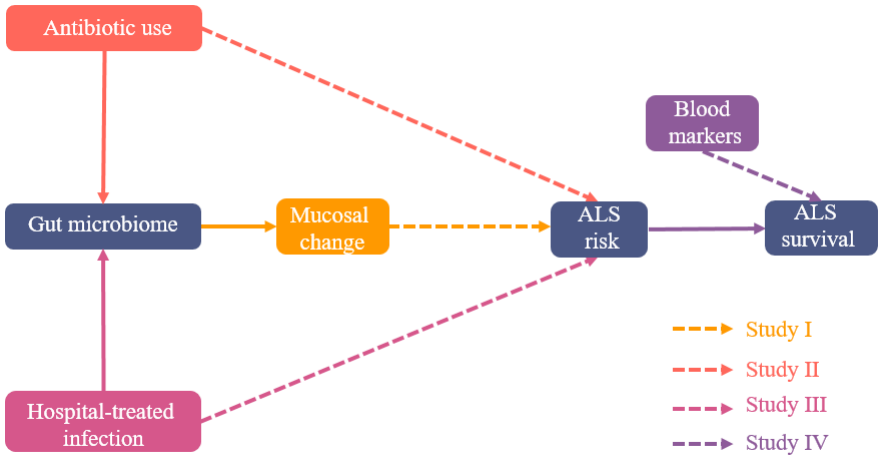

Figure 3.1 Overview of the thesis.

4 MATERIALS AND METHODS

4.1 OVERVIEW

The methods used in **Study I-IV** are summarized in Figure 4.1.

Study	Study design	Data source	Study population	Exposure & outcome	Statistical analyses
I	Matched cohort study	ESPRESSO; Swedish national healthcare registers	Individuals with a GI biopsy of normal mucosa or non-specific inflammation between 1965 and 2016 and their matched references	Exposure: normal mucosa or non-specific inflammation; Outcome: ALS risk and survival	Stratified Cox regression; Conditional logistic regression
II	Nested case-control study	Swedish national healthcare registers	2484 ALS patients diagnosed between July 2006 and December 2013 and their individually matched controls	Exposure: antibiotic use before diagnosis; Outcome: ALS risk	Conditional logistic regression
III	Nested case-control study	Swedish national healthcare registers	291,941 AD cases, 103,919 PD cases, and 10,161 ALS cases diagnosed between 1970 and 2016 and their individually matched controls	Exposure: hospital-treated infection; Outcome: risk of AD, PD or ALS	Conditional logistic regression
IV	Population-based cohort study	SCREAM; Swedish healthcare national registers	399 newly diagnosed ALS patients between 2006 and 2011 in Stockholm	Exposure: Eight commonly measured blood markers; Outcome: ALS survival	Cox regression; Joint model

Figure 4.1 Overview of methods used in **Study I-IV**.

4.2 DATA SOURCES

4.2.1 ESPRESSO

ESPRESSO contains all histopathological data in the GI tract, liver, gallbladder, and pancreas during 1965-2016 from all 28 pathology departments in Sweden (Figure 4.2). [123] A total of 2.1 million individuals with 6.1 million histopathological records were identified as one individual may have multiple biopsy records. The histopathological data include biopsy date, topography (where the biopsy was taken), morphology (biopsy appearance), and free text of the histopathology report. The morphology was assigned by the Swedish version of the Systematized Nomenclature of Medicine (SNOMED) system. Each individual with histopathological data was matched with up to five reference individuals randomly selected from the general population by age, sex, calendar year of biopsy, and county of residence. Then all study participants were linked to Swedish national healthcare registers to identify clinical diagnoses, dispensed medicine, migration status, and vital status. We identified the exposure information (i.e., the GI biopsy result of normal mucosa or non-specific inflammation, and biopsy location) from ESPRESSO in **Study I**.

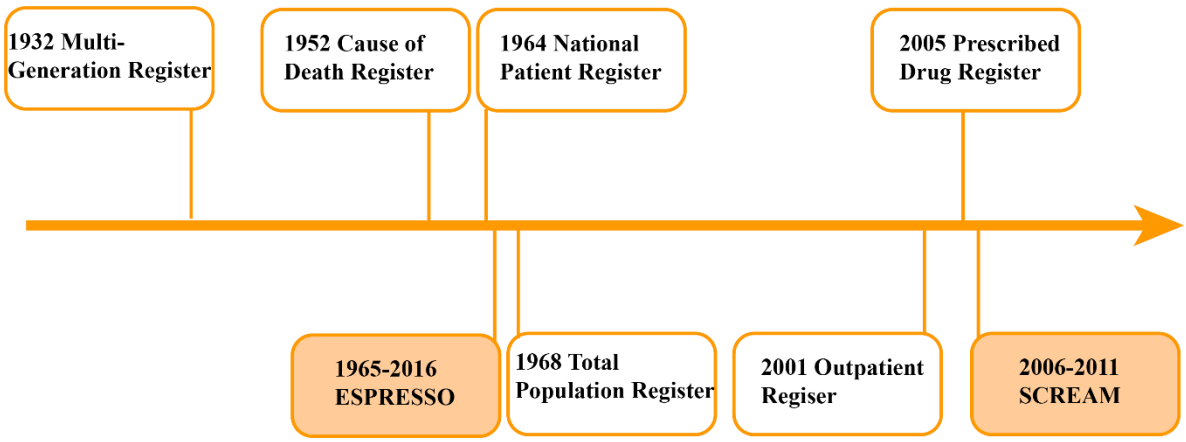

Figure 4.2 Time line of the data sources used in this thesis.

4.2.2 Swedish national healthcare registers

In **Study II** and **Study III**, we linked several Swedish national registers using the individually unique personal identity number, which is assigned to all residents staying at least one year in Sweden. [124] In **Study III**, we only enrolled individuals born after 1900 in Sweden whose parents were also born in Sweden.

Total Population Register (TPR), initiated in 1968 and maintained by Statistics Sweden, contains information on birth, family relationship, marital status, death, as well as immigration and emigration. [125] It is used to identify general population controls and to ascertain follow-up outcomes. Migration Register is a special register based in the TPR.

National Patient Register (NPR), maintained by the Swedish National Board of Health and Welfare (Socialstyrelsen, NBHW), covers inpatient care since 1964 (nationwide coverage since 1987) [126] and outpatient care since 2001 with a coverage of above 80%. Information on primary health care is not included in this register. The register includes information on dates of admission and discharge, primary and additional diagnoses, as well as procedures. This register was used to identify individuals with ALS (**Study I-IV**), number of healthcare visits, and clinical diagnoses of GI diseases (**Study I**).

Prescribed Drug Register (PDR), maintained by NBHW, covers information on all prescribed drugs dispensed in Sweden since July 2005, including date of prescription and dispensing, dispensed items (e.g., substance, brand name, and package), amount, dosage, and cost. [127] All medications are coded according to the Anatomical Therapeutic Chemical (ATC) classification system. However, it does not include vaccines and drugs used in the hospital or nursing home and does not include over-the-counter medications. This register was used to identify antibiotic use in **Study I** and riluzole use in **Study II**.

Cause of Death Register (CDR), initiated in 1952 and digitalized in 1961, contains information on date of death, underlying and contributing causes of death (based on ICD codes), place of death (e.g., hospital, nursing home), autopsy type (from 1992), and death

abroad (since 1987). [128] The register was used to define censoring in follow-up (**Study I-IV**) and define individuals died due to ALS (**Study I**).

Multi-Generation Register contains largely complete information on familial links for all individuals born in Sweden since 1932. [129] The register was used to identify parents and siblings of cases and their matched population controls in **Study III**.

4.2.3 SCREAM

SCREAM is repository of laboratory data of over 1.1 million adults with at least one creatinine test performed between 2006 and 2011 in Stockholm, Sweden, which represented 66% of the whole population in the region. [130] The coverage is higher for older adults (>75% for those aged 45-64 years, and >90% for ≥65 years), compared with the younger ones (50% for 18-44 years). All laboratory tests were performed by three different laboratories (Aleris, Unilabs and Karolinska), which provide over 90% of services to the Stockholm population. Information on date of test, test method, as well as unit of measurement was retrieved. SCREAM was also linked to Swedish national healthcare registers to identify clinical diagnoses, dispensed medicine, migration status, and vital status. SCREAM was used to identify levels of the eight commonly measured blood markers in **Study IV**.

4.3 STUDY DESIGN

4.3.1 Matched cohort study

Matching exposed to unexposed subjects in a constant ratio can eliminate confounding by the matching factors (Figure 4.3). [131] Matching can prevent an association between the exposure and matching factors at the initiation of follow-up. This original balance by matching however will be broken if the exposure and the matching factors have a role on censoring and outcome risk. Therefore, adjusting for matching factors in the analysis is necessary to obtain a valid effect estimate (e.g., risk difference or rate ratio).

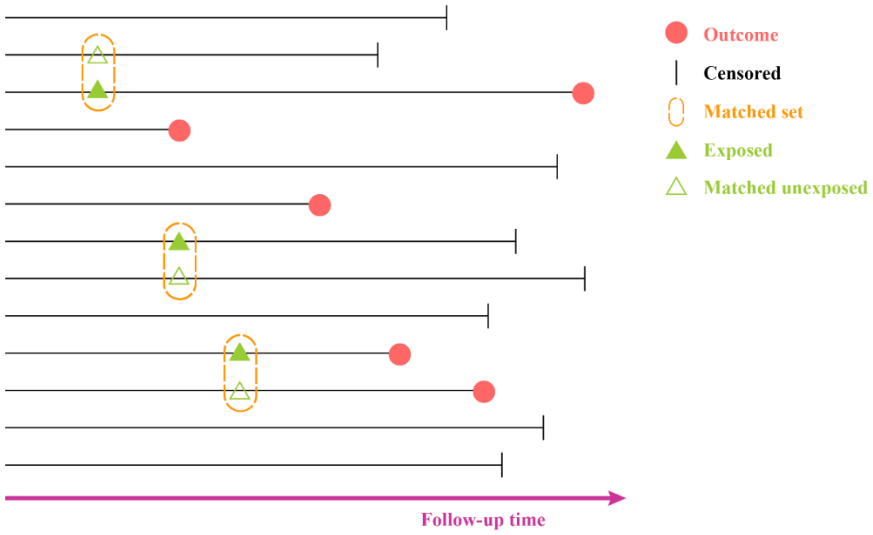

Figure 4.3 Illustration of the matched cohort study.

Study I: We performed a nationwide matched cohort study based on ESPRESSO. For individuals with a GI biopsy of normal mucosa (n=483,442) or non-specific inflammation (n=566,663) (exposed persons), we randomly selected up to five reference individuals (n=2,392,647 for normal mucosa; n=2,724,515 for non-specific inflammation) from the general population and individually matched them to the exposed person by age, sex, calendar year of biopsy, and county of residence. The reference individuals should be living in Sweden and have no GI biopsy record when being selected. Then, the exposed person and their matched references were followed from date of biopsy or selection until ALS diagnosis, emigration, death, or the end of study (end of 2016), whichever came first, through cross-linkages to the Swedish national healthcare registers.

4.3.2 Nested case-control study

Nested case-control study is a specific design conducted within an established cohort. It can produce the same findings with nearly the same precision compared with cohort study. In a nested case-control study, cases that developed the interested outcome at a given time point are matched to a random subset of participants who have not experienced the outcome at the time (controls). These controls may experience the interested outcome later in time and act as a control for other cases. Cases and controls are usually selected by incidence density sampling (risk-set sampling, Figure 4.4). [132] A nested case-control study design was applied in **Study II** and **Study III**.

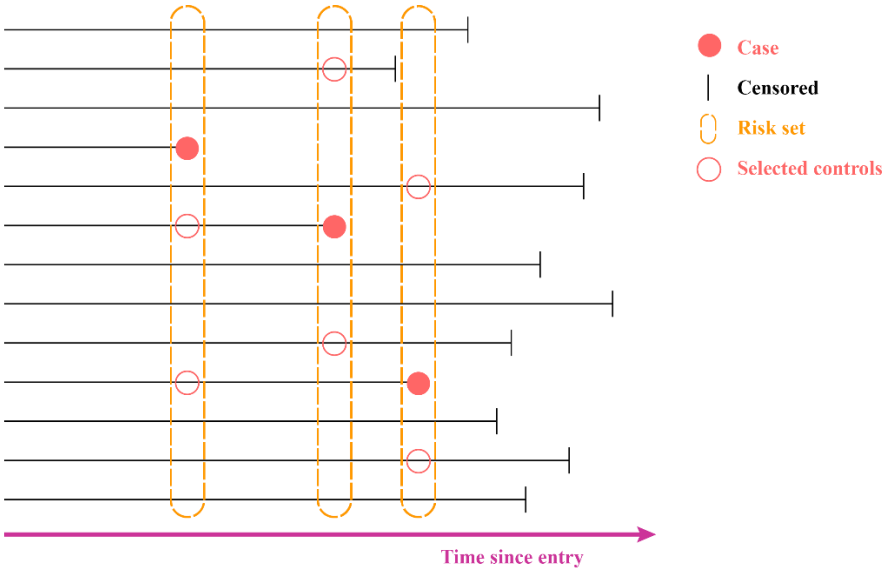

Figure 4.4 Illustration of the incidence density sampling in the nested case-control study.

Study II: Individuals who lived in Sweden on July 2006 were enrolled as the study population and followed from July 2006 until diagnosis of ALS, emigration, death, or December 2013, whichever came first, by cross-linkages to the Swedish national healthcare registers. Individuals with newly diagnosed ALS in the follow-up period were selected as the cases. For each case, five controls individually matched to the case by sex, year of birth, and area of residence were randomly selected from the general population via the method of incidence density sampling. Eligible controls were those who were alive and without ALS diagnosis

when being selected. The index date was defined as the date of ALS diagnosis for cases and date of selection for controls, respectively.

