The epidemiology of systemic sclerosis in the UK: An analysis of the Clinical Practice Research Datalink

John D Pauling [1,2] Anita McGrogan [1] Julia Snowball [1] & Neil J McHugh [1]

[1] Department of Pharmacy and Pharmacology, University of Bath, Bath, UK

[2] Royal National Hospital for Rheumatic Diseases (part of the Royal United Hospitals NHS Foundation Trust), Bath, UK

Corresponding Author:

Dr John D Pauling BMedSci PhD FRCP

Senior Lecturer & Consultant Rheumatologist

Department of Pharmacy and Pharmacology, University of Bath, Bath, UK

Royal National Hospital for Rheumatic Diseases, Upper Borough Walls, Bath, BA1 1RL

Tel: (0044) 1225 821630 jdp32@bath.ac.uk

Key Indexing Terms: Systemic sclerosis; Raynaud's phenomenon, Epidemiology, Incidence, Prevalence, Mortality, Diagnostic delay

Key Messages:

- A rigorous case ascertainment strategy is important in epidemiological studies of rare autoimmune rheumatic diseases
- A steady/falling incidence rate but trend for rising prevalence may indicate improved outcomes in SSc

- Diagnostic delay of systemic sclerosis in primary care remains an important unmet research need

Short running title: Epidemiology of SSc in UK

Sources of support: This work was funded with matched funding from Scleroderma Raynaud's UK and the Bath Institute for Rheumatic Diseases

Word Count: 3,568

Abstract

Objectives: We developed and tested a robust case ascertainment strategy within the Clinical Practice Research Datalink (CPRD) to assess the incidence, prevalence, mortality and delay in diagnosis of systemic sclerosis (SSc) in the UK.

Methods: A 2-stage case ascertainment strategy was devised and tested to establish a valid cohort of SSc cases within the CPRD. Incidence, prevalence and mortality statistics were analysed, alongside evaluation of the relationship between primary care codes for Raynaud's phenomenon (RP) and SSc to examine diagnostic delay.

Results SSc Read codes were identified in 3,123 patients (from study cohort of >10.1 million individuals). Of these, 1,757 cases of SSc were identified using our case ascertainment approach. The overall incidence rate of SSc over the period between 1999-2017 was 10.7/million/year (95% CI 9.9–11.4); being higher in females (17.69/million/year (95% CI 16.32–19.07)) than males (3.59/million/year (95% CI 2.97–4.21)). The overall prevalence of SSc in adults was 235.5/million (CI 207.2-245.7). The mean rate of mortality was 32/1000 person-years with an overall standardised mortality ratio of 3.51 (95%CI 3.19-3.84). Of those with an initial code of RP prior to a Read code of SSc, 191/854 (22.4%) had a lag period of >10 years.

Conclusion: We have developed and tested a robust case ascertainment strategy to examine the incidence, prevalence, mortality and diagnostic delay of SSc using primary care records of over 10 million UK residents. A significant lag between coding of RP and SSc in many patients suggests diagnostic delay in SSc remains an important unmet need.

Word count: 250

Introduction

Systemic sclerosis (SSc) is associated with high disease-related morbidity, healthcare utilisation and mortality. Epidemiological studies of SSc over the last 60 years suggest a marked temporal increase in the occurrence of SSc and significant geographical variation worldwide(1-3). Reported incidence rates of SSc in the US have risen from 0.6/million/year (between 1947-52) to 21/million/year (1989-1991)(4, 5). There has been a similar increase in estimates of prevalence; rising from 138/million (1950-79) to 276/million (1989-1991). Early epidemiological studies of SSc in the UK suggested a much lower prevalence. The first identified cases using hospital admission data, local membership records from major UK scleroderma charities, data from local primary care centres and requests to regional clinicians (e.g. rheumatologists, respiratory and renal physicians) to report cases of SSc seen over a 12-month period(6). Population estimates for the region were used to estimate SSc prevalence of 30.8/million with an incidence rate of 3.7/million/year(6). A subsequent study employed a similar approach to case identification, (with collaborative support from rheumatologists, dermatologists and renal physicians from within the region), and estimated a higher prevalence of SSc of 88/million; which remained significantly lower than contemporary reports from the US and Europe(4, 7, 8). A recent analysis of the UK's Health Improvement Network (THIN) primary care database reported a higher incidence rate of SSc of 20.2/million/year (95%CI 18.7–21.7) and a prevalence of 253.8/million (95%CI 236.8–271.6) at the end of 2012(9).

The Clinical Practice Research Datalink (CPRD) includes anonymized primary care records from a representative sample of around 8.5% of the UK population and includes medical diagnoses, prescriptions issued in primary care, test data recorded in primary care, referral information and demographics. A recent CPRD analysis estimated the UK prevalence of SSc to be 307/million (95%CI 290–323), with an overall incidence rate of 19.4/million person-years (95%CI 18.3–20.4) (10). The case ascertainment strategy for this study assumed Read codes of 'systemic sclerosis', 'scleroderma' and 'CREST' were sufficient to confirm SSc on the assumption that primary care physicians were unlikely to utilize relevant Read codes unless the diagnosis was confirmed by a hospital specialist(10). External validation of the case ascertainment process was not performed on the assumption that validation of other

autoimmune diseases (citing Wegener's granulomatosis and idiopathic thrombocytopenic purpura) in the CPRD have revealed positive predictive values (PPV) of > 90% (10). Confidence in the accuracy of case ascertainment is necessary to identify small but clinically relevant increased relative risks for co-morbidities such as cancer, thromboembolic and cardiovascular disease. The objectives of the present study were to devise and validate a robust case ascertainment strategy for SSc cases within the CPRD and report our own analysis of the incidence, prevalence, mortality and time to diagnosis of SSc in the UK.

