Environmental Triggers of Inflammatory Skin Disease

Anastasiya Muntyanu, Experimental Medicine, McGill University, Montreal

Aug 15, 2024

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Experimental Medicine

Table of Contents

Abstract

Over the last few decades, the healthcare system observed a steady increase in the incidence of autoimmune (*e.g.* systemic sclerosis (SSc)) skin diseases contributing to significant morbidity and mortality. Limited data is available about the role of the environment on immune system dysregulation and subsequent precipitation of these conditions. Given the skin is the largest organ in the human body, the large surface area in contact with the outside world, may have an impact on development of skin disease. The goal of this thesis is to comprehensively evaluate epidemiology and geographic distribution of SSc and psoriasis and subsequently explore the possible contributing risk factors to development of these conditions.

To understand the updated epidemiology of SSc, a populational database was used in Quebec, Canada between 1996-2019. There were 8,180 incident cases of SSc identified with a female to male ratio of 4:1 and average age at diagnosis of 57.3 ± 16.3 years. The age-standardized incidence rate (ASIR) was 4.14/100,000 person-years (95% CI 4.05–4.24) with an average annual percent change (AAPC) of ~4%. Analysis by age group and sex demonstrated the highest incidence among 60–79-year-olds for females and >80-year-olds for males. The largest AAPC of ~10% was seen in children. Standardized incidence ratios (SIR) based on Forward Sortation Area (FSA) varied between 0.52-1.64. The average prevalence over the study period was 28.96/100,000 persons (95% CI 28.72–29.20).

A geographic distribution study of 1,505 SSc cases from the Canadian Scleroderma Research Group (CSRG) registry between 2004-2019 also revealed significant regional variations in period prevalence. High rates were observed in clusters in urban centers including Hamilton, Ontario, and Montreal, Quebec which are characterized by substantial industrial and mining activities. Of the 19 FSAs with higher prevalence of 2-5.5-fold higher than expected, six were

found in rural regions, far from recruitment centers. Comparison to the 19 lowest-prevalence FSAs, the 19 highest-prevalence FSAs showed significantly higher industrial density ($p=0.0059$).

Using this registry, an analysis assessing the disease manifestations of SSc in patients with identified occupational exposures compared to non-exposed individuals was performed. Among 1,439 patients, 20.2% reported exposure to organic solvents and 6.6% reported exposure to silica. Organic solvent exposed patients exhibited a higher risk or renal crisis and severe gastrointestinal disease and a trend towards increasing mortality despite adjusting for multiple confounders. Multivariate regression, controlled for multiple confounders, showed that silica exposure was associated with a younger age at diagnosis and worse disease severity and mortality.

To demonstrate effectiveness of this approach in other inflammatory cutaneous diseases, a study using populational psoriasis data containing 43,663 patients was performed in Quebec between 1997-2015. Incidence rate varied across FSAs between 1.6-325.6/100,000 person-years. A machine learning model was used to evaluate environmental/neighbourhood factors and their contribution to high psoriasis incidence. The parsimonious model of the top 9 predictors for high psoriasis incidence had an area under the curve (AUC) of 0.77. Positive association with psoriasis incidence was noted for nighttime light brightness. Negative association was observed for ultraviolet radiation (UVR), maximum daily temperature, proportion of females, soil moisture, urbanization, and distance to expressways. Assessment of socioeconomic factors demonstrated that the middle-class neighbourhoods were more likely to be affected.

In conclusion, uneven geographic distribution was observed for SSc and psoriasis and possible contributing factors were identified in the environment/occupation. These findings can translate into improved distribution of medical resources to high-risk areas, enhanced patient education, and advocacy for preventative strategies by addressing risk factors.

Résumé

Au cours des dernières décennies, le système de santé a observé une augmentation constante dans l'incidence des maladies auto-immunes cutanées (par exemple, la sclérose systémique (SSc)), contribuant à une morbidité et une mortalité significatives. Les données disponibles sur le rôle de l'environnement dans la dysrégulation du système immunitaire et la survenue subséquente de ces conditions, sont limitées. Étant donné que la peau est le plus grand organe du corps humain, la vaste surface en contact avec l'environnement pourrait avoir un impact sur le développement des maladies cutanées. Le but de cette thèse est d'évaluer de manière exhaustive l'épidémiologie et la répartition géographique de la SSc et du psoriasis, puis d'explorer les déclencheurs possibles contribuant au développement de ces conditions.

Pour comprendre la tendance épidémiologique de la SSc, une base de données populationnelle a été utilisée au Québec, Canada, entre 1996 et 2019. Il y a eu 8,180 cas incidents de SSc identifiés, avec un ratio femme/homme de 4:1 et un âge moyen au diagnostic de 57.3 ± 16.3 ans. Le taux d'incidence standardisé selon l'âge (TISA) était de 4.14/100 000 personnes-années (IC à 95 % 4.05–4.24) avec une variation annuelle moyenne en pourcentage (VAMP) de ~4 %. L'analyse par groupe d'âge et sexe a montré la plus haute incidence chez les femmes de 60-79 ans et chez les hommes de plus de 80 ans. Le plus grand VAMP de ~10 % a été observé chez les enfants. Les ratios d'incidence standardisés basés sur les régions de tri d'acheminement (RTA) variaient entre 0.52 et 1.64. La prévalence moyenne sur la période d'étude était de 28.96/100 000 personnes (IC à 95 % 28.72–29.20).

Une étude de répartition géographique de 1,505 cas de SSc provenant du registre du Groupe de recherche canadien sur la sclérodermie entre 2004 et 2019 a également révélé des variations régionales significatives dans la prévalence périodique. Des taux élevés ont été observés dans des

regroupements situés dans des centres urbains tels que Hamilton, Ontario, et Montréal, Québec, caractérisés par des activités industrielles et minières considérables. Sur les 19 RTA avec des prévalences élevées de 2 à 5.5 fois supérieure à celle attendue, 6 se trouvaient dans des régions rurales, loin des centres de recrutement. Par rapport aux 19 RTA à basse prévalence, les 19 RTA à prévalence élevée montraient une densité industrielle significativement plus élevée (p=0.0059).

À l'aide de ce registre, une analyse a été réalisé en évaluant les manifestations de la SSc chez les patients avec ou sans des expositions professionnelles connues. Parmi les 1,439 patients inclus, 20.2 % ont signalé une exposition aux solvants organiques et 6.6 % ont signalé une exposition à la silice. Les patients exposés aux solvants organiques présentaient un risque plus élevé de crise rénale, de maladie gastro-intestinale plus sévère et une tendance vers une mortalité accrue malgré l'ajustement pour de nombreux facteurs confondants. La régression multivariée, contrôlée pour de multiples facteurs confondants, a montré que l'exposition à la silice était associée à un âge plus jeune au diagnostic, et à une sévérité et une mortalité plus importante face à la maladie.

Pour démontrer l'efficacité de cette approche dans d'autres maladies cutanées inflammatoires, une étude utilisant des données populationnelles sur le psoriasis, comprenant 43 663 patients a été réalisée au Québec entre 1997 et 2015. Le taux d'incidence variait selon les RTA entre 1.6 et 325.6/100 000 personnes-années. Un modèle d'apprentissage automatique a été utilisé pour évaluer les facteurs environnementaux/communautaire et leur contribution à une incidence élevée de psoriasis. Le modèle parcimonieux des 9 principaux prédicteurs d'une incidence élevée de psoriasis avait une aire sous la courbe (AUC) de 0.77. Une association positive avec l'incidence du psoriasis a été notée pour la luminosité nocturne. Une association négative a été observée pour les rayons ultraviolets, la température quotidienne maximale, la proportion de femmes, l'humidité

du sol, l'urbanisation et la distance aux autoroutes. L'évaluation des facteurs socio-économiques a démontré que les quartiers de classe moyenne étaient plus susceptibles d'être affectés.

En conclusion, une répartition géographique inégale a été observée pour la SSc et le psoriasis, et des facteurs contributifs possibles ont été identifiés dans l'environnement/occupation. Ces trouvailles peuvent se traduire par une meilleure répartition des ressources médicales dans les zones à haut risque, une éducation améliorée des patients et un plaidoyer pour des stratégies préventives en abordant les facteurs de risque.

Acknowledgements

This thesis would not have been possible without the support of those around me.

Firstly, I would like to extend my heartfelt thanks to my family, especially my parents, who have been a constant source of support and inspiration. Their own achievements in completing PhDs in Physics have instilled in me a deep sense of curiosity and passion for research. I am incredibly fortunate to have such remarkable role models.

I would like to express my sincere gratitude to my research supervisors, Dr. Elena Netchiporouk and Dr. Ivan Litvinov, for their unwavering encouragement, invaluable advice and continuous mentorship. Their extensive knowledge and experience have continued to inspire me in my academic research and daily life. Their support and guidance have been instrumental throughout my research journey.

I am also deeply grateful to the members of my thesis committee, Dr. Elizabeth O'Brien, Dr. Elham Rahme, and Dr. Stephane Laporte, for their insightful feedback and advice. Your contributions have been influential in shaping this work.

Special thanks to Dr. Raymond Milan, Dr. Mohammed Kaouache, and Dr. Elham Rahme for their exceptional teaching and guidance in statistics, which has been crucial for my research.

I am also grateful to Dr. Murray Baron and the Canadian Scleroderma Research Group for the opportunity to collaborate and complete my projects with them. Your support and collaboration have greatly enriched my research experience.

I am deeply grateful to Canadian Institute of Health Research (CIHR), Scleroderma Foundation, Canadian Dermatology Foundation (CDF) and the Association des médecins spécialistes dermatologues du Québec (AMSDQ) for the funding supporting data access.

Lastly, I would like to thank all those who have supported and encouraged me throughout this journey. Your belief in me has been a driving force behind this achievement.

Contribution to Originality

The work presented in this thesis is original contribution to the literature in providing 1) up to date, comprehensive epidemiology of SSc in Canada, 2) evaluating geographic distribution of SSc in Canada and assessing the contributing environmental and 3) occupational factors. Finally, this is the first study using a populational database to 4) comprehensively study predictive factors for high psoriasis incidence using machine learning algorithms.

Studies outlining epidemiology of SSc have been limited after 2013, around the time of the development of classification criteria and the work in this thesis contributes up to date information using populational data in the province of Quebec, Canada.

Our studies both at the provincial and national level for SSc showed an uneven geographic distribution which has not been previously described in Canada and only small number of studies globally reported these trends. We also showed that high industrial density and air pollution was associated with the high prevalence areas of SSc.

Our studies on occupational exposure to silica and organic solvents contributed to the literature by describing the demographics, disease manifestations, and mortality among exposed patients compared to non-exposed by using a large cohort of patients with comprehensive longitudinal patient specific information.

Finally, using this approach, a similar study was performed using populational data in Quebec for psoriasis. There was an uneven geographic distribution which has not been previously described on a provincial level. A machine learning approach was used to comprehensively study >350 environmental and neighbourhood factors. To our knowledge this is the first study to evaluate impact of greenness index, nighttime light pollution and more in psoriasis. This approach will

serve as a foundation for investigation of risk factors in other chronic diseases in dermatology and beyond.

I collaborated and published the following articles as first author (relating to the topics of this thesis):

1. **Muntyanu A,** Aw K, Kaouache M, Rahme E, Osman M, Baron M, Ghazal S, Netchiporouk E. Epidemiology of systemic sclerosis in Quebec, Canada: a population-based study. *Lancet Reg Health Am. 2024 Jun 8;35:100790. doi: 10.1016/j.lana.2024.100790*
2. **Muntyanu A**, Ouchene L, Zhou S, Hudson M, Rezaeian M, LaChance A, Litvinov IV, Baron M, Netchiporouk E; Canadian Scleroderma Research Group. Geographical distribution of systemic sclerosis in Canada: An ecologic study based on the Canadian Scleroderma Research Group. *J Am Acad Dermatol. 2022 Nov;87(5):1095-1097. doi: 10.1016/j.jaad.2021.12.055. Epub 2022 Jan 11.*
3. **Muntyanu A,** Milan R, Rahme E, LaChance A, Ouchene L, Cormier M, Litvinov IV, Hudson M, Baron M, Netchiporouk E; Canadian Scleroderma Research Group. Exposure to silica and systemic sclerosis: A retrospective cohort study based on the Canadian Scleroderma Research Group. *Front Med (Lausanne). 2022 Sep 29;9:984907. doi: 10.3389/fmed.2022.984907.*
4. **Muntyanu A,** Milan R, Rahme E, Baron M, Netchiporouk E; Canadian Scleroderma Research Group.Organic solvent exposure and systemic sclerosis: A retrospective cohort study based on the Canadian Scleroderma Research Group registry. *J Am Acad Dermatol. 2024 Mar;90(3):605-607. doi: 10.1016/j.jaad.2023.04.062.*

5. **Muntyanu A**, Milan R, Kaouache M, Ringuet J, Gulliver W, Pivneva I, Royer J, Leroux M, Chen K, Yu Q, Litvinov IV, Griffiths CEM, Ashcroft DM, Rahme E, Netchiporouk E. Tree-Based Machine Learning to Identify Predictors of Psoriasis Incidence at the Neighborhood Level: A Populational Study from Quebec, Canada. *Am J Clin Dermatol. 2024 May;25(3):497-508. doi: 10.1007/s40257-024-00854-3.*

Contributions of the Authors

This thesis is structured in a manuscript-based format. There are 9 chapters which comprise a literature review, methods, 5 published manuscripts, discussion and conclusion. The research questions have been developed in conjunction with my supervisors Dr. Netchiporouk and Dr. Litvinov. Below is a breakdown of author contributions per study (as indicated in the published manuscripts).