Study III: All individuals born after 1900 in Sweden whose parents were also born in Sweden were eligible for this study. We followed these individuals from January 1st 1970 until a diagnosis of neurodegenerative disease, emigration, death, or December 31st 2016, whichever came first, through cross-linkages to the Swedish national healthcare registers. Individuals with a newly diagnosed neurodegenerative disease, including AD, PD and ALS, were identified as the cases. Five controls per case, who were individually matched to the case by sex and year of birth, were randomly selected from the abovementioned study base using the method of incidence density sampling. Date of diagnosis and date of selection were used as index date for the cases and their matched controls, respectively. The controls should be alive and free of neurodegenerative disease at the index date.

4.3.3 Population based cohort study

In a cohort study, a disease-free study population is identified with their exposure status determined. The study population is then followed in time until the outcome of interest occurs, either prospectively or retrospectively. As a form of longitudinal study, the goal of a cohort is usually to measure and compare the incidence of the interested outcome between the exposed and unexposed populations. [131] A population-based retrospective cohort study was applied in **Study IV**.

Study IV: Individuals with a newly diagnosed ALS (n=399) in the SCREAM project were enrolled in a retrospective cohort study. The date when the diagnosis of ALS was issued was defined as the ALS diagnosis date. All patients were then followed from date of diagnosis until emigration out of Stockholm, death, or the end of 2011, whichever came first.

4.4 MEASUREMENTS

4.4.1 Exposure

Study I: The exposed group was defined as those with a GI biopsy result of normal mucosa or non-specific inflammation. We used the SNOMED codes M00110 and M00100 to identify normal mucosa, and M40000, M40400, M40460, M41000, M42000, M42100, M43000, M45000, and M47000 to identify non-specific inflammation. To minimize information bias, we treated GI biopsy result as a time-varying exposure (Figure 4.5). Namely, reference individuals contributed person-time to the reference group first, and then some of them contributed person-time to the exposed group if they received a GI biopsy during the follow-up.

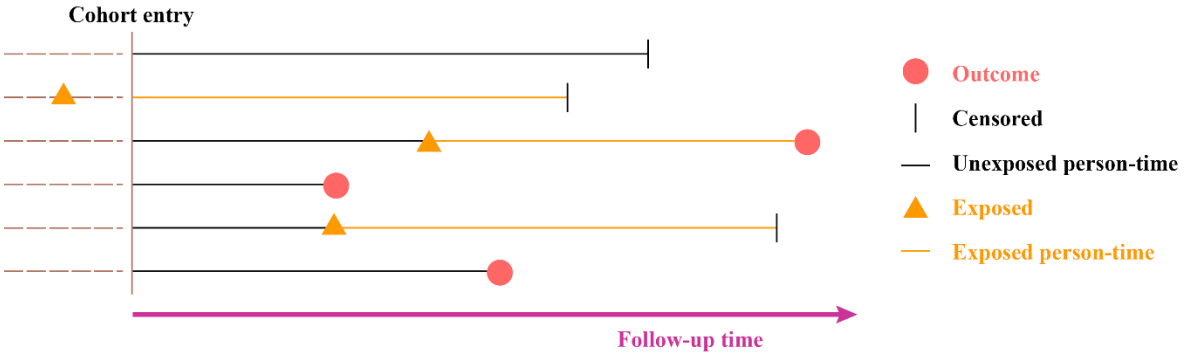

Figure 4.5 Illustration of the time-varying exposure of GI biopsy in **Study I**.

Study II: In this nested case-control study, the exposure is antibiotics use before the index date of cases and controls. Antibiotics use from July 2015 until the index date was ascertained by ATC codes (J01A-J01X) in the PDR. We studied any antibiotics use (classified as 0, 1, 2-3, or ≥4 prescriptions), antibiotics for respiratory infections and antibiotics for urinary tract or skin and soft tissue infections, as well as seven individual antibiotics as the exposure. The airway antibiotics included amoxicillin, penicillin, cephalosporin, and macrolides, while the urinary tract or skin and soft tissue antibiotics included pivmecillinam, sulfonamide, trimethoprim, norfloxacin, ciprofloxacin, nitrofurantoin, flucloxacillin, cloxacillin, and dicloxacillin. The seven individual antibiotics included tetracycline (J01AA), penicillin with extended spectrum (J01CA), beta-lactamase sensitive penicillin (J01CE), cephalosporin (J01DB and J01DC), trimethoprim (J01EA and J01EE), macrolides (J01FA), and fluoroquinolones (J01MA).

Study III: The exposure is the hospital-treated infection, either inpatient or outpatient, before the index date (Table 4.1). We studied any infection as a binary variable (the primary analysis) and then classified infection by type (bacterial, viral, or other infection) and site (CNS, gastrointestinal, genitourinary, respiratory, or skin infection). We additionally performed analysis by age at infection (<40, 40-59.9, or ≥60 years) and frequency of infections (0, 1, ≥2 events).

Study IV: The exposure is the serum or plasma levels of eight commonly measured blood markers at baseline and during the follow-up, including creatinine (lmol/l), albumin (g/l), haemoglobin (g/l), potassium (mmol/l), sodium (mmol/l), calcium (mmol/l), CRP (high-sensitivity type, mg/l), and glucose (mmol/l). Five markers were available for majority of the ALS patients, namely creatinine, haemoglobin, potassium, sodium, and CRP.

Table 4.1: The Swedish revisions of ICD codes for hospital-treated infection

	ICD-8 (1969-1986)	ICD-9 (1987-1996)	ICD-10 (1997-)
Bacterial	000, 001, 002, 003, 004, 005, 073, 076, 080, 081, 082, 083, 320, 362, 380, 381, 382, 383, 421, 461, 481, 482, 501, 510, 567, 590, 595, 597, 612, 613, 614, 616, 620, 622, 630, 635, 670, 678, 680, 681, 682, 684, 710, 720, 010, 011, 012, 013, 014, 015, 016, 017, 018, 019, 020, 021, 022, 023, 024, 025, 026, 027, 030, 031, 032, 033, 034, 035, 036, 037, 038, 039, 090, 091, 092, 093, 094, 095, 096, 097, 098, 099, 100, 101, 102, 103, 104, 390, 391, 392	001, 002, 003, 004, 005, 073, 076, 077, 078, 079, 080, 081, 082, 083, 320, 381, 382, 383, 383, 421, 461, 475, 481, 482, 510, 567, 590, 595, 597, 670, 730, 010, 011, 012, 013, 014, 015, 016, 017, 018, 020, 021, 022, 023, 024, 025, 026, 027, 030, 031, 032, 033, 034, 035, 036, 037, 038, 039, 040, 041, 090, 091, 092, 093, 094, 095, 096, 097, 098, 099, 100, 101, 102, 103, 104, 390, 391, 392, 614, 615, 616, 680, 681, 682, 683, 684, 685, 686	A00, A01, A02, A03, A04, A05, A15, A16, A17, A18, A19, A20, A21, A22, A23, A24, A25, A26, A27, A28, A30, A31, A32, A33, A34, A35, A36, A37, A38, A39, A40, A41, A42, A43, A44, A45, A46, A47, A48, A49, A50, A51, A52, A53, A54, A55, A56, A57, A58, A65, A66, A67, A68, A69, A70, A71, A72, A73, A74, A75, A76, A77, A78, A79, B95, B96, G00, G01, H60, H70, I00, I01, I02, I33, J01, J13, J14, J15, J36, J86, K65, L00, L01, L02, L03, L04, L05, L06, L07, L08, M00, M86, N30, N34, N70, N71, N72, N73, N74, N75, N76, N77, O23, O85, O86, P36
Viral	075, 360, 420, 422, 460, 464, 465, 466, 480, 040, 041, 042, 043, 044, 045, 046, 050, 051, 052, 053, 054, 055, 056, 057, 060, 061, 062, 063, 064, 065, 066, 067, 068, 070, 071, 072, 073, 074, 075, 076, 077, 078, 079, 470, 471, 472, 473, 474	070, 071, 072, 074, 075, 077, 078, 079, 372, 420, 422, 460, 464, 465, 466, 480, 487, 647, 711, 045, 046, 047, 048, 049, 050, 051, 052, 053, 054, 055, 056, 057, 060, 061, 062, 063, 064, 065, 066	A08, A60, A80, A81, A82, A83, A84, A85, A86, A87, A88, A89, A90, A91, A92, A93, A94, A95, A96, A97, A98, A99, B00, B01, B02, B03, B04, B05, B06, B07, B08, B09, B15, B16, B17, B18, B19, B20, B21, B22, B23, B24, B25, B26, B27, B28, B29, B30, B31, B32, B33, B34, B97, B99, H10, I30, I40, J00, J04, J05, J06, J10, J12, J20, J21, O98, P35, Z21

Others	084, 085, 086, 087, 088, 089, 112, 113, 114, 117, 120, 121, 122, 124, 125, 126, 128, 130, 131, 132, 133, 134, 135, 136, 363, 610, 611, 615, 732, 763, Y41	084, 085, 086, 087, 088, 112, 113, 114, 117, 118, 120, 121, 122, 124, 125, 126, 128, 130, 131, 132, 133, 134, 135, 136, 370, 675, 771, 137, 138, 139	V02, B37, B38, B45, B46, B47, B48, B49, B50, B51, B52, B53, B54, B55, B56, B57, B58, B60, B61, B62, B63, B64, B65, B66, B67, B68, B69, B70, B72, B73, B74, B75, B76, B77, B78, B79, B80, B83, B85, B86, B87 B88, B89, B90, B91, B92, B93, B94, H16, H32, M01, M02, M03, O91, P37, P38, P39, Z22, A59, A63, A64
CNS	013, 062, 063, 064, 065, 066, 071, 094, 292, 320, 323, 324, 390, 474, 040, 041, 042, 043, 044, 045, 046	013, 062, 063, 064, 071, 094, 320, 323, 326, 392, 045, 046, 047, 048, 049	A17, A80, A81, A82, A83, A84, A85, A86, A87, A88, A89, G00, G01, G02, G04, G05, I02
Gastrointestinal	014, 123, 127, 129, 540, 567, 000, 001, 002, 003, 004, 005, 006, 007, 008, 009	014, 123, 127, 129, 540, 567, 000, 001, 002, 003, 004, 005, 006, 007, 008, 009	A00, A01, A02, A03, A04, A05, A06, A07, A08, A09, B71, B81, B82, K35, K65, K67
Genitourinary	016, 590, 595, 597	016, 590, 595, 597	N30, N34, O23
Respiratory	010, 011, 012, 033, 034, 075, 115, 116, 490, 501, 503, 510, 460, 461, 462, 463, 464, 465, 466, 470, 471, 472, 473, 474, 480, 481, 482, 483, 484, 485, 486	010, 011, 012, 033, 034, 075, 115, 116, 473, 475, 487, 490, 510, 460, 461, 462, 463, 464, 465, 466, 480, 481, 482, 483, 484, 485, 486	A15, A16, A37, A38, B27, B39, B40, B41, B42, B44, B59, J00, J01, J02, J03, J04, J05, J06, J10, J12, J13, J14, J15, J16, J17, J18, J20, J21, J22, J32, J36, J40, J41, J42, J86, P23
Skin	110, 111, 050, 051, 052, 053, 054, 055, 056, 057, 680, 681, 682, 683, 684, 685, 686	110, 111, 050, 051, 052, 053, 054, 055, 056, 057, 680, 681, 682, 683, 684, 685, 686	B00, B01, B02, B03, B04, B05, B06, B07, B08, B09, B35, B36, B43, L00, L01, L02, L03, L04, L05, L06, L07, L08

4.4.2 Outcome

We mainly identified patients with ALS through the NPR by ICD codes (**Study I-IV**). In **Study II**, we additionally identified ALS through CDR and prescription of ALS medication (e.g., riluzole; ATC code: N07XX02). In **Study III**, we also identified AD and PD (Table 4.2).

Table 4.2. The Swedish revisions of ICD codes for neurodegenerative disease

Disease	ICD-8 (1969-1986)	ICD-9 (1987-1996)	ICD-10 (1997-)
AD	290	290A, 290B, 331A	F00, G30
PD	342	332A	G20
ALS	348,00	335C	G122

4.5 STATISTICAL METHODS

4.5.1 Cox proportional hazard model

Cox proportional hazard model has been widely applied in analyzing time-to-event data, to identify difference in survival due to a factor (e.g., treatment, intervention, or prognostic factor). [133] The model is a "semi-parametric" procedure as it does not make any assumption regarding the shape of the baseline hazard function. However, it requires the proportionality of the hazard, implying that the effect of a factor on the hazard is constant over time (i.e., hazards of exposed and non-exposed groups are proportional over time). The underlying time scale is automatically adjusted for when modeling a Cox model. The proportional hazard assumption however could be violated if time-dependent variables are adjusted for but without appropriate modeling. We applied the Cox proportional hazards model in **Study IV**.

Because we conducted a matched cohort study design in **Study I**, we applied the stratified Cox model to analyze the data. The stratified Cox model is a modification of the Cox proportional hazard model that enables to control for a variable that does not satisfy the proportional hazard assumption via "stratification". [134]

Study I: We applied the stratified Cox model to estimate the association between a GI biopsy result of normal mucosa or non-specific inflammation and risk of ALS, with attained age as the underlying time scale. We excluded the first two years of follow-up (in the main analysis) and first three or four years of follow-up (in the sensitivity analyses) from the analysis to alleviate the influence of surveillance bias and reverse causation caused by the known diagnostic delay of ALS.

Study IV: We applied the Cox proportional hazard model to assess the association of eight commonly measured blood marker levels at diagnosis with mortality risk among patients with ALS, with attained age as the underlying time scale.

4.5.2 Conditional logistic regression model

Conditional logistic regression model is same with logistic regression with the exception that its estimates are conditional on the matching set. Compared with logistic regression, there is no constant term in its output. The coefficient can be interpreted as the change in the log-odds of the outcome for every unit change in a variable while holding other covariates constant. [135] An odds ratio can be obtained by taking the exponential of the coefficient. We applied conditional logistic regression model in **Study I-III**.