Methods

Data source and ethics

The CPRD has collected anonymised patient data from 42 million patient lives (including 13 million currently registered patients) from a network of primary care practices across the UK. Approval was obtained from the CPRD Independent Scientific Advisory Committee (protocol 17_109R_18/7/2017). Case ascertainment was based on searches of CPRD records for relevant Read codes. Read codes are derived from a complex structured UK-based clinical classification system devised to provide uniformity within primary care electronic health records. The Read code system comprises disease names, alongside symptoms, relevant history, examination findings, therapeutic procedures, laboratory results, administrative data (e.g. referral to rheumatology) and social information (11).

Study cohort

To be eligible, patients needed to be permanently registered with their general practitioner and have >12 months of data between their left and right censor dates that was up to a standard suitable for research and within the time period for the study: from 01/01/1998 to 31/12/2017. Patients with SSc were initially identified using a combination of diagnostic Read codes for SSc: 'Scleroderma' [Medcode ID: 3670], 'CREST syndrome' [17675], 'Systemic sclerosis' [28417], 'Progressive systemic sclerosis' [44141], 'Myopathy due to scleroderma' [55601], 'Other forms of systemic sclerosis' [71763], 'Lung disease with systemic sclerosis' [94996], 'Acute scleroderma renal crisis' [105976] and 'Renal involvement in scleroderma' [107382].

Development and validation of case ascertainment strategy

We utilised our in-house browser to review individual potential SSc case records in a format consistent with the presentation of data to primary care physicians. An initial review of 101 primary care records by an experienced SSc clinician (JDP) identified inconsistencies around SSc coding and identified supportive information that improved confidence around true cases. We sought to derive a case ascertainment strategy that achieved a minimum acceptable PPV of 0.75.

A review of cases with a code for linear or circumscribed localised scleroderma/morphoea confirmed a high negative predictive value and were excluded. Relevant primary supporting information was sought from the patients' records as evidence of a diagnosis of SSc including:

a) a diagnosis of Raynaud's phenomenon (RP): This was only counted once irrespective of the number of times recorded for a patient.
b) two or more prescriptions on different days for proton pump inhibitors (PPI), calcium channel blockers including nifedipine, ACE inhibitors, or DMARDs (distinct dates of prescriptions were summed and added to the evidence count).
c) referral to rheumatology services in secondary care, digital ulcers, interstitial lung disease, pulmonary hypertension or thyroid disease as supporting comorbidities (distinct dates of diagnosis of these conditions were summed and added to the evidence count).
d) relevant autoantibody tests; a count of all of the individual dates that relevant autoantibody test results were recorded was included in the evidence count.

A preliminary review of cases with no additional supporting evidence confirmed a high negative predictive value and these were excluded (Figure 1). An aggregate evidence score of 2, 3, 4 or 5 (with points allocated for each of the above items) was devised, including an initial point for the SSc diagnostic code. The case notes of all 142 patients with an aggregate score of 2-4 were individually reviewed (JDP) to identify SSc cases. It was hoped this process would establish an aggregate score cut-off that would achieve the desired PPV threshold of 0.75 but this was not achieved. A random sample of 363 cases with an aggregate score of ≥5 were then reviewed (JDP) indicated that five pieces of evidence still did not achieve reach the desired threshold for an accurate outcomes study (PPV 0.55, CI 0.5-0.6).

Secondary evidence criteria were therefore devised for all remaining cases with an initial aggregate score of ≥5. Based on observations that certain features were more commonly associated with a SSc diagnosis, subjects were attributed an additional point if the following features were present: a) SSc or CREST medical code (rather than scleroderma alone); b) RP; c) hypothyroidism; d) rheumatology referrals, e) evidence of DMARDs or PPIs. Having identified useful negative predictive associations from the earlier case-note review, a point was deducted for each of the following criteria: a) medical codes referring to diabetes; b) when a diagnosis for SSc only appears early in the patient record; c) where there was a single diagnostic code for scleroderma only d) where no PPIs were prescribed. An additional 769 cases were individually reviewed (JDP) to determine the threshold number of pieces of secondary evidence necessary to achieve the desired PPV of >0.75 (Figure 1).

Covariates

Records for body mass index (BMI), smoking status and alcohol use at the date closest to the diagnosis of SSc were captured from the CPRD. BMI was categorised into bands (<18.5, 18.5-24, 25-29, 30-34, 35-39, 40+ kg/m^2); smoking status was defined as smoker, non-smoker or ex-smoker where smoking status in the patient's record had changed from smoker to non-smoker or the status was recorded as ex-smoker.

Statistical analysis

Incident cases were defined as those that occurred during the study period and where patients had not had a previous diagnosis of SSc. Incidence of SSc was calculated by year of diagnosis (1999-2017) and by fifteen-year age band at diagnosis using the rest of the population present on the CPRD with at least one year of up to standard data as the denominator. An incidence rate standardised to the 2013 European standard population was reported. Stratified incidence rate estimates were obtained by age and sex. Point prevalence rates at the mid-year point were evaluated using the mid-year SSc and population counts as the numerator and denominator respectively. Mortality rates were reported by year, age and sex and standardized mortality rates. The lag period between a primary care code for RP that preceded the first code for SSc was examined to explore potential delays to diagnosis.

Results

Epidemiology of SSc in UK

Case ascertainment

Patients with SSc were identified from a total study cohort of 10.1 million individuals. Of 3,123 patients with a Read code for possible SSc, we determined that only 1,757 (56.3%) of cases could be reliably considered SSc (694 of which were independently verified after case note review). A summary of the outcome of our case ascertainment strategy is presented in Figure 1. A total of 1,312 patient case-notes were individually reviewed as part of this process. Within these, the negative predictive value (NPV) was 0.95 for patients with a code for morphoea/linear scleroderma. Similarly, the NPV was >0.9 for patients with no supportive evidence beyond a Read code for possible SSc. When primary supporting evidence was used, the PPV rose from 0.16 for 2 pieces of evidence to 0.55 for 5 pieces of evidence. Using secondary evidence, an initial review of case-notes of individuals with 0-2 pieces of supportive evidence identified a PPV of 0.61 (95%CI 0.57-0.64). A decision was made to review each of these case notes individually.

