Phenome-Wide Association Study of Drugs and Co-morbidities Associated with Gastrointestinal Dysfunction in Systemic Sclerosis

Rory H. Maclean MBBS[1], Fiza Ahmed MBBS[1], Voon H. Ong FRCP[1], Charles D. Murray FRCP[2], Christopher P. Denton FRCP[1]

Affiliations

[1]Centre for Rheumatology, Royal Free Campus, Division of Medicine, University College London, U.K.

[2]Department of Gastroenterology, Royal Free London, U.K.

Correspondence to:

Professor Christopher P. Denton, Centre for Rheumatology, Division of Medicine, University College London, Royal Free Campus, Hampstead, London NW3 2PF

E-mail address: c.denton@ucl.ac.uk

Phone/Fax number: 44 20 7794 0432

Funding: This work was supported by The Royal Free Charity from a generous donation supporting the Dita Athanassoulia clinical fellowship in GI scleroderma.

Conflicts of interest: CPD reports personal fees or research grants to his institution from GlaxoSmithKline, Galapagos, Boehringer Ingelheim, Roche, CSL Behring, Corbus, Horizon, and Arxx Therapeutics; all outside the submitted work.

Other authors have no relevant disclosures.

Running title: PheWAS in systemic sclerosis

Key words: Systemic sclerosis; scleroderma; gastrointestinal diseases; drug toxicity

Abstract

Objective

To explore the causes of and contributors to gastrointestinal (GI) dysfunction in systemic sclerosis in a Phenome-Wide Association Study (PheWAS), using real-world clinical records data.

Methods

12,535 documented clinical assessments of 2,058 consented individuals with systemic sclerosis at the Royal Free Hospital (UK) were available for detailed phenotyping. Diagnoses and drugs were mapped to structured dictionaries of terms (Disease Ontology project and DrugBank Open Data, respectively). A PheWAS model was used to explore links between six important SSc-GI domains (constipation, diarrhoea, dysmotility, incontinence, reflux, and SIBO) and exposure to various co-morbidities and drugs. 'Hits' from the PheWAS model were confirmed and explored in a sub-cohort reporting quantitative GI symptom scores from the UCLA GIT 2.0 questionnaire (GIT 2.0).

Results

1,546 individuals were entered into the PheWAS analysis. 673 distinct diagnoses and 634 distinct drugs were identified in the dataset, as well as SSc-specific phenotypes such as antinuclear antibodies (ANA). PheWAS analysis revealed associations between drugs, diagnoses, and ANAs with six important SSc-GI outcomes: constipation, diarrhoea, dysmotility, incontinence, reflux, and SIBO. Subsequently, using GIT 2.0 symptom scores, links with SSc-GI were confirmed for 22 drugs, four diagnoses, and three ANAs.

Conclusion

Using a hypothesis-free PheWAS approach, we replicated known, and revealed potential novel, risk factors for SSc-GI dysfunction, including drug classes such as opioid, anti-muscarinic, and endothelin receptor antagonist, and ANA subgroup.

Introduction

Systemic sclerosis associated gastrointestinal dysfunction (SSc-GI) is a significant burden to many patients. It is among the most frequent manifestations of this multi-system disease[1]. It is believed that the central SSc disease mechanisms of vasculopathy, fibrosis, and inflammation are the drivers of GI dysfunction. However, at the patient level, there is significant heterogeneity in both the sites of the gut involved and the severity of GI symptoms[2]. Previous studies have identified that differences in SSc-related antinuclear antibodies between individuals may explain the heterogeneity in the manifestation of SSc[3]. For GI involvement, ACA and ARA have been linked to increased disease severity based on patient-reported outcomes measures[4].

The treatment of SSc-GI is focused on symptom control[1,5]. Proton pump inhibitors are used for gastroesophageal reflux, laxatives for constipation, and prokinetics for dysmotility[6,7]. Small intestinal bacterial overgrowth can be managed with antibiotic therapy. However, as a multi-system disease, individuals are appropriately prescribed drugs for other-organ indications, and off-target effects of such drugs may contribute to GI dysfunction. Furthermore, links between the organ-specific manifestations of SSc, and other apparently unrelated disease processes in the individual may reveal shared disease mechanisms. As such, the study of a breadth of drugs and co-morbidities in a population with a single disease may further understanding of the disease process.

The Phenome-Wide Association Study (PheWAS) was conceived to look for associations in the opposite direction to a Genome Wide Association Study, that is from a target genetic variant to multiple phenotypic traits[8]. However, the approach has been broadened from genetic variant to target disease, symptom, and even laboratory result vs. a breadth of phenotypes. PheWAS methodology typically uses electronic health record data to generate phenotypes, for example using coding schema such as ICD-10. In this study, we had access to detailed clinical records from a single centre, and were able to create additional detailed SSc-relevant phenotypes. Additionally, we have included drug exposures in the scope of the phenotypes studied.

In this study we explored links between SSc-GI dysfunction and a large number of drugs and co-morbidities using a hypothesis-free PheWAS approach. This was followed by confirmatory analysis using patient-reported GI symptom scores from the UCLA SCTC GIT 2.0 (GIT 2.0) questionnaire, a validated SSc-specific tool to measure the burden of GI symptoms[9].