Study I: As information on the indication for GI biopsy was unavailable in the study, we applied conditional logistic regression model to compare the diagnosed GI diseases during the five years prior to date of GI biopsy (exposed group) or date of selection (reference group).

Study II: Conditional logistic regression model was used in this nested case-control study to assess the relationship between previous antibiotics use and risk of ALS. We first explored the temporal pattern of the interested association by splitting the time before the index date into several periods: 0–1, 1–2, 2–3, 3–4, 4–5, and 5–9 years. Then we excluded antibiotics use within one (main analysis), or two or three (sensitivity analyses) years prior to the index date to decrease the influence of reverse causation due to diagnostic delay.

Study III: Conditional logistic regression model was applied to explore the association between hospital-treated infections and risk of neurodegenerative disease. Due to the known diagnostic delay of neurodegenerative disease, infections diagnosed during five years before the index date were excluded from the main analysis. The analyses were first performed for any infection and then by type and site of infection as well as by age at infection (<40 y, 40-59.9 y, and ≥60 y). Further, to examine a potential dose-response relationship within specific age windows, we also analyzed frequency of infections (0, 1, and ≥2 events) by age at infection. For any and specific infections, we stratified the analysis by sex (male or female) and age at index date (<60 y or ≥60 y) to assess whether the associations would differ between male and female or young and older individuals.

We also stratified the analysis for any infection by calendar period (1970-1986 vs 1987-2016) to estimate the impact of register coverage on the results. We further restricted the study sample to those without a family history of the disease to assess whether the associations would be modified by family history. To assess potential misclassification of neurodegenerative disease, we performed another analysis by defining neurodegenerative disease through at least two hospital visits concerning the same disease. Finally, to assess the robustness of the results to the 5-year lag time, we performed another sensitivity analysis by excluding the infections experienced during the ten years before the index date.

4.5.3 Joint model

Longitudinal study often includes repeated measurement and survival outcomes. Analyzing these data separately would be inefficient, as it does not effectively exploit the dependence between the process for repeated measurement and the hazard for survival, and may lead to an

biased result caused by measurement error. [136] Measurement error in a time-varying variable would bias the estimate from the survival analysis (e.g., Cox regression) towards null. However, the joint model, which allows the linear mixed effect model for repeated measurement data and Cox model for time-to-event data to be modeled together, can correct this bias. The longitudinal component usually characterizes the trajectory of a covariate over time with random effects. The random effects are then included in a survival model, so the relationship between the covariate trajectory and risk of time-to-event outcome can be estimated. [137] We applied the joint model in **Study IV**.

Study IV: A joint model was applied to estimate the association between the temporal change of eight blood markers after ALS diagnosis and risk of death. A linear mixed effect model was used for the repeated measurements, while a Cox model for the censored outcome.

4.5.4 Kaplan-Meier curve

Kaplan-Meier method is a way to calculate the survival probability in time-to-event data. We plotted the Kaplan-Meier curve in **Study I** and **Study IV**. We need to acknowledge however that Kaplan-Meier method only provides the unadjusted survival probability.

Study I: To investigate the role of a GI biopsy finding of normal mucosa or non-specific inflammation on the survival of patients with ALS, we compared the Kaplan-Meier curves between the exposed individuals and their matched references.

Study IV: We applied the Kaplan–Meier curve to compare the survival of ALS patients with a higher level of a specific marker at diagnosis with those with a lower level of that marker.

5 RESULTS

5.1 STUDY I

5.1.1 Population characteristics

Mean age at cohort entry was 44.44 and 50.82 years for individuals with a GI biopsy result of normal mucosa (≥60 years: 25.45%; female: 61.93%) and for individuals with a GI biopsy result of non-specific inflammation (≥60 years: 40.10%; female: 51.02%). During a mean follow-up time of ~10 years, a total of 367 and 442 individuals were diagnosed with ALS in these two exposed groups, 1,273 and 2,105 in their matched references, respectively (Table 5.1).

Table 5.1 Characteristics of cohort participants. Adapted from Table 1 in Paper I.

	NM	References for NM	NSI	References for NSI
N	483,442	2,392,647	566,663	2,724,515
Mean age	44.44	44.33	50.82	50.25
Aged ≥60, %	25.45	25.19	40.10	38.63
Female, %	61.93	61.85	51.02	50.87
Mean of follow-up years	9.97	11.40	9.26	12.18

5.1.2 Normal mucosa or non-specific inflammation and ALS risk or prognosis

To minimize the influence of surveillance bias and reverse causation, we excluded the first two years of follow-up after the biopsy and found that a GI biopsy result of normal mucosa was associated with a higher risk of ALS (HR=1.22; 95%CI: 1.04-1.42, *P*=0.0122). The increased risk was also noted in subgroup analysis by sex (male or female) and age at cohort entry (<60 or ⩾60 years). In contrast, no clear relationship was observed for a GI biopsy result of non-specific inflammation (Figure 5.1). Moreover, neither normal mucosa nor non-specific inflammation was associated with risk of death after ALS diagnosis.

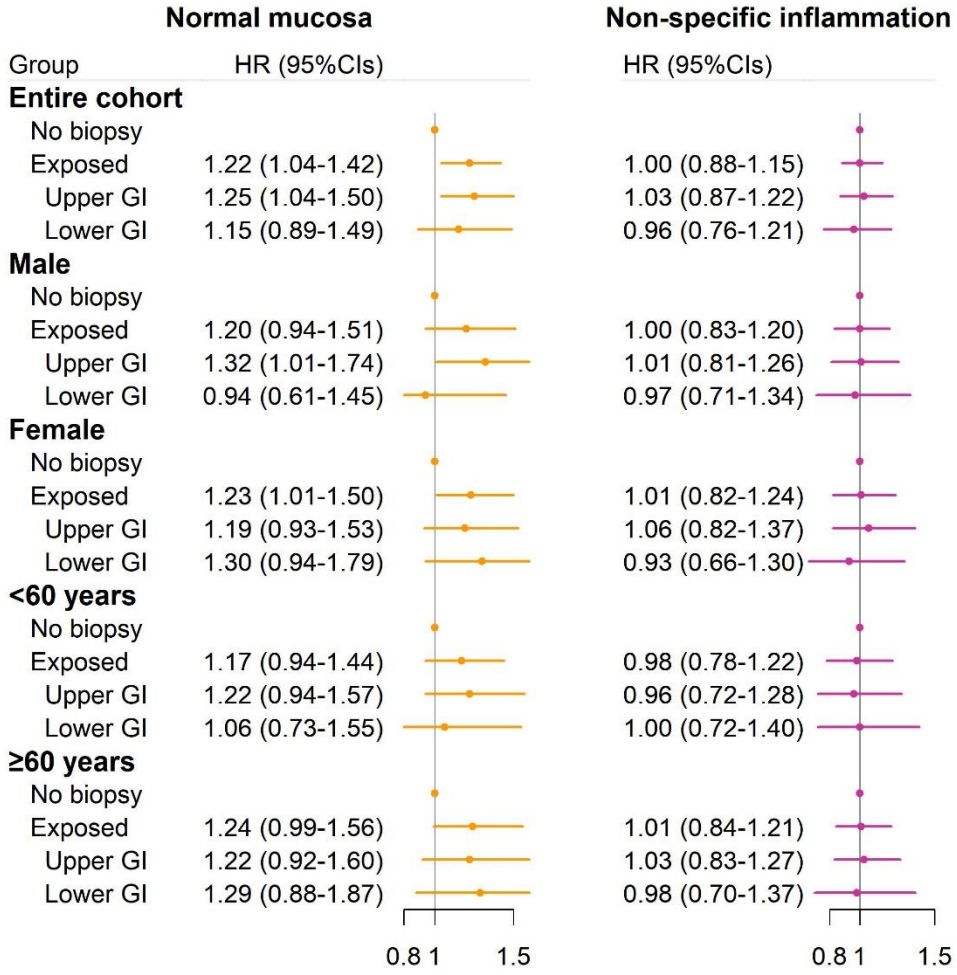

Figure 5.1 Risk of ALS among individuals with a GI biopsy result of normal mucosa or non-specific inflammation, compared with their matched reference individuals, adjusted for age at cohort entry, sex, county of residence, and number of healthcare visits (defined as number of healthcare visits from 2 years before to 1 year before biopsy, representing the regular frequency of healthcare visit). The first two years after the index date were excluded from the analysis. Figure was based on estimates from Table 2 and Table 3 in Paper I.

5.1.3 GI diagnoses before biopsy

Compared with their matched reference individuals, individuals with a GI biopsy result of either normal mucosa or non-specific inflammation had increased risks of GI diagnoses during five years before biopsy (Figure 5.2). Individuals with normal mucosa tended to get functional GI diagnoses, including irritable bowel syndrome, intestinal malabsorption, functional dyspepsia, and other functional intestinal disorders; while individuals with non-specific inflammation tended to get inflammatory GI diagnoses, such as esophagitis, gastritis and duodenitis, ulcerative colitis, Crohn's disease, peritonitis, and appendicitis.

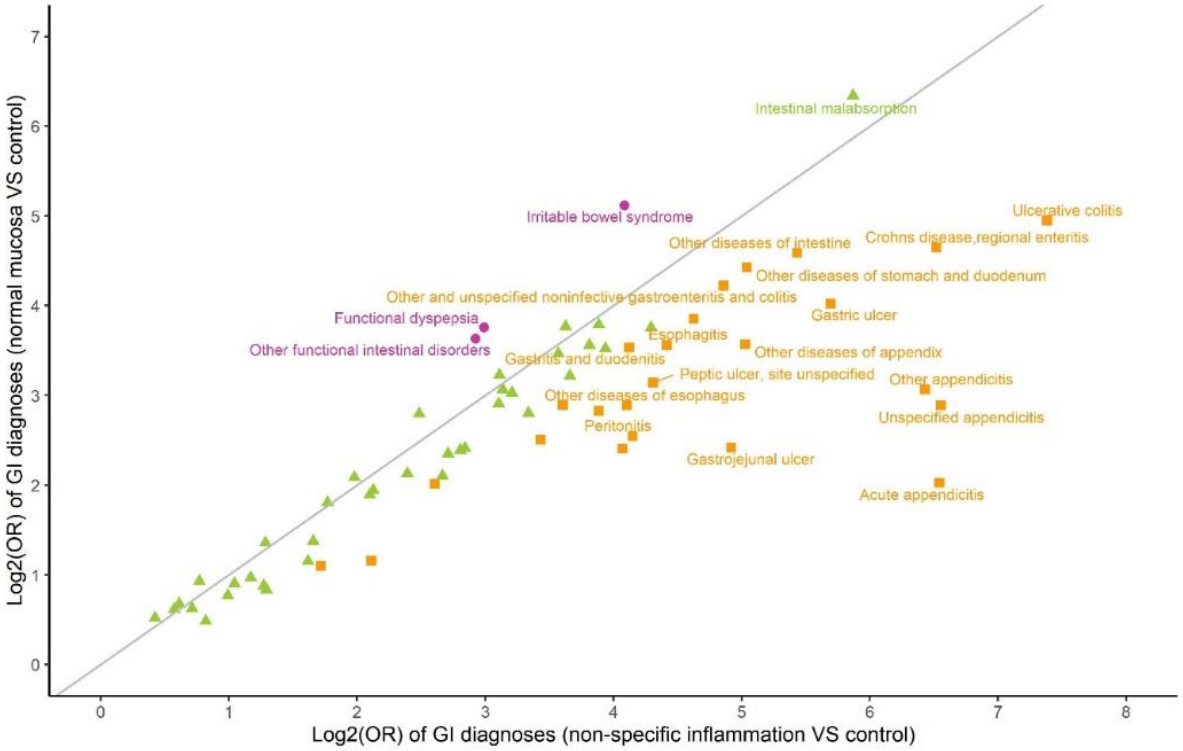

Figure 5.2 Scatter plot of log2 transformed odds ratios (ORs) of GI diseases during the five years prior to biopsy. ORs were obtained from conditional logistic regression model, and *P*-values for all ORs were less than 0.05. Purple circles: ORs are significantly greater for normal mucosa than for non-specific inflammation; green triangles: ORs are similar between normal mucosa and non-specific inflammation; brown squares: ORs are significantly smaller for normal mucosa than for non-specific inflammation. Adapted from Figure 2 in Paper I.

5.2 STUDY II

5.2.1 Population characteristics

A total of 2,484 patients and 12,420 matched controls were included in this nested case-control study. Male consisted 56.7% of all enrolled patients, and tended to be younger than female at diagnosis (Table 5.2).

Table 5.2. Characteristics of patients with ALS and their matched controls. Adapted from Table 1 in Paper II.

	Patients with ALS			Controls		
	Men	Women	Total	Men	Women	Total
N	1408	1076	2484	7040	5380	12420
Mean age at the index date, yrs [a]	67.9	69.7	68.6	67.9	69.7	68.6
Antibiotics use before the index date, %	62.9	67.9	65.1	58.6	65.9	61.8
Antibiotics use before the index date, % [b]	52.5	59.8	55.6	50.4	57.3	53.4

[a] Index date: date of diagnosis for patients with ALS and date of selection for controls.

[b] After excluding all antibiotics prescriptions within one year before the index date.

5.2.2 Antibiotic prescription before ALS diagnosis

Compared with their matched controls, patients with ALS were more likely to receive antibiotic prescriptions, which could be tracked back as early as six to eight years prior to diagnosis. Even larger difference was observed during the year before ALS diagnosis (Figure 5.3).