Specific factors greatly increased the likelihood of a patient being categorised as SSc if they also had 5 pieces of primary evidence. For example, patients with a diagnosis of interstitial lung disease (ILD) gave a PPV of 0.98 but only 47 patients had this diagnosis. A code for Raynaud's in patients with SSc (and 5 pieces of primary evidence) also had a high PPV (0.82). Many patients with SSc (508/1757), however, did not have a Read code for Raynaud's, although cases were also identified that were not suggestive of SSc despite the presence of a Read code for Raynaud's (311/1366). Suspected SSc (or later excluded SSc) was one reason why a patient could have Read codes for both RP and 'SSc' and be considered as unlikely to have SSc. Patients with 5 pieces of primary evidence taking a disease modifying anti-rheumatic disease drug or proton pump inhibitor had a PPV of 0.63 and patients with hypothyroidism had a PPV of 0.61. Patients with five pieces of primary evidence and a secondary evidence score of ≥3 reached our threshold target, with a PPV of 0.85. Using this, a further 1063 patients were identified for inclusion in the SSc cohort using an algorithm of medical codes and supporting information (Figure 1).

SSc cohort

The patient characteristics for the 1757 SSc patients included in the cohort are summarised in Table 1, separated into incident and prevalent cases. Consistent with the expected demographics from previously published patient registry analyses, the majority of SSc patients were female (>83%), aged between 55 and 69 years at diagnosis and with a low rate of obesity (~15%). As anticipated, SSc in childhood or young adults was uncommon (only ~10% under 40yrs).

Other clinical features provided confidence of the veracity of our final SSc cohort. Of these, the presence of a Read code for RP at any stage (present in 1249 [71.1%]) was the most common. Known co-morbidities of SSc were also prevalent. For example, 412 (23.4%) had a Read code of hypothyroidism (with hypothyroidism predating the diagnosis of SSc in 264 [15% of total] of cases). Accepting such information may be incompletely documented within primary care records, Read codes for typical SSc organ-specific complications included lung fibrosis in 214 (12.2%) cases, pulmonary arterial hypertension in 158 (9%) cases, digital ischaemic lesions in 89 (5.1%) cases, telangiectasia in 56 (3.2%) cases and calcinosis in 84 (4.8%) cases. Medication usage (at any stage in disease course) was also consistent with an underlying diagnosis of SSc with 1567 (89.2%) prescribed PPIs, 1339 (76.2%) prescribed calcium channel blockers (e.g. nifedpine/amlodipine) and 386 (22.0%) prescribed immunomodulatory agents such as methotrexate, mycophenolate mofetil and/or D-penicillamine.

Prevalence of SSc in UK

The overall prevalence of SSc within the adult population of the UK (>18yrs) was 235.5/million (95%CI 215.45-255.55) in 2016. For the under 18s the prevalence in 2016 was much lower at 1.74/million (95%CI 1.67-5.15). A summary of the prevalence of SSc by age-band and sex in 2016 is presented in Table 2. The prevalence rises with age with the highest prevalence being reported in those aged 70-84yrs (543/million), before falling in those ≥ 85 years. The prevalence of SSc has increased over the period 1998-2017 (Figure 2).

Incidence of SSc in UK

The overall incidence rate was 10.7/million person-years (95%CI 9.9–11.4) (Table 2). As anticipated, the highest incidence was identified in females aged 55-59 years (43.4/million person-years [95%CI 38.2–48.4]). We identified a mildly declining incidence over the period 1999-2017 (Figure 3). The apparent increase in incidence in the final full year of data collection (for 2016) could be an artefact in the data and future work should be undertaken to determine any further trends in incidence over time. The aforementioned rising prevalence rate between 1999-2017, despite a steady/falling incident rate, may indicate improved survival in SSc.

Mortality in SSc in UK

Mortality has fluctuated over time with a mean rate of 32/1000 person-years and an overall standardised mortality ratio of 3.51 (95%CI 3.19–3.84). For patients with an incident diagnosis of SSc in the CPRD, there were 150 deaths in total. Over half of incident SSc patients who had died (83/150, 55.3%), died within 1-5 years from diagnosis and a further third (49/150) died with 5-10 years of diagnosis. Only 12% of the incident SSc patients who had died before the end of the period of study (18/150) lived for >10yrs after diagnosis with SSc.

Relationship between RP and SSc Read codes in CPRD

A total of 1249 patients (71.1%) had a Read code for RP. Of these, 80 (6.4%) received their first 'Read code' for RP and SSc on the same day. This likely represents a secondary care diagnosis resulting in both diagnostic labels being flagged to primary care services contemporaneously. The Read code for RP followed their code for SSc in 315 patients (25.2%). Similarly, it is unlikely that RP symptoms occurred following the onset of SSc but that this diagnostic code was first inputted into the primary care record following the diagnosis of SSc. In the remaining 854 (68.4%) subjects, the Read code for RP pre-dated their code for SSc, suggesting the patient likely visited their primary care provider for RP symptoms before subsequently being provided with a diagnosis of SSc in secondary care. Table 3 provides a summary of the lag period between the RP and SSc codes. Whilst this period was less than 1 year for 243 (28.5%), almost half of these patients (420 [49.25%]) had a lag of between 1 and 10 years. Almost a quarter of subjects (n=191, 22.4%) of those with RP code before SSc had a

lag period of >10 years from their initial code of RP to a Read code of SSc being recorded, suggesting diagnostic delay remains a major problem in SSc.

