Incidence, prevalence and co-occurrence of autoimmune disorders, trends over time and by age, sex and socioeconomic status. A population-based study in 22 million individuals.

Nathalie Conrad[1] PhD, Shivani Misra[2] PhD, Jan Y Verbakel[1, 3], Prof. Geert Verbeke[4] PhD, Prof. Geert Molenberghs[4] PhD, Prof. Peter N Taylor[5] MD, Prof. Justin Mason[6] PhD, PhD, Prof. Naveed Sattar[7] MD, Prof. John J. V. McMurray[8] MD, Prof. Iain B. McInnes[7] PhD, Prof. Kamlesh Khunti[9] MD, Prof. Geraldine Cambridge[10] PhD

1 Department of Public Health and Primary Care, KU Leuven, Belgium

2 Faculty of Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, United Kingdom

3 Nuffield Department of Primary Care Health Sciences, University of Oxford, United Kingdom

4 Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Hasselt University and KU Leuven, Belgium

5 School of Medicine, Cardiff University, United Kingdom

6 Faculty of Medicine, National Heart & Lung Institute, Imperial College London, United Kingdom

7 College of Medical Veterinary and Life Sciences, University of Glasgow, United Kingdom

8 Institute of Cardiovascular and Medical Sciences, University of Glasgow, United Kingdom

9 Diabetes Research Centre, University of Leicester, United Kingdom

10 Division of Medicine, Dept Rheumatology, University College London, United Kingdom

Correspondence to:

Dr. Nathalie Conrad

Department of Public Health and Primary Care

Kapucijnenvoer 33 - 3000 Leuven – Belgium

Tel: +44 74 70 421 005

E-mail: nathalie.conrad@kuleuven.be

Word count abstract: 417

Word count manuscript: 3 650

Abstract

Background: A rise in the incidence of selected autoimmune disorders has been described, however, contemporary estimates of the overall incidence of autoimmune diseases and trends over time, are scarce and inconsistent.

Methods: We used linked primary and secondary electronic health records of 22 million individuals from the Clinical Practice Research Datalink (CPRD), a cohort that is representative of the UK population in terms of age and sex. We calculated incidence and prevalence of 19 autoimmune disorders (AID) from 2000 to 2019 and used negative binomial regression models to investigate temporal trends and variation by age, sex, socioeconomic status, season of onset and geographical region. To characterise co-occurrence of autoimmune diseases, we calculated incidence rate ratios, comparing incidence rates of comorbid autoimmune disease among patients with a first (index) autoimmune disease with incidence rates in the general population, using negative binomial regression models, adjusted for age and sex.

Findings: Among the 22,009,375 individuals included in the study, we identified a total of 978 872patients with a new diagnosis of at least one autoimmune disease between 2000 and 2019 (mean (SD) age: 54.0 (21.4) years, 64% women). Over the study period, age-standardised incidence rates of autoimmune diseases increased by 4%, similarly for men and women. The largest increases were seen in Graves' disease, coeliac disease and Sjogren's syndrome, for which incidences have doubled over the past two decades. Two conditions exhibited a significant decrease in incidence (Hashimoto's thyroiditis and pernicious anaemia). Taken together the 19 autoimmune disorders examined affected 10.2% of the population over the study period (13.1% of women, 7.4% of men). A socioeconomic gradient was evident across several diseases, including Graves' disease, pernicious anaemia, rheumatoid arthritis, and systemic lupus erythematosus. Seasonal variations were observed for type 1 diabetes (more commonly diagnosed in winter) and vitiligo (more commonly diagnosed in summer), and regional variations were observed for a range of conditions. Autoimmune disorders were commonly associated with each other, particularly Sjogren's, systemic lupus erythematosus and systemic sclerosis. Patients with type 1 diabetes also had significantly higher rates of Addison's, coeliac, and thyroid diseases, and multiple sclerosis stood out as having low rates of co-occurrence with other autoimmune diseases.

Interpretation: Autoimmune diseases affect about one in ten individuals. Their burden continues to increase over time, albeit at varying rates across individual diseases. The socioeconomic, seasonal, and regional disparities observed among several autoimmune disorders, implicate environmental factors in disease pathogenesis. The interrelations between autoimmune diseases are commensurate with shared pathogenetic mechanisms or predisposing factors, particularly among connective tissue diseases and among endocrine diseases.

Funding: Research Foundation Flanders (FWO).

Keywords: autoimmune disorders, immune mediated inflammatory diseases, incidence, prevalence, co-occurrence, cohort study, epidemiology, CPRD.

Research in Context

Evidence before this study

We searched Pubmed and Embase for reports published between 1 January 2000 to 30 July 2022 related to "autoimmune disorders" (any of the 19 individual conditions investigated) and "incidence", reviewed references of clinical practice guidelines and consulted with experts for relevant studies. Most studies investigated one autoimmune disorder at a time, and generally the more common autoimmune disorders, such as type 1 diabetes or psoriasis.

Studies generally relied on a small number of cases and presented different designs, case identification and diagnostic methods, rendering adequate synthesis and the calculation of pooled estimates and temporal trends difficult. Evidence was particularly scarce for rarer autoimmune disorders. We found no study that reported large-scale disease incidence and temporal trends of autoimmune disorders as a group of conditions

Added value of this study

We present standardised incidence rates derived from a large, representative, general population cohort, setting a baseline for international comparison, monitoring of prevention strategies and for the design of public health policies. Temporal trends across a broad range of autoimmune diseases do not support the idea of an epidemic of autoimmunity at present day, and provide valuable reference rates for future studies investigating population-level impact of newly-introduced risk factors, such as the covid-19 pandemic.

We provide robust evidence of socioeconomic, seasonal, and regional disparities for several autoimmune diseases (particularly Graves' disease, pernicious anaemia, rheumatoid arthritis, systemic lupus erythematosus, and type 1 diabetes). Such variations are unlikely to be attributable to genetic differences alone and suggest involvement of potentially modifiable risk factors in the pathogenesis of autoimmune diseases.

We further demonstrate important interrelations between many autoimmune diseases, and confirm that co-occurrence of autoimmune disease is common, yet orders of magnitude differ widely between diseases. Associations were highest amongst connective tissue diseases, for patients with type 1 diabetes and Addison's disease, coeliac and thyroid diseases. More generally, increased risk of developing Addison's disease was observed following almost every autoimmune disease investigated. Multiple sclerosis stood out as having low rates of co-occurrence with other autoimmune diseases, suggesting a distinct pathophysiology.

Implications of all the available evidence

The burden of autoimmune disorders appears higher than previous estimates, and continues to increase over time, albeit at varying rates across individual diseases.

Socioeconomic, seasonal, and regional disparities in disease incidence point to potentially preventable factors involved in the pathogenesis of autoimmune diseases. Co-occurrences of diseases point to common genetic and environmental risk factors that interact and operate variably across these diseases.

Introduction

Autoimmune diseases arise when immune dysregulation causes host tissue damage.[1] A wide range of autoimmune diseases are described that present with variable age of onset, tissue distribution and clinical and functional impact.[1] Most of these diseases are incurable and require lifelong treatment.

Adequate public health and service delivery planning requires reliable information about contemporary population-level disease incidence. However, estimates of autoimmune disease incidence rates and their temporal trends, even in high-income countries, are scarce and inconsistent.[1] Selected autoimmune disorders, such as type 1 diabetes, are reported to have increased over the past several decades, raising the question as to whether the overall incidence of autoimmune disorders is on the rise, driven perhaps by common environmental factors or behavioural changes. Even for type 1 diabetes, whose incidence is among the best studied within autoimmune diseases, reports rely on relatively small cohorts,[2,3] and estimates vary by a factor 10 between studies in Europe alone.[4,5] For many other autoimmune diseases, evidence concerning disease incidence and prevalence is more limited. The relatively modest absolute numbers of patients affected by individual autoimmune diseases presents as a major challenge to investigators, and hinders adequate synthesis across studies.[6] As a result, reliable estimates of disease incidence and how they evolve over time, particularly as pertains to autoimmune diseases as a group, are not available.