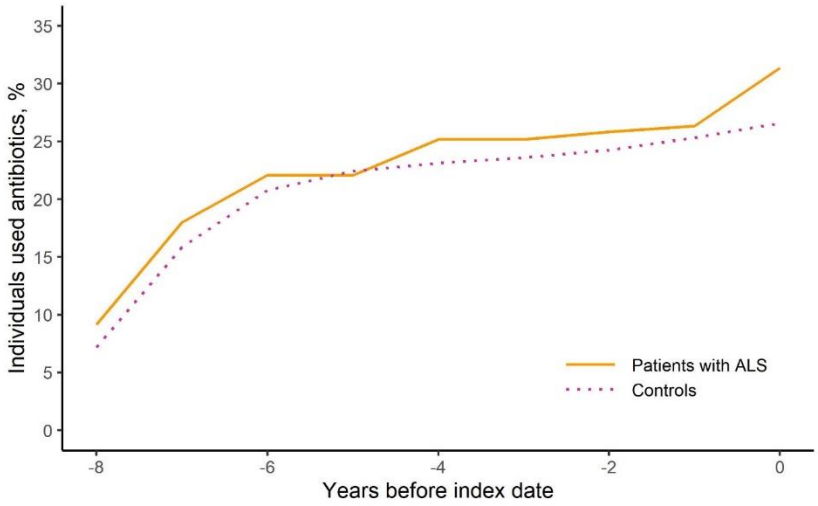

Figure 5.3 Percentage of individuals with antibiotic use among patients with ALS and their matched controls during the eight years before the index date. Adapted from Figure 1 in Paper II.

5.2.3 Antibiotic prescription and risk of ALS

Without considering the antibiotic prescriptions within one year before diagnosis, we observed a dose-response association between number of antibiotic prescriptions and risk of ALS (Figure 5.4, *P* for trend = 0.0069). The OR (95% CIs) were 1.06 (0.94–1.19), 1.13 (1.00–1.28), and 1.18 (1.03–1.35) for 1, 2-3, or $\geqslant$4 prescriptions of any antibiotics. Similar results were also observed when excluding antibiotic prescriptions within two or three years before diagnosis. A significant association was found for beta-lactamase sensitive penicillin as well (OR=1.28; 95% CI: 1.10–1.50, for more than two prescriptions).

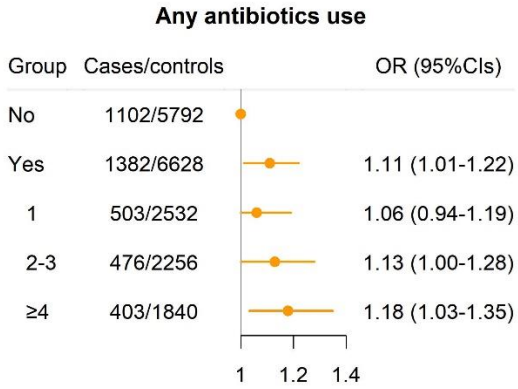

Figure 5.4 Association between any antibiotics use and risk of ALS after conditioned on the matching factors (age, sex, and area of residence). Any antibiotic prescriptions during the year before the index date were excluded.

5.3 STUDY III

5.3.1 Population characteristics

We enrolled a total of 291,941 AD cases, 103,919 PD cases and 10,161 ALS cases in the study. The mean age at diagnosis was 67.7, 71.5, and 67.7 for AD, PD, and ALS, and male accounted for 46.6%, 55.1%, and 56.8% of the cases, respectively. Compared with controls, individuals with neurodegenerative disease had a higher percentage of family history of neurodegenerative disease but lower comorbidity (Table 5.3).

Table 5.3 Characteristics of patients with neurodegenerative disease and their matched controls.

	AD		PD		ALS	
	Case	Control	Case	Control	Case	Control
N	291941	1459705	103919	519595	10161	50805
Mean age at the index date	67.7	67.7	71.5	71.5	67.7	67.7
Male, %	46.6	46.6	55.1	55.1	56.8	56.8
Family history of the disease, %	6.3	3.7	2.3	1.4	1.8	0.3
History of comorbidity [a], %	14.0	16.4	15.7	16.0	13.1	13.5

Index date: date of diagnosis for cases and date of selection for controls.

[a] Measured by the Charlson comorbidity index.

5.3.2 Hospital-treated infection and neurodegenerative disease

There was a slightly higher percentage of individuals with hospital-treated infections among individuals with AD or PD, but not ALS, during the 20 years before diagnosis, compared with their matched controls.

After excluding infections diagnosed during five years before the index date, we observed that an event of hospital-treated infection was associated with a higher risk of AD (OR=1.16; 95%CI: 1.15-1.18) and PD (OR=1.04; 95%CI: 1.02-1.06), but not ALS (OR=0.97; 95%CI: 0.92-1.03) (Figure 5.5). The associations for AD and PD were observed for bacterial, viral, or other infections as well as for CNS, gastrointestinal, and genitourinary infections. For AD and

PD, the associations were primarily limited to individuals diagnosed younger than 60 and a dose-response relationship was observed by number of infections at early age. Individuals with more than two events of infection before 40 had the highest risk of AD (OR=2.62; 95%CI: 2.52-2.72) and PD (OR=1.41; 95%CI: 1.29-1.53).

After excluding infections experienced during 10 years before the index date, we still observed positive associations between any and specific hospital-treated infections and risk of AD and PD. For example, for any infection, the OR of AD and PD was 1.18 (95%CI: 1.16-1.19) and 1.04 (95%CI: 1.01-1.06), respectively.

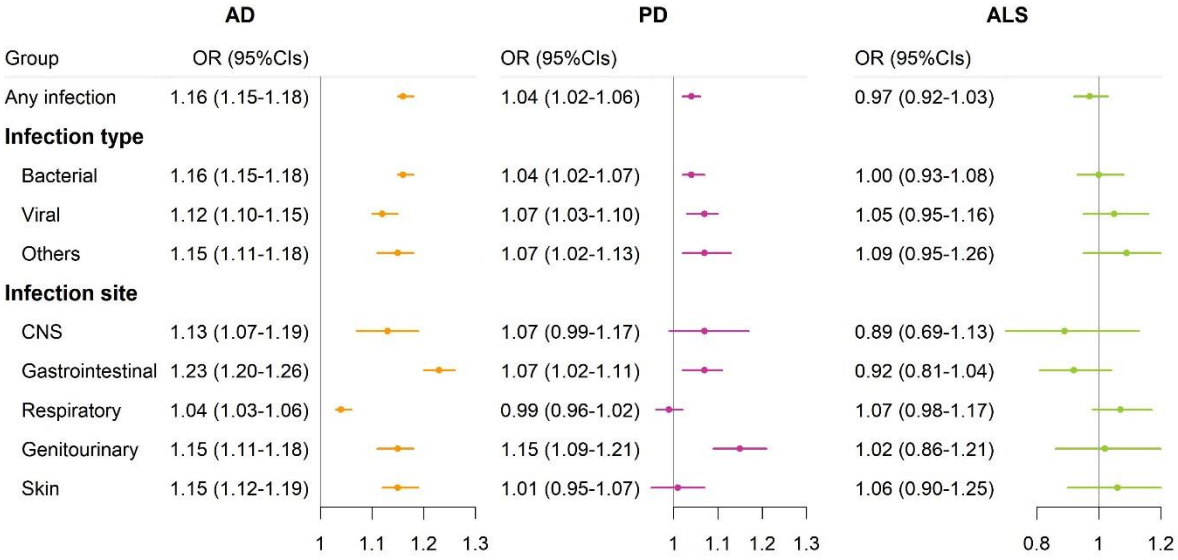

Figure 5.5 Associations between hospital-treated infections and the consequent risks of neurodegenerative diseases. Infections diagnosed during five years before the index date were excluded to alleviate the potential influence of reverse causation due to diagnostic delay.

5.4 STUDY IV

5.4.1 Population characteristics

A total of 399 individuals with a newly diagnosed ALS between 2006 and 2011 in Stockholm were included in this retrospective cohort study. Mean age at diagnosis was 66.25 years, and 231 of the patients were men (57.89%). During a mean of follow-up of 2.36 years, around 60% (239 patients) died (Table 5.4).

Table 5.4. Characteristics of patients with ALS. Adapted from Table 1 in Paper IV.

Characteristics		Median of first measurement	Mean (SD) of all measurements
No. of patients	399		
Age at diagnosis, mean (SD)	66.25 (12.47)		
Gender, n (%)			
Male	231 (57.89)		

	Female	168 (42.11)	
Biomarker			
	Serum creatinine, μmol/L, (n=399)	63.00	60.19 (28.16)
	Albumin, g/L, (n=269)	37.00	32.70 (6.03)
	Haemoglobin, g/L, (n=395)	140.00	128.84 (17.78)
	Potassium, mmol/L, (n=365)	4.00	3.99 (0.42)
	Sodium, mmol/L, (n=353)	140.00	139.24 (3.81)
	Calcium, mmol/L, (n=251)	2.32	2.28 (0.12)
	Log-CRP (mg/L), (n=329)	1.39	2.85 (1.70)
	Glucose, mmol/L, (n=284)	5.75	6.35 (1.95)

SD: standard deviation

5.4.2 Blood markers measured at baseline and mortality risk

We found that patients with a lower than median level of serum creatinine (HR=1.67; 95%CI: 1.31-2.12) or albumin (HR=1.49; 95%CI: 1.13-1.96) had an increased risk of mortality, compared with other patients. So did the patients with a higher than median level of log(CRP) (HR=1.33; 95%CI: 1.04-1.71) or glucose (HR=1.34; 95% CI: 1.01–1.78) at baseline. However, no clear association was found for haemoglobin, sodium, potassium, or calcium (Figure 5.6, left).

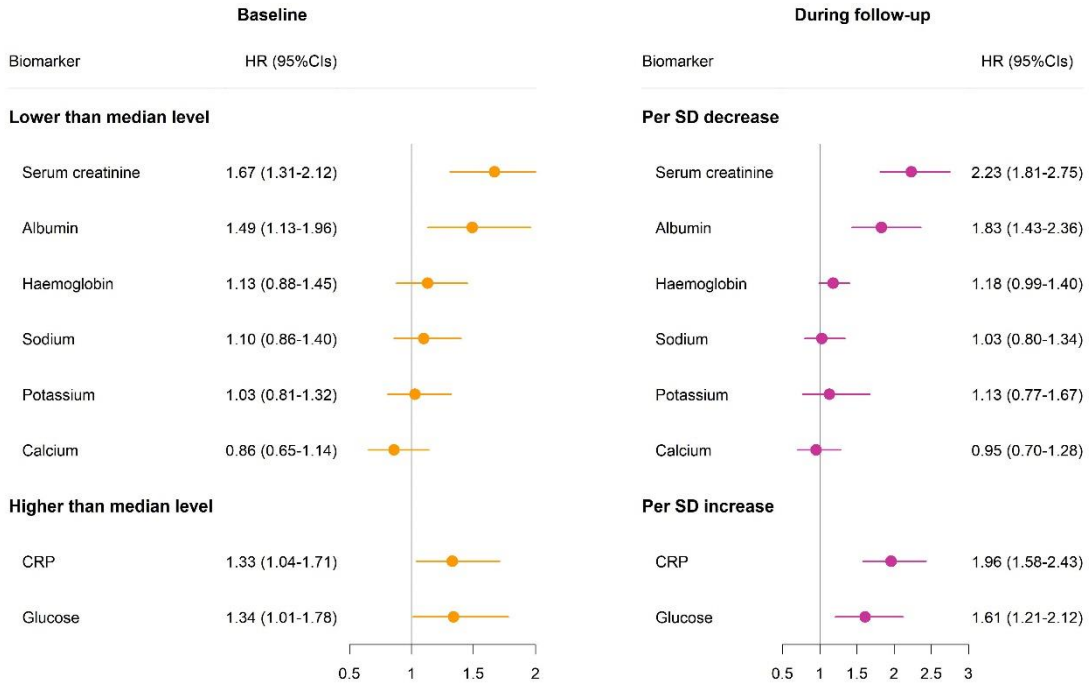

Figure 5.6 Associations of blood markers at baseline and during follow-up with the risk of mortality after ALS diagnosis. Adapted from Figure 1 and Figure 2 in Paper IV.

5.4.3 Temporal change of blood markers and mortality risk

By applying the joint model, we found that per SD decrease in serum creatinine (HR=2.23; 95% CI: 1.81-2.75), albumin (HR=1.83; 95% CI: 1.43–2.36), or haemoglobin (HR=1.18; 95%

CI: 0.99-1.40) after ALS diagnosis was related to an increased mortality risk; this was also a similar pattern for per SD increase in log(CRP) (HR=1.96; 95% CI: 1.58-2.43) and glucose (HR=1.61; 95% CI: 1.21-2.12) (Figure 5.6, right).

5.4.4 Change pattern of blood markers prior to death

Compared with patients with a slow progression (survived ≥ 3 years after diagnosis), a greater decline in the level of serum creatinine was observed among patients in the very fast (died within one year) or medium (survived 1-3 years) progression group. A similar change pattern was also found for albumin and haemoglobin. In contrast, a greater increase in the level of log(CRP) and glucose was observed during the months preceding death. No clear change pattern was found for potassium, sodium and calcium (Figure 5.7).