Discussion

A comprehensive understanding of the epidemiology of SSc in the UK is desirable. Worldwide geographic differences in prevalence may be a consequence of a number of important factors including differences in genetic, ethnic, socioeconomic and environmental factors causing SSc and also differences in healthcare service provision. Understanding changes in disease incidence may also identify relevant aetiological drivers of SSc, whilst a more accurate assessment of UK prevalence of SSc will have implications for service planning and accurate modelling of specialty services and high-cost drug use. Using a robust algorithm for the identification of patients with SSc in the CPRD, we have identified an overall incidence rate of 17.69/1,000,000/year (95%CI 16.32–19.07) in females and 3.59/1,000,000/year (95%CI 2.97–4.21) in males (between 1999-2017). Incidence and prevalence peaked in those aged 55-69 years; although prevalence remained high in those aged 70-84 years. Incidence has dropped since 2003, whereas overall prevalence increased steadily until 2014, with some fluctuation of overall mortality between 1999-2017. Our study benefits from access to the primary care records of over 10 million UK patients. The chief limitation of our study is the inability to perform linked hospital medical record reviews to confirm the diagnosis. CPRD data can be linked with Hospital Episode Statistics (HES) data but this only provides reliable data for hospital inpatient admissions, which for many people with SSc is not necessary. We also did not have access to death certificates which would have allowed a more detailed assessment of mortality trends. We could not verify our findings with hospital records because the CPRD only captures data on ~1/5 of UK patients and many patients travel to tertiary SSc centres for their care.

Our estimates of the incidence/prevalence of SSc reasonably align with estimates in European studies undertaken in Sweden (incidence 19/million/year and prevalence 305/million) (12), Spain (23/million/year and 277/million respectively) (8) and Italy (43/million/year and 341/million respectively) (13), suggesting a similar burden of SSc across European nations. There have been reports suggesting a lower prevalence of SSc in European nations including

Norway (99/million)(14), France (158/million)(15), Croatia (156/million)(16) and Greece (154/million) (17) that may reflect differences in methodological approaches.

Advances in disease classification (leading to improved recognition of early and limited forms of SSc) and different approaches to case ascertainment (e.g. hospital records and population surveys) may be one factor influencing the differences in observed prevalence. The broader repertoire of antigen targets that can be identified using modern serological testing, greater clinician awareness of SSc and the incorporation of nailfold capillaroscopic changes in updated classification criteria for SSc(18, 19) may have supported earlier recognition and diagnosis of SSc. Prolonged life expectancy in general, along with specific advances in the management of SSc e.g. lower mortality secondary to scleroderma renal crisis may have influenced crude prevalence rates, but might mask a more complex pattern of epidemiological trends that could shed important light on possible pathogenic drivers such as environmental exposure.

Our analyses suggest a much higher incidence/prevalence of SSc than that derived from early attempts to study this in the UK(6, 7), but a significantly lower incidence/prevalence of SSc than was reported in a more recent CPRD analysis(10). There were methodological limitations to early rudimentary attempts to describe the basic epidemiology of SSc in the UK that may have resulted in underestimates in disease prevalence (6, 7). Furthermore, these studies pre-dated current classification approaches and access to extended serological testing, which enable clinicians to identify SSc in patients with very early disease (often in the absence of skin thickening). Indeed, previous epidemiological studies that have compared different SSc classification approaches to carefully phenotyped SSc cases have identified an increase in incidence/prevalence of SSc of ~30% when applying more recent classification approaches (18, 19) compared with the earlier 1980 American Rheumatology Association classification criteria (20)for SSc (12) .

The main difference between the our study and the recently reported CPRD study was the extensive validation work we have undertaken to establish a robust case ascertainment strategy and exclude false positive cases. Reviewing potential SSc cases in the inhouse

browser to check diagnoses and supporting information proved to be valuable in this regard. Whilst others have anticipated that a medical diagnosis code for a condition as serious as SSc would not be used incorrectly(10), the opportunity to review the patient's primary care records indicated the appearance of SSc codes despite a low index of suspicion for true SSc. Indeed, 43.7% of the identified 3,123 cases with at least one Read code for SSc or CREST or scleroderma were excluded on the basis of our final case ascertainment strategy; devised following individual case-note review of 1,312 individual primary care health records. We did not record the reasons for exclusion of individual cases but it was noted that many had evidence of alternative sclerosing skin disorders related to diabetes, linear scleroderma/morphoea, hypergammaglobulinaemia, lipodermatosclerosis, chronic venous insufficiency, post-surgical scarring or other skin related conditions (particularly those with a code of 'scleroderma' alone). Occasionally medical diagnoses for CREST or SSc were used in anticipation of a diagnosis before presumed subsequent exclusion following referral to rheumatology (with no subsequent prescriptions, referrals or diagnoses to indicate a SSc diagnosis). Occasionally, a SSc code was inserted in the medical record of patients with established tuberous *sclerosis* or multiple *sclerosis* as apparent coding errors. The approach we have taken may have resulted in the exclusion of true cases and led to an under-estimate of the actual incidence/prevalence of SSc. Nonetheless, we consider our case ascertainment approach has resulted in a more accurate epidemiological analysis of SSc in the UK and our estimates of incidence and prevalence align with those derived from a smaller study undertaken using the THIN primary care database (that also incorporated a case validation exercise)(9). This study also suggested the rising prevalence of SSc was not related to changes in disease incidence(9). Our study benefits from examining a much larger database and our robust case ascertainment approach shall be valuable for planned work to examine the relationship between SSc and cancer, venous thromboembolic events and cardiovascular disease.