Muntyanu A, Aw K, Kaouache M, Rahme E, Osman M, Baron M, Ghazal S, Netchiporouk E. Epidemiology of systemic sclerosis in Quebec, Canada: a population-based study. *Lancet Reg Health Am. 2024 Jun 8;35:100790. doi: 10.1016/j.lana.2024.100790*

I contributed to conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing - original draft, writing - review & editing. Katie Aw contributed to data curation, writing - review & editing. Dr. Kaouache contributed to data curation, formal analysis, investigation, methodology, review & editing. Dr. Rahme contributed to conceptualization, methodology, writing - review & editing. Dr. Osman contributed to conceptualization, validation, writing - review & editing. Dr. Baron contributed to conceptualization, funding acquisition, validation, writing - review & editing. Dr. Ghazal contributed to methodology, writing - review & editing. Dr. Netchiporouk contributed to conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing - original draft, writing - review & editing.

Muntyanu A, Ouchene L, Zhou S, Hudson M, Rezaeian M, LaChance A, Litvinov IV, Baron M, Netchiporouk E; Canadian Scleroderma Research Group. Geographical distribution of systemic

sclerosis in Canada: An ecologic study based on the Canadian Scleroderma Research Group. *J Am Acad Dermatol. 2022 Nov;87(5):1095-1097. doi: 10.1016/j.jaad.2021.12.055. Epub 2022 Jan 11.*

In this study, I performed the literature review, mapping, performed aspects of the statistical analysis, and wrote and edited the manuscript. Dr. Ouchene contributed to the literature review and performing the Monte Carlo simulation. Dr. Zhou provided her expertise and guidance on the statistical analysis. Dr. Rezaeian provided guidance and expertise on spatiotemporal analysis. Dr. LaChance contributed her clinical expertise for study conceptualization, review and editing of the manuscript, and findings from a similar study completed in Massachusetts, United States. Dr. Litvinov contributed to the study conceptualization, provided feedback, and contributing to reviewing the manuscript. Dr. Baron provided his expertise and guidance on the project as well as access to the CSRG database. Dr. Netchiporouk participated in conceptualization, data analysis and study planning, and writing and editing the manuscript.

Muntyanu A, Milan R, Rahme E, LaChance A, Ouchene L, Cormier M, Litvinov IV, Hudson M, Baron M, Netchiporouk E; Canadian Scleroderma Research Group. Exposure to silica and systemic sclerosis: A retrospective cohort study based on the Canadian Scleroderma Research Group. *Front Med (Lausanne). 2022 Sep 29;9:984907. doi: 10.3389/fmed.2022.984907.*

I participated in designing the study collecting and organizing data, preliminary statistical analysis, and writing and editing the manuscript. Dr. Milan helped perform regression analysis and oversee the statistics on the study. Dr. Rahme provided her expertise on the data and statistical analysis. Dr. LaChance contributed her clinical expertise of running a large connective tissue disease clinic and review and editing the manuscript. Dr. Ouchene contributed to the literature review and statistical analysis. Dr. Cormier, Dr. Litvinov, Dr. Hudson and Dr. Baron provided feedback on the design, analyses, and manuscript. Dr. Baron provided his expertise and guidance

on the project as well as access to the CSRG database. Dr. Netchiporouk guided and supervised the study including with conceptualization, methodology, analysis, and review and editing of the manuscript. All authors have read and approved the final version of the manuscript.

Muntyanu A, Milan R, Rahme E, Baron M, Netchiporouk E; Canadian Scleroderma Research Group.Organic solvent exposure and systemic sclerosis: A retrospective cohort study based on the Canadian Scleroderma Research Group registry. *J Am Acad Dermatol. 2024 Mar;90(3):605-607. doi: 10.1016/j.jaad.2023.04.062.*

I participated in designing the study collecting and organizing data, preliminary statistical analysis, and writing and editing the manuscript. Dr. Milan helped perform regression analysis and oversee the statistics on the study. Dr. Rahme provided her expertise on the data and statistical analysis. Dr. Baron provided his expertise and guidance on the project as well as access to the CSRG database. Dr. Netchiporouk guided and supervised the study. All authors have read and approved the final version of the manuscript.

Muntyanu A, Milan R, Kaouache M, Ringuet J, Gulliver W, Pivneva I, Royer J, Leroux M, Chen K, Yu Q, Litvinov IV, Griffiths CEM, Ashcroft DM, Rahme E, Netchiporouk E. Tree-Based Machine Learning to Identify Predictors of Psoriasis Incidence at the Neighborhood Level: A Populational Study from Quebec, Canada. *Am J Clin Dermatol. 2024 May;25(3):497-508. doi: 10.1007/s40257-024-00854-3.*

I contributed to conceptualization, methodology, data access and curation, writing original draft, review and editing, creating tables and figure including mapping. Dr. Milan contributed to methodology, statistical analysis. Data sources and curation, review and editing. Dr. Kaouache contributed to methodology, statistical analysis, data curation and review and editing. Dr. Ringuet

and Dr. Gulliver contributed to methodology, sharing clinical expertise, and review and editing. Dr. Pivneva, Dr. Royer, Dr. Leroux, and Dr. Chen are part of the Analysis Group and contributing to building machine learning models and methodology used in the study. Dr. Yu contributed to methodology, sharing her expertise in machine learning, and review and editing. Dr. Litvinov, Dr. Griffiths, Dr. Ashcroft and Dr. Rahme provided their expertise and feedback, and review and editing. Dr. Netchiporouk conceptualized the study, contributed to the methodology, and writing.

List of Figures

Figures of Supplementary Material for Manuscript 1

Supplementary Figure 1. Female to male ratio for Incidence and Prevalence between 1996-2019.

Figures of Manuscript 2

Figure 1. (a) High-prevalence Forward Sortation Areas for Systemic Sclerosis in Hamilton, Ontario. (b) Expected prevalence of Systemic Sclerosis from the Monte Carlo Simulation based on age and sex alone.

Figures of Supplementary Material for Manuscript 2

Supplementary Figure 1. Geographic mapping of SSc patients in Canada (from 2004-2019) using the CSRG database.

Supplementary Figure 2. High prevalence of SSc FSAs in Montréal, Québec.

Supplementary Figure 3. Expected prevalence of SSc, from Monte Carlo simulation, in Montreal, Quebec.

Figures of Manuscript 3

Figure 1. Kaplan-Meier curve evaluating mortality over time in silica exposed vs. unexposed groups

Figures of Supplementary Material for Manuscript 4

Supplementary Figure 1. Kaplan-Meier curve evaluating mortality of systemic sclerosis (SSc) patients over time in organic solvent exposed vs. unexposed groups.

Figures of Manuscript 5

Figure 1. (a) Geographic distribution of psoriasis incidence in Quebec over the study period. **(b)** Montreal Metropolitan area and (c) Eastern Quebec.

Figure 2. Parsimonious plots demonstrating the change in the area under the curve with an increasing number of included variables.

Figure 3. Partial dependence plot showing directionality of the association between psoriasis incidence and the highest ranked features. (a) Vitamin D level, (b) Temperature, (c) Social Factor Score, and (d) proportion of females.

Figures of Supplementary Material for Manuscript 5

Supplementary Figure 1. Study design.

Supplementary Figure 2. Psoriasis incidence in Quebec during the study period.

Supplementary Figure 3. Moving to the right on the x-axis corresponds to a higher soil moisture and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 4. Moving to the right on the x-axis corresponds to more deprivation in a neighbourhood and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 5. Moving to the right on the x-axis corresponds to a higher concentration of visible minorities and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 6. Moving to the right on the x-axis corresponds to more nighttime light brightness exposure and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 7. Moving to the right on the x-axis corresponds to increased proportion of land covered by mid- and high-rise buildings in an open arrangement and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 8. Moving to the right on the x-axis corresponds to a a longer distance to expressways and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 9. Moving to the right on the x-axis corresponds to a higher concentration of PM2.5 and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 10. Moving to the right on the x-axis corresponds to a higher concentration of O3 and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 11. Moving to the right on the x-axis corresponds to a higher concentration of SO2 and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 12. Moving to the right on the x-axis corresponds to a higher concentration of NO2 and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 13. Moving to the right on the x-axis corresponds to a higher level of instability in a neighbourhood and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 14. Moving to the right on the x-axis corresponds to a higher level of material factor deprivation and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 15. Moving to the right on the x-axis corresponds to a higher level of dependency in a neighbourhood and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 16. Moving to the right on the x-axis corresponds to a higher median age at a Forward Sortation Area (FSA) and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 17. Moving to the right on the x-axis corresponds to a higher greenness level in a 1000m area and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 18. Moving to the right on the x-axis corresponds to increased proportion of land covered by mid- and high-rise buildings in a dense arrangement and moving up on the y-axis corresponds to increasing psoriasis incidence.

Supplementary Figure 19. (a) Distribution of the social factor score quintile between the high (top 10%) and low (other 90%) FSAs by psoriasis incidence. Higher values indicate lower socioeconomic status. **(b)** Distribution of principal component factor score for deprivation of the Canadian Marginalization Index between the high (top 10%) and low (other 90%) FSAs by psoriasis incidence. Higher values indicate more deprivation.

List of Tables

Table 2. Univariate and multivariate logistic regression model for factors associated with exposure to silica among all study sample.

Table 3. Univariate and multivariate logistic regression model for factors associated with exposure to silica among male and female patients.

Table 4. Association between health outcomes and exposure to silica (yes vs. no).

Tables of Manuscript 4

Table 1. Baseline characteristics of patients with systemic sclerosis.

Table 2. Association between health outcomes and exposure to organic solvents (yes vs. no) among patients with systemic sclerosis.

Tables of Supplementary Material for Manuscript 4

Supplementary Table 1. Mortality case count and incidence rate per 1,000 person years for systemic sclerosis (SSc).

Tables of Manuscript 5

Table 1. Characteristics of patients with psoriasis included in the study.

Table 2. Ranking of all the environmental features analyzed in the Gradient Boosting Machine Learning Algorithm.