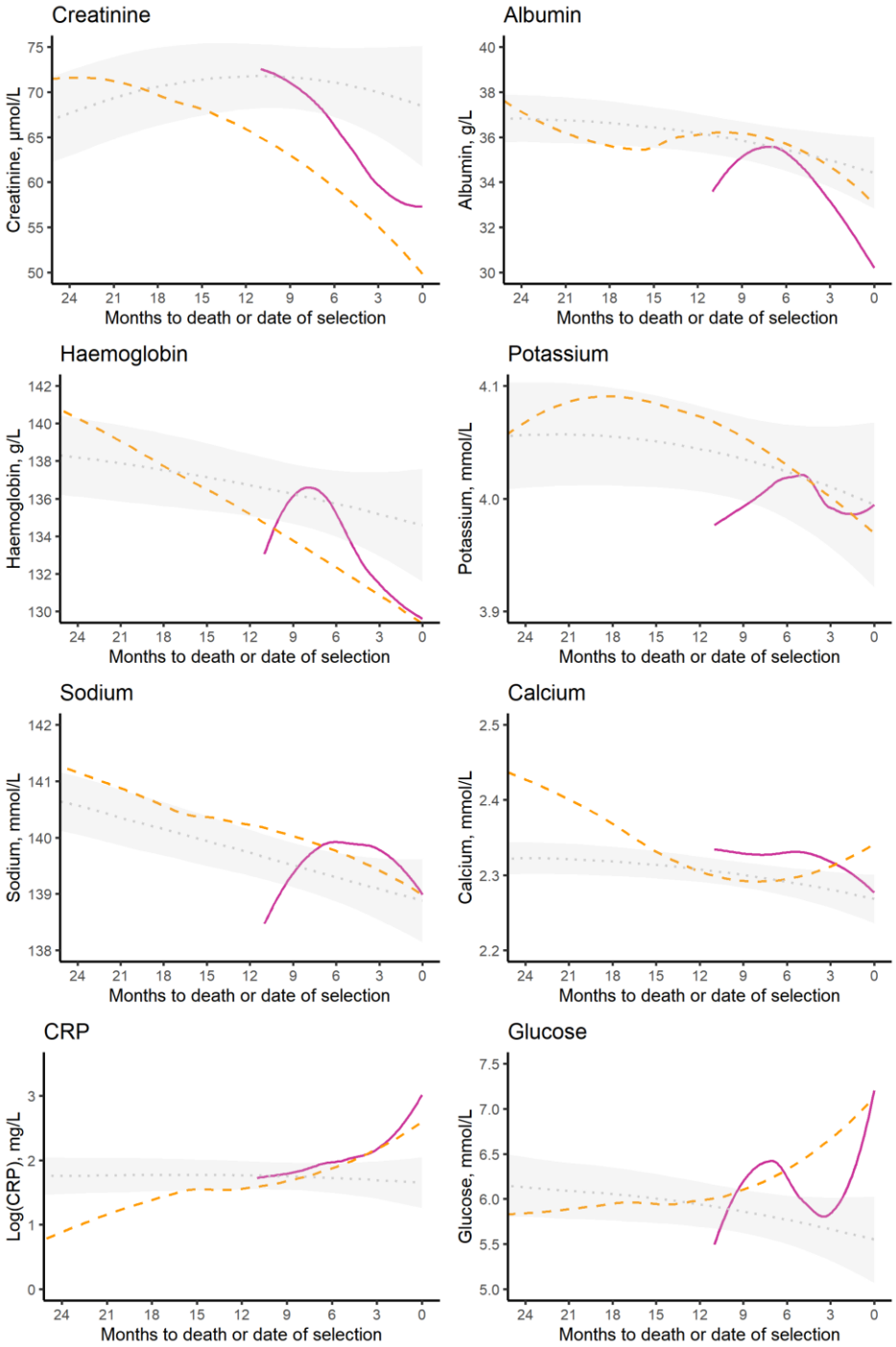

Figure 5.7 Profiles of blood markers prior to death for ALS patients with very fast progression (solid cerise curve, died within 1 year after diagnosis), medium progression (dashed orange curve, died within 1-3 years after diagnosis), or slow progression (dotted grey curve with 95% confidence interval, survived ≥ 3 years after diagnosis).

6 DISCUSSION

6.1 INTERPRETATION OF THE MAIN FINDINGS

6.1.1 GI biopsy and ALS

In **Study I**, we explored the association between GI biopsy result of normal mucosa or non-specific inflammation and ALS risk for the first time and found that a GI biopsy finding of normal mucosa was related to an increased subsequent ALS risk, whereas this was not the case for non-specific inflammation. Besides, either normal mucosa or non-specific inflammation had a role on survival after ALS diagnosis.

Our finding provides new, albeit weak, evidence for the involvement of GI dysfunction in ALS, either as a prodromal non-motor symptom or as a disease mechanism. [138] Previous studies have proposed several hypotheses to elucidate the involvement of GI dysfunction in neurodegenerative diseases, in terms of microbial dysbiosis and aggregations of misfolded proteins in ENS. [138]

We have little acknowledge about the reasons for the contrasting results between normal mucosa and non-specific inflammation. Chance finding could be possible. Therefore, validation for our findings from future studies are needed. If future studies from an independent population or studies for other neurodegenerative diseases verify our finding, we may hypothesize that a GI biopsy of normal mucosa, as a potentially distinct type of GI dysfunction without severe inflammation, might be specifically involved in neurodegenerative diseases (e.g., ALS). In a sensitivity analysis, we indeed noted that individuals with a GI biopsy of normal mucosa were more likely to receive a functional GI diagnoses during the five years before biopsy, while individuals with a GI biopsy result of non-specific inflammation had more inflammatory GI diagnoses.

6.1.2 Antibiotic prescription and ALS

Study II was the first population-based study to assess the association between antibiotics use and the consequent risk of ALS, and noted a dose-response relationship between them. The observed association did not differ between antibiotics used for respiratory infection and antibiotics used for urinary tract or skin and soft tissue infection, implying that the underlying mechanisms for the noted association are unlikely to be specific to certain organ systems.

This might provide evidence to the involvement of the altered gut microbiome in neurodegenerative diseases. Antibiotics have been suggested to greatly affect the microbial composition with a lasting effect and result in dysbiosis. [139] Broad-spectrum antibiotics can alter the abundance of 30% of the gut bacteria and trigger low microbial richness and diversity. [140] As suggested previously, [101, 141, 142] altered gut microbiome, together with its metabolites, could indeed contribute to neurodevelopmental and neurodegenerative diseases.

However, we have to acknowledge that disentangling the effect of antibiotics from the effect of the underlying indications for antibiotics use is almost impossible; therefore, this finding is more of an association rather than a causal relationship.

6.1.3 Hospital-treated infection and neurodegenerative disease

To our knowledge, **Study III** is the first to date to comprehensively assess the associations of hospital-treated infections with the risk of the three most common neurodegenerative diseases in the same study population. We found that hospital-treated infections were associated with an increased risk of AD and PD, especially cases diagnosed before age 60. The positive associations were observed across infection types and sites but were stronger for infections – especially repeated infections - in early or mid-life. These findings are novel and potentially important. We hypothesize that infectious events may serve as a trigger for disease onset and lead to a diagnosis of neurodegenerative disease at relatively early age (e.g., below 60), especially among individuals with underlying disease predisposition. [143-145]

Our finding has support from previous studies, both in animal and human research. Human studies have however mostly examined the role of a specific infection, e.g., herpesvirus for AD, [146] influenza, [147] hepatitis C virus, [148] and *Helicobacter pylori* [149] for PD, and the results are inconclusive. [1, 146, 150-153] Given that studies addressed different infections and neurodegenerative diseases, it is difficult to disentangle methodological drawbacks from real biological differences. A comprehensive examination of different infections, by infection types and sites, across different neurodegenerative diseases in a single study population helps to allay this concern, as we did in **Study III**.

This study did however not support an association of hospital-treated infections with the risk of ALS. As we used a specific definition of infections in the study, namely infections requiring inpatient or outpatient care, this null finding does not rule out the possibility that milder infections not attended by specialist care might still be of importance. Previous studies have indeed suggested that infections might contribute to protein aggregation and mislocalization as well as glutamate excitotoxicity, known pathological processes of ALS. [153] Enterovirus sequences have also been shown in the CNS of ALS patients, [154] whereas disturbed gut microbiome composition [1] and increased use of antibiotics [155] have also been suggested among patients with ALS.

6.1.4 Commonly measured blood markers as predictors of ALS prognosis

In **Study IV**, we explored the ability of eight commonly measured blood markers in predicting disease prognosis after ALS diagnosis and found that serum creatinine, albumin, CRP and glucose, either measured at the time of diagnosis or their temporal patterns after diagnosis, were associated with ALS prognosis. As majority of patients with ALS die within 1 to 3 years after diagnosis, our finding might help to guide patient care in clinical practice.

Our finding on creatinine is in line with previous studies, indicating that either baseline or longitudinal measurement of creatinine was associated with ALS prognosis. [51, 156, 157] One

study also found a correlation of creatinine with ALSFRS-R score and muscle strength. [157] Compared with previous findings, [51, 158] our study additionally supported a link between fast decline in albumin level after ALS diagnosis and a higher mortality risk. The finding that declining level of haemoglobin was related to an increased mortality risk supports the importance of maintaining nutritional status in the patients with ALS. The fast increasing level of CRP during the months prior to death of ALS patients supports the presence of altered immune responses and inflammation, [159, 160] which may be caused by respiratory infections in the later stage of ALS.

Our finding of glucose corroborates previous findings, [161, 162] supporting the involvement of altered glycolytic metabolism in ALS, including perhaps insulin resistance and glucose intolerance. [86, 163] Such dysregulation may be more notable in patients with worse prognosis. However, as type 2 diabetes might be related to a lower ALS risk or later onset of ALS, [164, 165] future studies are warranted to elucidate the underlying reasons for such contradiction.

This study provides new evidence for the importance of monitoring serum creatinine, albumin, CRP, and glucose in ALS care, especially given their easy access.

6.2 METHODOLOGICAL CONSIDERATIONS

Due to the nature of observational studies, three methodological considerations are discussed below.

6.2.1 Selection bias

Selection bias arises when the parameter of interest (e.g., prevalence for descriptive measures or risk ratio for effect measures) in a target population differs from the parameter in the subset of individuals from the target population that is available for analysis. [166]

In **Study I**, for the exposed individuals with a GI biopsy result of normal mucosa or non-specific inflammation, we randomly selected up to five reference individuals from the general Swedish population and individually matched them by age, sex, calendar year of biopsy, and county of residence. In **Study II** (antibiotics use) and **Study III** (hospital-treated infection), we applied the incidence density sampling method to select the controls from the general Swedish population that were individually matched to cases. As these matching factors (e.g., sex or year of birth) are unlikely affected by the exposure and the outcome, or intermediates between the exposure and the outcome, these selection methods are unlikely to introduce selection bias. [167] However, we have to acknowledge that the random selection method only enables us to generalize the effect measure to the population from which the references or controls are selected. In **Study IV**, as we only included individuals with a newly diagnosed ALS in Stockholm, the effect measure was only generalize to the Stockholm population who aged ≥18 years and had at least one measurement of serum creatinine from 2006 to 2011.

6.2.2 Measurement bias

Measurement error in either the exposure or the outcome is a common problem in observational studies. [168] It may be nondifferential if the measurement of the exposure does not depend on the true outcome conditional on the true exposure and vice versa. Otherwise, it might be differential. [169] If the exposure and the outcome are binary, then the independent nondifferential measurement error in the exposure and the outcome will bias the effect towards the null; this is however not the case if the exposure is polytomous, where the direction of the trend of an exposure could be changed. [170] The differential measurement error can bias the estimate to varying directions and distort the shape of the dose-response curve. [131]

Although the four constituent studies in this thesis should not be greatly affected by differential measurement error, we have to knowledge that nondifferential measurement error is a concern. The misclassification in the exposure and the outcome cannot be ruled out from these studies. In **Study I**, individuals with a GI biopsy result of normal mucosa or non-specific inflammation may simultaneously have other biopsy results. In **Study II**, misclassification error could be introduced as the PDR does not have information on antibiotics used before July 2005 and in the hospital, nursing home, and over-the-counter. In **Study III**, because of the incomplete coverage of inpatient care data before 1987 and lack of outpatient care data before 2001 in NPR, some individuals with hospital-treated infections may have been misclassified as not having infection. In addition, although register-based definitions have been shown to have satisfactory specificity for neurodegenerative diseases, the positive predictive values are low: 57% for AD, [171] 71% for PD, [172] and 91% for ALS, [164] suggesting that not all patients were captured in the register. Similar to above, such misclassification is also likely to dilute the associations toward null.

6.2.3 Confounding

Confounding is one of the central challenges in observational studies, especially when causal inference is the goal. [173] It could be caused by measured or unmeasured confounders. Confounders are factors that explain or produce all or part of the difference between the association and the effect that would be obtained in a counterfactual situation. [131] We acknowledge that confounding exists in these studies.

In **Study I**, **Study II** and **Study III**, due to the nature of register-based studies, although we adjusted for several confounders (e.g., age, sex, and county of residence), other potential risk or protective factors for ALS (e.g., smoking, BMI) were not adjusted for. In **Study IV**, the lack of clinical characteristics and genetic information in patients with ALS, some of which are known to affect disease progression, precluded the possibility to investigate whether the noted associations were independent of such factors.

6.3 ETHICAL CONSIDERATIONS

Source datasets used in the four studies were de-identified. The identity of individual study participant could therefore not be traced. According to Swedish regulations, informed consent from the participants was waived in studies based on national registers. The potential benefits of conducting these studies should believably outweigh possible hazards to personal or population integrity, through improving our understanding of ALS, a devastating disease.

Throughout the thesis work, datasets were handled according to the national legislations and General Data Protection Regulations, as well as supplementary legislations at the Karolinska Institutet (KI). All datasets were stored at secured networks at KI and protected by network firewall. Because the KI network is continuously scanned by an intruder detection system and all laptops in KI are fully encrypted, the source datasets were always under strict protection. It was unlikely that the research of the thesis work has caused secondary harm to the participants. Further, documentation for data extraction, analysis, and result interpretation was all archived according to KI rules to guarantee reproducibility of the study results.

The constituent studies in this thesis were approved by the Regional Ethical Review Board in Stockholm, Sweden (Dnr, 2014/1287-31/4; Dnr, 2011/917-31/2; Dnr, 2012/1814-31/4; Dnr, 2011/1730-31/2).