Methods

*Royal Free observational systemic sclerosis cohort (Sclero**derMA CohoRT; SMART)***

The first entry in the SMART cohort was 2013-07-03 and the last was 2020-03-19. The archive of documented clinical assessments spanned 2002-2021. The SMART cohort includes individuals with a diagnosis of systemic sclerosis, morphoea, or Raynaud's disease, and healthy control subjects. In this study we only included individuals with a confirmed diagnosis of systemic sclerosis.

The SMART study (UK Scleroderma Cohort (SMART) database and tissue bank - Version 7, 2020-02-17) was approved by University College London Joint Research Office and by London-Fulham Research Ethics Committee (REC reference: 20/LO/0404, IRAS project ID: 279682). Participants gave written informed consent to join the study having read the patient information sheet. The SMART dataset includes demographic information and clinical information (diagnosis, disease onset, internal organ involvement, and test results).

Patient-reported GI symptoms sub-cohort

During the GIT 2.0 sub-study period, 2018-2020, consecutive SSc patients fulfilling the 2013 ACR/EULAR criteria were recruited. Participants completed the UCLA Scleroderma Clinical Trials Consortium Gastrointestinal 2.0 (GIT 2.0) GI symptoms questionnaire, a patient-reported outcome measure (PROM) focused on gastrointestinal symptoms in SSc[9]. In this study, we used the GIT 2.0 symptom scores for reflux (8 items), distension/bloating (4 items), faecal soilage (1 item), diarrhoea (2 items), and constipation (4 items). The total GI score also includes social functioning (6 items) and emotional wellbeing (9 items) domains.

Data extraction and processing

Documented clinical assessments were extracted from the electronic archive using patient identifiers from the SMART cohort. Semi-structured text data was extracted from the documented clinical assessments, including dates, numbered lists of diagnoses, and numbered lists of drugs; negated terms were excluded.

A human disease ontology was accessed from the 'Human Disease Ontology' website, under CC0 license[10], and extracted with R package OntologyIndex. The dictionary was enriched for SSc-specific vocabulary, and regular expressions to match variations in syntax (e.g. anti-ACA vs. ACA).

The DrugBank drug vocabulary was downloaded from DrugBank under CC0 license[11] and synonyms were listed for each drug entity. This drug dictionary was enriched for SSc-relevant drugs including probiotics, trial drugs, nutritional supplements, total parenteral nutrition (TPN), and enteral feeds. Three-letter or shorter names were excluded.

Mapping to drugs and diagnoses

The diagnostic and drug vocabularies were used to map the extracted text to standardised names, dealing with variations in naming and formatting. For each study subject, the exposure to drugs and labelling with diagnoses over time is summarised as 'ever-labelled'.

Quality control

A subset of 90 cases were manually labelled with significant diagnoses (GAVE, renal crisis, cardiac scleroderma, RA, ILD, PH, cancer, IA, and myositis) by experienced clinicians with access to the full clinical notes. Manual and programmatic labels were compared using sensitivity, specificity, and accuracy metrics.

Statistical analysis

PheWAS

Six SSc-GI outcomes were explored, based on the burden of SSc-GI disease assessed with GI symptom scores from PROMs. The selected outcomes were constipation, diarrhoea, dysmotility, incontinence, reflux, and small intestinal bacterial overgrowth (SIBO).

The PheWAS analysis involved multiple univariate logistic regressions (LR). Each of the six key SSc-GI outcomes was taken in turn and regressed against every diagnosis and drug in multiple univariate models.

The p value threshold for PheWAS 'hits' was set using the Benjamini-Hochberg False Discovery Rate method. This was applied for each of the six key SSc-GI outcomes separately. An adjusted p value < 0.05 was the threshold for a PheWAS 'hit'.

GIT 2.0 linear models of symptom scores

Six outcomes were explored from the GIT 2.0 GI symptoms score: total GI symptoms score, reflux, bloating, diarrhoea, constipation, and soilage (incontinence). For each GIT 2.0 outcome, the PheWAS hits were entered as predictors in univariate linear models. Using the GIT 2.0 symptom scores, PheWAS hits were confirmed if both the 95% confidence interval did not include zero, and the effect direction was concurrent with the initial hit.

Analysis was conducted in R version 4.1.3[12].

Results

Characteristics of the systemic sclerosis observational cohort

From the digital patient records of SSc cases within the SMART cohort, 12,535 documented clinical assessments from the clinic visits of 2,058 consented participants at the Royal Free Hospital (UK) were extracted into a dataset. Of these participants, 1,546 had at least three documented clinical assessments on record, and were included in the analysis.

Sub-cohort data for GIT 2.0

In the GIT 2.0 sub-cohort, 370 participants completed the GIT 2.0 symptoms questionnaire, and this data was used to confirm and explore the PheWAS hits. The clinical and demographic characteristics of the study cohorts are summarised in Table 1.

Quality control of labelling

A subset of diagnostic labels was manually curated in 86 cases by clinicians with full access to the notes. This revealed a high accuracy of programmatic labelling (accuracy 0.91-0.98) and high specificity. Although sensitivity was lower, especially for rarer labels (e.g. cardiac involvement), the rate of mislabelling was low.