List of Abbreviations

ACA	Anti-centromere Antibodies
ACE inhibitors	Angiotensin-Converting Enzyme inhibitors
ACR	American College of Rheumatology
ALE	Active Living Environment
ANA	Anti-Nuclear Antibodies
ASIR	Age-Standardized Incidence Rate
ATA	Anti-TopoisomeraseI/Scl-70 Antibodies
AUC	Area Under the Curve
BDCU	Banque de données communes des urgences
BE	Built Environment
BSA	Body Surface Area
CAD	Canadian currency
CAI	Commission d'Accès à l'Information
CANJEM	CANadian Job-Exposure Matrix
CANUE	Canadian Urban Environmental Health Research Consortium
CI	Confidence Interval
CSRG	Canadian Scleroderma Research Group
DALYs	Disability Adjusted Life Years
dcSSc	Diffuse Cutaneous SSc
DLQI	Dermatology Life Quality Index
EPOI	Enhanced Point of Interest
ESRI	Environmental Systems Research Institute

EULAR	European League Against Rheumatism
FSA	Forward Sortation Area
GBD	Global Burden of Disease
HRCT	High Resolution CT scan
HRQoL	Health-Related Quality of Life
ICD	International Classification of Diseases
IGA	Investigator Global Assessment
ILD	Interstitial Lung Disease
ISQ	Institute de la statistique du Québec
JAAD	Journal of the American Academy of Dermatology
lcSSc	Limited Cutaneous SSc
lncRNAs	Long non-coding RNAs
MED-ECHO	Maintenance et exploitation des données pour l'étude de la clientèle hospitalière
OR	Odds Ratio
PAH	Pulmonary Artery Hypertension
PASI	Psoriasis Area and Severity Score
PGA	Physician Global Assessment
PM2.5	Particulate Matter 2.5
PPV	Positive Predictive Value
PVS	Provincial Vital Statistics
RAMQ	Régie de l'assurance maladie du Québec
RR	Relative Risk
SARDs	Systemic Autoimmune Rheumatic Diseases

SES	Socioeconomic Status
SIR	Standardized Incidence Ratio
SMR	Standardized Mortality Ratio
SSc	Systemic Sclerosis
U1-RNP	U1-ribonucleoprotein
USD	United States Dollar
UVR	Ultraviolet Radiation
VEDOSS	Very Early Diagnosis of SSc
WHO	World Health Organization

Introduction

Over the last few decades, the healthcare system observed a steady increase in the incidence of autoimmune diseases[1]. A study in the United States demonstrated that the prevalence of antinuclear antibody (ANA) positivity, a marker of autoimmunity, increased from 11.0% (in 1988-1991) to 11.5% (in 1999-2004) to 15.9% (in 2011-2012), even with adjustment for multiple confounders and improved detection rates[2]. These findings suggest the contribution of an external factor, such as the environment, in the initiation and the potential progression of autoimmune skin diseases.

The skin is the largest organ in the human body and has a critical function of protecting the body from noxious external substances, whether it is ultraviolet radiation (UVR), toxic chemicals, or air pollutants[3-5]. Given the constant skin contact with external antigens and pathogens, it is not surprising that environmental pollutants may elicit host cell damage. This damage leads to an influx of new epitopes or exposure of epitopes that mimic a self-protein, therefore triggering autoimmunity[6]. As a large surface area of the skin is in contact with the outside world, the environment can have an impact on any dermatologic disease.

Investigating high incidence geographic clusters for a given disease is often an initial step for cause identification. Historically, this led to the discovery of asbestos as the main cause of mesothelioma[7] and benzene exposure as a cause of pediatric lymphoma in Massachusetts[8]. Determining the uneven geographic distribution of a disease and subsequent identification of its risk factors could, in the future, improve screening and diagnosis, inform the distribution of medical resources to significantly affected communities, and build foundation for future research regarding modifiable environmental factors in dermatological disease.

In this thesis, investigations on two immune mediated cutaneous diseases, systemic sclerosis (SSc) and psoriasis, will be discussed. SSc is an autoimmune fibrosing condition that leads to significant morbidity and mortality for patients due to skin fibrosis and internal organ involvement[9]. Psoriasis is an inflammatory skin condition and is also associated with numerous comorbidities, including metabolic syndrome and erosive arthritis[10]. For both conditions the proposed pathogenesis includes genetic predisposition and an exogenous trigger leading to disease development[11,12]. Although there have been substantial therapeutic advancements, little is known about the role of environment on immune system dysregulation and subsequent development of SSc or psoriasis. While different environmental triggers are thought to be involved in these conditions, exploring the impact of the environment may help strengthen this concept and support further research on environmental risk factors leading to preventative counselling. Evaluating the impact of the environment in multiple conditions could also help identify common pathways that drive autoimmunity and inflammation and may uncover broader insights into the interplay between environmental factors and immune system dysregulation in chronic disorders.

Given the focus on two parallel dermatologic diseases, SSc and psoriasis, the discussion about these conditions will be addressed separately.

Systemic Sclerosis

SSc has a typical onset between 20-50 years of age and is up to eight times more common in females than males[9]. It is chronic and irreversible[13]. The most affected organs include the skin, gastrointestinal tract and lungs. Interestingly these are the organs with the most contact to the outside world[13]. Mortality is most often due to pulmonary disease, followed by cardiac and kidney involvement[14]. Skin fibrosis affecting the hands is universal in all subtypes of SSc and is the leading cause of disability[15]. Considering a prevalence of 74.4 cases per 100,000 females and 13.3

cases per 100,000 males (Canadian SSc prevalence data, 2003), approximately 2,500 males and 14,000 females are affected by this debilitating condition in Canada[16]. Most SSc treatments target specific end-organ manifestations, instead of the disease pathogenesis, and have significant adverse effects[17]. The pathogenesis is not fully understood but thought to be induced by an environmental trigger in a genetically predisposed host[18]. The low concordance rate between monozygotic twin patients provides evidence for the importance of an environmental trigger[19]. The nature of such trigger and other factors accounting for the variability in SSc clinical severity and prognosis remain unknown. Proposed factors include viruses, environmental exposures such as air pollution, and occupational exposures such as silica and organic solvents[20]. To date, unfortunately, most of the hypotheses are based on case series and observational studies focusing on one to four exposures[20]. As a result, the 10-year survival rate has remained unchanged for over four decades and is below 65% [21]. Therefore, there is an urgent need to evaluate environmental triggers of SSc to be able to provide more information for patients, improve patient counselling with regards to risk factor avoidance and help stimulate new therapeutic advances.

Analyzing geographic distribution of SSc may allow the identification of high incidence areas and possible triggering factors. Internationally, geographic clustering of SSc cases has been demonstrated in a small Italian village near Rome, select London boroughs, and the Choctaw Native reserve in Oklahoma[20]. In Canada, high prevalence patient clusters were noted in Woodstock (Ontario) [22] and in a Native community, Kahnawake (First Nations Reserve, Quebec)[23], suggesting the role of local environmental exposures as potential triggers of disease. To date, no comprehensive province or nation-wide mapping studies have been performed for SSc. Thus, it is not known whether similar disease clustering occurs in Quebec or elsewhere in Canada.

Psoriasis

Psoriasis is a common immune-mediated inflammatory skin disease which affects 60 million people world-wide[24] and approximately 1-5% of the population in North America[25]. The rates in males and females are similar[25]. There is a bimodal age distribution of disease development, with peaks at 20-30-years-old and 50-60-years-old[26]. In almost a third of cases, the disease develops in the first two decades of life and takes a chronic and persistent course through a patient's lifetime, leading to significant morbidity[27]. Similarly to SSc, the disease is thought to be due to an environmental trigger in a genetically predisposed individual. In psoriasis, contributing factors include UVR exposure, temperature, viral infections, medications, smoking, and stress[28]. Although effective treatments are now available for psoriasis to achieve significant disease control, they are reserved for moderate to severe cases only, due to their high cost[29]. Hence, active identification of risk factors and subsequent preventative strategies is of significant importance.

For psoriasis, incidence varies geographically and ranges 31.4-521.1/100,000 people-years in adults[24]. Higher incidence is observed in industrialized countries and lower rates are seen in countries closer to the equator[24,30]. Variations in psoriasis epidemiology on a country level were also reported where prevalence was higher in areas with colder, drier weather, and with a lower solar irradiance[31,32]. As UVR exposure is therapeutic for patients, it is plausible that climate may also modulate the risk of developing psoriasis[33]. These findings are especially important in the context of Quebec's diverse environment and needs to be further explored as a source of prevention and management of psoriasis.

To date, limited comprehensive provincial or national studies have been done to systematically evaluate the influence of environmental factors on the incidence of SSc and psoriasis. If contributing factors are identified, this knowledge could help educate patients and

physicians, enhance screening and treatment, initiate more in-depth evaluation of environmental triggers in the high incidence areas, and stimulate future research of modifiable environmental factors in dermatologic disease.

Objectives

The overall purpose of this thesis is to address important gaps in knowledge regarding the epidemiology of SSc in Canada and the province of Quebec, to study the geographic distribution of the incidence cases and evaluate environmental and occupational risk factors which could contribute to a higher risk of developing the disease or a more severe disease phenotype and help explain the uneven geographic distribution. Subsequently, to perform a proof-of-concept analysis, using populational data for psoriasis, another inflammatory disease where gene and environment interaction is important. To complete these studies, populational health administrative database from Quebec, Canada was used for SSc and psoriasis, in addition to using a nationwide Canadian Scleroderma Research Group (CSRG) registry. The objectives of the study were as follows:

1. Determine the incidence, prevalence, and mortality of SSc using populational data in Quebec, Canada.
2. Using the CSRG national registry, determine the geographic distribution of SSc and evaluate the contribution of environmental triggers.
3. Explore the impact that occupational exposure to a) silica and b) organic solvents played on clinical manifestations of SSc, disease severity, and mortality.
4. Evaluate the epidemiology of psoriasis using populational data in Quebec, Canada and comprehensively evaluate the environmental factors that could be contributing to high psoriasis incidence.

Chapter 1: Literature Review

Over the last few decades, the healthcare system observed a steady increase in the incidence of autoimmune diseases, which are associated with significant morbidity and mortality for patients[1]. A large study investigating the 19 most common autoimmune diseases in 22 million individuals demonstrated that the age-standardized incidence rate (ASIR) of autoimmune diseases increased over time (incidence rate ratio in 2017/19 vs. 2000/02 of 1.04; 95% CI 1.00–1.09) and that 10.2% of the population had an autoimmune disease[34]. The proposed factors for these trends included foods, xenobiotics, air pollution, infections, personal lifestyles, stress, and climate change[35].

The skin, largest organ in the human body, has a critical function of protecting the body from noxious external substances, whether it is UVR, toxic chemicals, or air pollutants[3-5]. Because of its large surface area in contact with the outside world, the environment can be an important contributor to the development of autoimmune and inflammatory skin disease, such as SSc and psoriasis. In addition to exposures in the environment, characteristics of the neighbourhood such as nighttime light brightness, greenness index, and active living environment (ALE) may also play a role. Association of these factors with metabolic syndrome has been previously reported, and patients with psoriasis frequently have metabolic comorbidities[36,37]. While different environmental triggers are thought to be involved in SSc compared to psoriasis, exploring the environmental impact, in parallel for both aspects, will help strengthen this concept and support further focus on environmental risk factors.

Current research methodology focuses mainly on the person and time variables, whereas spatial analysis is often underutilized[38]. Along with assessment of individual level data and longitudinal changes over time, the consideration of a person's workplace, home, and outdoor

environment can be useful to identify environmental triggers of disease and may reveal new etiologic factors. Geospatial analysis, a method used for data visualization, provides information on geographic distribution of a disease, and allows subsequent evaluation of risk factors in the identified high-risk areas[39]. This may demonstrate interesting findings that would be more difficult to conclude from assessing only numerical data. Confounding factors such as age, sex, and socioeconomic status (SES) which may vary geographically, are important to consider. Using geospatial analysis, these factors could be considered when determining high incidence/prevalence areas. Additionally, knowledge of the high-risk areas could help inform resource distribution and education, prevention, and treatment programs. In the present day, access is available to many large comprehensive health information databases across the world and sophisticated technologies, such as geographical information systems. These are useful tools to drive the study and knowledge of environmental risk factors.

Ecological study designs can be used to analyze the relationship between disease incidence and exposures and serve as a first step in the study of etiologic factors. Thus, large comprehensive patient health databases can be utilized to identify a cohort of patients with a certain diagnosis. For rare diseases, such as SSc, there are no national reporting guidelines, but several multi-institutional registries recruiting patients followed in clinics exist and patients can also be identified using International Classification of Diseases (ICD)-9 and ICD-10 codes from large populational health administrative databases. The CSRG, a multi-institutional patient registry, was developed in 2004 to increase the capacity to perform high impact research on SSc in Canada and help improve patient outcomes for this debilitating disease[13,40]. This longitudinal registry contains detailed demographic, clinical, laboratory, and imaging data providing information for the evaluation of potential environmental triggers. However, this institutional registry is subject to recruitment bias

as patients are included from only 15 centers across the country and miss those in rural communities. To confirm findings and provide more detailed evidence, complete populational data is required. The Régie de l'assurance maladie du Québec (RAMQ), is a provincial health database in Quebec, used by physicians to document services and treatments provided for each patient. Patients with a certain diagnosis can be identified using specific diagnostic codes and additional information can be retrieved on comorbidities, medication use, and mortality to study disease manifestations and severity. While some limitations exist in every registry, such as incorrectly classified cases, missing cases of true disease, and limited access to individual level confounder data, registry-based studies allow robust initial evaluation of risk factors and serve as a foundation for many prospective cohort studies.