7 CONCLUSIONS

This thesis investigated the roles of biomedical factors on the risk and prognosis of ALS. The main conclusions of the four constituent studies were summarized below:

Study I: Individuals with a GI biopsy of normal mucosa were more likely to get functional GI diagnoses during the five years before biopsy; while individuals with a GI biopsy of non-specific inflammation tended to have more diagnoses of inflammatory GI diseases. A GI biopsy of normal mucosa was associated with an increased ALS risk, whereas no clear association was noted for non-specific inflammation. Neither normal mucosa nor non-specific inflammation was correlated with ALS prognosis.

Study II: Patients with ALS were more likely to have antibiotic prescriptions before diagnosis. Antibiotics use, especially repeated use, was associated with an increased risk of ALS. Positive association was also observed for beta-lactamase sensitive penicillin.

Study III: Individuals with neurodegenerative diseases were more likely to experience hospital-treated infections during the 20 years before diagnosis. After excluding infections experienced within five years before diagnosis to decrease the potential bias due to diagnostic delay, individuals with hospital-treated infections, especially in early and mid-life, were at an increased risk of developing AD and PD, especially those diagnosed below 60 years of age. No association was observed for ALS.

Study IV: Lower levels of serum creatinine and albumin whereas higher levels of CRP and glucose at diagnosis, as well as decreasing levels of creatinine and albumin and increasing levels of CRP and glucose after diagnosis, were associated with worse prognosis in ALS patients.

8 FUTURE PERSPECTIVE

Although these studies are based on the Swedish national healthcare registers, low incidence of ALS has led to a limited sample size in some analyses, making it difficult to differentiate a real signal form noise. Future studies with pooled data from different studies or countries should be encouraged.

Although contrasting result was noted between a GI biopsy of normal mucosa and non-specific inflammation in **Study I**, it is worthy of further investigation to understand whether this finding is caused by chance or also applies to other neurodegenerative diseases. Studies with information on the indications of GI biopsies may help to elucidate such contrasting finding. Because evidence has accumulated to suggest GI dysfunction as a prodromal symptom in different neurodegenerative diseases, [174, 175] it would be interesting to explore whether other GI diseases (e.g., inflammatory bowel disease, irritable bowel syndrome) are associated with ALS risk, as has been suggested in AD and PD. [176-180]

Future studies with longer study periods are needed to validate the influence of repeated antibiotics use on the ALS risk due to the relative short follow-up period in the Swedish PDR (from July 2006 to December 2013). Besides, efforts of using different study designs and elaborated measurements are warranted to disentangle the effect of antibiotics from the effect of underlying indications.

Future studies in independent populations are needed to validate our finding that infectious events might serve as a trigger for disease onset and lead to a diagnosis of neurodegenerative disease at relatively early age among individuals with underlying disease predisposition. If confirmed, discussion should be initiated to understand how we can best monitor or treat infections among such high-risk individuals. Future studies are also needed to better understand the roles of specific infectious agents, duration of infections, and treatments of infections (e.g., antibiotics, known to affect microbial environment and lead to dysbiosis) on the link between infections and neurodegenerative disease.

The value of adding commonly measured blood markers in prognostic prediction of ALS to the routine care of ALS also warrants further investigation due to their low cost and easy access. How to integrate such finding to support clinical decision-making and disease monitoring has also to be further investigated.

As the four constituent papers are observational, which are prone to measurement bias and confounding, future studies with the goal of causal inference such as Mendelian randomization are needed.

9 ACKNOWLEDGEMENTS

If I have seen further, it is by standing upon the shoulders of giants.

----Isaac Newton

This project derives from so many relationships, discussions, and mistakes over my PhD adventure. Many people have contributed to it. Without their help and support, I would not have accomplished this exciting, rewarding, and challenging journey.

I am deeply grateful to my principal supervisor: **Fang Fang**, who has encouraged me and provided invaluable support along the way. While our discussions were primarily about epidemiology and biostatistics and tried to discover the most reasonable study design, you have also taught me a lot and about how to think and act from a long-term perspective. Thank you for your understanding and strong support during my hardest period, I have never worried about the research work at that time because I have you. Thank you for making time for me whenever I had a question and supporting me whenever I wanted to attend a course or conference. Thank you for your persistent trust in me! I look forward to our future collaborations!

To my co-supervisor **Yudi Pawitan**. Every conversation we have had, either scientific or non-scientific, has given me food for thought. I have enjoyed them so much. Your experience, insight, and openness always inspire me and I am so lucky to have you! Thank you for taking me to taste the delicious food around Stockholm and having me in your (and **Marie Reilly's**) retirement ceremony. I am grateful that you love the snacks and noodles that I brought to you. Great thanks to your caring and warmth when I was sick. Your talk and messages cheered me up. Thank you for your smiling, big, and warm hug whenever we met. You are my super academic idol!

To my co-supervisor **Ulrika Zagai**. Thank you for all your valuable support, comments and suggestions for this project. It is a great pleasure to collaborate with you!

To my co-supervisor **Caroline Ingre**. I have truly enjoyed our collaboration and your clinical experience, commitment, and insight are valuable for this project. You are a great physician and I have learned a lot from you!

To **Hong Qiu**, the best mentor that I have hoped for. I appreciate your encouragement and feedback along the way. It is a great pleasure to have you here and when I worked in the pharmaceutical company.

To my co-authors: **Jonas F. Ludvigsson, Bjorn Roelstraete, Yiqiang Zhan, Daniela Mariosa, Henrik Larsson, Catarina Almqvist, Fredrik Piehl, Karin Wirdefeldt, Weimin Ye, Juan Jesus Carrero, Marie Evans, Tingting Huang, and Justine W. Debelius**. Thank you for your contributions to the papers as well as for providing valuable comments.

Thanks to **the Chinese Scholarship Council, the European Research Council, the Swedish Research Council, the US CDC, and Karolinska Institutet** for the financial support.

To my warm and brilliant colleagues in the research group: **Daniela Mariosa, Elisa Longinetti, Solmaz Yazdani, Can Cui, Yihan Hu, Donghao Lu, Ruoqing Chen, Qing Shen, Qianwei Liu, Kejia Hu, Shifeng Lian, Qian Yang, Lu Pan, Yihui Yang, Christina Seitz, Weiyao Yin, Anikó Lovik, Jacob Bergstedt, Emma Bränn, Fan Tian, Emily Joyce, Charilaos Chourpiliadis, Juan Gonzalez Hijon, Katarina Lyberg, Ulf Kläppe, Katja Fall, John Andersson, Yang Cao, Katarina Markus, Elgeta Hysaj, Jenny Deng, and Hang Yu**. Thank you for your contributions in making our unit such a great place to work. Your care, friendship and support for my PhD journal have been of benefit to me.

I am thankful to the "Big Data Analytics" book group: **Yunzhang Wang, Zheng Ning, Yinxi Wang, Xia Li, Lin Li, and Weiwei Bian**. Without your tireless and selfless contribution, our book that integrating four software (i.e., SAS, R, Stata, and Python) and comparing them in data cleaning, management, and analyses would have not been published in China. I hope the readers will find our book useful in their work.

Thanks to **Ruoqing Chen** for bringing me to KI. Thanks to **Shihua Sun** for sharing your experience and knowledge whenever I have encountered problems in my PhD journey. Special thanks to **Zheng Chang** for agreeing to chair my defense.

Thanks to my friends at KI for making this journey so much fun: **Ashley Tate, Bojing Liu, Bowen Tang, Cen Chen, Chen Wang, Chenxi Qin, Dang Wei, Erwei Zeng, Fen Yang, Ge Bai, Haomin Yang, Hong Xu, Hua Chen, Jet Termorshuizen, Ji Zhang, Jiangnan Liu, Jiayao Lei, Jie Song, Jingru Yu, Jingyan He, Le Zhang, Linghua Kong, Marco Trevisan, Marica Leone, Nanbo Zhu, Philippe Weitz, Ping Yuan, Qi Wang, Qingyang Xiao, Ruyue Zhang, Shengxin Liu, Shuyang Yao, Tong Jiao, Tyra Lagerberg, Wei He, Wenjiang Deng, Xia Shen, Xiaoying Kang, Xinhe Mao, Xu Chen, Yufeng Chen, Yuqi Zhang, and Yun Du**.

Yongbing Xiang, my main supervisor in the master program, I am grateful to your guidance and dedication to my education. Thanks for bringing the scientific world to me.

This book would never have been written had I not received support and encouragement from my beloved family. Thanks, **Mom** and **Dad**, for your continuous and selfless love, spoiling me and instilling persistence, resilience and patience in me. Without your encouragement, I am afraid that I would not have the courage to go abroad and follow my dreams. My **Gege** and **Jiejie**, thanks for taking care of and supporting me, as well as taking care of our parents when I am absent. **Lin's parents**, a big thanks for your love and unwavering support to us.

As always, I am most deeply indebted to **Lin** for your endless support and encouragement and for taking care of me. I am super lucky to have you as my closest comrade in life! **Ethan**, my greatest gift ever, all your smiles, hugs and kisses have kept me going. Nothing will be worth without you and **Lin**. Love you all!

10 REFERENCES

1. Sun, J., et al., *Gut microbiome and amyotrophic lateral sclerosis: A systematic review of current evidence.* J Intern Med, 2021. **290**(4): p. 758-788.
2. Taylor, J.P., R.H. Brown, Jr., and D.W. Cleveland, *Decoding ALS: from genes to mechanism.* Nature, 2016. **539**(7628): p. 197-206.
3. Brown, R.H. and A. Al-Chalabi, *Amyotrophic Lateral Sclerosis.* N Engl J Med, 2017. **377**(2): p. 162-172.
4. van Es, M.A., et al., *Amyotrophic lateral sclerosis.* Lancet, 2017. **390**(10107): p. 2084-2098.
5. Al-Chalabi, A., et al., *Amyotrophic lateral sclerosis: moving towards a new classification system.* Lancet Neurology, 2016. **15**(11): p. 1182-1194.
6. Hardiman, O., et al., *Amyotrophic lateral sclerosis.* Nat Rev Dis Primers, 2017. **3**: p. 17071.
7. Brooks, B.R., et al., *El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis.* Amyotroph Lateral Scler Other Motor Neuron Disord, 2000. **1**(5): p. 293-9.
8. de Carvalho, M., et al., *Electrodiagnostic criteria for diagnosis of ALS.* Clin Neurophysiol, 2008. **119**(3): p. 497-503.
9. Masrori, P. and P. Van Damme, *Amyotrophic lateral sclerosis: a clinical review.* Eur J Neurol, 2020. **27**(10): p. 1918-1929.
10. Zoccolella, S., et al., *Riluzole and amyotrophic lateral sclerosis survival: a population-based study in southern Italy.* Eur J Neurol, 2007. **14**(3): p. 262-8.
11. Abe, K., et al., *Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial.* Lancet Neurology, 2017. **16**(7): p. 505-512.
12. Marin, B., et al., *Variation in worldwide incidence of amyotrophic lateral sclerosis: a meta-analysis.* Int J Epidemiol, 2017. **46**(1): p. 57-74.
13. Longinetti, E. and F. Fang, *Epidemiology of amyotrophic lateral sclerosis: an update of recent literature.* Curr Opin Neurol, 2019. **32**(5): p. 771-776.
14. Leighton, D.J., et al., *Changing epidemiology of motor neurone disease in Scotland.* J Neurol, 2019. **266**(4): p. 817-825.
15. Arthur, K.C., et al., *Projected increase in amyotrophic lateral sclerosis from 2015 to 2040.* Nat Commun, 2016. **7**: p. 12408.
16. Nowicka, N., et al., *Risk Factors and Emerging Therapies in Amyotrophic Lateral Sclerosis.* Int J Mol Sci, 2019. **20**(11).
17. Chia, R., A. Chiò, and B.J. Traynor, *Novel genes associated with amyotrophic lateral sclerosis: diagnostic and clinical implications.* The Lancet Neurology, 2018. **17**(1): p. 94-102.

18. Ingre, C., et al., *Risk factors for amyotrophic lateral sclerosis.* Clin Epidemiol, 2015. **7**: p. 181-93.

19. Mathis, S., et al., *Genetics of amyotrophic lateral sclerosis: A review.* J Neurol Sci, 2019. **399**: p. 217-226.

20. Kametani, F., et al., *Mass spectrometric analysis of accumulated TDP-43 in amyotrophic lateral sclerosis brains.* Sci Rep, 2016. **6**: p. 23281.

21. Kwiatkowski, T.J., Jr., et al., *Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis.* Science, 2009. **323**(5918): p. 1205-8.

22. Selvaraj, B.T., et al., *C9ORF72 repeat expansion causes vulnerability of motor neurons to Ca(2+)-permeable AMPA receptor-mediated excitotoxicity.* Nat Commun, 2018. **9**(1): p. 347.

23. Lee, Y.B., et al., *Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic.* Cell Rep, 2013. **5**(5): p. 1178-86.

24. Lee, Y.B., et al., *C9orf72 poly GA RAN-translated protein plays a key role in amyotrophic lateral sclerosis via aggregation and toxicity.* Hum Mol Genet, 2017. **26**(24): p. 4765-4777.

25. Dickerson, A.S., et al., *Population-based study of amyotrophic lateral sclerosis and occupational lead exposure in Denmark.* Occup Environ Med, 2019. **76**(4): p. 208-214.

26. Fang, F., et al., *Blood Lead, Bone Turnover, and Survival in Amyotrophic Lateral Sclerosis.* Am J Epidemiol, 2017. **186**(9): p. 1057-1064.

27. Peters, S., et al., *Associations of Electric Shock and Extremely Low-Frequency Magnetic Field Exposure With the Risk of Amyotrophic Lateral Sclerosis.* Am J Epidemiol, 2019. **188**(4): p. 796-805.

28. Seelen, M., et al., *Long-Term Air Pollution Exposure and Amyotrophic Lateral Sclerosis in Netherlands: A Population-based Case-control Study.* Environ Health Perspect, 2017. **125**(9): p. 097023.