Commonalities and differences between individual diseases also remain poorly understood and continue to be subject to much research. While emerging evidence has suggested that autoimmune diseases tend to co-occur within individuals, large-scale investigations across a broad spectrum of autoimmune diseases that could provide clues about shared pathogenesis and risk factors are not currently available.[7,8]

To address these knowledge gaps, we analysed a large longitudinal database of primary and secondary care records that provides information on millions of individuals' diagnoses with several years of follow-up.[9,10] We investigated the incidence and prevalence for 19 of the most common autoimmune diseases, assessed trends over time, by sex, age, socioeconomic status, season and region, and examined rates of co- occurrence among autoimmune diseases.

Methods

Data source

We used electronic health records from the Clinical Practice Research Datalink (CPRD, GOLD and AURUM datasets) from 1 January 1985 to 30 June 2019. The CPRD database contains anonymised patient data from approximately 20% of the current UK population and is broadly representative in terms of age, sex and ethnicity. CPRD is one of the largest databases of longitudinal medical records from primary care in the world and has been validated for epidemiological research for a broad range of conditions.[9] Primary care records from CPRD were linked to secondary care records from Hospital Episodes Statistics (HES admitted patient care and HES outpatient) data. Linkage was available for a subset of English practices from 1 January 1998 onwards, covering approximately 50% of all CPRD records. Scientific approval for this study was given by the CPRD Independent Scientific Advisory Committee (ISAC).

Case identification

We investigated 19 of the most common autoimmune disorders (AID): Addison's disease; ankylosing spondylitis; coeliac disease; childhood-onset type 1 diabetes; Graves' disease; Hashimoto's thyroiditis; inflammatory bowel disease (Crohn's disease or ulcerative colitis); multiple sclerosis; myasthenia gravis; pernicious anaemia; polymyalgia rheumatica; primary biliary cholangitis; psoriasis; rheumatoid arthritis; Sjogren's syndrome; systemic lupus erythematosus; systemic sclerosis; vasculitis; and vitiligo.[11–13] Diseases were considered individually and as a composite outcome of all AID combined. For the combined analyses, we calculated primary incidence (first recorded AID, reflecting the number of patients affected by AIDs) and cumulative incidence (all recorded AIDs, reflecting the cumulative number of AID diagnoses).

Although some of these diseases remain debated in terms of their autoimmune aetiology and may be more appropriately described as 'immune mediated inflammatory diseases', to assist readability we refer to this group of conditions as 'autoimmune diseases'.

For each condition, algorithms for identification of diagnoses from electronic health records were defined based on diagnostic codes from hospital/death (International Classification of Diseases, tenth revision (ICD-10)) and primary care (Read[14]) coding schemes, and, for selected conditions, prescriptions of certain drugs (**appendix**). Specifically, individuals with childhood-onset type 1 diabetes were identified as those with at least one diagnostic code referring to type 1 or insulin-dependent diabetes, at least 1 insulin prescription, and aged 19 years or less at first diagnosis. Individuals with Hashimoto thyroiditis were identified as those with at least one levothyroxine prescription and no history of hyperthyroidism, pituitary disease, thyroid surgery, or thyroid-altering medication (amiodarone, lithium, sodium valproate, carbimazole, propylthiouracil, thalidomide, sunitinib). When conflicting arthritis diagnoses (ankylosing spondylitis and rheumatoid arthritis) were recorded in the same individual, the last recorded diagnosis was used. Incident diagnoses were defined as the first record of that condition in primary or secondary care records from any diagnostic position.

Study population

Included in the study were men and women with records labelled as 'acceptable', approved for HES and ONS linkage, and registered with their general practice for at least 12 months during the study period (01/01/1998 to 30/06/2019). For incidence calculations, we excluded all individuals who had a diagnosis of the disease of interest prior to study start date (1 January 2000), or within the first 12 months of registration with their general practice.

Covariates

Smoking status and body mass index (BMI) were abstracted from electronic health records as the most recent measurement within 2 years prior to diagnosis. BMI was categorised into 'underweight' (<18.5 kg/m^2), 'normal' (18.5-24.9 kg/m^2), 'overweight' (25-29.9 kg/m^2), and 'obese' (>=30 kg/m^2). Socioeconomic status was defined as the Index of Multiple Deprivation (IMD) 2015 quintile,[15] a composite measure of relative deprivation at a small area level, covering an average population of 1500 people, ranked in ascending order of deprivation score and grouped in equal fifths, with quintiles 1 and 5 representing the least and most deprived areas, respectively. Ethnicity was extracted from both primary and secondary care records. When ethnicity differed between primary and secondary care records, secondary care data was used. To assist readability, ethnicity was grouped into two categories, 'white' and 'other'.

Statistical analyses

Baseline characteristics are presented as frequencies (%) for categorical data, medians and interquartile range (IQR) for non-normally distributed continuous data, or means and standard deviation (SD) for normally distributed continuous data, over the whole autoimmune disease cohort and stratified by sex, socioeconomic quintile, and period of diagnosis. Number and percentage of records with missing data are displayed for variables with missing entries.

Observed incidence rates were computed by dividing the number of incident cases by the number of patient-years in the cohort. Category-specific rates were computed separately for subgroups of age, sex, socioeconomic status, region, calendar year of diagnosis and season of diagnosis. Winter was defined as the period from January to March, and Summer as June to August. Time at risk was restricted to days alive and registered with a general practice for over 12 months. Observed prevalence rates were computed considering all patients ever diagnosed with autoimmune disease (numerator) among patients alive and registered with a general practitioner on June 30th in each year (denominator). To allow comparison with other studies, we further calculated incidence and prevalence rates of childhood onset type 1 diabetes restricting the denominator to those aged 19 or less (**appendix**).

Standardised rates were computed by applying direct age and sex standardisation[16] to the 2013 European Standard Population[17] using 5-year age bands up to 90 years of age.

Negative binomial regression models were used to examine overall and category-specific incidence rate ratios (IRR) and corresponding 95% confidence intervals (CI). Models were adjusted for calendar year, age (categorised into five years age-bands), sex, socioeconomic status and region.

To characterise co-occurrence of autoimmune diseases, we calculated incidence rate ratios, comparing incidence rates of comorbid autoimmune disease among patients with a first (index) autoimmune disease with incidence rates in the general population, using negative binomial regression models adjusted for age and sex. We performed separate analyses for each pair and sequence of autoimmune diseases, following methods by Somer et al.[7] For these analyses, time at risk started at the latest of patient registration plus 12 months or diagnosis of the index disease, and stopped at the earliest of the incidence of the comorbid autoimmune disease, death, the patient ceasing registration with the general practice, or the end of the study.

Study findings are reported in accordance with the REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) recommendations.[18] Statistical analyses were performed in R, version 3.6.2 (R Foundation for Statistical Computing, Vienna, Austria).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

A total of 22 009 375 individuals contributed data between 01/01/2000 and 31/12/2019 with 135 691 152 patient-years of follow-up. Among those, we identified 1 123 789 new diagnoses of autoimmune diseases, affecting a total of 978 872 patients. The mean (SD) age at AID diagnosis was 54.0 (21.4) years, and 64% of patients were women (**Table 1**).

Temporal trends

The number of patients newly diagnosed with one (or more) autoimmune diseases increased only modestly over time (IRR comparing 2017-2019 vs 2000-2002: 1.04 [1.00, 1.09]). The number of new autoimmune disease diagnoses increased by 22%, largely due to an increasing number of secondary autoimmune disease diagnoses among patients already affected by a first autoimmune disease (IRR comparing 2017-2019 vs 2000-2002: 1.22 [1.18, 1.28]). Coeliac, Graves', and Sjogren's syndrome showed the greatest increases, whereas Hashimoto's thyroiditis and pernicious anaemia declined modestly over the study period (**Figure 1**). The observed increase in coeliac disease, Graves' disease, and Sjogren's syndrome was largely driven by a higher number of diagnoses in women. The increase in Graves' disease was largely driven by the very old (80 years and older), whereas the increase in coeliac disease was largely driven by diagnoses in the very young (up to 40 years of age) (**appendix**).