Systemic Sclerosis (SSc)

Diagnosis and Clinical Manifestations

SSc is an autoimmune fibrosing disease that affects the skin and internal organs with a typical onset between 20-50 years of age and a female predominance of 8-12:1[13]. It is chronic, irreversible, and leads to significant morbidity and mortality[13]. It is classified into limited cutaneous (lcSSc) (the most common form), diffuse cutaneous (dcSSc), sine scleroderma and overlap disease based on the extent of skin fibrosis and the pattern of internal organ involvement[40]. Most commonly affected organs are those with direct exposure to the environment such as the skin, gastrointestinal tract, and lungs.[14] In all forms, and especially in dcSSc, there is significant extracutaneous involvement with considerable morbidity and mortality, with a 10-year likelihood of survival of less than 65% [14], which has remained unchanged for the last four decades.[21] While SSc is strikingly more common in females, inverse gender predisposition has been reported in North East England (5 males to 1 female) and Tokyo (14 males to 1 female)[41]. Compared to

females, SSc in males is more likely to present with a more severe phenotype including dcSSc, cardiomyopathy, interstitial lung disease (ILD) and scleroderma renal crisis[41]. Whether these differences are related to sex-hormones or acquired (e.g. occupational/environmental) risk factors require further investigation.

The diagnosis of SSc is based on the American College of Rheumatology (ACR)/European League against Rheumatism (EULAR) classification criteria, which includes clinical disease manifestations, such as bilateral, symmetric skin thickening proximal to the metacarpophalangeal joints, presence of SSc-related abnormalities (*e.g.* Raynaud's phenomenon, fingertip lesions, telangiectasias, SSc-specific antibodies, abnormal nailfold capillaroscopy) and internal organ involvement (*e.g.* ILD, pulmonary artery hypertension) **(Table 1)**[42]. In the last few decades, the availability of nailfold capillaroscopy, showing dilated loops and drop out sign, and SSc-specific autoantibodies, have enabled an earlier diagnosis of SSc. These specific criteria are part of very early diagnosis of SSc (VEDOSS) which was established in 2011[43]. Early diagnosis helps with initiating treatment and minimizing end organ damage thereby improving patient outcomes.

Table 1. ACR/EULAR classification criteria for SSc. Patients with a total score of ≥9 are being classified as having definite SSc. Adapted from van de Hoogen *et al*.[42]

Item	Sub Item	Weight/Score
Skin thickening of the fingers of both hands extending proximal to the metacarpophalangeal joints (sufficient criterion)		9
Skin thickening of the fingers (only count the highest score)	Puffy fingers	2
	Sclerodactyly of the fingers (distal to MCP but proximal to the PIPs)	4
Fingertip lesions (only count the highest score)	Digital Tip Ulcers	2
	Fingertip Pitted Scars	3
Telangiectasia		2

Abnormal nailfold capillaries		2
Lung Involvement (*Maximum score is 2)	Pulmonary arterial hypertension (PAH)	2
	Interstitial lung Disease (ILD)	2
Raynaud's phenomenon		3
Scleroderma related antibodies (*Maximum score is 3)	ACA	3
	ATA	3
	anti-RNA polymerase III (RNAP3)	3

MCP – metacarpophalangeal joint, PIP - proximal interphalangeal joint

There are several specific autoantibodies that help with the diagnosis, clinical presentation, prognosis, and the exclusion of other conditions presenting with skin thickening. ANA is positive in ~95% of patients[9,44,45]. SSc-specific-autoantibodies include anti-centromere antibodies (ACA), which is associated with lcSSc and pulmonary artery hypertension (PAH), anti-topoisomeraseI/Scl-70 antibodies (ATA), associated with dcSSc and interstitial lung disease (ILD), and anti-RNApolymerase (RNAP)-III antibodies, associated with malignancy and renal crisis **(Table 2)**[9]. Other autoantibody profiles include Polymyositis/Scl, To/Th ribonucleoprotein, U1-ribonucleoprotein (U1-RNP), U3-RNP/fibrillarin[9] and a combination of autoantibodies can also help predict survival, timing, risk, and incidence of systemic complications[9,46].

Table 2. Summary of autoantibody profiles in SSc and corresponding key systemic associations. Adapted from Jerjen *et al*[9].

Antibody	**Estimated Prevalence**	**Subtype of SSc**	**Main Systemic Associations**
ACA	20-25%	lcSSc	• PAH (15-20%) • Esophageal dysmotility and gastrointestinal dysfunction • Low risk of ILD, cardiac and renal disease
ATA	20-30%	dcSSc >lcSSc	• High risk of ILD (early) • PAH • Scleroderma Renal Crisis • Cardiac • Myositis

Polymyositis (PM) /Scl	2-4%	Overlap Polymyositis/ lcSSc	• Myositis • ILD (50% by 15 years) • PAH (about 36% by 15 years) • Cardiac • Renal
To/Th ribonucleoprotein	<5%	lcSSc	• ILD (45%) • PAH (25%)
RNA polymerase III (RNAP-III)	1-22%	dcSSc (rapidly progressive)	• Scleroderma Renal Crisis (early) • Increased risk of malignancy within 3 years of diagnosis • Moderate risk of ILD • PAH (later) • Gastric antral vascular ectasia • Myositis • Less cardiac involvement
U1-ribonucleoprotein (U1-RNP)	5-10%	lcSSc, overlap syndromes (mixed connective tissue disease)	• Myositis • PAH • ILD
U3-ribonucleoprotein (U3-RNP) / Fibrillarin	4-10%	dcSSc	• Early severe organ involvement: • PAH (highest risk) • ILD • Scleroderma Renal Crisis • Cardiac • Small bowel dysmotility • Myositis

Systemic manifestations include pulmonary (ILD or PAH), renal (*i.e.* scleroderma renal crisis or renal vasculopathy), cardiac (*i.e.* heart failure, arrythmias, pericardial effusion and valve sclerosis), gastrointestinal (*i.e.* gastroesophageal reflux disease, impaired motility, gastric antral vascular ectasia), and urogenital (*i.e.* sexual dysfunction) involvement as well as increased malignancy risk (*i.e.* cutaneous, breast, bladder, lung, liver, and hematological)[9]. Regular screening for lung involvement is mandatory and consists of pulmonary function tests as well as High Resolution CT scan (HRCT)[47]. Cardiac involvement is more common in older patients and in those with ATA antibodies[9,48]. Scleroderma renal crisis is a severe manifestation but less common now

given the knowledge that high dose prednisone can precipitate it, as well as preventative options available such as angiotensin-converting enzyme (ACE) inhibitors[49,50]. It is typically manifests within the first 5 years of SSc diagnosis. The most common internal organ system to be involved is the gastrointestinal tract and most patients with SSc are affected to some degree[51].

Epidemiology

Based on the Canadian SSc prevalence data in 2003, there are approximately 74.4 cases per 100,000 females and 13.3 cases per 100,000 males, equating to 2,500 males and 14,000 females affected by this debilitating condition[16]. The highest point prevalence observed globally is 47/100,000 persons in the Canadian First Nation residents. Internationally, the prevalence of SSc has been reported to be between 7 to 489 cases per million individuals[52]. SSc is diagnostically challenging due to heterogeneous clinical presentations and limited understanding of the pathogenesis, hence resulting in the wide discrepancy of prevalence reported. Higher case numbers are reported in North America and Australia compared to continental Europe, United Kingdom, and Japan[52]. In the United States, estimated period prevalence is 50/100,000 persons and the age-sex adjusted annual incidence is 5.6/100,000 person-years (years 2003-2008)[53,54]. In Europe, a north–south gradient is observed with increased prevalence in southern countries[52]. A small number of case reports or observational studies have been published to date reporting uneven geographic distribution. In 1990, Silman *et al*. described the first cluster of SSc patients in boroughs near international airports in London, United Kingdom[55]. Later, Valesini *et al*. found five cases in a rural community of 572 individuals near Rome[56]. In this region, ten additional individuals had clinical features suggestive of SSc and eight had SSc-related serological abnormalities, evidenced by positive ANA and ACA antibodies[56]. The prevalence of SSc was also documented to be 7-fold higher in Libby, Montana, where mining activities are predominant[57].

Other reported clusters include the Choctaw Native reserve, United States[58], Woodstock, Canada[59], and Kahnawake First Nations Community, Quebec,Canada[60]. These studies suggest a non-random distribution of SSc cases and highlight the need to further investigate the epidemiology and environmental risk factors for this disease.

Healthcare burden

This debilitating condition is associated with significant healthcare and economic burden. Within the first year of diagnosis, SSc patients experience at least double the risk of hospital admission, emergency room and outpatient visits. Additionally, they experience an adjusted annual difference in direct and indirect healthcare cost of $12,820 USD and $3,103 USD, respectively, compared to matched controls[61]. In Canada, the average annual societal cost per SSc patient was estimated to be $18,453 CAD in 2009[62]. Hence, SSc leads to significant medical and economic burden[16]. Therapies for SSc target end organ manifestations rather than pathogenesis and have significant adverse effects, thus further contributing to morbidity and mortality[17]. Elucidating preventive strategies is therefore of significant importance.

Pathogenesis

The pathogenesis of SSc is not fully understood, but is thought to be due to an environmental trigger in a genetically susceptible host[18]. A three-step hypothesis has been proposed which includes endothelial cell dysfunction/vasculopathy, inflammation due to disbalance of innate and adaptive immune system, and finally abnormal collagen deposition leading to skin thickening/fibrosis **(Figure 1)**[9]. Genetic predisposition has been demonstrated with an increased risk of disease development in family members. The adjusted relative risk (RR) of SSc in a first- relative of a patient with SSc was 13.23 (95 % CI 3.11–56.33), with siblings

demonstrating a RR of 81.22 (95% CI 11.40–578.72) and a RR of 9.17 (95% CI 1.30–65.86) for children[63]. However, evidence of geographic clustering and low concordance rates in monozygotic twin studies strongly suggest the importance of environmental influences in disease initiation and warrants further investigation[19]. To further support this, several long non-coding RNAs (lncRNAs) implicated in modulation of gene-environment interaction have been recently identified in SSc[64].

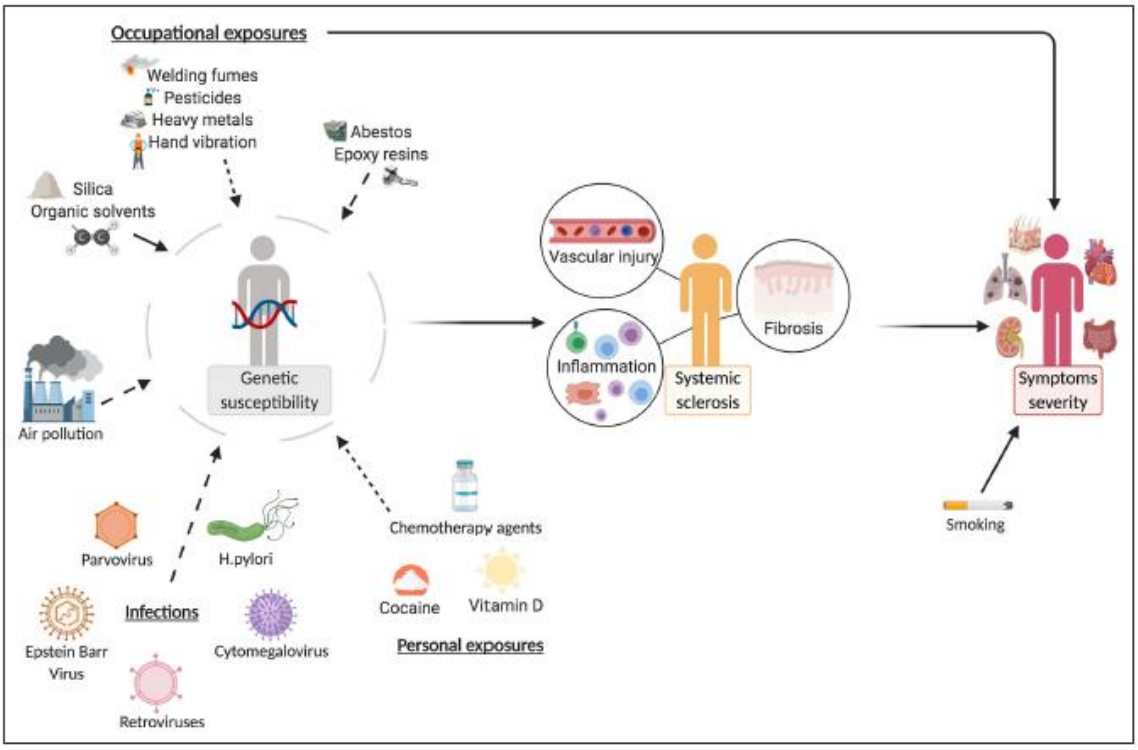


Figure 1. Pathogenesis of SSc and the role of occupational, environmental, and personal exposures to disease development. H. pylori - Helicobacter pylori. Thick lines—exposures increasing the incidence and/or severity of SSc risk based on observational studies. Dashed lines—exposures with probable increased incidence and/or severity of SSc (limited data). Dotted lines—exposures with anecdotal/unproven data only. Adapted from Ouchene L, Muntyanu A, *et al*[22].