29. Tai, H., et al., *Military service and the risk of amyotrophic lateral sclerosis: A meta-analysis.* Journal of Clinical Neuroscience, 2017. **45**: p. 337-342.

30. Blecher, R., et al., *Contact Sports as a Risk Factor for Amyotrophic Lateral Sclerosis: A Systematic Review.* Global Spine J, 2019. **9**(1): p. 104-118.

31. Nakken, O., et al., *High BMI is associated with low ALS risk: A population-based study.* Neurology, 2019. **93**(5): p. e424-e432.

32. Lacorte, E., et al., *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**: p. 61-79.

33. Nieves, J.W., et al., *Association Between Dietary Intake and Function in Amyotrophic Lateral Sclerosis.* JAMA Neurol, 2016. **73**(12): p. 1425-1432.

34. Morozova, N., et al., *Diet and amyotrophic lateral sclerosis.* Epidemiology, 2008. **19**(2): p. 324-37.

35. Huisman, M.H., et al., *Effect of Presymptomatic Body Mass Index and Consumption of Fat and Alcohol on Amyotrophic Lateral Sclerosis.* JAMA Neurol, 2015. **72**(10): p. 1155-62.

36. Pupillo, E., et al., *Amyotrophic lateral sclerosis and food intake.* Amyotroph Lateral Scler Frontotemporal Degener, 2018. **19**(3-4): p. 267-274.

37. de Jong, S.W., et al., *Smoking, alcohol consumption, and the risk of amyotrophic lateral sclerosis: a population-based study.* Am J Epidemiol, 2012. **176**(3): p. 233-9.

38. D'Ovidio, F., et al., *Association between alcohol exposure and the risk of amyotrophic lateral sclerosis in the Euro-MOTOR study.* J Neurol Neurosurg Psychiatry, 2019. **90**(1): p. 11-19.

39. Hemani, G., et al., *The MR-Base platform supports systematic causal inference across the human phenome.* Elife, 2018. **7**.

40. Yavorska, O.O. and S. Burgess, *MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data.* Int J Epidemiol, 2017. **46**(6): p. 1734-1739.

41. Zhan, Y. and F. Fang, *Smoking and amyotrophic lateral sclerosis: A mendelian randomization study.* Ann Neurol, 2019. **85**(4): p. 482-484.

42. Bandres-Ciga, S., et al., *Shared polygenic risk and causal inferences in amyotrophic lateral sclerosis.* Ann Neurol, 2019. **85**(4): p. 470-481.

43. Zeng, P. and X. Zhou, *Causal effects of blood lipids on amyotrophic lateral sclerosis: a Mendelian randomization study.* Hum Mol Genet, 2019. **28**(4): p. 688-697.

44. Chio, A., et al., *Prognostic factors in ALS: A critical review.* Amyotroph Lateral Scler, 2009. **10**(5-6): p. 310-23.

45. Westeneng, H.-J., et al., *Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model.* The Lancet Neurology, 2018. **17**(5): p. 423-433.

46. Spataro, R., et al., *Tracheostomy mechanical ventilation in patients with amyotrophic lateral sclerosis: clinical features and survival analysis.* J Neurol Sci, 2012. **323**(1-2): p. 66-70.

47. Dreyer, P., et al., *Survival in ALS with home mechanical ventilation non-invasively and invasively: a 15-year cohort study in west Denmark.* Amyotroph Lateral Scler Frontotemporal Degener, 2014. **15**(1-2): p. 62-7.

48. Benjaminsen, E., et al., *Amyotrophic lateral sclerosis in Nordland county, Norway, 2000-2015: prevalence, incidence, and clinical features.* Amyotroph Lateral Scler Frontotemporal Degener, 2018. **19**(7-8): p. 522-527.

49. Hayashi, N., et al., *Prognosis of amyotrophic lateral sclerosis patients undergoing tracheostomy invasive ventilation therapy in Japan.* J Neurol Neurosurg Psychiatry, 2020.

50. Vijayakumar, U.G., et al., *A Systematic Review of Suggested Molecular Strata, Biomarkers and Their Tissue Sources in ALS.* Front Neurol, 2019. **10**: p. 400.

51. Chio, A., et al., *Amyotrophic lateral sclerosis outcome measures and the role of albumin and creatinine: a population-based study.* JAMA Neurol, 2014. **71**(9): p. 1134-42.

52. Rosenbohm, A., et al., *Association of Serum Retinol-Binding Protein 4 Concentration With Risk for and Prognosis of Amyotrophic Lateral Sclerosis.* JAMA Neurol, 2018. **75**(5): p. 600-607.

53. Lunetta, C., et al., *Serum C-Reactive Protein as a Prognostic Biomarker in Amyotrophic Lateral Sclerosis.* JAMA Neurol, 2017. **74**(6): p. 660-667.

54. Gaiani, A., et al., *Diagnostic and Prognostic Biomarkers in Amyotrophic Lateral Sclerosis: Neurofilament Light Chain Levels in Definite Subtypes of Disease.* JAMA Neurol, 2017. **74**(5): p. 525-532.

55. Thompson, A.G., et al., *Cerebrospinal fluid macrophage biomarkers in amyotrophic lateral sclerosis.* Ann Neurol, 2018. **83**(2): p. 258-268.

56. Cryan, J.F., et al., *The Microbiota-Gut-Brain Axis.* Physiol Rev, 2019. **99**(4): p. 1877-2013.

57. Human Microbiome Project, C., *A framework for human microbiome research.* Nature, 2012. **486**(7402): p. 215-21.

58. Proctor, L., *Priorities for the next 10 years of human microbiome research.* Nature, 2019. **569**(7758): p. 623-625.

59. Lynch, S.V. and O. Pedersen, *The Human Intestinal Microbiome in Health and Disease.* N Engl J Med, 2016. **375**(24): p. 2369-2379.

60. Rooks, M.G. and W.S. Garrett, *Gut microbiota, metabolites and host immunity.* Nat Rev Immunol, 2016. **16**(6): p. 341-52.

61. Neuman, H., et al., *Microbial endocrinology: the interplay between the microbiota and the endocrine system.* FEMS Microbiol Rev, 2015. **39**(4): p. 509-21.

62. Lloyd-Price, J., et al., *Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases.* Nature, 2019. **569**(7758): p. 655-662.

63. Franzosa, E.A., et al., *Gut microbiome structure and metabolic activity in inflammatory bowel disease.* Nat Microbiol, 2019. **4**(2): p. 293-305.

64. Vatanen, T., et al., *The human gut microbiome in early-onset type 1 diabetes from the TEDDY study.* Nature, 2018. **562**(7728): p. 589-594.

65. Zhou, W., et al., *Longitudinal multi-omics of host–microbe dynamics in prediabetes.* Nature, 2019. **569**(7758): p. 663-671.

66. Zhi, C., et al., *Connection between gut microbiome and the development of obesity.* Eur J Clin Microbiol Infect Dis, 2019. **38**(11): p. 1987-1998.

67. Turnbaugh, P.J., et al., *An obesity-associated gut microbiome with increased capacity for energy harvest.* Nature, 2006. **444**(7122): p. 1027-31.

68. Hao, W.Z., et al., *A review of antibiotics, depression, and the gut microbiome.* Psychiatry Res, 2019: p. 112691.

69. Golofast, B. and K. Vales, *The connection between microbiome and schizophrenia.* Neurosci Biobehav Rev, 2020. **108**: p. 712-731.

70. Hsiao, E.Y., et al., *Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders.* Cell, 2013. **155**(7): p. 1451-63.

71. Bull-Larsen, S. and M.H. Mohajeri, *The Potential Influence of the Bacterial Microbiome on the Development and Progression of ADHD.* Nutrients, 2019. **11**(11).

72. Sochocka, M., et al., *The Gut Microbiome Alterations and Inflammation-Driven Pathogenesis of Alzheimer's Disease-a Critical Review.* Mol Neurobiol, 2019. **56**(3): p. 1841-1851.

73. Jangi, S., et al., *Alterations of the human gut microbiome in multiple sclerosis.* Nat Commun, 2016. **7**: p. 12015.

74. Sampson, T.R., et al., *Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease.* Cell, 2016. **167**(6): p. 1469-1480 e12.

75. Blacher, E., et al., *Potential roles of gut microbiome and metabolites in modulating ALS in mice.* Nature, 2019.

76. Skelly, A.N., et al., *Mining the microbiota for microbial and metabolite-based immunotherapies.* Nat Rev Immunol, 2019. **19**(5): p. 305-323.

77. Kroemer, G. and L. Zitvogel, *Cancer immunotherapy in 2017: The breakthrough of the microbiota.* Nat Rev Immunol, 2018. **18**(2): p. 87-88.

78. Benakis, C., et al., *Commensal microbiota affects ischemic stroke outcome by regulating intestinal gammadelta T cells.* Nat Med, 2016. **22**(5): p. 516-23.

79. Jie, Z., et al., *The gut microbiome in atherosclerotic cardiovascular disease.* Nat Commun, 2017. **8**(1): p. 845.

80. Urban, R.J., et al., *Altered Fecal Microbiome Years after Traumatic Brain Injury.* J Neurotrauma, 2020.

81. Bravo, J.A., et al., *Communication between gastrointestinal bacteria and the nervous system.* Curr Opin Pharmacol, 2012. **12**(6): p. 667-72.

82. Carabotti, M., et al., *The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems.* Ann Gastroenterol, 2015. **28**(2): p. 203-209.

83. Tilocca, B., et al., *Gut-Brain Axis and Neurodegeneration: State-of-the-Art of Meta-Omics Sciences for Microbiota Characterization.* Int J Mol Sci, 2020. **21**(11).

84. Rhee, S.H., C. Pothoulakis, and E.A. Mayer, *Principles and clinical implications of the brain-gut-enteric microbiota axis.* Nat Rev Gastroenterol Hepatol, 2009. **6**(5): p. 306-14.

85. Vandoorne, T., K. De Bock, and L. Van Den Bosch, *Energy metabolism in ALS: an underappreciated opportunity?* Acta Neuropathol, 2018. **135**(4): p. 489-509.

86. Dupuis, L., et al., *Energy metabolism in amyotrophic lateral sclerosis.* Lancet Neurol, 2011. **10**(1): p. 75-82.

87. Pagani, M., et al., *Functional pattern of brain FDG-PET in amyotrophic lateral sclerosis.* Neurology, 2014. **83**(12): p. 1067-74.

88. Allen, S.P., et al., *C9orf72 expansion within astrocytes reduces metabolic flexibility in amyotrophic lateral sclerosis.* Brain, 2019. **142**(12): p. 3771-3790.

89. Fan, Y. and O. Pedersen, *Gut microbiota in human metabolic health and disease.* Nat Rev Microbiol, 2021. **19**(1): p. 55-71.

90. Niedzielska, E., et al., *Oxidative Stress in Neurodegenerative Diseases.* Mol Neurobiol, 2016. **53**(6): p. 4094-4125.

91. Smith, E.F., P.J. Shaw, and K.J. De Vos, *The role of mitochondria in amyotrophic lateral sclerosis.* Neurosci Lett, 2019. **710**: p. 132933.

92. Greco, V., et al., *Crosstalk Between Oxidative Stress and Mitochondrial Damage: Focus on Amyotrophic Lateral Sclerosis.* Adv Exp Med Biol, 2019. **1158**: p. 71-82.

93. Saint-Georges-Chaumet, Y., et al., *Targeting microbiota-mitochondria inter-talk: Microbiota control mitochondria metabolism.* Cell Mol Biol (Noisy-le-grand), 2015. **61**(4): p. 121-4.

94. Golko-Perez, S., et al., *A Novel Iron Chelator-Radical Scavenger Ameliorates Motor Dysfunction and Improves Life Span and Mitochondrial Biogenesis in SOD1(G93A) ALS Mice.* Neurotox Res, 2017. **31**(2): p. 230-244.

95. Ryu, D., et al., *Urolithin A induces mitophagy and prolongs lifespan in C. elegans and increases muscle function in rodents.* Nat Med, 2016. **22**(8): p. 879-88.

96. Zhao, Y., V. Jaber, and W.J. Lukiw, *Secretory Products of the Human GI Tract Microbiome and Their Potential Impact on Alzheimer's Disease (AD): Detection of Lipopolysaccharide (LPS) in AD Hippocampus.* Front Cell Infect Microbiol, 2017. **7**: p. 318.

97. Quigley, E.M.M., *Microbiota-Brain-Gut Axis and Neurodegenerative Diseases.* Curr Neurol Neurosci Rep, 2017. **17**(12): p. 94.

98. Zhan, X., et al., *Gram-negative bacterial molecules associate with Alzheimer disease pathology.* Neurology, 2016. **87**(22): p. 2324-2332.

99. Loffredo, L., et al., *Oxidative Stress and Gut-Derived Lipopolysaccharides in Neurodegenerative Disease: Role of NOX2.* Oxid Med Cell Longev, 2020. **2020**: p. 8630275.

100. Zeng, Q., et al., *The alteration of gut microbiome and metabolism in amyotrophic lateral sclerosis patients.* Sci Rep, 2020. **10**(1): p. 12998.

101. Sharon, G., et al., *The Central Nervous System and the Gut Microbiome.* Cell, 2016. **167**(4): p. 915-932.

102. Thevaranjan, N., et al., *Age-Associated Microbial Dysbiosis Promotes Intestinal Permeability, Systemic Inflammation, and Macrophage Dysfunction.* Cell Host Microbe, 2017. **21**(4): p. 455-466 e4.

103. Alvarez, C.S., et al., *Outer Membrane Vesicles and Soluble Factors Released by Probiotic Escherichia coli Nissle 1917 and Commensal ECOR63 Enhance Barrier Function by Regulating Expression of Tight Junction Proteins in Intestinal Epithelial Cells.* Front Microbiol, 2016. **7**: p. 1981.