Age at diagnosis

Autoimmune disorders developed over the whole life spectrum, from the first to the 95th year of life. Median age at first autoimmune disease presentation varied greatly among individual autoimmune disease. For many conditions, incidence increased with age; this was the case for Graves' disease, pernicious anaemia, and rheumatoid arthritis. Only six conditions were commonly diagnosed before the age of five, these were: Addison's, coeliac disease, type 1 diabetes, psoriasis, vitiligo and vasculitis (largely due to Henoch-Schonlein purpura, Kawasaki disease, and glomerulonephritis). For others, such as multiple sclerosis, psoriasis, and lupus, the incidence peaked in the middle age. Finally, some conditions presented a bi-modal age-distribution with a peak in childhood or early adulthood and another one later in life; this was the case for coeliac disease, inflammatory bowel disease and vasculitis (**Figure 2**). Generally, these distributions were consistent with prior studies and add to the validity of our current search approach.

Incidence by sex

Most autoimmune disorders were more common in women than men (IRR for women compared to men: 1.74 [1.72, 1.77] for all diseases combined). Thyroid disorders, Sjogren's syndrome, lupus, and systemic sclerosis

had the highest incidence rate ratio in women compared to men. Only three diseases were more common in men than women, namely ankylosing spondylitis; type 1 diabetes; and myasthenia gravis (**Figure 3**).

Incidence by socioeconomic status

Overall, the most deprived socioeconomic groups had a higher incidence of autoimmune diseases (IRR for the most deprived compared to least deprived quintile: 1.14 [1.11, 1.16] for all autoimmune diseases combined). A marked socioeconomic gradient was visible across several individual diseases, including Graves' disease, pernicious anaemia, rheumatoid arthritis, and lupus. For other diseases, such as Hashimoto's thyroiditis and inflammatory bowel disease, no difference was observed among socioeconomic groups despite relatively large number of cases; and for coeliac and polymyalgia rheumatica, disease incidence was highest in the least deprived group (**Figure 4**).

Regional differences

Overall, variation by geographic region was limited. Notable exceptions were polymyalgia rheumatica, type 1 diabetes, and coeliac disease, which were considerably more common outside of the capital city, as well as pernicious anaemia, which appeared to be more common in northern regions. Sjogren's syndrome, lupus and vitiligo, on the other hand, presented lower incidence rates outside of London (**Figure S4**).

Seasonal differences

Most autoimmune diseases were diagnosed throughout the year with no significant differences between winter and summer months. Seasonal variation was observed only for type 1 diabetes, which was more commonly diagnosed in the winter months (January to March) compared to summer months (June to August), and for vitiligo, which was more likely to be diagnosed during the summer (**Figure S5**).

Co-occurrence of autoimmune diseases

Autoimmune disorders were commonly associated with each other. Increased risk of developing a second autoimmune disorder was seen across many autoimmune diseases, but orders of magnitude differed widely between diseases. Associations were generally highest among connective tissue diseases, particularly between Sjogren's, systemic lupus erythematosus and systemic sclerosis. Patients with type 1 diabetes also had significantly higher rates of Addison's, coeliac and thyroid diseases. More generally, Addison's disease occurred at a considerably higher incidence among patients with pre-existing autoimmune disease than in the general population. Multiple sclerosis stood out as having low rates of co-occurrence with other autoimmune diseases and even showed an inverse association with some autoimmune disorders (**Figure 5, Table S1**).

Prevalence trends over time

Together the 19 autoimmune disorders investigated in this study affected 10.2% of the population over the study period (13.1% of women, 7.4% of men). Age- and sex- standardised prevalence increased over time from 7.7% in 2000-2002 to 11.0% in 2017-2019 (RR comparing 2017-2019 vs 2000-2002: 1.41 [1.37, 1.44]).

Discussion

Our large-scale population-based study provides several novel insights into the burden of autoimmune disorders, its variation over time, by individual diseases, and patient subgroups of age, sex and socioeconomic status. Our findings confirm and extend evidence from previous studies demonstrating an increasing incidence of several autoimmune disorders[19], and shows that the increase was particularly pronounced for Graves' disease, coeliac disease, and rheumatic disorders. Yet, overall and in consideration of the increasing awareness for some of these conditions, improved coding practices, and earlier recognition of these conditions over the past two decades, the observed increase remains modest.

The epidemiology of type 1 diabetes is perhaps the best studied of all. Previous studies had reported varying estimates and trends over time,[4,5] and several surveys from Scandinavian countries have reported steep increases since the 1950s with a plateau since mid 2000.[20,21] In our study, we observed only a modest increase in disease incidence over the past two decades. Overall estimates were comparable with estimates from a recent simulation study on the incidence of type 1 diabetes[22] in Western Europe and similar to those reported by the International Diabetes Federation Atlas for the UK,[23,24] but lower than those recently reported in

northern European countries.[25] Another key finding was the reduction in Hashimoto's thyroiditis over time, which could be due to more careful initiation of levothyroxine in older persons with subclinical hypothyroidism following trials demonstrating limited benefits in this population.[26,27] Pernicious anaemia also showed an apparent decline in incidence over time, a decline that appears to coincide with increased use of dietary supplements over the same period, and possibly a more widespread recognition of other causes of vitamin B12 deficiency such as helicobacter pylori, although no causal inference can be made from our data.[28]

The observed increase in incidence of rheumatic diseases will likely have important implications for health services and already substantial medication expenditures linked to biologics. Improved coding practices during the study period, the introduction, in 2013, of a quality audit in primary care (so-called Quality and Outcomes Framework, QOF) rewarding general practitioners for maintaining a register and evaluating cardiovascular/fracture risks in patients with rheumatoid arthritis[29], and novel classification criteria for ankylosing spondylitis, are all likely to have contributed to the observed trend. The increase in incidence of axial spondyloarthritis in women starting around the time of publication of ASAS (Assessment of SpondyloArthritis international Society)[30] classification criteria - introducing the concept of non-radiographic axial spondyloarthritis - is consistent with other studies indicating that whilst ankylosing spondylitis is male-predominant, non-radiographic axial spondyloarthritis is similarly common in women and men.[31,32]

Our stratified analyses by socioeconomic status, region, and seasonal variations in disease incidence provide further insights into the possible role of environmental factors in the development of autoimmune diseases. Four diseases - Graves' disease, pernicious anaemia, rheumatoid arthritis, and systemic lupus erythematosus - presented a clear socioeconomic gradient with those in the most deprived group up to 50 % more likely to develop the disease than their affluent counterpart. Such socioeconomic disparities could indicate that diet, smoking, obesity, air pollution, or other currently unrecognised environmental exposures might play a role in the development of these diseases. Two conditions – coeliac disease and polymyalgia rheumatica – presented an inverse socioeconomic gradient, a phenomenon rarely observed in public health research, and which could be linked to increased awareness and testing for these diseases in more affluent populations or indeed lifestyle differences.

Seasonal variations on the other hand were limited. Vitiligo was more commonly diagnosed in the summer, perhaps due to increased visibility of depigmented skin areas in summer months, and type 1 diabetes was more commonly diagnosed in the winter and outside of the capital region, a finding compatible with hypotheses of viral triggers, diet, higher weight, or ethnicity playing a role in the disease's pathogenesis.[33–35] Other regional variations, such as coeliac disease more common outside of the capital, remain unexplained for now. While numerous reports have linked smoking to the incidence of certain autoimmune diseases such as rheumatoid arthritis, multiple sclerosis, or psoriasis[36], we did not observe a decline in these diseases' incidences despite considerable reduction in smoking prevalence over the same period[37]. This could be due to a parallel increase in other risk factors, such as obesity, over the same period, confounding via socioeconomic status, or currently unknown reasons.

Finally, co-occurrences of diseases provide valuable insights into a possibly common aetiology across some of these conditions. For example, we found high rates of co-occurrence among a range of connective tissue diseases, particularly Sjogren's syndrome, systemic lupus erythematosus, and systemic sclerosis, regardless of diagnostic sequence and after accounting for age and sex. Interestingly polymyalgia rheumatica did not have much association with other connective tissue disorders, except for vasculitis. Myasthenia gravis and multiple sclerosis also tended to co-occur. However, the association was weak and given the similarity in the symptoms of both conditions and the nature of our data, which is based on routine clinical practice data, we cannot fully rule out the possibility of inaccurate diagnoses or coding.