Prevalence of drugs

We identified 634 distinct drugs in the SMART observational cohort. The most prevalent drugs were omeprazole (866/1546) 56%, losartan (758/1546) 49%, lansoprazole (649/1546) 42%, and mycophenolate mofetil (MMF) (618/1546) 40%.

Prevalence of diagnoses

We identified 673 distinct diagnoses in the SMART observational cohort. The most prevalent diagnoses were limited SSc (1005/1546) 65%, Raynaud's (773/1546) 50%, interstitial lung disease (ILD) (557/1546) 36%, and overlap 27% (417/1546).

PheWAS analysis

PheWAS analysis demonstrated 88 hits across the six SSc-GI domains (56 drugs, 26 diagnoses, 6 ANAs), involving 37 distinct drugs, 18 distinct diagnoses, and 3 distinct ANAs (Fig. 1).

Constipation

For constipation, 12 drugs, 4 diagnoses and 1 ANA were significantly associated. The largest effect sizes were prucalopride (OR=50.5, p=1.3e-19), bisacodyl (OR=36.7, p=1.15e-7), docusate (OR=23.1, p=1.29e-4), migraine (OR=13.8, p=1.29e-4), and aspiration (OR=14.8, p=0.00527). Of interest, amitriptyline (OR=3.31, p=0.0127), nicorandil (OR 12.8, p=0.0489), fentanyl (OR=11.5, p=0.00192), ANA negative (OR=4.84, p=0.0461), and fexofenadine (OR=5.09, p=0.0136) were also linked to constipation.

Diarrhoea

For diarrhoea, 4 drugs, 4 diagnoses and 1 ANA were significantly associated. The largest effect sizes were cimetidine (OR=11.6, p=0.0166), rifaximin (OR=10.1, p=0.0053), ambrisentan (OR=9.51), loperamide (OR=8.45, p=7.74e-9), and constipation (OR=7.74, p=5.57e-4). Of interest, pulmonary hypertension (OR=3.81, p=0.00133), and ACA (OR=3.04, p=0.0118) were also linked to diarrhoea.

Dysmotility

For dysmotility, 14 drugs, 4 diagnoses, and 1 ANA were significantly associated. The largest effect sizes were oxycodone (OR=12.2, p=4.4e-4), barium (OR=9.43, p=0.0495), docusate (OR=9.07, p=0.0251), bisacodyl (OR=7.04, p=0.044), and tamsulosin (OR=6.92, p=0.00594). Of interest, iloprost (OR=2.55, p=0.0026), Raynaud's (OR=1.97, p=0.0495), ANA homogenous (OR=6.33, p=0.0495), SIBO (OR=2.99, p=0.0495), zoledronic acid (OR=3.27, p=0.0495), and thiamine (OR=6.33, p=0.0495) were also linked to dysmotility.

Incontinence

For incontinence, 4 drugs and 5 diagnoses were significantly associated. The largest effect sizes were rectal prolapse (OR=19.8, p=1.59e-9), ambrisentan (OR=8.57, p=0.00113), loperamide (OR=7.87, p=8.47e-11), gamolenic acid (evening primrose oil) (OR=5.47, p=0.0337), and constipation (OR=5.33,

p=0.00976). Of interest, ACA (OR=3.91, p=2.1e-5), limited SSc (OR=3.85, p=0.018), mycophenolate (OR=0.273, p=0.0125) were also linked to incontinence – noting that mycophenolate had a decreased odds of incontinence.

Reflux

For reflux, 9 drugs, 7 diagnoses, and 1 ANA were significantly associated. The largest effect sizes were famotidine (OR=13.1, p=0.00102), ANA homogenous (OR=7.71, p=0.00677), aspiration (OR=6.88, p=0.0349), diarrhoea (OR=4.62, p=4.44e-5), and dysmotility (OR=4.27, p=1.46e-8). Of interest, ILD (OR=1.69, p=0.00892), Raynaud's (OR=1.61, p=0.0239), digital ulcers (OR= 1.84, p=0.00936), doxycycline (OR= 2.86, p=0.0153), domperidone (OR= 2.39, p=1.68e-6), and metoclopramide (OR= 2.69, p=3.45e-4) were also linked to reflux.

SIBO

For SIBO, 13 drugs, 2 diagnoses, and 1 ANA were significantly associated. The largest effect sizes were rifaximin (OR=26.3, p=3.41e-9), ciprofloxacin (OR=17.5, p=5.24e-22), metronidazole (OR=17.4, p=8.12e-11), total parenteral nutrition (OR=10.7, p=0.0062), and docusate (OR=8.91, p=0.0358). Of interest, ACA (OR= 2.46, p=0.0154), anaemia (OR= 2.93, p=0.0196), and osteoporosis (OR= 2.9, p=0.0152) were also linked to SIBO.