We have identified a significant delay in diagnosis of SSc in patients previously being labelled as having 'Raynaud's'. The availability of a convenient 'diagnosis' in the form of 'Raynaud's' may contribute to delayed diagnosis of SSc. A previous meta-analysis reported ~12.6% of people initially diagnosed with primary RP followed up in secondary care services eventually

transitioned to a defined autoimmune rheumatic disease(21). Studies indicate that features of autoimmune rheumatic disease are typically present at an earlier stage that could have supported earlier diagnosis. Indeed, the presence of RP, nailfold capillary abnormalities and a SSc-associated antibody (without any overt skin thickening) are sufficient for the classification of SSc(19). The average time between the onset of RP and the emergence of the first non-RP symptom in women with limited cutaneous SSc is >5 years(22). Over a quarter of women with SSc are not diagnosed until >10 years after RP symptoms first emerged(23). Our data confirms that delayed diagnosis of SSc in primary care represents an important unmet research need. Over 70% of our SSc cases had a Read code for RP in their primary care record, and in over 2/3 of these cases, the Read code for RP pre-dated that of SSc; up to 20 or more years for some. Nearly 40% of those with documented Read code of RP before a subsequent diagnosis of SSc, had >5 years before a SSc diagnosis. Many patients have other potential red flags that could also be better exploited to alert primary care physicians of the possibility of SSc. Approximately half of SSc diagnoses, meanwhile, are made without any prior documentation of RP within the primary care record. Previous community-based studies have identified low healthcare utilization for RP symptoms (24). UK-based studies suggest only ~20% of people with RP had consulted their GP about RP symptoms(25, 26). Raising awareness amongst the general public of the potential importance of RP symptoms may help avoid diagnostic delay in SSc.

Conclusions

We have utilised primary care records of >10 million people to explore the incidence, prevalence, mortality and diagnostic delay of SSc in the UK. We highlight the need to develop robust case ascertainment strategies when working with large primary care datasets. We have also identified potentially important areas of unmet need around diagnostic delay of SSc and appropriate recognition of potential red flags such as RP/hypothyroidism etc. in patients attending primary care services that may increase the pre-test probability of SSc.

Disclosures

John Pauling has received research grants, honoraria and undertaken consultancy work for Actelion pharmaceuticals. John Pauling has undertaken consultancy work for Boehringher-Ingelheim and Sojournix Pharma.

CRediT Author Statement

John Pauling: Conceptualization, Methodology, Investigation, Data Curation, Writing -Original Draft, Visualisation, Supervision and Project Administration; Anita McGrogan: Methodology, Investigation and Writing-Review & Editing; Julia Snowball: Methodology, Investigation and Writing-Review & Editing; Neil J McHugh: Writing- Reviewing and Editing.

References

1. Barnes J, Mayes MD. Epidemiology of systemic sclerosis: incidence, prevalence, survival, risk factors, malignancy, and environmental triggers. Curr Opin Rheumatol. 2012;24(2):165-70.
2. Nikpour M, Stevens WM, Herrick AL, Proudman SM. Epidemiology of systemic sclerosis. Best Pract Res Clin Rheumatol. 2010;24(6):857-69.
3. Valter I, Saretok S, Maricq HR. Prevalence of scleroderma spectrum disorders in the general population of Estonia. Scand J Rheumatol. 1997;26(6):419-25.
4. Mayes MD, Lacey JV, Jr., Beebe-Dimmer J, Gillespie BW, Cooper B, Laing TJ, et al. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. Arthritis Rheum. 2003;48(8):2246-55.
5. Medsger TA, Jr., Masi AT. Epidemiology of systemic sclerosis (scleroderma). Ann Intern Med. 1971;74(5):714-21.
6. Silman A, Jannini S, Symmons D, Bacon P. An epidemiological study of scleroderma in the West Midlands. Br J Rheumatol. 1988;27(4):286-90.
7. Allcock RJ, Forrest I, Corris PA, Crook PR, Griffiths ID. A study of the prevalence of systemic sclerosis in northeast England. Rheumatology (Oxford). 2004;43(5):596-602.
8. Arias-Nunez MC, Llorca J, Vazquez-Rodriguez TR, Gomez-Acebo I, Miranda-Filloy JA, Martin J, et al. Systemic sclerosis in northwestern Spain: a 19-year epidemiologic study. Medicine (Baltimore). 2008;87(5):272-80.
9. Garcia Rodriguez LA, Gonzalez-Perez A, Michel A, Saez ME. Contemporary epidemiology of systemic sclerosis: A population-based cohort study in the United Kingdom. Semin Arthritis Rheum. 2019;49(1):105-11.
10. Royle JG, Lanyon PC, Grainge MJ, Abhishek A, Pearce FA. The incidence, prevalence, and survival of systemic sclerosis in the UK Clinical Practice Research Datalink. Clinical rheumatology. 2018;37(8):2103-11.
11. Chisholm J. The Read clinical classification. BMJ. 1990;300(6732):1092.
12. Andreasson K, Saxne T, Bergknut C, Hesselstrand R, Englund M. Prevalence and incidence of systemic sclerosis in southern Sweden: population-based data with case ascertainment using the 1980 ARA criteria and the proposed ACR-EULAR classification criteria. Ann Rheum Dis. 2014;73(10):1788-92.