Our study further provides robust evidence for an increased incidence of a range of autoimmune diseases, particularly coeliac disease and thyroid disorders, among patients with type 1 diabetes. Such observations have been reported before and may point to overlapping genetic or environmental risk factors as well as people with a first autoimmune disease being more likely to undergo screening for these conditions.[39] While we did not observe associations between childhood-onset diabetes and diseases typically occurring at older age groups, our data was limited by follow-up duration, so that we cannot exclude that such associations might exist. Similarly, in our study, psoriasis presented limited co-occurrence with autoimmune diseases,

possibly because of the large proportion of mild cases typically observed among cohorts based on primary care data.[40]

Another consideration is whether treatment of an index autoimmune disease affects risk of developing a comorbid disease. Although nonspecific immunosuppressives could in theory decrease risk of a comorbid autoimmune disease, targeted therapies could have differential effects. For example, an intriguing finding was the high incidence of Addison's disease following almost every other autoimmune disorder, which could perhaps be related to glucocorticoid induced adrenal insufficiency (which is typically recorded as Addison's disease in the UK).[41] Such a mechanism might also explain the higher rates of cardiovascular diseases observed in patients with Addison's disease in this same cohort.[42] Nevertheless, given the relatively small number of patients with Addison's disease, these results must be interpreted with caution. Overall, this research confirms that some autoimmune diseases co-occur with one another at a rate greater than expected by chance or surveillance bias alone, but it reveals that this phenomenon is not generalized across all autoimmune diseases.

A major strength of this study is the selection of a statistically powerful data source with over 130 million person-years of data to investigate the incidence, prevalence, and co-occurrence of autoimmune disorders. The very large size of our cohort allowed us to perform stratified analyses of unprecedented granularity, over a broad spectrum of conditions, as well as allowing examination of the influence of age, sex, and socioeconomic status, as well as trends over 20 years. The use of routinely reported diagnoses also captures the burden of disease as experienced by physicians and health services, and likely increases the generalizability of findings. One of the key limitations of our study was the limited diversity in ethnic backgrounds in our cohort and the unavailability or significant missingness of additional variables potentially relevant to autoimmune disease pathogenesis, such as smoking, body mass index or blood biomarkers such as vitamin D deficiency. Research using electronic health records is also reliant on the accuracy of clinical coding carried out during consultations and hospital admissions. The validity of diagnoses underlying our study has been carefully assessed and was considered appropriate in light of the over two hundred independent studies that have investigated the validity of diagnoses recorded in CPRD which reported an average positive predictive value of about 90% for a broad range of conditions.[43] These include recent studies on several of the diseases studied here, which have demonstrated that algorithms based on diagnostic codes perform well at identifying patients with these conditions in primary care records.[32,40] Finally, a large number of tests and subgroup analyses within one study must be interpreted within that context and with adequate caution.

Our findings present an important new piece in the puzzle of autoimmune disease aetiology, a group of conditions that are apparent in near 10% of the population and which consume considerable health resources. The socioeconomic, seasonal, and regional disparities observed among several autoimmune disorders, point towards involvement of environmental factors in the pathogenesis of selected autoimmune diseases. The interrelations between many autoimmune diseases further suggest a shared pathogenesis, particularly among connective tissue diseases as well as between diabetes, coeliac and thyroid disorders. To this day, the exact causes of many of the autoimmune diseases studied here remain unknown and require further research.

Contributions

NC, GC and JCM conceived and designed the study. NC designed the statistical analysis plan and performed the statistical analysis. NC, JV, GV, and GM contributed to acquiring the data. All authors contributed to interpreting the results, drafting the manuscript and the revisions. NC and GC had full access to the data in the study and had final responsibility for the decision to submit for publication. All authors gave final approval of the version to be published.

Declarations of interest

This project has received funding from the Research Foundation Flanders (grant number 12ZU922N).

SM is personally funded by a Wellcome Trust career development fellowship and supported by an institutional Biomedical Research Council award. IM declares honoraria from AbbVie; grant support paid to his university from AstraZeneca and Eli Lilly; participation on data safety monitoring boards/advisory boards of AstraZeneca, BMS, Eli Lilly, Novartis, Janssen, GSK, AbbVie, Cabaletta, Compugen, Causeway, Gilead,

Moonlake, Reflexion, UCB, XinThera; patents from Novartis; leadership roles with Evelo, Versus Arthritis, and Greater Glasgow and Clyde Health Board; and stock or stock options with Evelo, Compugen, and Cabaletta. JJVM has received funding to his institution from Amgen and Cytokinetics for his participation in the Steering Committee for the ATOMIC-HF, COSMIC-HF, and GALACTIC-HF trials and meetings and other activities related to these trials; has received payments through Glasgow University from work on clinical trials, consulting and other activities from Alnylam, Amgen, AstraZeneca, Bayer, Boehringer Ingelheim, Bristol Myers Squibb, Cardurion, Dal-Cor, GlaxoSmithKline, Ionis, KBP Biosciences, Novartis, Pfizer, and Theracos; has received personal lecture fees from the Corpus, Abbott, Hikma, Sun Pharmaceuticals, Medscape/Heart.Org, Radcliffe Cardiology, Servier Director, and Global Clinical Trial Partners (GCTP). NS declares consulting fees and/or speaker honoraria from Abbott Laboratories, Afimmune, Amgen, AstraZeneca, Boehringer Ingelheim, Eli Lilly, Hanmi Pharmaceuticals, Janssen, Merck Sharp & Dohme, Novartis, Novo Nordisk, Pfizer, and Sanofi; and grant support paid to his university from AstraZeneca, Boehringer Ingelheim, Novartis, and Roche Diagnostics. KK is supported by the National Institute for Health Research (NIHR) Applied Research Collaboration East Midlands (ARC EM) and the NIHR Leicester Biomedical Research Centre (BRC). KK has also acted as a consultant, speaker or received grants for investigator-initiated studies for Astra Zeneca, Bayer, Novartis, Novo Nordisk, Sanofi-Aventis, Lilly and Merck Sharp & Dohme, Boehringer Ingelheim, Oramed Pharmaceuticals and Applied Therapeutics.

The views expressed are those of the authors and not necessarily those of the funders.

Acknowledgements

We would like to thank Ms. Hilary Shepherd, Dr. Sonia Coton, and Dr. Eleanor L Axson from the Clinical Practice Research Datalink for their support and expertise in preparing the dataset underlying these analyses.

Prof Justin Mason died before publication of this work. He was widely recognised as an outstanding clinician and academic, a gifted leader, and a friend to many. Justin was an international authority on large vessel vasculitis and his care for his patients and contributions to the vasculitis community will continue to make a significant difference for many years to come.

Data sharing

Access to CPRD data is subject to a license agreement and protocol approval process that is overseen by CPRD's Independent Scientific Advisory Committee (ISAC). A guide to access is provided on the CPRD website.

References:

1 American Autoimmune Related Diseases Association (AARDA). The Cost Burden of Autoimmune Disease: The Latest Front in the War on Healthcare Spending. 2011.

2 Ehehalt S, Blumenstock G, Willasch AM, Hub R, Ranke MB, Neu A. Continuous rise in incidence of childhood Type 1 diabetes in Germany. *Diabetic Medicine* 2008; **25**: 755–7.

3 Newhook LA, Penney S, Fiander J, Dowden J. Recent incidence of type 1 diabetes mellitus in children 0-14 years in Newfoundland and Labrador, Canada climbs to over 45/100,000: a retrospective time trend study. *BMC research notes* 2012; **5**: 628.

4 Mobasseri M, Shirmohammadi M, Amiri T, Vahed N, Hosseini Fard H, Ghojazadeh M. Prevalence and incidence of type 1 diabetes in the world: a systematic review and meta-analysis. *Health promotion perspectives* 2020; **10**: 98–115.