PheWAS hits were confirmed and explored using GI symptom scores (GIT 2.0)

The 29 distinct PheWAS hits (drugs, diagnoses, and ANAs) were confirmed and explored formally using the patient-reported GI symptom scores (GIT 2.0) across GI domains: reflux, bloating, diarrhoea, constipation, soilage (incontinence), and the total score. This allowed a more detailed analysis of associations as summarised in Fig. 2, including both the direction and the magnitude of the effect on patient-reported SSc-GI symptoms (Tables 2 and 3).

Confirmed associations with drugs

22 drugs were significantly associated with GI symptom scores. Many drugs were associated with multiple symptom domains (including the total score): ranitidine (six), domperidone (five),

omeprazole (five), amitriptyline (four), and loperamide (four), for example. The largest increase in symptom scores were seen with TPN (total, bloating, diarrhoea, and reflux), docusate (total), prucalopride (bloating), and oxycodone (reflux, total, bloating, and constipation). Drugs not indicated for GI symptoms included: amitriptyline (bloating, reflux, total score, and constipation), fentanyl (total score), fexofenadine (constipation), iloprost (reflux and total score), and oxycodone (reflux, total score, bloating, and constipation). For PheWAS domains with a matching GIT 2.0 domain, associations were confirmed with greater certainty: fexofenadine, prucalopride, lactulose, amitriptyline, domperidone, and ranitidine (constipation); loperamide (diarrhoea); loperamide (incontinence); ondansetron, ranitidine, metoclopramide, pantoprazole, domperidone, and omeprazole (reflux).

Confirmed associations with diagnoses

Four diagnoses were significantly associated with GI symptom scores. Aspiration was associated with increased bloating, total score, and incontinence. Rectal prolapse was associated with increased incontinence, total score, bloating, reflux, and constipation. Dysmotility was associated with increased reflux and total score. Raynaud's was associated with increased constipation. For PheWAS domains with a matching GIT 2.0 domain, associations were confirmed with greater certainty: rectal prolapse (incontinence); dysmotility (reflux).

Confirmed associations with ANAs

Three ANAs were significantly associated with GI symptom scores. ANA homogenous pattern and ANA negative were associated with increased constipation. ACA was associated with incontinence. For PheWAS domains with a matching GIT 2.0 domain, associations were confirmed with greater certainty: ANA negative (constipation), and ACA (incontinence).

Discussion

In this study, we explored GI involvement in systemic sclerosis, mapping drug exposures, co-morbidities, and antinuclear antibodies. Using real-world data, we revealed both the expected and appropriate prescribing of drugs to treat GI symptoms, but also potential causes and contributors to GI dysfunction in systemic sclerosis. We confirmed and explored initial associations ('hits') with detailed patient-reported GI symptom scores, confirming multiple hits and exploring the effects on GI symptoms.

The PheWAS analysis included 1,546 individuals; the majority were limited SSc subtype, and ANA subgroups reflected a typical SSc population. The median follow up was long (16.5 years) , supporting the ascertainment of SSc-GI outcomes which can develop late in the disease. The confirmatory GIT 2.0 cohort (n=370) was slightly enriched for diffuse SSc.

Almost 700 distinct diagnoses were identified in the clinical records. This permitted a hypothesis-generating approach, with the diagnostic space not limited to prior associations with SSc. However, as expected, the most frequent diagnoses and co-morbidities were those directly related to SSc, including limited SSc, Raynaud's, and ILD. However, diagnoses not specific to SSc such as migraine, osteoporosis, anaemia, and breast cancer were associated with SSc-GI in PheWAS analysis, underlining the scope of this approach.

Over 600 distinct drugs were identified in the clinical records. As expected, the top drugs identified were those indicated for the management of SSc manifestations, including PPIs, angiotensin receptor blockers, and mycophenolate. However, most medicines were prescribed for a non-SSc indication, and as such we were able to assess the potential contribution of such drugs to SSc-GI dysfunction.

Eighty-eight hits were generated by the PheWAS model; by domain: constipation 17, diarrhoea 9, dysmotility 19, incontinence 10, reflux 17, and SIBO 16. Interestingly, despite a similar total number of distinct drugs and diagnoses identified, most of the hits were for drugs (56), followed by

diagnoses (26), followed by ANAs (6). Three of the six ANA hits were for ACA, associated with diarrhoea, incontinence, and SIBO. This fits with prior work demonstrating a higher burden of SSc-GI disease with ACA[4].

Altogether, 29 of the 88 hits were confirmed using the GIT 2.0 patient-reported symptom scores; certain hits were significantly associated with multiple domains of the GIT 2.0. For PheWAS hits with a matching domain of the GIT 2.0 symptom score (constipation, diarrhoea, incontinence, and reflux) we were able to confirm these associations with a greater degree of certainty.

Bringing together related PheWAS hits, Raynaud's, digital ulcers, iloprost, ACA, and migraine were all linked to increased GI disease. This may suggest that, between the canonical disease mechanisms in SSc, vasculopathy is a prominent driver of GI disease. Among these hits linked to vasculopathy, Raynaud's, iloprost, and ACA were confirmed to increased GI symptoms on the GIT 2.0.

Hits were discovered linking reflux with aspiration and ILD, which may reflect a causal pathway leading from reflux, to aspiration, to ILD. Although the most significant drivers of ILD might be inflammation and fibrosis, reflux is known to contribute to the progression of ILD[13]. As such, this association serves to emphasise the importance of controlling reflux, especially in those at risk of ILD progression.