104. Wu, S., et al., *Leaky intestine and impaired microbiome in an amyotrophic lateral sclerosis mouse model.* Physiol Rep, 2015. **3**(4).

105. Zhang, Y.G., et al., *Target Intestinal Microbiota to Alleviate Disease Progression in Amyotrophic Lateral Sclerosis.* Clin Ther, 2017. **39**(2): p. 322-336.

106. Braniste, V., et al., *The gut microbiota influences blood-brain barrier permeability in mice.* Sci Transl Med, 2014. **6**(263): p. 263ra158.

107. Liu, Z., et al., *Peripheral and Central Nervous System Immune Response Crosstalk in Amyotrophic Lateral Sclerosis.* Front Neurosci, 2020. **14**: p. 575.

108. Beers, D.R. and S.H. Appel, *Immune dysregulation in amyotrophic lateral sclerosis: mechanisms and emerging therapies.* The Lancet Neurology, 2019. **18**(2): p. 211-220.

109. Erny, D., et al., *Host microbiota constantly control maturation and function of microglia in the CNS.* Nat Neurosci, 2015. **18**(7): p. 965-77.

110. Parada Venegas, D., et al., *Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases.* Front Immunol, 2019. **10**: p. 277.

111. Schulthess, J., et al., *The Short Chain Fatty Acid Butyrate Imprints an Antimicrobial Program in Macrophages.* Immunity, 2019. **50**(2): p. 432-445 e7.

112. Figueroa-Romero, C., et al., *Temporal evolution of the microbiome, immune system and epigenome with disease progression in ALS mice.* Dis Model Mech, 2019. **13**(2).

113. Burberry, A., et al., *C9orf72 suppresses systemic and neural inflammation induced by gut bacteria.* Nature, 2020. **582**(7810): p. 89-94.

114. McAlary, L., et al., *Amyotrophic Lateral Sclerosis: Proteins, Proteostasis, Prions, and Promises.* Front Cell Neurosci, 2020. **14**: p. 581907.

115. Rao, M. and M.D. Gershon, *The bowel and beyond: the enteric nervous system in neurological disorders.* Nat Rev Gastroenterol Hepatol, 2016. **13**(9): p. 517-28.

116. Niesler, B., et al., *Disorders of the enteric nervous system - a holistic view.* Nat Rev Gastroenterol Hepatol, 2021.

117. Poewe, W., et al., *Parkinson disease.* Nat Rev Dis Primers, 2017. **3**: p. 17013.

118. Liu, B., et al., *Vagotomy and Parkinson disease: A Swedish register-based matched-cohort study.* Neurology, 2017. **88**(21): p. 1996-2002.

119. Svensson, E., et al., *Vagotomy and subsequent risk of Parkinson's disease.* Ann Neurol, 2015. **78**(4): p. 522-9.

120. Sampson, T.R., et al., *A gut bacterial amyloid promotes alpha-synuclein aggregation and motor impairment in mice.* Elife, 2020. **9**.

121. Vucic, S., *Sensory and autonomic nervous system dysfunction in amyotrophic lateral sclerosis.* Neuropathol Appl Neurobiol, 2017. **43**(2): p. 99-101.

122. Pfeiffer, R.F., *Gastrointestinal Dysfunction in Parkinson's Disease.* Curr Treat Options Neurol, 2018. **20**(12): p. 54.

123. Ludvigsson, J.F. and M. Lashkariani, *Cohort profile: ESPRESSO (Epidemiology Strengthened by histoPathology Reports in Sweden).* Clin Epidemiol, 2019. **11**: p. 101-114.

124. Ludvigsson, J.F., et al., *The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research.* Eur J Epidemiol, 2009. **24**(11): p. 659-67.

125. Ludvigsson, J.F., et al., *Registers of the Swedish total population and their use in medical research.* Eur J Epidemiol, 2016. **31**(2): p. 125-36.

126. Ludvigsson, J.F., et al., *External review and validation of the Swedish national inpatient register.* BMC Public Health, 2011. **11**: p. 450.

127. Wettermark, B., et al., *The new Swedish Prescribed Drug Register--opportunities for pharmacoepidemiological research and experience from the first six months.* Pharmacoepidemiol Drug Saf, 2007. **16**(7): p. 726-35.

128. Brooke, H.L., et al., *The Swedish cause of death register.* Eur J Epidemiol, 2017. **32**(9): p. 765-773.

129. Ekbom, A., *The Swedish multi-generation register*, in *Methods in biobanking*. 2011, Springer. p. 215-220.

130. Runesson, B., et al., *The Stockholm CREAtinine Measurements (SCREAM) project: protocol overview and regional representativeness.* Clinical Kidney Journal, 2016. **9**(1): p. 119-127.

131. Lash, T.L., et al., *Modern epidemiology. 4th ed*. 2021: Philadelphia: Wolters Kluwer.

132. Essebag, V., et al., *The nested case-control study in cardiology.* American Heart Journal, 2003. **146**(4): p. 581-590.

133. Bellera, C.A., et al., *Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer.* BMC Med Res Methodol, 2010. **10**: p. 20.

134. Kleinbaum, D.G. and M. Klein, *The Stratified Cox Procedure.* 2012: p. 201-240.

135. Kuo, C.L., Y. Duan, and J. Grady, *Unconditional or Conditional Logistic Regression Model for Age-Matched Case-Control Data?* Front Public Health, 2018. **6**: p. 57.

136. Asar, O., et al., *Joint modelling of repeated measurement and time-to-event data: an introductory tutorial.* Int J Epidemiol, 2015. **44**(1): p. 334-44.

137. Miller, R.G., et al., *Hemoglobin A1c Level and Cardiovascular Disease Incidence in Persons With Type 1 Diabetes: An Application of Joint Modeling of Longitudinal and Time-to-Event Data in the Pittsburgh Epidemiology of Diabetes Complications Study.* Am J Epidemiol, 2018. **187**(7): p. 1520-1529.

138. Ambrosini, Y.M., et al., *The Gut-Brain Axis in Neurodegenerative Diseases and Relevance of the Canine Model: A Review.* Front Aging Neurosci, 2019. **11**: p. 130.

139. Clemente, J.C., et al., *The impact of the gut microbiota on human health: an integrative view.* Cell, 2012. **148**(6): p. 1258-70.

140. Lozupone, C.A., et al., *Diversity, stability and resilience of the human gut microbiota.* Nature, 2012. **489**(7415): p. 220-30.

141. Willyard, C., *How gut microbes could drive brain disorders.* Nature, 2021. **590**(7844): p. 22-25.

142. Fang, P., et al., *The Microbiome as a Modifier of Neurodegenerative Disease Risk.* Cell Host Microbe, 2020. **28**(2): p. 201-222.

143. Balestrino, R. and A.H.V. Schapira, *Parkinson disease.* Eur J Neurol, 2020. **27**(1): p. 27-42.

144. Sims, R., M. Hill, and J. Williams, *The multiplex model of the genetics of Alzheimer's disease.* Nat Neurosci, 2020. **23**(3): p. 311-322.

145. Mejzini, R., et al., *ALS Genetics, Mechanisms, and Therapeutics: Where Are We Now?* Front Neurosci, 2019. **13**: p. 1310.

146. Sait, A., et al., *Viral Involvement in Alzheimer's Disease.* ACS Chem Neurosci, 2021. **12**(7): p. 1049-1060.

147. Cocoros, N.M., et al., *Long-term Risk of Parkinson Disease Following Influenza and Other Infections.* JAMA Neurol, 2021.

148. Wang, H., et al., *Bacterial, viral, and fungal infection-related risk of Parkinson's disease: Meta-analysis of cohort and case-control studies.* Brain Behav, 2020. **10**(3): p. e01549.

149. Dardiotis, E., et al., *H. pylori and Parkinson's disease: Meta-analyses including clinical severity.* Clin Neurol Neurosurg, 2018. **175**: p. 16-24.

150. Smeyne, R.J., et al., *Infection and Risk of Parkinson's Disease.* J Parkinsons Dis, 2021. **11**(1): p. 31-43.

151. Itzhaki, R.F., et al., *Do infections have a role in the pathogenesis of Alzheimer disease?* Nat Rev Neurol, 2020. **16**(4): p. 193-197.

152. Warren-Gash, C., et al., *Human herpesvirus infections and dementia or mild cognitive impairment: a systematic review and meta-analysis.* Sci Rep, 2019. **9**(1): p. 4743.

153. Lotz, S.K., et al., *Microbial Infections Are a Risk Factor for Neurodegenerative Diseases.* Front Cell Neurosci, 2021. **15**: p. 691136.

154. Celeste, D.B. and M.S. Miller, *Reviewing the evidence for viruses as environmental risk factors for ALS: A new perspective.* Cytokine, 2018. **108**: p. 173-178.

155. Sun, J., et al., *Antibiotics use and risk of amyotrophic lateral sclerosis in Sweden.* Eur J Neurol, 2019. **26**(11): p. 1355-1361.

156. Lanznaster, D., et al., *Plasma creatinine and amyotrophic lateral sclerosis prognosis: a systematic review and meta-analysis.* Amyotroph Lateral Scler Frontotemporal Degener, 2019. **20**(3-4): p. 199-206.

157. van Eijk, R.P.A., et al., *Monitoring disease progression with plasma creatinine in amyotrophic lateral sclerosis clinical trials.* J Neurol Neurosurg Psychiatry, 2018. **89**(2): p. 156-161.

158. Ong, M.L., P.F. Tan, and J.D. Holbrook, *Predicting functional decline and survival in amyotrophic lateral sclerosis.* PLoS One, 2017. **12**(4): p. e0174925.

159. Murdock, B.J., et al., *The dual roles of immunity in ALS: injury overrides protection.* Neurobiol Dis, 2015. **77**: p. 1-12.

160. Murdock, B.J., et al., *Correlation of Peripheral Immunity With Rapid Amyotrophic Lateral Sclerosis Progression.* JAMA Neurol, 2017.

161. Wuolikainen, A., et al., *Multi-platform mass spectrometry analysis of the CSF and plasma metabolomes of rigorously matched amyotrophic lateral sclerosis, Parkinson's disease and control subjects.* Mol Biosyst, 2016. **12**(4): p. 1287-98.

162. Gray, E., et al., *The longitudinal cerebrospinal fluid metabolomic profile of amyotrophic lateral sclerosis.* Amyotroph Lateral Scler Frontotemporal Degener, 2015. **16**(7-8): p. 456-63.

163. Kirk, S.E., et al., *Biomarkers of Metabolism in Amyotrophic Lateral Sclerosis.* Front Neurol, 2019. **10**: p. 191.

164. Mariosa, D., et al., *Blood biomarkers of carbohydrate, lipid, and apolipoprotein metabolisms and risk of amyotrophic lateral sclerosis: A more than 20-year follow-up of the Swedish AMORIS cohort.* Ann Neurol, 2017. **81**(5): p. 718-728.

165. Kioumourtzoglou, M.A., et al., *Diabetes Mellitus, Obesity, and Diagnosis of Amyotrophic Lateral Sclerosis: A Population-Based Study.* JAMA Neurol, 2015. **72**(8): p. 905-11.

166. Hernan, M.A., *Invited Commentary: Selection Bias Without Colliders.* Am J Epidemiol, 2017. **185**(11): p. 1048-1050.

167. Mansournia, M.A., M.A. Hernan, and S. Greenland, *Matched designs and causal diagrams.* Int J Epidemiol, 2013. **42**(3): p. 860-9.

168. Hernan, M.A. and S.R. Cole, *Invited Commentary: Causal diagrams and measurement bias.* Am J Epidemiol, 2009. **170**(8): p. 959-62; discussion 963-4.

169. VanderWeele, T.J. and M.A. Hernan, *Results on differential and dependent measurement error of the exposure and the outcome using signed directed acyclic graphs.* Am J Epidemiol, 2012. **175**(12): p. 1303-10.

170. Dosemeci, M., S. Wacholder, and J.H. Lubin, *Does nondifferential misclassification of exposure always bias a true effect toward the null value?* Am J Epidemiol, 1990. **132**(4): p. 746-8.

171. Rizzuto, D., et al., *Detection of Dementia Cases in Two Swedish Health Registers: A Validation Study.* J Alzheimers Dis, 2018. **61**(4): p. 1301-1310.

172. Feldman, A.L., et al., *Accuracy and sensitivity of Parkinsonian disorder diagnoses in two Swedish national health registers.* Neuroepidemiology, 2012. **38**(3): p. 186-93.

173. VanderWeele, T.J., M.A. Hernan, and J.M. Robins, *Causal directed acyclic graphs and the direction of unmeasured confounding bias.* Epidemiology, 2008. **19**(5): p. 720-8.

174. Schapira, A.H.V., K.R. Chaudhuri, and P. Jenner, *Non-motor features of Parkinson disease.* Nat Rev Neurosci, 2017. **18**(7): p. 435-450.

175. Skjaerbaek, C., et al., *Gastrointestinal Dysfunction in Parkinson's Disease.* J Clin Med, 2021. **10**(3).

176. Zingel, R., J. Bohlken, and K. Kostev, *Association Between Inflammatory Bowel Disease and Dementia: A Retrospective Cohort Study.* J Alzheimers Dis, 2021.

177. Zhang, B., et al., *Inflammatory bowel disease is associated with higher dementia risk: a nationwide longitudinal study.* Gut, 2021. **70**(1): p. 85-91.

178. Lee, H.S., et al., *Inflammatory bowel disease and Parkinson's disease: common pathophysiological links.* Gut, 2021. **70**(2): p. 408-417.

179. Villumsen, M., et al., *Inflammatory bowel disease increases the risk of Parkinson's disease: a Danish nationwide cohort study 1977-2014.* Gut, 2019. **68**(1): p. 18-24.

180. Liu, B., et al., *Irritable bowel syndrome and Parkinson's disease risk: register-based studies.* NPJ Parkinsons Dis, 2021. **7**(1): p. 5.

11 APPENDIX