5 EURODIAB ACE Study Group. Variation and trends in incidence of childhood diabetes in Europe. EURODIAB ACE Study Group. *Lancet (London, England)* 2000; **355**: 873–6.

6 McLeod DSA, Cooper DS. The incidence and prevalence of thyroid autoimmunity. *Endocrine* 2012; **42**: 252–65.

7 Somers EC, Thomas SL, Smeeth L, Hall AJ. Autoimmune Diseases Co-occurring Within Individuals and Within Families A Systematic Review. 2006. DOI:10.1097/01.ede.0000193605.93416.df.

8 Somers EC, Thomas SL, Smeeth L, Hall AJ. Are Individuals With an Autoimmune Disease at Higher Risk of a Second Autoimmune Disorder? *American Journal of Epidemiology* 2009; **169**: 749–55.

9 Herrett E, Gallagher AM, Bhaskaran K, *et al.* Data Resource Profile: Clinical Practice Research Datalink (CPRD). *International journal of epidemiology* 2015; **44**: 827–36.

10 Wolf A, Dedman D, Campbell J, *et al.* Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *International Journal of Epidemiology* 2019; published online March 11. DOI:10.1093/ije/dyz034.

11 Cooper GS, Bynum MLK, Somers EC. Recent insights in the epidemiology of autoimmune diseases: Improved prevalence estimates and understanding of clustering of diseases. *Journal of Autoimmunity* 2009; **33**: 197–207.

12 Osborne D, Sobczyńska-Malefora A. Autoimmune mechanisms in pernicious anaemia and thyroid disease. *Autoimmunity Reviews* 2015; **14**: 763–8.

13 Dean LE, Jones GT, MacDonald AG, Downham C, Sturrock RD, Macfarlane GJ. Global prevalence of ankylosing spondylitis. *Rheumatology* 2014; **53**: 650–7.

14 Chisholm J. The Read Clinical Classification: The NHS Has Acquired A Coding System Designed For The Computer Age on JSTOR. *British Medical Journal* 1990; **300**: 1092.

15 Department for Communities and Local Government (DCLG). The English Index of Multiple Deprivation 2015: Guidance. 2015.

16 Kirkwood B, Sterne J. Essential medical statistics. 2010.

17 Eurostat's task force. Revision of the European Standard Population Report. 2013.

18 Benchimol EI, Smeeth L, Guttmann A, *et al.* The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLOS Medicine* 2015; **12**: e1001885.

19 Lerner A, Jeremias P, Matthias T. The World Incidence and Prevalence of Autoimmune Diseases is Increasing. *International Journal of Celiac Disease* 2016; **3**: 151–5.

20 Parviainen A, But A, Siljander H, *et al.* Decreased Incidence of Type 1 Diabetes in Young Finnish Children. *Diabetes Care* 2020; **43**: 2953–8.

21 Berhan Y, Waernbaum I, Lind T, Möllsten A, Dahlquist G, Group* for the SCDS. Thirty Years of Prospective Nationwide Incidence of Childhood Type 1 Diabetes. *Diabetes* 2011; **60**: 577–81.

22 Ward ZJ, Yeh JM, Reddy CL, *et al.* Estimating the total incidence of type 1 diabetes in children and adolescents aged 0–19 years from 1990 to 2050: a global simulation-based analysis. *The Lancet Diabetes & Endocrinology* 2022; published online Nov 10. DOI:10.1016/S2213-8587(22)00276-5.

23 Ogle GD, James S, Dabelea D, *et al.* Global estimates of incidence of type 1 diabetes in children and adolescents: Results from the International Diabetes Federation Atlas, 10th edition. *Diabetes Research and Clinical Practice* 2022; **183**: 109083.

24 Harvey JN, Hibbs R, Maguire MJ, O'Connell H, Gregory JW, Brecon Group (The Wales Paediatric Diabetes Interest Group). The changing incidence of childhood-onset type 1 diabetes in Wales: Effect of gender and season at diagnosis and birth. *Diabetes Res Clin Pract* 2021; **175**: 108739.

25 Waernbaum I, Lind T, Möllsten A, Dahlquist G. The incidence of childhood-onset type 1 diabetes, time trends and association with the population composition in Sweden: a 40 year follow-up. *Diabetologia* 2022; published online Oct 20. DOI:10.1007/s00125-022-05816-0.

26 Jonklaas J, DeSale S. Levothyroxine prescriptions trends may indicate a downtrend in prescribing. *Therapeutic Advances in Endocrinology and Metaboli* 2020; **11**: 1–14.

27 Stott DJ, Rodondi N, Kearney PM, *et al.* Thyroid Hormone Therapy for Older Adults with Subclinical Hypothyroidism. *New England Journal of Medicine* 2017; **376**: 2534–44.

28 Mikulic M, ,. OTC vitamins & minerals sales value, Great Britain, 2009-2021. https://www.statista.com/statistics/521713/otc-vitamins-and-minerals-sales-value-great-britain/ (accessed Dec 17, 2022).

29 Stokes T. Rheumatoid arthritis is a new clinical domain for QOF 2013/14. Guidelines in Practice. https://www.guidelinesinpractice.co.uk/musculoskeletal-and-joints-/rheumatoid-arthritis-is-a-new-clinical-domain-for-qof-2013/14/337862.article (accessed Nov 16, 2022).

30 Rudwaleit M, Heijde D van der, Landewé R, *et al.* The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Annals of the Rheumatic Diseases* 2009; **68**: 777–83.

31 Wright GC, Kaine J, Deodhar A. Understanding differences between men and women with axial spondyloarthritis. *Semin Arthritis Rheum* 2020; **50**: 687–94.

32 Scott IC, Whittle R, Bailey J, *et al.* Rheumatoid arthritis, psoriatic arthritis, and axial spondyloarthritis epidemiology in England from 2004 to 2020: An observational study using primary care electronic health record data. *The Lancet Regional Health – Europe* 2022; **23**. DOI:10.1016/j.lanepe.2022.100519.

33 Honkanen H, Oikarinen S, Nurminen N, *et al.* Detection of enteroviruses in stools precedes islet autoimmunity by several months: possible evidence for slowly operating mechanisms in virus-induced autoimmunity. *Diabetologia* 2017; **60**: 424–31.

34 Versini M, Jeandel P-Y, Rosenthal E, Shoenfeld Y. Obesity in autoimmune diseases: Not a passive bystander. *Autoimmunity Reviews* 2014; **13**: 981–1000.

35 Ji J, Hemminki K, Sundquist J, Sundquist K. Ethnic Differences in Incidence of Type 1 Diabetes among Second-Generation Immigrants and Adoptees from Abroad. *The Journal of Clinical Endocrinology & Metabolism* 2010; **95**: 847–50.

36 Perricone C, Versini M, Ben-Ami D, *et al.* Smoke and autoimmunity: The fire behind the disease. *Autoimmunity Reviews* 2016; **15**: 354–74.

37 Office for National Statistics. Adult smoking habits in England. 2020 https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/datasets/adultsmokinghabitsinengland (accessed Nov 19, 2022).

38 Dehbashi S, Hamouda D, Shanina E. Co-occurrence of multiple sclerosis and myasthenia gravis: A case report and review of immunological theories. *Multiple Sclerosis and Related Disorders* 2019; **34**: 135–6.

39 Cohn A, Sofia MA, Kupfer SS. Type 1 Diabetes and Celiac Disease: Clinical Overlap and New Insights into Disease Pathogenesis. *Curr Diab Rep* 2014; **14**: 517.

40 Seminara N m., Abuabara K, Shin D b., *et al.* Validity of The Health Improvement Network (THIN) for the study of psoriasis. *British Journal of Dermatology* 2011; **164**: 602–9.

41 Prete A, Bancos I. Glucocorticoid induced adrenal insufficiency. *BMJ* 2021; **374**: n1380.

42 Conrad N, Verbeke G, Molenberghs G, *et al.* Autoimmune diseases and cardiovascular risk: a population-based study on 19 autoimmune diseases and 12 cardiovascular diseases in 22 million individuals in the UK. *Lancet (London, England)* 2022; **0**. DOI:10.1016/S0140-6736(22)01349-6.