13. Lo Monaco A, Bruschi M, La Corte R, Volpinari S, Trotta F. Epidemiology of systemic sclerosis in a district of northern Italy. Clin Exp Rheumatol. 2011;29(2 Suppl 65):S10-4.
14. Hoffmann-Vold AM, Midtvedt O, Molberg O, Garen T, Gran JT. Prevalence of systemic sclerosis in south-east Norway. Rheumatology (Oxford). 2012;51(9):1600-5.
15. Le Guern V, Mahr A, Mouthon L, Jeanneret D, Carzon M, Guillevin L. Prevalence of systemic sclerosis in a French multi-ethnic county. Rheumatology (Oxford). 2004;43(9):1129-37.
16. Radic M, Martinovic Kaliterna D, Fabijanic D, Radic J. Prevalence of systemic sclerosis in Split-Dalmatia county in Southern Croatia. Clin Rheumatol. 2010;29(4):419-21.
17. Alamanos Y, Tsifetaki N, Voulgari PV, Siozos C, Tsamandouraki K, Alexiou GA, et al. Epidemiology of systemic sclerosis in northwest Greece 1981 to 2002. Semin Arthritis Rheum. 2005;34(5):714-20.
18. LeRoy EC, Medsger TA, Jr. Criteria for the classification of early systemic sclerosis. J Rheumatol. 2001;28(7):1573-6.
19. van den Hoogen F KD, Fransen J, Johnson SR, Baron M, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. Arthritis & Rheumatism. 2013;65(11):2737-47.
20. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Arthritis Rheum. 1980;23(5):581-90.
21. Spencer-Green G. Outcomes in primary Raynaud phenomenon: a meta-analysis of the frequency, rates, and predictors of transition to secondary diseases. Arch Intern Med. 1998;158(6):595-600.
22. Walker UA, Tyndall A, Czirjak L, Denton C, Farge-Bancel D, Kowal-Bielecka O, et al. Clinical risk assessment of organ manifestations in systemic sclerosis: a report from the EULAR Scleroderma Trials And Research group database. Ann Rheum Dis. 2007;66(6):754-63.
23. Delisle VC, Hudson M, Baron M, Thombs BD, And The Canadian Scleroderma Research Group A. Sex and time to diagnosis in systemic sclerosis: an updated analysis of 1,129 patients from the Canadian scleroderma research group registry. Clin Exp Rheumatol. 2014;32(6 Suppl 86):S-10-4.
24. Silman A, Holligan S, Brennan P, Maddison P. Prevalence of symptoms of Raynaud's phenomenon in general practice. BMJ (Clinical research ed). 1990;301(6752):590-2.
25. Heslop J, Coggon D, Acheson ED. The prevalence of intermittent digital ischaemia (Raynaud's phenomenon) in a general practice. J R Coll Gen Pract. 1983;33(247):85-9.
26. Palmer KT, Griffin MJ, Syddall H, Pannett B, Cooper C, Coggon D. Prevalence of Raynaud's phenomenon in Great Britain and its relation to hand transmitted vibration: a national postal survey. Occup Environ Med. 2000;57(7):448-52.

Figure 1. Summary of case ascertainment strategy and determination of final cohort of SSc patients.