ANA negative was confirmed to be associated with constipation in the GIT 2.0. Prior work has identified a link between ANA negative subtype and increased lower GI disease, specifically malabsorption[14].

Downstream of SSc-GI disease, several likely consequences of GI dysfunction were identified. For example, SIBO was linked to osteoporosis and anaemia in the PheWAS, highlighting the consequences of GI disease. Malnutrition is SSc is multifactorial[15], including decreased appetite, upper GI dysmotility, but SIBO is certainly a driver of malabsorption. As such, treatments for SIBO not only improve the symptom burden (bloating, flatulence, discomfort) but may also improve nutritional status.

SSc-GI and drugs

The associations between drugs and SSc-GI should be considered in two groups. Firstly, drugs prescribed for symptom-control (appropriate prescribing)[16] and, secondly, drugs prescribed for another indication (e.g. opioids for pain) with an off-target effect of GI disease.

The PheWAS hit linking amitriptyline and constipation was confirmed with the GIT 2.0 constipation symptom score. Amitriptyline has significant anticholinergic activity, and side effects include dry mouth, constipation, and nausea; paralytic ileus is a rare side effect. Autoantibodies to the muscarinic-3 receptor autoantibodies have been identified in those with severe SSc-GI involvement[17]. Additionally, there is observational evidence that cholinesterase inhibitor pyridostigmine is effective for SSc-GI symptoms, in particular constipation[6]. Taken together, this may prompt clinicians to review with a high anticholinergic burden where possible.

Oxycodone, initially a hit for dysmotility, was confirmed to be associated with reflux, bloating, constipation, and the total GI symptom scores. Opioids have well-recognised GI side effects and previous work has identified opioids as a risk factor for intestinal pseudo-obstruction in systemic sclerosis[18]. Additionally, fentanyl was confirmed to increase the total GI symptom score, after a PheWAS hit for constipation, supporting the hypothesis.

Fexofenadine is a highly specific histamine H1 receptor reverse agonist, binding to and stabilising the inactive form of the receptor. It is reported to have low off-target effects. The only gastrointestinal side effect reported for the product is nausea which is common. In this study, fexofenadine was a hit for constipation in the PheWAS which was confirmed with increased constipation symptom score in the GIT 2.0 model. As such, there is a higher degree of confidence that this is a real association. However, it is possible that causally the indication for fexofenadine (allergy/atopic conditions) is a confounder of the potential causal effect, and further investigation is warranted.

Ambrisentan was a hit for diarrhoea; this was a recognised side effect in clinical trials[19]. Use of ambrisentan may also be a surrogate marker for severe vasculopathy, which itself might be linked to more severe GI disease.

Tamsulosin was a hit for dysmotility. As an alpha-1 receptor antagonist, it leads to the relaxation of smooth muscle. Constipation and diarrhoea are reported to be uncommon side effects, however the plausibility of the mechanism supports this hypothesis.

Iloprost was a hit for dysmotility, and increased reflux and the total GI scores in the GIT 2.0 analysis. This relationship could be confounded by vasculopathy, making a direct effect of iloprost on GI dysfunction less likely.

The hit for mycophenolate suggested a protective effect for incontinence (OR=0.27, p=0.0125). Whilst this could be a treatment effect, targeting inflammatory-driven GI dysfunction, it is possible that – as mycophenolate prescription is linked with diffuse SSc, and ACA/limited disease is linked to GI disease, especially incontinence – the relationship is confounded by skin subset, which was not adjusted for in the PheWAS analysis. An observational study of a US healthcare insurance claims database suggested that the real-world prescribing of immunomodulatory therapy for SSc was not enriched for certain organ involvement, including GI disease[20]. However, in that study, the prevalence of GI involvement at one year was 22% which is low compared to cohort studies[4].

In this study, we mapped the links between co-morbidities and SSc-GI disease. As well as demonstrating expected, mechanistic, associations between diagnoses (e.g. reflux and aspiration), our approach yielded links that may suggest novel disease mechanisms (e.g. migraine being linked to constipation in SSc). Regarding drugs, we highlight the adverse effects of drugs. In addition, mapping drug effects in the context of SSc may shed light on SSc disease mechanisms. Real-world data including drugs prescribed for an orthogonal indication to SSc-GI disease may demonstrate the disease-specific effects of drug target perturbation, which could be explored further with the aim of drug repurposing.

There are several strengths to the present study. These include a long duration of follow up of the cohort, and the granularity of detail for individual clinical records available for phenotyping. The use of a patient-reported outcome measure, the GIT 2.0, to confirm initial PheWAS hits strengthens our conclusions. Many of the PheWAS hits and those confirmed with the GIT 2.0 replicate well-known associations between drugs and diagnoses in SSc, particularly the most common treatments for SSc-GI manifestations: PPIs, H2 receptor antagonists, antibiotics for SIBO, laxatives, prucalopride, anti-emetics, and pro-motility agents. This supports the validity of the novel associations uncovered.