43 Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *British journal of clinical pharmacology* 2010; **69**: 4–14.

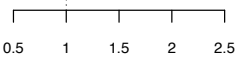
Table1: Baseline characteristics of patients with incident autoimmune disease during 2000-2019.

	All patients (N = 978 872)	**Sex**		**Socioeconomic status quintile**		**Time period**	
		Women (N = 625 879)	**Men** (N = 352 993)	**Least deprived** (N = 220 047)	**Most deprived** (N =170 856)	**2000-2004** (N = 245 323)	**2015-2019** (N = 222 060)
Age at diagnosis (years)							
Mean (SD)	54.0 (21.4)	54.4 (21.1)	53.3 (22.0)	55.5 (21.0)	51.0 (22.0)	55.6 (20.8)	52.6 (22.0)
Women, N (%)	625 879 (63.9%)			139 694 (63.5%)	110 322 (64.6%)	162 647 (66.3%)	138 598 (62.4%)
Ethnicity							
White	824 573 (88.3%)	533 414 (88.6%)	291 159 (87.7%)	189 790 (91.5%)	137 470 (83.1%)	210 062 (92.6%)	181 080 (84.0%)
Other	109 241 (11.7%)	68 391 (11.4%)	40 850 (12.3%)	17 648 (8.5%)	27 952 (16.9%)	16 825 (7.4%)	34 387 (16.0%)
Missing	45 058 (4.6%)	24 074 (3.8%)	20 984 (5.9%)	12 609 (5.7%)	5 434 (3.2%)	18 436 (7.5%)	6 593 (3.0%)
Socioeconomic status quintile							
1 (least deprived)	220 047 (22.5%)	139 694 (22.3%)	80 353 (22.8%)			55 431 (22.6%)	50 631 (22.8%)
2	204 731 (20.9%)	129 945 (20.8%)	74 786 (21.2%)			52 273 (21.3%)	45 861 (20.7%)
3	197 109 (20.1%)	125 845 (20.1%)	71 264 (20.2%)			49 740 (20.3%)	44 327 (20.0%)
4	186 129 (19.0%)	120 073 (19.2%)	66 056 (18.7%)			46 500 (19.0%)	42 053 (18.9%)
5 (most deprived)	170 856 (17.5%)	110 322 (17.6%)	60 534 (17.1%)			41 379 (16.9%)	39 188 (17.6%)
Number of autoimmune disorders							
1	833 157 (85.1%)	522 095 (83.4%)	311 062 (88.1%)	187 435 (85.2%)	145 738 (85.3%)	196 603 (80.1%)	200 578 (90.3%)
2	124 742 (12.7%)	87 678 (14.0%)	37 064 (10.5%)	28 024 (12.7%)	21 449 (12.6%)	40 314 (16.4%)	19 323 (8.7%)
3 or more	20 973 (2.1%)	16 106 (2.6%)	4 867 (1.4%)	4 588 (2.1%)	3 669 (2.1%)	8 406 (3.4%)	2 159 (1.0%)
Body mass index (kg/m2)							
Mean (SD)	27.8 (6.33)	27.8 (6.69)	27.7 (5.55)	27.1 (5.75)	28.5 (6.99)	27.4 (5.93)	28.1 (6.66)
Missing (%)	509 139 (52.0%)	316 940 (50.6%)	192 199 (54.4%)	120 581 (54.8%)	82 763 (48.4%)	165 750 (67.6%)	102 834 (46.3%)
Smoking status							
Current smoker	126 240 (20.7%)	77 046 (19.7%)	49 194 (22.6%)	18 463 (14.1%)	33 855 (30.3%)	22 325 (23.0%)	26 683 (18.9%)
Former smoker	167 829 (27.6%)	92 501 (23.6%)	75 328 (34.7%)	37 139 (28.4%)	28 250 (25.2%)	22 641 (23.3%)	41 881 (29.6%)
Never smoker	314 512 (51.7%)	221 793 (56.7%)	92 719 (42.7%)	75 190 (57.5%)	49 784 (44.5%)	52 174 (53.7%)	72 703 (51.5%)
Missing (%)	370 291 (37.8%)	234 539 (37.5%)	135 752 (38.5%)	89 255 (40.6%)	58 967 (34.5%)	148 183 (60.4%)	80 793 (36.4%)

Socioeconomic status was defined as the Index of Multiple Deprivation (IMD) 2015 quintile, with 1 referring to the most affluent and 5 to the most deprived socioeconomic quintile. Number and percentage of records with missing data are displayed for variables with missing entries. For variables with missing entries, summary statistics present observed values alongside the percentage of missing values. Category percentages refer to complete cases.

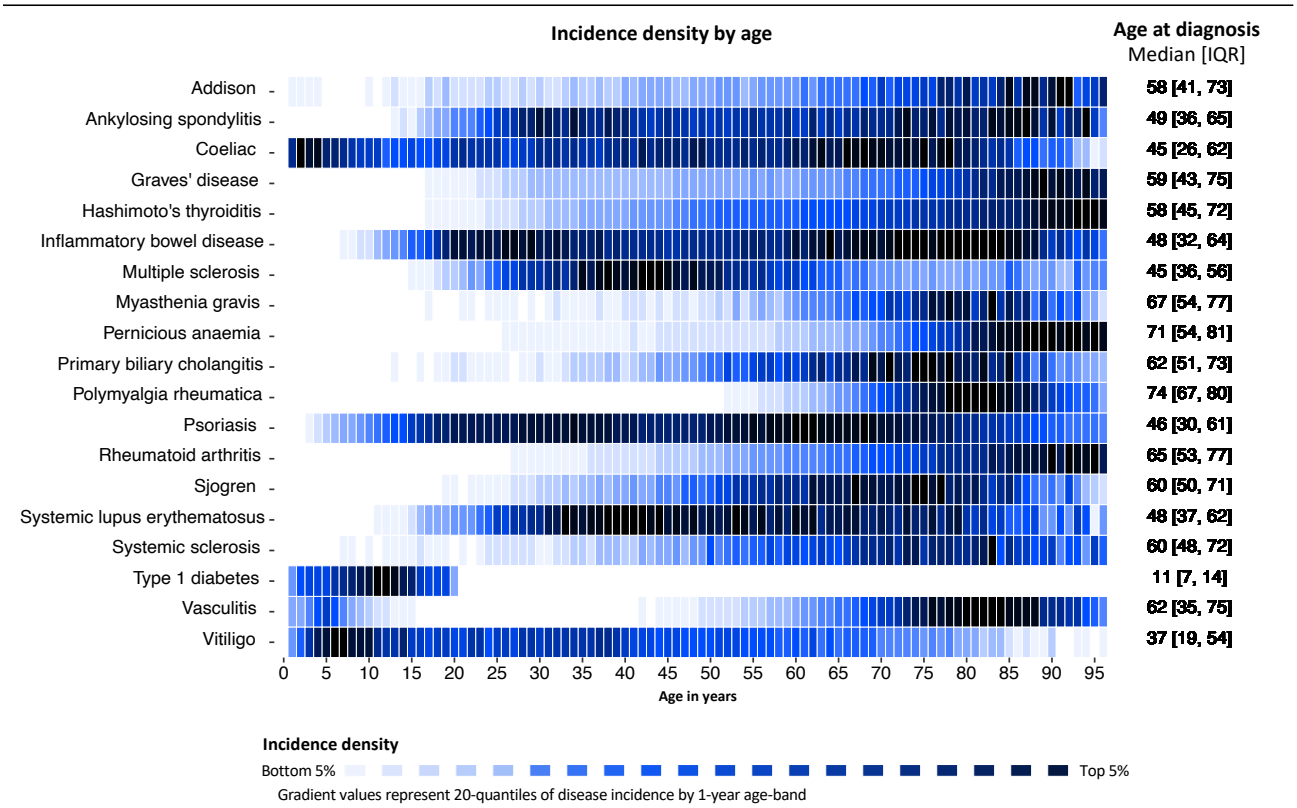
Figure 1: Incidence of autoimmune disorders over time from 2000-2019.