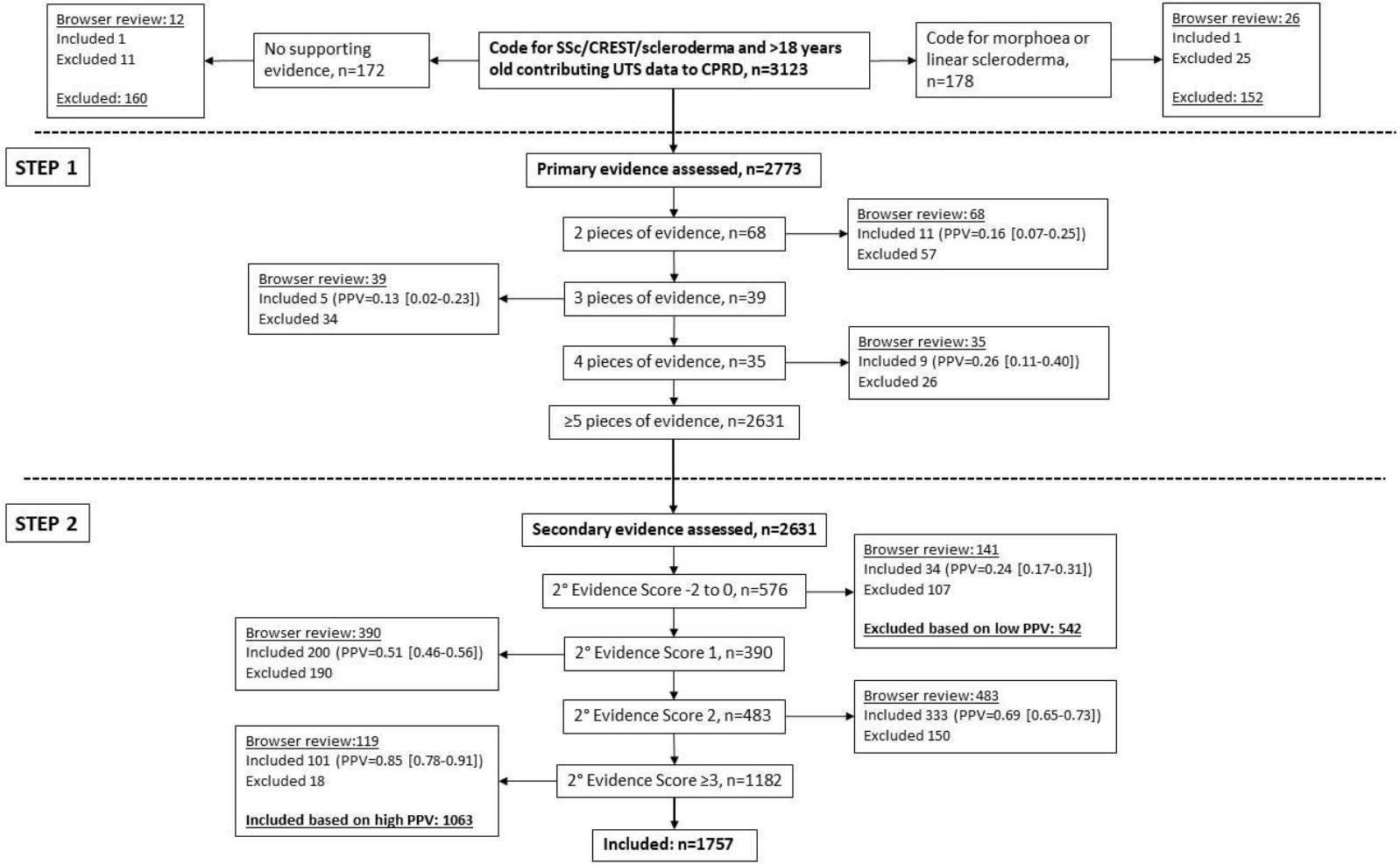

Table 1: Patient demographics of incident and prevalent cases of SSc

		Incident		Prevalent	
		N	**%**	**N**	**%**
Sex	**Female**	649	83.4	854	87.2
	Male	129	16.6	125	12.8
Age at index	**<15**	0	0	0	0
	15-39	70	9.0	104	10.6
	40-54	194	25.0	295	30.1
	55-69	338	43.4	385	39.3
	70-84	165	21.2	177	18.1
	85+	11	1.4	18	1.8
Smoking	**Y**	159	20.5	206	21.0
	N	412	53.0	510	52.1
	Ex	203	26.1	253	25.8
	Unknown	-	-	10	1.0
BMI	**<18.5**	35	4.5	33	3.4
	18.5-24.9	266	34.2	335	34.2
	25-29.9	261	33.6	203	20.7
	30-34.9	93	12.0	62	6.3
	35-39.9	25	3.2	18	1.8
	40+	19	2.5	-	-
	Unknown	79	10.2	325	33.2

Table 2: Incidence (1999-2017) and prevalence (during 2016) of SSc in UK by sex and age-band

	Total		Female		Male	
Age	**Number**	**Incidence/million/year**	**Number**	**Incidence/million/year**	**Number**	**Incidence/million/year**
15-39	77	3.5 (2.7 – 4.3)	63	5.9 (4.4 – 7.3)	14	1.2 (0.6 – 1.9)
40-54	195	12.1 (10.4 – 13.8)	156	19.7 (16.6 – 22.8)	39	4.8 (3.3 – 6.3)
55-69	332	26.0 (23.26 – 29.0)	278	43.4 (38.2 – 48.5)	54	8.5 (6.2 – 10.7)
70-84	152	20.0 (16.8 – 23.1)	131	31.0 (25.7 – 36.4)	21	6.2 (6.5 – 8.8)
≥ 85	9	4.9 (1.7 – 8.2)	9	7.3 (2.5 – 12.0)	0	0
Total	**765**	**10.7 (9.9 – 11.4)**	637	**17.7 (16.3 – 19.1)**	**128**	**3.6 (3.0 – 4.2)**
Age	**Number**	**Prevalence/million**	**Number**	**Prevalence/million**	**Number**	**Prevalence/million**
15-39	27	31.1 (19.4-42.9)	26	60.1 (37.0-83.2)	1	2.3 (0-6.8)
40-54	99	162.3 (130.3-194.2)	83	275.1 (215.9-334.3)	16	51.9 (26.5-77.3)
55-69	227	455.9 (396.6-515.2)	192	766.2 (657.9-874.6)	35	141.5 (94.6-188.4)
70-84	160	542.7 (458.7-626.8)	146	917.8 (769.0-1066.6)	14	103.2 (49.1-157.2)
≥ 85	18	241.6 (130.0-353.2)	16	332.2 (169.4-494.9)	-	75.9 (0-181.1)
Total	531	226.5 (207.2-245.7)	463	463 (353.0 - 423.7)	68	59 (45.0 - 73.0)