There are also some clear limitations. This study used real-world data, and as such there are possible biases. The documented clinical assessments are written in a semi-structured format, and we would expect a degree of variation between clinicians and over time; we included individuals with at least three separate documented clinical assessments across time to mitigate this variation. The GIT 2.0 sub-cohort was cross-sectional, at a time point towards the end of the SMART observational cohort window. Whilst this was adequate for the purpose of confirming PheWAS hits, longitudinal GIT 2.0 data would allow the examination of the relationships between risk factor and SSc-GI over time.

Conclusions and future work

We have used a novel analytical approach in a large single-centre observational cohort to explore association of drug treatments and disease characteristics with significant GI manifestations in SSc and the associated symptom burden. Our findings have face validity and reflect previous studies but also highlight the relevance of treatments for non-GI complications. These findings highlight the importance of careful and integrated multi-disciplinary care for SSc that includes specialist pharmacist input and routine assessment of GI symptom severity using validated patient reported outcome tools. Future work using real-world data covering drugs, diagnoses, and disease-related outcomes might look at polypharmacy in complex autoimmune diseases, especially the co-occurrence of drug-disease pairs and appropriate, insufficient, and problematic polypharmacy.

References

1. Volkmann ER, McMahan Z. Gastrointestinal involvement in systemic sclerosis: pathogenesis, assessment and treatment. *Curr Opin Rheumatol.* 2022 Aug 22. doi: 10.1097/BOR.0000000000000899.

3. Nihtyanova SI, Sari A, Harvey JC, Leslie A, Derrett-Smith EC, Fonseca C, Ong VH, Denton CP. Using Autoantibodies and Cutaneous Subset to Develop Outcome-Based Disease Classification in Systemic Sclerosis. *Arthritis Rheumatol. Hoboken NJ* **72**, 465–476 (2020).

4. Ahmed F, Maclean RH, Nihtyanova SI, Ong VH, Murray CD, Denton CP. Autoantibody predictors of gastrointestinal symptoms in systemic sclerosis. *Rheumatol. Oxf. Engl.* **61**, 781–786 (2022).

5. Cheah, J. X., Khanna, D., McMahan, Z. H. Management of scleroderma gastrointestinal disease: Lights and shadows. *J. Scleroderma Relat. Disord.* **7**, 85–97 (2022).

6. Ahuja, N. K., Mische, L., Clarke, J. O., Wigley, F. M. & McMahan, Z. H. Pyridostigmine for the treatment of gastrointestinal symptoms in systemic sclerosis. *Semin. Arthritis Rheum.* **48**, 111–116 (2018).

7. Dein, E. J., Wigley, F. M. & McMahan, Z. H. Linaclotide for the treatment of refractory lower bowel manifestations of systemic sclerosis. *BMC Gastroenterol.* **21**, 174 (2021).

8. Bastarache, L., Denny, J. C. & Roden, D. M. Phenome-Wide Association Studies. *JAMA* **327**, 75–76 (2022).

9. Khanna D, Hays RD, Maranian P, et al. Reliability and validity of the University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract Instrument. *Arthritis Rheum.* **61**, 1257–1263 (2009).

10. Schriml LM, Mitraka E, Munro J, et al. Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res.* **47**, D955–D962 (2019).

11. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668-672 (2006).

12. R Core Team. R: A language and environment for statistical computing. (2022).

13. Christmann, R. B., Wells, A. U., Capelozzi, V. L. & Silver, R. M. Gastroesophageal reflux incites interstitial lung disease in systemic sclerosis: clinical, radiologic, histopathologic, and treatment evidence. *Semin. Arthritis Rheum.* **40**, 241–249 (2010).

14. Salazar GA, Assassi S, Wigley F, et al. Antinuclear Antibody Negative Systemic Sclerosis. *Semin. Arthritis Rheum.* **44**, 680–686 (2015).

15. Harrison, E., Herrick, A. L., McLaughlin, J. T. & Lal, S. Malnutrition in systemic sclerosis. *Rheumatol. Oxf. Engl.* **51**, 1747–1756 (2012).

16. Hansi N, Thoua N, Carulli M, et al. Consensus best practice pathway of the UK scleroderma study group: gastrointestinal manifestations of systemic sclerosis. *Clin. Exp. Rheumatol.* **32**, S-214-221 (2014).

17. Kawaguchi Y, Nakamura Y, Matsumoto I, et al. Muscarinic-3 acetylcholine receptor autoantibody in patients with systemic sclerosis: contribution to severe gastrointestinal tract dysmotility. *Ann. Rheum. Dis.* **68**, 710–714 (2009).

18. Dein E, Kuo PL, Hong YS, Hummers LK, Mecoli CA, McMahan ZH. Evaluation of risk factors for pseudo-obstruction in systemic sclerosis. *Semin. Arthritis Rheum.* **49**, 405–410 (2019).

19. Galiè N, Barberà JA, Frost AE, et al. Initial Use of Ambrisentan plus Tadalafil in Pulmonary Arterial Hypertension. *N. Engl. J. Med.* **373**, 834–844 (2015).