Exposures

Possible extrinsic triggers that have been studied in SSc include infectious agents, chemicals, occupational or environmental pollutants, and medications/supplements, all of which could modulate immune response and elicit microvascular damage and inflammation, ultimately leading to fibrosis at various degrees[65]. Focusing on the role of environmental pollutants is critical. Air pollution, a frequently studied environmental factor, is known to adversely impact many diseases, and was designated by the World Health Organization (WHO) as the primary environmental threat to humanity. It accounts for at least 7 million deaths globally every year[66]. Currently, the highest level of evidence for extrinsic triggers of SSc is due to occupational or environmental exposures to silica and organic solvents, both of which contribute to increasing incidence and severity[22]. Specifically, a summary of cohort studies, including 247,563 patients, demonstrated that workers exposed to silica had an 18-fold increased incidence of SSc[6]. Occupations with a high risk of exposure to silica are shown in **Table 3.** Silica exposure was associated with lower survival rates and more severe disease phenotype such as dcSSc, digital ulcers, ILD, myocardial dysfunction, and positive ATA antibodies[6]. Analysis of 14 studies demonstrated that organic solvent exposure led to 2-4-fold elevated risk of SSc development, with increased disease severity and mortality[6,67]. Other exposures proposed to increase SSc risk include epoxy resins, asbestos, and particulate air pollution[6].

Table 3. Occupations with the highest risk of crystalline silica exposure. Only the occupations with >50% probability are shown. Adapted from Ouchene L, Muntyanu A, *et al*[22].

Occupation	**Probability**	**Intensity**	**Frequency (h)**	**FWI**
Coremaker (hand)	100	High	40	25
Bench molder (metal)	93	High	40	25
Terrazzo worker	80	Medium	40	5.4
Mineral-crushing- machine operator	71	Medium	40	7.6

Stone cutter and finisher	71	High	40	25
Coremaker (machine)	67	High	40	15
Gem cutter and polisher	67	Medium	40	5
Other miners and quarrymen	60	Medium	40	5.4
Concrete-mixer operator	60	Medium	40	7.8
Miner (general)	59	Medium	40	5.5

Abbreviations: h, hours per week; FWI, frequency-weighted intensity. FWI is calculated as: exposure intensity x frequency of exposure in hours worked per week/40 hours and is used to estimate the overall importance of occupational exposure to the given agent. FWI=25 is equivalent to 40 h per week at high intensity. Jobs included in the table were judged to have definite risk of exposure to silica, at least 5 hours of weekly exposure and medium or high intensity of exposure based on CANadian Job-Exposure Matrix estimates (CANJEM)[68].

Exposure to silica and solvents is not only work-related[69] but could be due to indoor inhalation of benzene from cooking practices or smoking, outdoor ambient air pollution and industrial emissions[69]. A recent study in the United States showed that within 10 km of a silica mining operation, ambient air silica levels were consistently above the national average[70]. A Canadian study demonstrated an association between ambient air pollution with fine particulate matter ($PM_{2.5}$) and the development of systemic autoimmune rheumatic diseases (SARDs), including SSc, in Calgary[71]. The study revealed that particulate pollution was more prevalent in urban settings and so was the prevalence of SARDs, highlighting the need to further investigate the epidemiology and environmental triggers of SSc.

Canada is a leading mining country with industries ranging from exploration, mining, and downstream processing and manufacturing[72]. Our country is the primary producer of potash and is in the top five producers of many other minerals and metals including gold, cobalt, nickel, and platinum[72]. Overall the minerals and metals industry has over 200 mines and approximately 6,500 sand or gravel or stone quarries[72]. Furthermore, Canada accounts for 31 nonferrous smelters and refineries in Newfoundland and Labrador, New Brunswick, Québec, Ontario, Manitoba, Alberta

and British Columbia, which are the downstream industry of mineral processing[73]. Hence, Canada is a particularly suitable country to study the impact of environmental pollution on SSc.

Management

No curative treatments exist for SSc and available therapies may only have some affect on slowing disease progression, with significant side effects[17]. Treatment is focused on the disease manifestations present. For skin fibrosis, immunosuppressive agents such as mycophenolate mofetil and methotrexate are typically used, with newer agents such as tocilizumab and rituximab showing early promise in clinical trials[47,74]. In recent years, autologous hematopoietic stem cell transplantation provided a lot of hope[75]. This procedure reduces the aberrant immune cells and allows re-population with a self-tolerant immune system. While several studies have shown a decrease in skin involvement, improvement of organ function, quality of life measures, and overall survival[76], this procedure has significant risks limiting its use to certain subsets of patients.

Psoriasis

Diagnosis and Clinical Manifestations

Psoriasis is a common immune-mediated inflammatory skin disease. It presents with well-defined erythematous plaques on the skin with classic silvery scale[26]. The diagnosis is typically clinical, although a skin biopsy can be done and shows regular acanthosis of the epidermis, neutrophils in the stratum corneum, suprapapillary thinning, dilated vessels in the dermal papillae, and absence of a granular layer[77].

There is a bimodal peak in disease development: 20-30-years-old and 50-60-years old[26]. In almost a third of cases, disease onset is in the first two decades of life and takes a chronic and

persistent course with high cumulative disability over a lifetime[27]. Early onset psoriasis is defined when the disease starts before 40 years of age[78].

Patients affected by psoriasis often experience pain and pruritus. Severity is graded based on body surface area involvement (BSA), Investigator Global Assessment (IGA), Physician Global Assessment (PGA), Psoriasis Area and Severity Score (PASI), and Dermatology Life Quality Index (DLQI), as discussed below [79]. The visible nature of the disease and involvement of sensitive areas (*e.g.* palms/soles, genitals) often translates to significant psychological burden[80] and stigmatization[81-84]. Stigma may lead to poor health-related quality of life (HRQoL), increased risk of medical and mental health comorbidities and an accentuated barrier to treatment[82]. In 2013, the WHO recognized psoriasis as a major global health problem[30]. A recent Global Burden of Disease (GBD) study ranked psoriasis as the second contributor to all skin-related Disability Adjusted Life Years (DALYs)[30]. The GBD study highlighted an increase in prevalence and morbidity (with worse trends seen in males) of psoriasis world-wide, and especially in North America and Europe. Thereby, psoriasis is associated with significant individual and societal impact.

- BSA determines the percent of the body surface that is involved. In clinical practice, BSA is most commonly estimated using the palm method where one palm of the patient's hand corresponds to 1% of the whole BSA.
- IGA is assessed by the investigator at the clinic visit at one point in time. This is similar to the PGA, which is completed by a physician (not a study investigator). It is graded as follows: (0) = clear; (1) = almost clear; (2) = mild; (3) = moderate; and (4) = severe.
- PASI evaluates three clinical domains of the disease including thickness, redness and scaling together with the body surface involved by the psoriasis and is a marker of severity.

This score assesses four areas of the body such as the head and neck, upper extremities, trunk and lower extremities.

Psoriasis is a multisystem disease and an independent risk factor for development of inflammatory (*e.g.* psoriatic arthritis, inflammatory bowel disease), metabolic (*e.g.* diabetes, obesity, sleep apnea), cardiovascular/cerebrovascular, neoplastic, mental health and other disorders[10,25]. Hence, patients with psoriasis require multidisciplinary care and screening for co-occurring comorbidities.

Epidemiology

Psoriasis affects >100 million people world-wide[30] and ~1-5% of the population in North America (with similar rates in males and females)[25,85]. A study using health administrative data from Ontario, Canada, identified 273,238 patients with psoriasis out of 10,774,802 individuals ≥20-years-old residing in Ontario in 2015[85]. Hence the cumulative prevalence estimate was 2.54%[85]. There was a gradual increase in age and sex standardized cumulative prevalence noted over the time period from 1.74% in 2000 to 2.32% in 2015. However, there was a decreasing trend for annual incidence rates between 2008-2015. Another study in southwestern Ontario comprising of 7,935 patients with psoriasis out of 325,618 recorded patients between 2008-2012 indicated a prevalence of 2.44%[86].

Psoriasis incidence varies geographically and ranges from 31.4-521.1/100,000 people-years in adults[24]. Higher incidence is seen in industrialized countries and lower rates are seen in countries closer to equator[24,30]. Variations in psoriasis epidemiology within a country were also reported, with a higher prevalence (no incidence data) in areas with colder, drier weather and

lower solar irradiance[31,32]. As UVR exposure is therapeutic for patients, it is plausible that climate may also modulate the risk of developing psoriasis[33].

Healthcare Burden

There are safe and very effective medications available to treat psoriasis such as biologics and targeted therapies. However, these agents are associated with a significant cost which can be a barrier to treatment initiation for the patient and a significant burden on the healthcare system. In 2008, it was estimated that the mean cost of psoriasis was 1.7$ billion dollars annually to the Canadian healthcare system[87]. The average annual cost per patient is calculated at $7,999[87]. In this study, the average lost productivity costs were reported as $3,442 per patient, making up 43% of the average annual cost. Indirect costs of psoriasis, such work-related productivity loss, totals 749$ million dollars, or up to 2,270$ in mean wage loss per person yearly, for all moderate and severe psoriasis patients in Canada[88].

Pathogenesis

The overall pathogenesis of psoriasis is thought to be due to genetic predisposition and an environmental trigger. A study evaluating 69,828 patients diagnosed with psoriasis as part of National Health Insurance in 2010 found that the adjusted RR for individuals with an affected first-degree relative was 5.50 (95% CI, 5.19–5.82) and affected second-degree relative 2.54 (95% CI, 2.08–3.12)[89]. When compared to the general population, individuals with one type of affected first-degree relative had a RR of 5.19 (95% CI, 4.9–5.51) and those with two or more had a RR of 27.42 (22.07–34.08) for psoriasis[89]. While genetics play an important predisposing role, family history of psoriasis is present in ~30% of patients and is unlikely to account for rising prevalence/morbidity, suggesting that potentially modifiable risk factors such as environmental, SES, behavioral and/or other exposures could be contributing.

Exposures

Numerous studies have previously demonstrated that triggers may include infections, psychological stress, medications, skin trauma, smoking, unhealthy lifestyle factors as well as dysbiosis of the skin and gut microbiome, dysregulated lipid metabolism and sex hormones[28]. Limited studies exist on evaluating environmental/neighbourhood characteristics. In a review completed by our team, four studies focused on air pollution, two on rurality and two on social and material deprivation[90]. An increased risk of psoriasis flare was suggested in association with particulate air pollution ($PM_{2.5}$ and PM_{10}) in four studies. In psoriatic arthritis, air pollution was associated with biologic failure[20]. While no data was identified in regards to air pollution and psoriasis incidence, ambient air pollution is one of the leading environmental health threats[21] and a well-established risk factor for psoriatic comorbidities[20–25]. Unfortunately, data on other potentially important environmental risk factors that could be integrated into the prevention and management of psoriasis are absent in the literature. These include noise pollution (associated with higher risk of metabolic/mental health diseases), nighttime light pollution (circadian rhythm disrupter), and vegetation index (greenness mitigates harmful exposures and promotes healthy lifestyle)[36,37,91-93].