Figure 2: Changing overall prevalence of SSc in UK over period 1999-2017

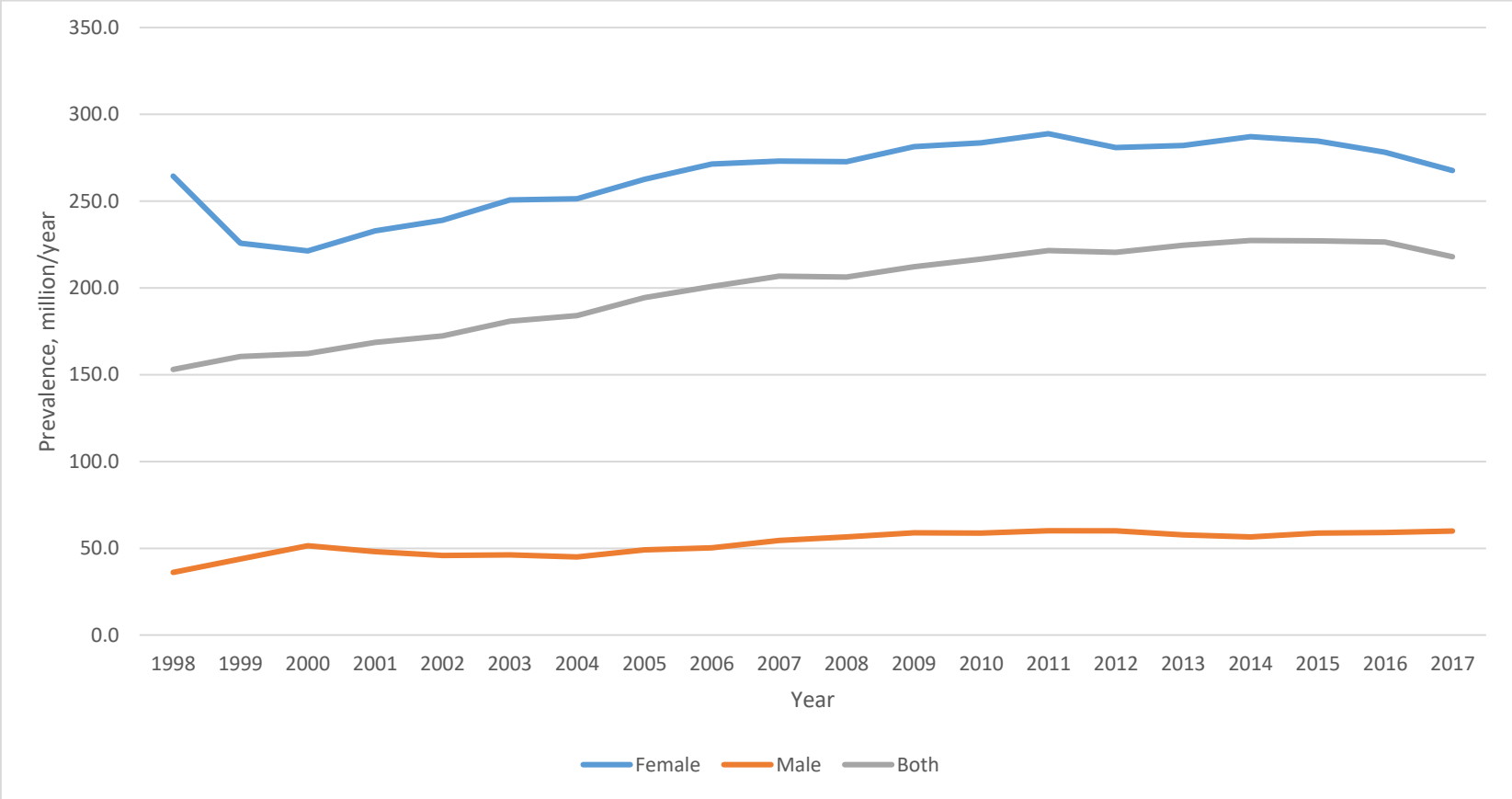

Figure 3: Changing incidence of SSc in UK over period 1999-2017

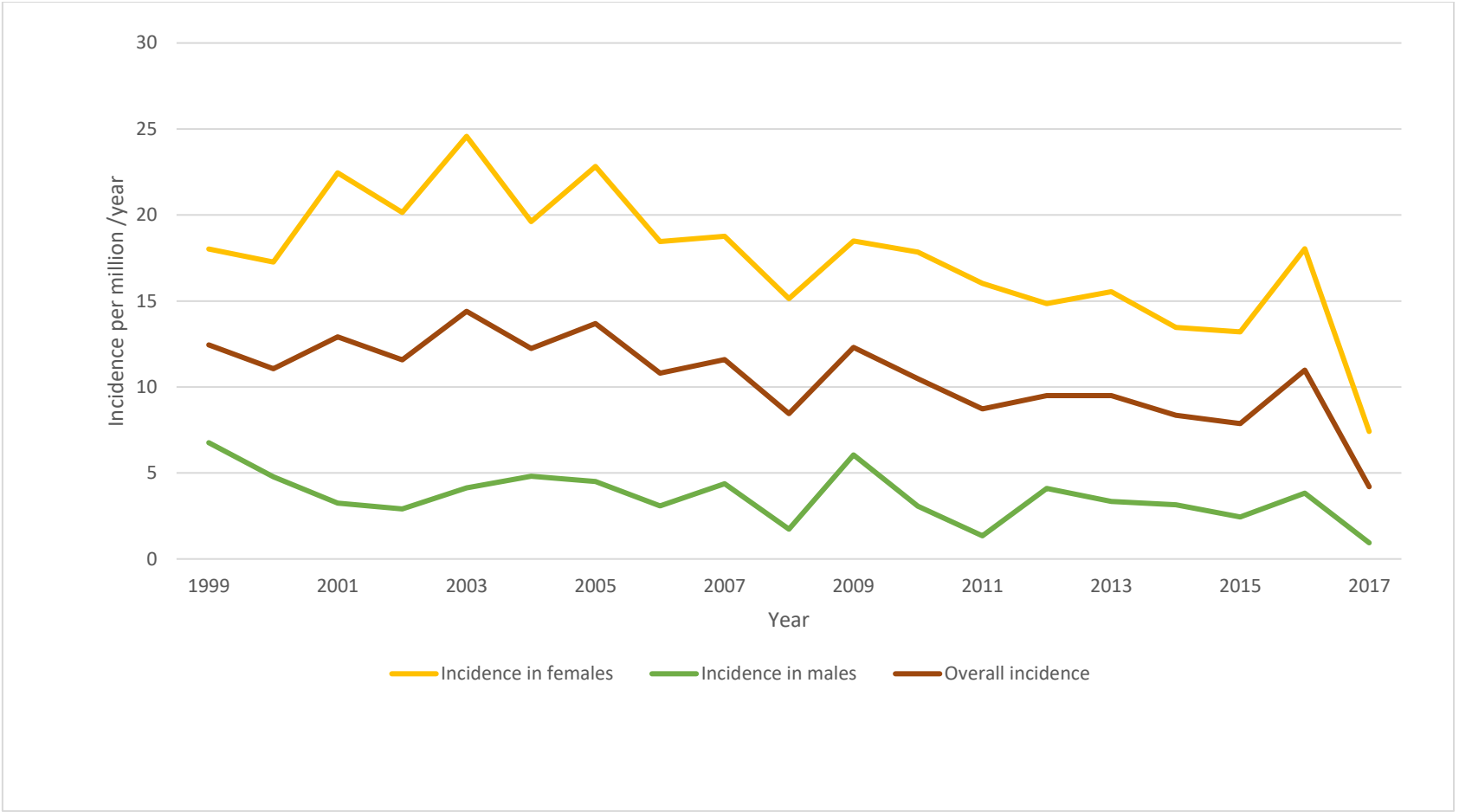

Table 3. Time between the first read code for Raynaud's and the first read code for SSc

Time from Raynaud's to SSc diagnosis	Count of patients (%)	Count of patients (%)
<1 month	33 (3.9)	243 (28.5)
1-3 months	62 (7.3)	
3-6 months	57 (6.7)	
6 months - 1 year	91 (10.7)	
1-3 years	178 (20.8)	278 (32.6)
3-5 years	100 (11.7)	
5-10 years	142 (16.6)	333 (39)
10 years - 15 years	82 (9.6)	
15 years - 20 years	55 (6.4)	
> 20 years	54 (6.3)	
Total	**854**	