20. Gale SL, Trinh H, Mathew N, Jahreis A, Lin CJF, Sarsour K. Characterizing Disease Manifestations and Treatment Patterns Among Adults with Systemic Sclerosis: A Retrospective Analysis of a US Healthcare Claims Population. *Rheumatol. Ther.* **7**, 89–99 (2020).

Table and figure legends

Table 1. Description of SMART observational cohort and GIT 2.0 confirmatory sub-cohort

Table 2. Confirmed hits with matching GIT outcomes (PheWAS outcome: GIT outcome)

Table 3. Confirmed hits with the GIT total score

Figure 1. PheWAS plot showing key clinical and drug associations with gastrointestinal complications of systemic sclerosis

PheWAS analysis generated 'hits' with drugs, diagnoses, and ANAs across six SSc-GI outcomes. Key: ANAs red; diseases green; drugs blue; red line adjusted p values < 0.05; certain associations are labelled, odds ratio in brackets. Constipation was linked to fentanyl, amitriptyline, migraine, ANA negative, aspiration, and incontinence. Diarrhoea was linked to ambrisentan and ACA. Dysmotility was linked to ANA homogenous, tamsulosin, fentanyl, and oxycodone. Incontinence was linked to ACA, ambrisentan, and inversely linked to mycophenolate. Reflux was linked to ILD and ANA homogenous. SIBO was linked to ACA, total parenteral nutrition, osteoporosis, and anaemia.

Figure 2. PheWAS hits confirmed and explored with GIT 2.0 symptom scores

PheWAS hits were confirmed and explored with GIT 2.0 symptom scores: total score, reflux, bloating, diarrhoea, constipation, and incontinence. Estimates with 95% CIs are presented; red $p < 0.05$. Square – drugs; triangle – diagnoses; circle – ANAs.

Tables

Table 1. Description of SMART observational cohort and GIT 2.0 confirmatory sub-cohort

SMART observational cohort	**n=1546**	**Missing**	
Female	1296	0	84%
Limited SSc	1030	0	67%
Diffuse SSc	516	0	33%
Disease onset (year)	1489	57	Range: 1959-2019
Disease duration to end of study window (years)	1489	57	Median (range): 16.5 (1.9-62)
ACA	394	0	25.5%
ATA	376	0	24.3%
ARA	169	0	10.9%
PM-Scl	71	0	4.6%
U1RNP	89	0	5.8%
U3RNP	61	0	3.9%
ThT0	8	0	0.5%
Confirmatory cohort (GIT 2.0)	**n=370**	**Missing**	
Female	308	0	83%
Limited SSc	228	0	62%
Diffuse SSc	140	0	38%
Juvenile SSc	2	0	0.5%
Disease onset (year)	360	10	Range: 1962-2018
Disease duration (years)	360	10	Median (range): 13.7 (1-57)
ANA	345	4	94%
ENA	303	3	83%
ACA	115	0	31%
ATA	83	0	22%
ARA	43	0	12%
PM-Scl	18	0	4.9%
U1RNP	22	0	5.9%
U3RNP	14	0	3.8%

Table 2. Confirmed associations with matching PheWAS outcome and GIT domain

			PheWAS Model		GIT 2.0 Linear Model			
Outcome	**Type**	**Predictor**	**Odds Ratio**	**Adjusted p value**	**Estimate**	**Lower CI**	**Upper CI**	**P value**
Constipation	ANA	ANA negative	4.84	0.0461	0.397	0.0228	0.77	0.0393
Constipation	Drug	Amitriptyline	3.31	0.0127	0.279	0.113	0.444	0.00111
Constipation	Drug	Domperidone	3.28	0.00777	0.285	0.14	0.43	1.53e-04
Constipation	Drug	Fexofenadine	5.09	0.0136	0.354	0.0567	0.652	0.0208
Constipation	Drug	Lactulose	5.64	0.00192	1.03	0.556	1.5	2.69e-05
Constipation	Drug	Prucalopride	50.5	1.60e-19	0.885	0.52	1.25	3.26e-06
Constipation	Drug	Ranitidine	3.25	0.00799	0.248	0.12	0.376	1.83e-04
Diarrhoea	Drug	Loperamide	8.45	7.75e-09	0.381	0.156	0.606	0.00108
Reflux	Diagnosis	Dysmotility	4.27	1.46e-08	0.429	0.158	0.699	0.00215
Reflux	Drug	Domperidone	2.39	1.68e-06	0.462	0.284	0.641	6.86e-07
Reflux	Drug	Metoclopramide	2.69	3.45e-04	0.894	0.616	1.17	9.42e-10
Reflux	Drug	Omeprazole	2.22	2.80e-05	0.217	0.0658	0.369	0.00543
Reflux	Drug	Ondansetron	4.11	0.00302	1.09	0.503	1.68	3.47e-04
Reflux	Drug	Pantoprazole	2.47	0.0162	0.477	0.145	0.809	0.00538
Reflux	Drug	Ranitidine	3.82	1.85e-17	0.666	0.518	0.814	5.87e-17
Incontinence	ANA	ACA	3.91	2.10e-05	0.296	0.0928	0.5	0.00479
Incontinence	Diagnosis	Rectal prolapse	19.8	1.59e-09	1.61	1.01	2.21	2.70e-07
Incontinence	Drug	Loperamide	7.87	8.47e-11	0.96	0.644	1.28	7.14e-09