The association of behavioral risk factors and lifestyle (*e.g.* physical activity, weight management, smoking, alcohol) with psoriasis incidence/severity, risk of developing comorbidities and response to treatment is well established[94,95]. However, it is increasingly recognized that unhealthy lifestyle associated with psoriasis disease spectrum is not only driven by individual factors (where the interventions focusing on the individual often fail), but rather arises in the context of larger social, cultural, economic and environmental determinants of health[96]. For example, encouraging a patient with psoriasis who also has increased body weight

and a sedentary lifestyle to engage in lifestyle modification may simply not be feasible on an individual level if they live in a precarious neighborhood with no access to recreation facilities, groceries stores within walking distance, or public transportation. Population-health research in chronic diseases associated with psoriasis (*e.g.* diabetes, obesity, hypertension) highlights the importance of the built environment (BE) (*i.e.* man-made buildings and spaces) as a critical element to address population-level health differences, and as an intervention to reduce chronic disease rates. However, data on such population-level interventions to address psoriasis is exceedingly limited. We identified only one study focusing on the deprivation index (high deprivation was associated with psoriasis prevalence), one on income quintiles (lower incomes were associated with increased psoriatic prevalence) and two studies on urban vs. rural residence (with conflicting results). All came with moderate to high risk of bias as they addressed one or few factors, without adjusting for confounders and did not study psoriatic incidence as outcome. Incidence is more informative than prevalence to establish risk factors. Other features of the BE that support a healthy lifestyle (*e.g.* ALE, safety, green space, grocery stores) and were shown to be protective against some of the effects of the social determinants of health, such as low SES, were not studied in psoriasis to date[97]. A thorough understanding of the social determinants of health in psoriasis is essential to reduce health disparities and promote equity.

Management

While safe and effective treatments (*e.g.* biologics, targeted molecules) are available and indicated for recalcitrant moderate-severe psoriasis, their high cost precludes routine use beyond indication. Whether they reduce/prevent associated comorbidities needs further research.

Chapter 2: Methods

Geographic Distribution

In Canada, postal codes are defined by six-character alphanumeric string associated with one or more mail delivery points, each one accounting for 19 households on average[98]. Forward Sortation Areas (FSAs) represent the first three characters of a postal code and correspond to a postal facility from which mail delivery originates and serves approximately 8,000 households (range 0 to >60,000)[98]. For example, the Montreal General Hospital has a postal code of H3G 1A4. Hence, the FSA to which it belongs is H3G. In 2011, the midpoint year for our study period, there were 1,621 FSAs in Canada. FSAs were categorized as urban or rural, defined by population density of less or more than 400 habitants/km^2, according to Statistics Canada[99].

Figure 2. Example of demarcation of FSAs in a region of Canada

Data Source

Objective 1, 4: *(1) Determine the incidence, prevalence, and mortality of SSc using populational data in Quebec, Canada. (4) Evaluate the epidemiology of psoriasis using populational data in Quebec, Canada and comprehensively evaluate the environmental factors that could be contributing to high psoriasis incidence.*

Québec health services administrative databases are linkable by a unique patient identifier and access is facilitated by the Institute de la Statistique du Québec (ISQ). All citizens in the province of Quebec are covered by the provincial health plan. The database encompasses all individuals, which is approximately 8 million citizens. ISQ data includes the RAMQ physician billing claims database, Maintenance et exploitation des données pour l'étude de la clientèle hospitalière (MED-ECHO) hospitalization data, Provincial Vital Statistics (PVS) mortality data, the RAMQ pharmaceutical claims for prescribed medication database, and emergency department encounters from the Banque de données communes des urgences (BDCU) database. The ISQ coordinates the approval of data release through the Commission d'accès à l'information (CAI) and allows data access on its premises with results of analyses accessible remotely as well. Unfortunately, the ISQ does not provide individual level data on sociodemographic features (*e.g.* ethnicity, education, income, or smoking).

Case Identification

For this study, patient data between 1996-2019 was used. An incident case of SSc was identified using the following definition: ≥2 ICD-9 (7101) and/or ICD-10 (L943, M340, M348, M349) diagnostic codes for SSc in RAMQ database at least 2 months apart and within 2 years; ≥1 billing code by a specialist (rheumatologist); or ≥1 billing code with a primary or secondary

diagnosis during a hospitalization (MED-ECHO database). Therefore, all patients in the province with a diagnosed case of SSc during the study period were included in the study. This definition for identifying SSc patients was previously validated in the Quebec population. Diagnostic codes for morphea (localized scleroderma) are distinct and hence were not included in this cohort.

With regards to the psoriasis cohort, the same database was used to identify patients with psoriasis using the following definition: ≥2 ICD-9 (696.1) and/or ICD-10 (L40.x) codes by any physician within 2 years, or one or more hospitalization with psoriasis as the primary or secondary diagnosis (MED-ECHO database). This definition was not validated in the RAMQ database. The study period for this study was 1997-2015.

Objective 2, 3a/b: *(2) Using the CSRG national registry, determine the geographic distribution of SSc and evaluate the contribution of environmental triggers. (3) Explore the impact that occupational exposure to a) silica and b) organic solvents played on clinical manifestations of SSc, disease severity, and mortality.*

Data Source

A national registry of SSc patients was used to determine the geographic distribution of SSc across the country and evaluate the role of occupational exposures in disease manifestations and severity. The CSRG is a multi-center SSc patient registry, capturing approximately 10% of all prevalent cases[16]. It contains comprehensive demographic, medical, and imaging data, allowing for assessment of disease severity and clinical course. Data was extracted from the date of the database inception (2004) until 2019 and originates from patients recruited at Canadian rheumatology centers located in Montreal, Sherbrooke, Québec City, London, Hamilton,

Newmarket, Winnipeg, Halifax, Moncton, Calgary, and Saskatoon. SSc diagnosis was verified by an experienced rheumatologist.

Statistical Analyses

The first 3 digits of a patient's postal code were available for each patient at recruitment into the registry. In Canada, of the 1,621 separate FSAs in 2011, only 665 contained SSc patients. The period prevalence was calculated between 2004-2019, per FSA, using the Canadian census data for 2011, the midpoint year, as the denominator. Prevalence per FSA was used to map the cases with geographic information systems software (ArcMap 10.7.1 from Environmental Systems Research Institute [ESRI], Redlands, California) as previously described[72,73,100,101]. The base map of postal codes was obtained from Statistics Canada[102].

A Monte Carlo simulation was performed to predict the distribution of the CSRG patients across the available FSAs using MATLAB. Population size, age and sex distributions for each geographic region were obtained from the 2011 Census from Statistics Canada and to added into the model. The simulation was repeated 100,000 iterations to obtain expected SSc counts with 95% CIs. Observed CSRG period prevalence of SSc cases per FSA was compared to expected prevalence, obtained from the simulation using binomial probability test in MATLAB. Significance level was set to $p<0.05$. This allowed determination of the high prevalence areas, defined as a statistically higher prevalence rate than expected from the simulation.

Data on the patient's age and sex, dcSSc or lcSSc subtypes, mortality[103], disease duration, and reported occupational exposures to silica and organic solvents were compared between areas of high prevalence and average CSRG prevalence to control for potential confounders. This was

performed using a Chi-squared test for categorical variables, and Student's t-test for continuous variables in GraphPad Prism version 8.

Objectives 2, 4: *(2) Using the CSRG national registry, determine the geographic distribution of SSc and evaluate the contribution of environmental triggers. (4) Evaluate the epidemiology of psoriasis using populational data in Quebec, Canada and comprehensively evaluate the environmental factors that could be contributing to high psoriasis incidence.*

Identification of Exposures

Data on >350 neighborhood and environmental variables including air/noise/light pollution measures, greenness indicators, climate metrics, SES (e.g. material/social deprivation, instability, marginalization) and BE features (*e.g.* urbanization, proximity measures, ALE) were obtained from the Canadian Urban Environmental Health Research Consortium (CANUE). This data was available per 6-digit postal code. Data was aggregated into FSAs to use for data analysis by taking into account the land area occupied by each 6-digit postal code as described in the manuscripts.

Statistics Canada sources were used to identify the proportion of females in an FSA and the median age in each FSA for each census year (*i.e.* 2001, 2006, 2011).

ArcGIS was used to assess the presence of industrial facilities as well as industrial and mining lands from the Enhanced Point of Interest (EPOI) database[104]. The EPOI file is a national database of over one million Canadian businesses and recreational points of interest. Industrial facilities, which are reported in the EPOI per FSA, were added to the base map in ArcGIS. The industries of interest selected in objective 2, included paper, petroleum and coal products, chemical, plastic and rubber products, non-metallic mineral product, primary metal, machinery,

and transportation manufacturing, along with mining land. Considering that previous studies have documented that most relevant exposures occur within a 5 km radius from one's place of residence, industries located within this distance from high incidence FSAs were marked as potentially relevant[70,105]. Proximity buffers in ArcGIS were used to determine the number of manufacturing industries and mining sites which fell within a 5 km area of the FSA of interest. This value was divided by the land area, obtained from the buffer function, to determine industry density.

Chapter 3: Epidemiology of SSc in Quebec, Canada

Preamble to Manuscript 1

There have been a limited number of studies comprehensively documenting the epidemiology of SSc, especially since the development of the VEDOSS criteria in 2011. The introduction of this criteria highlighted the importance of evaluating the impact on incidence, prevalence and survival. Regarding the latter, some studies have suggested an improved survival in SSc patients over the last several decades, however, conflicting evidence exists, especially when comparing the differences between males and females. In terms of geographic distribution, several studies in Europe and only two cohort studies in North America outlined an uneven distribution.

To better understand these trends, we used populational data from Quebec, Canada to determine age-standardized, as well as age and sex specific trends in incidence, prevalence, mortality and geographic distribution over a 24-year period (1996-2019).

In this study, we observed increasing trends in incidence and prevalence over time. Surprisingly, the greatest increase in incidence was among children. These increasing trends emphasize the importance of awareness and ongoing surveillance. Mortality was still higher in the SSc group compared to the general population, but declining trends were noted over the study period. This provides a hopeful outlook and suggests that early detection and improved treatment strategies may lead to reduced mortality. The analysis of geographic distribution identified a non-uniform geographic distribution of SSc incidence in Quebec, supporting the possible role of exogenous agents (*e.g.* environmental, occupational) influencing disease risk. This comprehensive epidemiological analysis acts as a foundation for designing future studies, increasing disease awareness, and counselling regarding preventative measures and resource allocation in relation to SSc.

This study was published in Lancet Regional Health Americas.

Muntyanu A, Aw K, Kaouache M, Rahme E, Osman M, Baron M, Ghazal S, Netchiporouk E. Epidemiology of systemic sclerosis in Quebec, Canada: a population-based study. *Lancet Reg Health Am. 2024 Jun 8;35:100790. doi: 10.1016/j.lana.2024.100790*

This manuscript received attention from McGill Newsroom, Institutional Communications. https://www.mcgill.ca/newsroom/channels/news/systemic-sclerosis-rise-quebec-especially-children-study-finds-357927

Epidemiology of Systemic Sclerosis in Quebec, Canada: a population-based study

Anastasiya Muntyanu[a,b], Katherine Aw[c], Mohammed Kaouache[d], Elham Rahme[e,f], Mohamed Osman[g], Murray Baron[h], Stephanie Ghazal[i], Elena Netchiporouk[i,*]

[a] Division of Experimental Medicine, McGill University Health Centre, Montreal, Quebec, Canada

[b] Division of Dermatology, University of Toronto, Toronto, Ontario, Canada

[c] Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada

[d] The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada

[e] Centre for Outcomes Research and Evaluation, Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada

[f] Division of Clinical Epidemiology, Department of Medicine, McGill University Health Centre, Montreal, Quebec, Canada

[g] Division of Rheumatology, Department of Medicine, University of Alberta, Edmonton, Alberta, Canada

[h] Division of Rheumatology, Department of Medicine, Jewish General Hospital, Montreal, Quebec, Canada.

[i] Division of Dermatology, Department of Medicine, McGill University Health Centre, Montreal, Quebec, Canada.

***Corresponding Author:** Dr. Elena Netchiporouk; elena.netchiporouk@mcgill.ca Division of Dermatology, Department of Medicine, McGill University Health Centre, 1650 Cedar Ave, Room L8 210, Montreal, Quebec, H3G 1A4, Canada

Phone: (514) 934-1934 ext: 34999

Summary

Background*:* Systemic sclerosis (SSc) is a systemic life-threatening autoimmune rheumatic disease. We aimed to assess the incidence, prevalence, mortality and spatiotemporal trends of SSc in Quebec, Canada with stratification by sex and age.

Methods: SSc cases were identified from Quebec populational databases from 1989 to 2019. Negative Binomial (NB) Generalized Linear Models were used for age-standardized incidence (ASIR) analyses and NB random walk for prevalence and mortality. A Poisson Besag-York-Mollié regression model was used for spatial analysis.