Disease	N	Standardised incidence per 100,000 person-years			Temporal trend
	All years 2000-2019	All years 2000-2019	2000-2002	2017-2019	IRR [95% CI] 2017-2019 vs. 2000-2002
Addison	4 319	3.1	2.0	4.0	1.87 [1.61, 2.16]
Ankylosing spondylitis	12 663	8.3	6.9	11.8	1.78 [1.61, 1.97]
Coeliac	31 447	20.5	13.1	29.6	2.19 [2.05, 2.35]
Graves' disease	75 524	48.8	33.6	66.0	2.07 [1.92, 2.22]
Hashimoto's thyroiditis	284 755	195.8	216.0	161.5	0.81 [0.75, 0.86]
Inflammatory bowel disease	66 173	42.9	40.7	49.0	1.21 [1.14, 1.28]
Multiple sclerosis	15 634	9.9	10.6	10.4	1.01 [0.91, 1.13]
Myasthenia gravis	3 655	2.5	2.1	2.8	1.49 [1.27, 1.76]
Pernicious anaemia	62 582	42.3	40.6	34.9	0.79 [0.72, 0.86]
Polymyalgia rheumatica	90 853	62.9	59.9	62.5	1.07 [0.94, 1.23]
Primary biliary cholangitis	5 683	3.6	3.0	3.6	1.10 [0.95, 1.28]
Psoriasis	236 027	154.5	144.3	149.2	1.04 [0.99, 1.10]
Rheumatoid arthritis	107 559	71.6	58.4	94.0	1.64 [1.52, 1.78]
Sjogren	12 292	8.1	6.0	10.7	2.09 [1.84, 2.37]
Systemic lupus erythematosus	14 164	9.1	8.8	9.4	1.15 [1.03, 1.28]
Systemic sclerosis	4 880	3.1	2.6	3.3	1.23 [1.06, 1.43]
Type 1 diabetes	9 022	5.9	5.3	5.9	1.13 [0.90, 1.41]
Vasculitis	58 919	39.1	36.2	42.5	1.20 [1.13, 1.28]
Vitiligo	27 638	17.6	13.6	18.4	1.28 [1.19, 1.38]
Any autoimmune disorder	**978 872**	**702.2**	**681.6**	**700.8**	**1.04 [1.00, 1.09]**



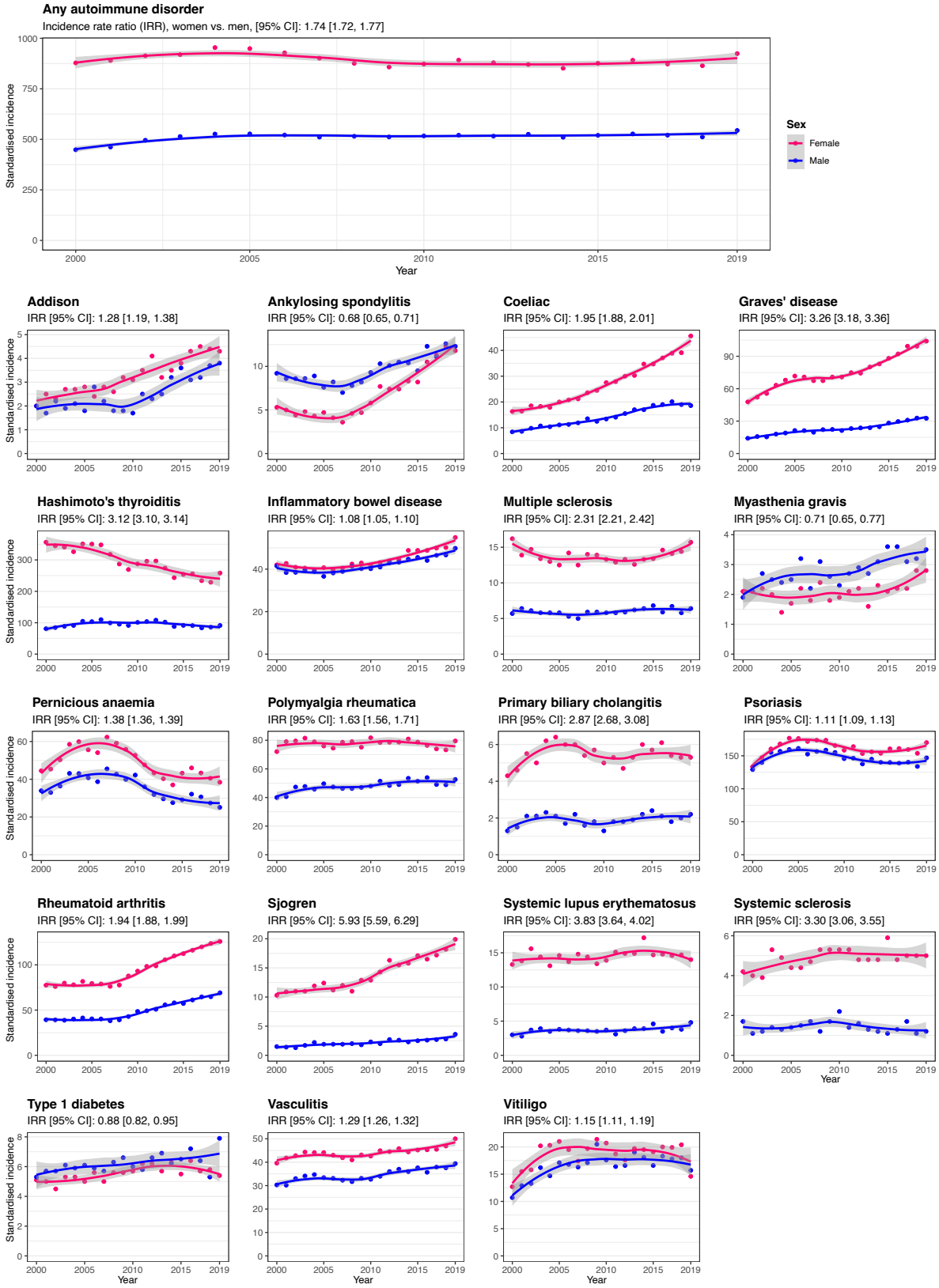
Incidence rates are presented as incidence rates per 100 000 person-years at risk and are age-sex-standardised to the 2013 European Standard Population. 'Any autoimmune disorder' refers to the primary incidence of the 19 autoimmune disorders investigated in this study (that is the number of patients first diagnosed with one autoimmune disease). 'N' refers to the number of patients newly diagnosed with autoimmune disease during the study period.