Table 3. Confirmed associations with GIT total score

Type	Predictor	PheWAS Model			GIT 2.0 Linear Model				
		Outcome	Odds Radio	Adjusted p value	Outcome	Estimate	Lower	Upper	P value
Disease	Aspiration	Constipation	14.8	0.00527	Total score	1.26	0.545	1.97	6.45E-04
Drug	Amitriptyline	Constipation	3.31	0.0127	Total score	0.293	0.117	0.469	0.00128
Drug	Docusate	Constipation	23.1	1.29e-04	Total score	1.52	0.277	2.76	0.0176
Drug	Domperidone	Constipation	3.28	0.00777	Total score	0.444	0.293	0.595	2.17E-08
Drug	Fentanyl	Constipation	11.5	0.00192	Total score	1.52	0.277	2.76	0.0176
Drug	Prucalopride	Constipation	50.5	1.60e-19	Total score	0.738	0.344	1.13	2.93E-04
Drug	Ranitidine	Constipation	3.25	0.00799	Total score	0.536	0.408	0.664	4.87E-15
Drug	Loperamide	Diarrhoea	8.45	7.75e-09	Total score	0.478	0.245	0.71	7.71E-05
Drug	Rifaximin	Diarrhoea	10.1	0.0053	Total score	0.79	0.167	1.41	0.0139
Disease	Aspiration	Reflux	6.88	0.0349	Total score	1.26	0.545	1.97	6.45E-04
Disease	Dysmotility	Reflux	4.27	1.46e-08	Total score	0.349	0.117	0.581	0.00352
Drug	Domperidone	Reflux	2.39	1.68e-06	Total score	0.444	0.293	0.595	2.17E-08
Drug	Metoclopramide	Reflux	2.69	3.45e-04	Total score	0.641	0.4	0.882	3.49E-07
Drug	Omeprazole	Reflux	2.22	2.80e-05	Total score	0.232	0.103	0.361	4.98E-04
Drug	Ondansetron	Reflux	4.11	0.00302	Total score	0.76	0.252	1.27	0.00377
Drug	Pantoprazole	Reflux	2.47	0.0162	Total score	0.288	0.00263	0.574	0.0498
Drug	Ranitidine	Reflux	3.82	1.85e-17	Total score	0.536	0.408	0.664	4.87E-15
Disease	Rectal prolapse	Soilage	19.8	1.59e-09	Total score	0.943	0.507	1.38	3.12E-05
Drug	Loperamide	Soilage	7.87	8.47e-11	Total score	0.478	0.245	0.71	7.71E-05

Figure 1. PheWAS plot

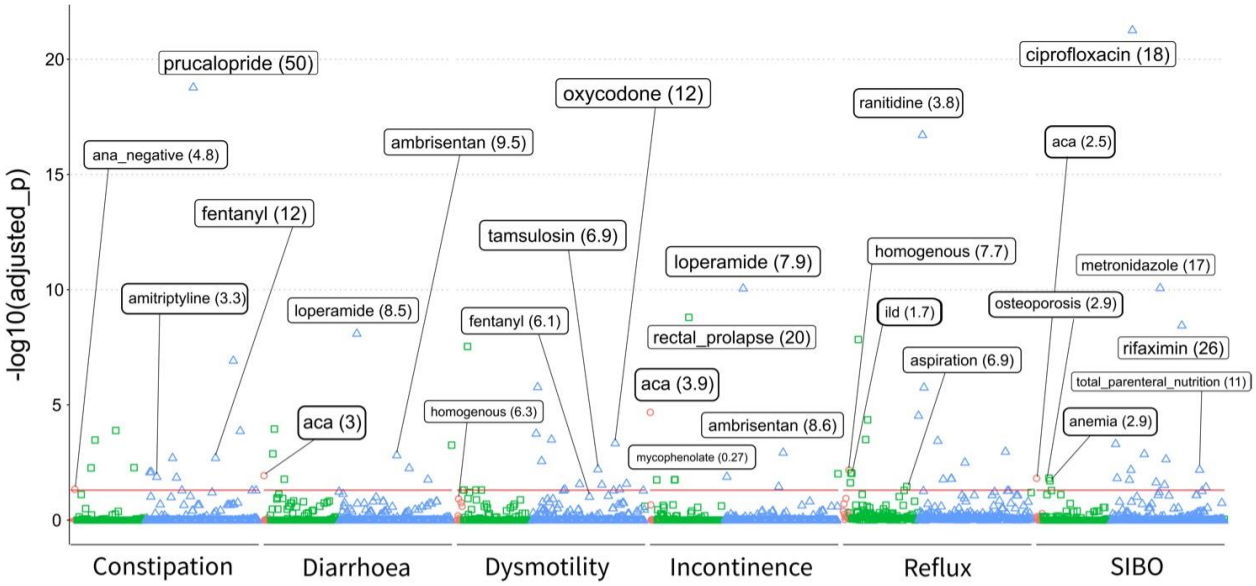

Figure 2. PheWAS hits confirmed and explored with GIT 2.0 symptom scores