Findings: 8,180 incident SSc cases were identified between 1996 and 2019 with an average age of 57.3±16.3 years. The overall ASIR was 4.14/100,000 person-years (95%, Confidence Interval (CI) 4.05-4.24) with a 4:1 female predominance. ASIR increased steadily over time with an Average Annual Percent Change (AAPC) of 3.99% (95% CI 3.50-4.49). While the highest incidence rates were in those aged 60-79 years old among females and >80 years old among males, the highest AAPC (~10%) was seen in children. ASIRs varied geographically with standardized incidence ratios of 0.52-1.64. The average prevalence was 28.96/100,000 persons (95% CI28.72-29.20). The Standard Mortality Ratio (SMR) decreased from 4.18 (95% CI 3.64-4.76) in 1996 to 2.69 (95% CI 2.42-2.98) in 2019. Females had a greater SMR until 2007 and males thereafter. The highest SMR was in children and young adults [31.2 (95% CI 8.39-79.82) in the 0–19-year age group].

Interpretation: We showed an increasing trend in SSc incidence and prevalence and a decline in SMR over a 25-year period in Quebec. An uneven geographic distribution of SSc incidence was demonstrated.

Funding: National Scleroderma Foundation, Canadian Dermatology Foundation/Canadian

Institutes of Health Research.

Keywords: systemic sclerosis, incidence, prevalence, mortality, epidemiology, populational

Role of the funding source

The study sponsors were not involved in the design of the study; the collection, analysis, and interpretation of data; the writing of the report; or the decision to submit the paper for publication.

Research in Context

Evidence before this study

Systemic sclerosis (SSc) is a chronic, autoimmune disease characterised by fibrosis of the skin and internal organs. The diagnosis is usually established based on characteristic clinical manifestations and SSc-specific antibodies. The introduction of the American College of Rheumatology (ACR) and European League Against Rheumatism (EULAR) classification criteria in 2013 and the preliminary criteria for the very early diagnosis of SSc (VEDOSS) in 2011 further refined the diagnostic process enabling earlier identification and treatment initiation.

We searched PubMed for epidemiological studies assessing incidence, prevalence, or mortality of SSc patients for articles published before March 2023 without language restriction. We used the following search terms: ((“systemic sclerosis” OR “scleroderma”) AND (“incidence” OR “prevalence” OR “mortality” OR “epidemiology”). Additional articles from internet searches (Google) and reference searches of identified papers were included along with the authors’ own clinical knowledge.

Recent systematic reviews demonstrated higher SSc incidence estimates in North America. Few assessed trends in incidence over time, especially after 2013, when classification criteria were developed. In terms of geographic distribution, several studies in Europe and only two cohort

studies in North America outlined an uneven distribution. Hence, further assessment based on populational data in North America is warranted. Furthermore, limited studies suggested an improved survival in SSc patients in the last decades however, conflicting evidence exists regarding differences between males and females.

Added value of this study

Using populational data from Quebec, Canada, spanning over two decades (1996-2019), our study provides an in-depth analysis of the incidence, prevalence, and mortality rates of SSc, stratified by age and sex. Notably, we observed increasing trends in both incidence and prevalence, with the greatest increase in incidence over time noted among children. Mortality, while still higher in SSc patients compared to the general population (especially in younger age groups), has shown a decline during the study period, suggesting possible benefits from early detection and improved treatment strategies. Additionally, our spatial analysis demonstrated a non-uniform geographic distribution of SSc incidence in Quebec, suggesting the potential role of environmental or regional factors influencing disease risk.

Implications of all the available evidence

The increasing trends in SSc incidence and prevalence emphasize the need for heightened awareness and ongoing surveillance. The higher incidence rates in certain age groups, notably children, call for specialized care and research to understand the underlying causes. Further studies on environmental risk factors are warranted to identify possible contributing factors to the rising incidence and geographic disparities. Our findings on declining mortality provide a hopeful outlook, underscoring the potential benefits of early diagnosis and intervention. This comprehensive epidemiological analysis serves as a foundation for future research, disease awareness, and healthcare planning related to SSc.

Introduction:

Systemic sclerosis (SSc) is a life-threatening fibrosing Systemic Autoimmune Rheumatic Disease (SARD).[1] SSc is usually diagnosed clinically in patients with skin thickening proximal to the metacarpophalangeal joints and the presence of SSc-related abnormalities (*e.g.* Raynaud's phenomenon, SSc-specific antibodies, abnormal nailfold capillaroscopy).[1] Several advancements made in the last decade, such as the use of nailfold capillaroscopy and SSc-specific autoantibodies, have enabled an earlier diagnosis of SSc.[1] This led to the development of the new SSc classification criteria by the American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR) in 2013 as well as the preliminary criteria for the very early diagnosis of SSc (VEDOSS) in 2011.[1,2] VEDOSS criteria aim to diagnose SSc before fibrosis onset, enabling early intervention to reduce disease damage and enhance patient outcomes.[3]

Recent systematic reviews on SSc epidemiology highlighted growing incidence and prevalence trends worldwide, especially in North/South Americas and Oceania.[4,5] However, data post-2013 (EULAR classification criteria) and spatial analyses are scarce and hence, more recent and demographically diverse data are necessary. Furthermore, limited studies suggested an improved survival in SSc patients in the last decades, however additional, as well as age- and sex-specific data are needed.[6]

This study used data from 1989-2020 in Quebec, Canada, to delve into the incidence, prevalence, and mortality rates of SSc, breaking it down by sex and age. Spatial analyses were conducted to study geographic variability in incidence rates on a jurisdictional level.

Methods:

Study design and data reporting were performed in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist.[7]

Data Sources

Quebec, Canada, with a population of approximately 8.5 million, provides its citizens with universal health services. The health data is captured in provincial databases which are linked by a unique patient's identifier. Specifically, the *Fichier d'inscription des personnes assurées* (FIPA) includes age and sex (assigned at birth) of the registered individual, physician billing codes (based on the International Classification of Diseases 9th and 10th revision, ICD-9 and ICD-10 codes) are recorded in the *Régie de l'Assurance Maladie du Québec* (RAMQ), hospitalization data in the *Maintenance et Exploitation des Données pour l'Étude de la Clientèle Hospitalière* (MED-ÉCHO) and mortality statistics in the Provincial Vital Statistics (PVS) database. Data on the provincial population distribution of Quebec by age and sex for all available years was sourced from RAMQ database (***Supplementary Table* 1**).

Identification of Patients with Systemic Sclerosis

The following, validated, case identification algorithm was used: ≥2 ICD-9 (710.1) and/or ICD-10 (M34) diagnostic codes for SSc at least 2 months apart and within 2 years; or ≥1 billing code (ICD-9 or ICD-10) for SSc in the RAMQ database by a specialist (rheumatologist); or ≥1 hospitalization with a primary or secondary diagnosis for SSc in MED-ECHO (ICD-9 or ICD-10).[8] Morphea (*i.e.* localized scleroderma) codes were excluded to avoid misclassification.

Statistical Analysis

Incidence

A 7-year washout was applied to remove prevalent cases. Crude incidence rate trends over time and stratified by age (0-19, 20-39, 40-59, 60-79, ≥80) and sex were estimated using Negative Binomial Generalized Linear Models (NBGLM) as it accommodates for data overdispersion.[9] These age groups were selected for consistency as there would be too few cases of pediatric SSc

if smaller age intervals were used. Annual sex-specific age-standardized incidence rates (ASIR) from 1996 to 2019 were calculated using the provincial population distribution in the corresponding years as the standard population. Trends in sex specific ASIR and Average Annual Percent Change (AAPC) were assessed using NBGLM.[9]

Prevalence

To identify prevalent cases, a 7-year lookback window was used. Similarly, trends in crude prevalence rates from 1996 to 2019 were estimated using flexible negative binomial random walk (NBRW) models for the entire population and by sex. This model had a better fit compared to the NBGLM.

Mortality

Trends in sex-stratified annual age-specific Standardized Mortality Ratios (SMR) from 1996 to 2019 were estimated using the flexible NBRW model with the general Quebec population as the standard population (denominator, for the complete list refer to ***Supplementary Table* 1**). The models only controlled for age and sex.

Geographic Distribution

Age and sex Standardized Incidence Ratios (SIR) were computed for each Forward Sortation Area (FSA, a 3-digit of a postal code) from 1996 to 2019. SIR represents the observed ASIR divided by the expected ASIR if the incidence rate of SSc in the FSA was not different from that of the general population during the study period (1996-2019). To protect patients' confidentiality, as per RAMQ/MED-ECHO rules, FSAs with fewer than a total of 5 residents of a certain age group and certain year were excluded from the analysis. The SIR per FSA was modeled using a Poisson Besag-York-Mollié (BYM) regression model with a spatially correlated random effect and smoothing.[10] Maps showing the geographic distribution of the SIR over the study period

were created using ArcMap 10.1.

Software and Statistical Models

Negative binomial GLM, random walk models and the BYM spatial model were fitted using the Integrated Nested Laplace Approximation (INLA) implemented in the R-INLA package. INLA is a method for approximate Bayesian inference that represents an efficient alternative to other Markov chain Monte Carlo methods. For all model parameters and hyperparameters, we used the default non-informative prior distributions in R-INLA. SMRs over the full study period were estimated using the epi.smr function in the epiR R package and confidence intervals for SMR were computed using the default Byar's approximation method. Additional details regarding the statistical models used can be found in the ***Supplementary Material***.

Results

Incidence

In total, 8,180 individuals received a new diagnosis of SSc from 1996 to 2019. Most (80.3%; 6,565/8180) were females. The median age (interquartile range, IQR) at diagnosis was similar for females [58.0 (22.0)] and males [59.0 (22.0)]. Of the study individuals, 2.3% were <20 years, 11.6% (20-39 years), 39.1% (40-59 years), 40.1% (60-79 years) and 7.0% (≥80 years).
The ASIR over the study period was 4.14/100,000 person-years (PYs) (95% Confidence Interval, CI 4.05-4.24) and was higher among females 6.61/100,000 PYs (95% CI 6.45-6.78) compared to males 1.63/100,000 PYs (95% CI 1.55-1.72). Among females, ASIR increased by more than 2.4-fold from 3.62/100,000 PYs (95% CI 3.38-3.88) in 1996 to 8.90/100,000 PYs (95% CI 8.40-9.45) in 2019 **(Figure 1)**. The AAPC for ASIR was 3.99% (95% CI 3.50-4.49) over the study period. Among males, a similar increasing trend was observed in ASIR from 0.87/100,000 PYs (95% CI 0.76-1.00) in 1996 to 2.06/100,000 PYs (95% CI 1.85-2.30) in 2019 and a similar AAPC of 3.98%

(95% CI 2.87-4.71). Additionally, ASIR F:M ratio increased over time (***Supplementary Figure* 1**).

A steady increase in the incidence rate of SSc over time was also observed for each age group for both sexes (**Figure 2**). The highest incidence was consistently observed in females ages 60-79 years old with the highest rate of 19.85/100,000 PYs (95% CI 17.87-22.01) in 2019 (**Figure 2**). The AAPC in this group was 3.3% (95% CI 2.41-4.17). However, the largest AAPC of 9.43% (95% CI 6.29-12.74) was observed among children (0-19 years old). For males, the greatest incidence rate was observed among the ≥ 80 years old group reaching the highest rate in 2019 with an incidence rate of 6.39/100,000 PYs (95% CI 4.46-8.86) and an average AAPC in this group of 4.49% (95% CI 1.02-8.04). The largest AAPC in males was also seen in children (0-19 years old) with an AAPC of 6.62% (95% CI 1.91-11.57).

Prevalence

The average prevalence over the study period was 28.96 per 100,000 persons (95% CI 28.72-29.20). Higher average prevalence was noted in females with 47.79 (95% CI 47.36-48.23) *vs.* 9.82 (95% CI 9.63-10.03) per 100,000 persons in males. A steady increase in prevalence of SSc was noted for both sexes with rates varying from 23.26/100,000 (95% CI 22.20-24.39) in 1996 to 81.59/100,000 persons (95% CI 79.15-84.09) in 2019 in females **(Figure 3)** and 4.96/100,000 (95% CI 4.53-5.42) in 1996 to 16.93/100,000 persons (95% CI 15.99-17.90) in 2019 in males. The AAPC from 1997-2019 was 5.6% for females (95% CI 5.36-5.88) and 5.5% for males (95% CI 5.04-5.97).

The prevalence in females increased at a faster rate compared to males until 2008 **(*Supplementary Figure* 1)**, at a similar rate from 2009 to 2016, and at a lower rate thereafter. The highest prevalence for both sexes combined was observed in the 60-79-year-old group (61.91/100,000 persons; 95% CI 61.07-62.75) (**Figure 4**). The same age group also had the highest

prevalence for females (98.68/100,000 persons; 95% CI 97.22-100.14) and males (21.01/100,000 persons; 95% CI 20.31-21.73).