Figure 2: Incidence of autoimmune disorders by age



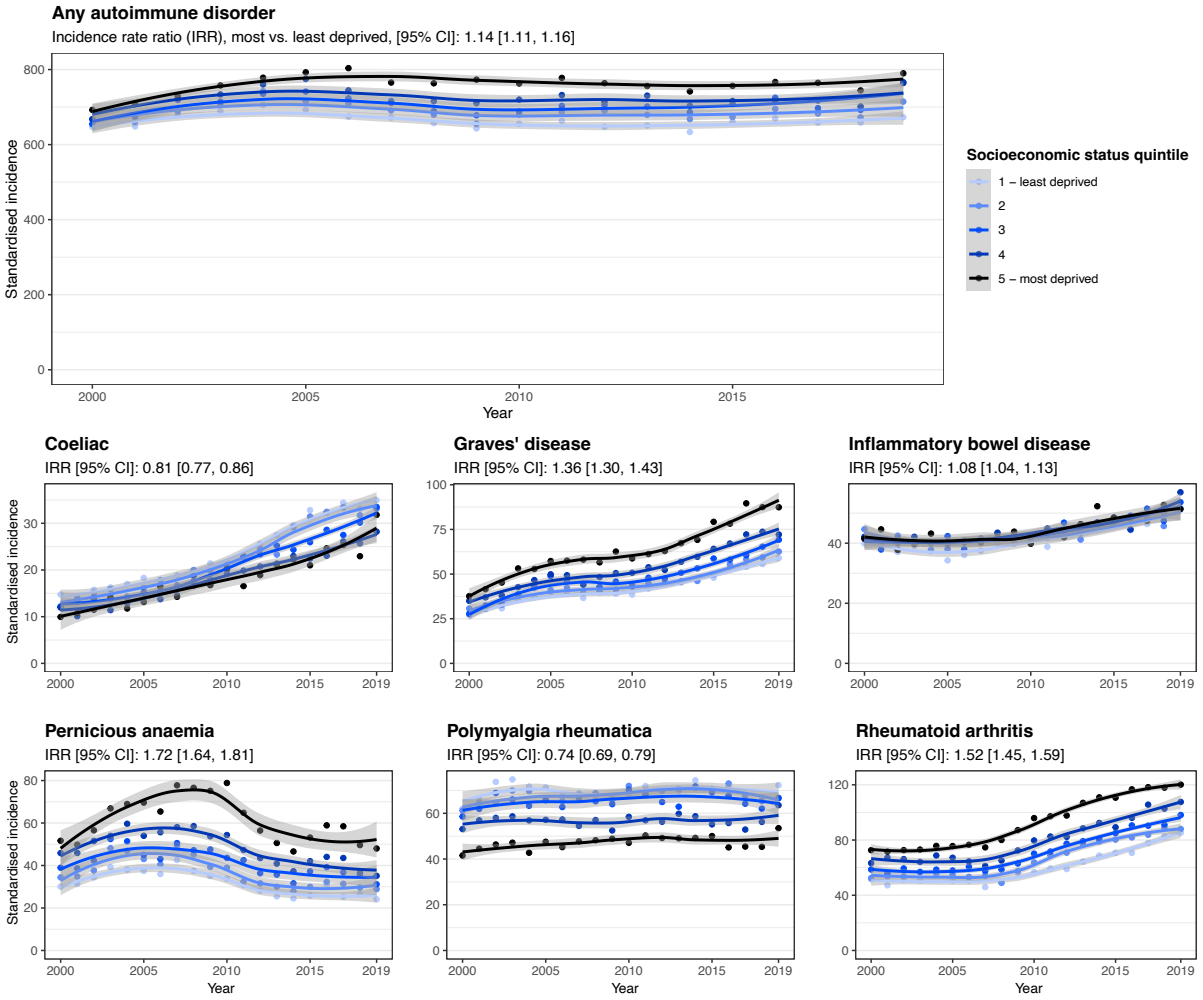
Incidence rates were calculated per 1-year age band and divided into a colour-gradient of 20-quantiles to reflect incidence density by age. The definition of type 1 diabetes used in this study refers to childhood-onset type 1 diabetes and hence to individuals aged 19 years or less at the time of diagnosis. IQR = interquartile range.

Figure 3: Incidence of autoimmune disorders by sex and over time.



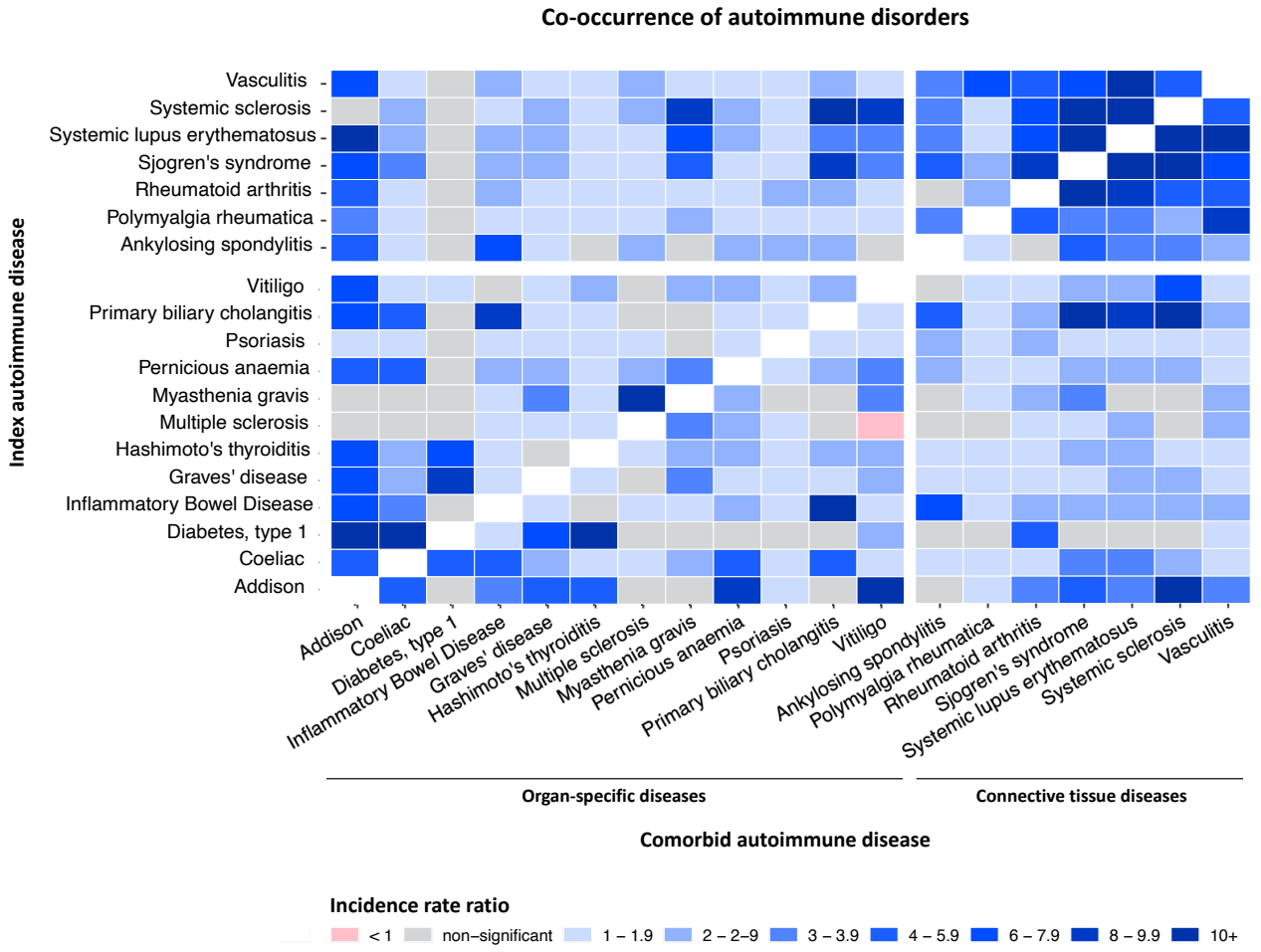
Age-standardised incidence rates by sex. 'Any autoimmune disorder' refers to the primary incidence of the 19 autoimmune disorders investigated in this study. The definition of type 1 diabetes used in this study refers to childhood-onset type 1 diabetes. Denominator populations include all age groups. Yearly incidence estimates were smoothed using loess (locally estimated scatterplot smoothing) regression lines. IRR refers to the incidence rate ratio of women compared to men and CI to the confidence interval.

Figure 4: Incidence of autoimmune disorders by socioeconomic quintile for selected disorders.



Age-sex-standardised incidence rates by socioeconomic status quintile (Index of Multiple Deprivation 2015) for selected autoimmune disorders. 'Any autoimmune disorder' refers to the primary incidence of the 19 autoimmune disorders investigated in this study. Yearly incidence estimates were smoothed using loess (locally estimated scatterplot smoothing) regression lines. IRR refers to the incidence rate ratio of the most deprived socioeconomic quintile compared to the least deprived quintile and CI to the confidence interval. Individual plots for each of the diseases investigated are presented in the ***appendix****.*

Figure 5: Incidence rate ratios for development of comorbid autoimmune disease among index populations with autoimmune disease compared with the general population



Incidence rate ratios were calculated for development of a second (comorbid) autoimmune disease among index populations with pre-existing autoimmune disease compared with the general population using negative binomial regression models adjusted for age and sex. Type 1 diabetes refers to childhood-onset type 1 diabetes, that is among people aged 20 years or less at the time of diagnosis.