Mortality

There were 2,190 deaths reported in the cohort (26.8%). The median age at death (IQR) was 72.0 (17.0) [70.0 (18.0) for males and 73.0 (17.0) for females]. The overall SMR over the time period was 3.31 (95% CI 3.18-3.45) with a decrease from 4.18 (95% CI 3.64-4.76) in 1996 to 2.69 (95% CI 2.42-2.98) in 2019. The SMR for females over the study period was 3.29 (95% CI 3.14-3.45) with a peak SMR of 4.21 (95% CI 3.61-4.87) in 1996 and a decreasing trend over time to an SMR of 2.63 (95% CI 2.34-2.95) in 2019. The SMR for males was 3.40 (95% CI 3.10-3.71) with a peak at 4.14 (95% CI 3.20-5.25) in 1996 with reducing trends over time reaching an SMR of 2.94 (95% CI 2.41-3.55) in 2019 (**Figure 5**). While initially females had higher SMR than males, this trend reversed in 2007 and until the end of the study, males had higher SMR compared to females. SMR was higher for younger age groups with the highest SMR of 31.2 (95% CI 8.39-79.82) in the 0–19-year age group. Distribution by age for males, females, and combined is shown in **Figure 6**.

Geographic Distribution

There were 401 FSAs in Quebec included in the study timeline. Spatial analysis revealed an uneven geographic distribution of SIRs over Quebec **(Figure 7a,b)**. Higher SIR areas are observed both within the Greater Montreal Area (urban) as well as in more rural areas including G0K, G0J, G0E (**Figure 7b**). Several high SIR FSAs were also found clustered in the southern part of Quebec. The overall SIRs with spatial modelling spanned generally between 0.52 to 1.64. The top 10 and lowest 10 FSAs are shown in ***Supplementary Table* 2**.

Discussion

Using populational data for the largest province in Canada by land area, Quebec, we demonstrated increasing incidence and prevalence rates, overall and by age and sex, of SSc from 1996 to 2019. However, SMR decreased steadily over time, indicating improved survival likely due to earlier detection and better treatments. There was an uneven geographic distribution observed for the SIR in the province, the factors for which are not yet known and warrant further investigation.

We used a validated case definition with an estimated sensitivity of 80.5% and specificity was 94.9%.[8] Similarly, other studies assessing concordance of incidence estimates based on 2013 ACR criteria *vs.* ICD codes revealed similar results and a study from Denmark revealed a positive predictive value of 94% compared to ACR/EULAR 2013 criteria as reference.[5,11]

Over the study period (1996-2019), the observed ASIR was 4.14/100,000 PYs (95% CI 4.05-4.24) with 4-fold greater incidence in females consistent with the annual incidence observed in North America (1.4–5.6/100,000 PYs),[12] but higher than pooled global estimates of 1.4/100,000 PYs.[5] ASIR of SSc gradually increased in our study between 1996-2019 with the peak ASIR reaching 8.90/100,000 PYs for females and 2.06/100,000 PYs for males in 2019 with ~4% annual increase. The increasing incidence over time has also been reported in other studies focusing on SSc or autoimmunity in general.[5] A US study demonstrated that the prevalence of antinuclear antibody positivity (a hallmark of autoimmunity) increased from 11.0% (years 1988-1991) to 11.5% (1999-2004) to 15.9% (2011-2012) despite adjusting for multiple confounders and accounting for improved detection rates.[13] A study of over 22 million individuals with the 19 most common autoimmune diseases showed that ASIR of any autoimmune diseases increased over time (incidence rate ratio in 2017/19 *vs.* 2000/02 of 1.04; 95% CI 1.00–1.09) and that 10.2% of the

population had an autoimmune disease.[14] The hypothesized causes for this included alterations in our foods, xenobiotics, air pollution, infections, personal lifestyles, stress, and climate change.[15] Although the increase in incidence identified in our study could also in part be attributed to better diagnostic criteria and awareness of the disease, and hence more frequent diagnosis, in the context of rising autoimmune disease rates worldwide, it is likely that there is a true increase in SSc incidence.

In our study, the average age of SSc diagnosis was 57 years for both males and females. Among age groups, the highest incidence occurred in 60-79 years-old females and >80 years-old males. While there is very limited age-specific incidence data, in 2003, Mayes *et al.* similarly reported a peak SSc incidence in 65-74-year-old White females and 75-84-year-old White males.[16] A study in the UK found that 55-69 years-old had the highest crude incidence rate 1994-2013.[17] Previous studies have shown a higher incidence of SSc in African American individuals compared to Caucasians and higher association with diffuse cutaneous SSc and more severe disease manifestations/increased mortality.[18,19] The populational data used in our study, does not contain information on race or ethnicity as there is no standard reporting required and hence, we were not able to control for this factor in the analysis.

Children represented only 2.3% of all incident SSc cases in our study with an overall incidence of 0.43 per 100,000 PYs. While this is slightly higher than the previously reported incidence rates for pediatric SSc ranging from 0.03 to 0.29 per 100,000 children-year, all of the previous studies in children were conducted prior to 2016.[20,21] We showed that the incidence of pediatric SSc had the highest AAPC of almost 10% which could account for this difference. To our knowledge, this increase in incidence hasn't been previously reported in children and needs to be investigated further in future studies and other populations.

The average prevalence in our study was 28.96/100,000 persons consistent with the reported rates in North America (13.5–44.3/100,000).[5,12] In fact, in North America, high prevalence estimates were seen in the majority of studies, despite considerable methodological variations among them, which indicates the occurrence of SSc is amongst the highest in the world.[5] Globally the data is heterogeneous with broad ranges of prevalence between 3.1-144.5/100,000 individuals with a pooled prevalence of 17.6/100,000 (95% CI 15.1-20.5).[5] An increasing prevalence over the study period is likely attributed to increasing incidence (as above) as well as improved survival over time.[22,23]

In our study, separation by sex revealed a prevalence rate of 58.7/100,000 in females and 12.2/100,000 persons in males (4.8-fold difference) consistent with previous research.[5] The highest prevalence was observed in the 60-79 year old group for both males and females. This is similar to other studies that found the highest prevalence in the 70-84 year old group.[17] Another study, however, found the highest prevalence in the 51-60-year-old group with 61-70 closely following.[24]

The SMR reports the relative risk of death in patients with SSc compared to the general population accounting for age and sex. We observed a declining trend in SMR for both males and females over the study period and a decreasing risk towards older age groups. Mortality in females also appeared to decrease at a faster rate and hence by the end of the study period (2019) was lower compared to males. This corresponds with findings from a population-based study in the US over a 48-year period where initially, there was an increase in mortality between 1968 and 2000, followed by a decline between 2001 and 2015.[25] This trend was hypothesized to be attributed to multiple factors, including improved SSc recognition, with formalized classification criteria and autoantibody profiles being proposed in the 1980s.[16,26] The introduction of VEDOSS and a better

understanding of the connection between autoantibody profiles and prognosis, has allowed for earlier diagnosis and treatment, leading to a decline in mortality.[3] Additionally, the recognition of adverse effects associated with the use of systemic steroids is SSc (*i.e.*, precipitation of renal crisis), discovery of new therapeutic agents and management strategies and consequent practice changes, have likely contributed to the improved survival seen in more recent decades.[6] Despite this improvement, mortality in patients with SSc remains higher than in the general population.[25] Notably, in our study, since 2009, the SMR was greater in males. This is supported by the literature, which suggests that males are diagnosed later (as in our study), have more severe disease phenotype, and hence, worse outcomes with increased mortality. However, conflicting results have been published regarding mortality, where some of the rates/SMR in males and females are near identical, while in others males have increased mortality (as seen in our findings).[27] A study in Italy reported a combined SMR of 2.8 (95% CI 1.9–3.8), with an SMR of 3.8 (95% CI 2.9–5.1) in males and 2.6 (95% CI 1.8–3.6) in females.[21] A New Zealand study showed an overall SMR of 2.59 (95% CI 1.67-4.01) which was higher in males (4.17, 95% CI 1.74-10.02 *vs.* 2.30, 95% CI 1.39-3.81).[28] Additionally, a study compared SMR between the inception cohort (recruited within 4 years of SSc disease onset) *vs.* prevalence cohort (all patients irrespective of disease duration) conducted in 3 registries including Canadian, Australian, and Spanish.[27] In Canada, the SMR for the inception cohort was found to be 5.1, compared to 3.8 in the prevalent cohort. Separation by sex showed that males had higher SMR in both inception and prevalent cohort (8.6 and 5.9, respectively) compared to females (4.4 *vs.* 3.4 respectively).[27] Prior to 2009, females had a higher SMR than males. Several factors could have contributed to this higher mortality in females in the 90s and early 2000s including lower disease awareness leading to a later diagnosis and limited treatment options.

Interestingly, we observed the highest SMR in the 0-19-year age group. Given the relatively lower prevalence of SSc in children, mortality data in this specific age group are limited in the literature. However, a Danish cohort study found the highest SMR in their 5-34 year age group, with an SMR of 13 (95% CI 2.7-37).[29] Determinants of mortality in pediatric SSc need to be further studied.

We observed an uneven geographic distribution of the SIR, indicating certain FSAs have an increased ASIR compared to the national average. Our previous study using the Canadian Scleroderma Research Group identified uneven geographic distribution of prevalence across Canada with higher than expected prevalence FSAs correlating with higher industrial density and increased levels of air pollution.[30] Another study from Massachusetts, USA, similarly identified non-uniform distribution of prevalent cases, correlated with proximity to hazardous waste facilities and oil release or disposal sites.[31] Otherwise, no other studies examined the spatial epidemiology of SSc in North America to our knowledge. In Europe, higher prevalence was found in Italy, Spain and Sweden compared to France, Netherlands, and Norway[5]. In our study, a visually higher incidence was seen in Northern and Eastern Quebec. While previous studies hypothesized the reasons for observed gradients or uneven distribution, there is a lack of objective evidence for the exact contributing factors. When occupational factors have been studied, only about 30% of patients reported occupational exposures as contributory, hence assessing risk factors at a populational level is of high interest and should be addressed in the future.[32]

This study provided an update on epidemiology of SSc in Quebec, Canada based on a populational database over a 23-year period. Our results should be interpreted within the study characteristics whereas we conducted a populational study using administrative databases. As SSc case ascertainment was based on ICD 9/10 codes, this may result in misidentification or missed cases. However, previous studies have validated this approach, including Quebec data, and the

sensitivity and specificity are high.[8] Trends over time may be influenced by billing/coding practices among physicians which would not be accounted for. Based on the nature of the data, it is not possible to analyze clinical features such as subtypes of SSc, internal organ involvement, autoantibody profiles, and specific disease manifestations. Unfortunately, we did not have data on race/ethnicity for our population and we acknowledge this limitation. As this was a populational study, it was not possible to assess individual risk factors.

In conclusion, in this observational study, we identified increasing trends in incidence and prevalence of SSc in Quebec for both sexes and age groups with the highest incidence noted for females aged 60-79 years and > 80-year-old males with the steepest incidence increase seen in children. While there was a progressive decrease in SMR over time, mortality remains higher than in the general population, in particular for the younger age groups. We also report an uneven geographic distribution of the incidence which could prompt future studies on risk factors.

Contributors

AM contributed to conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing - original draft, writing - review & editing. KA contributed to data curation, writing - review & editing. MK contributed to data curation, formal analysis, investigation, methodology, review & editing. ER contributed to conceptualization, methodology, writing - review & editing. MO contributed to conceptualization, validation, writing - review & editing. MB contributed to conceptualization, funding acquisition, validation, writing - review & editing. SG contributed to methodology, writing - review & editing. EN contributed to conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing - original draft, writing - review & editing.

Data sharing statement

The data used in this study originates from the © Government of Quebec (2022). All data access and usage were in accordance with © Government of Quebec guidelines, and all necessary precautions were taken to ensure the confidentiality and privacy of the information. Data obtained is used exclusively for the purpose specified in the authorization. Any other use requires additional permission. Hence, if a data-sharing request is submitted to the study authors (EN), it will be subject to the © Government of Quebec data-sharing permission request.

Declaration of interests

The authors have no declaration of interest to disclose.

Acknowledgements

We would like to thank the National Scleroderma Foundation and Canadian Dermatology Foundation/Canadian Institutes of Health Research for funding support.

Editor's note

The Lancet Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

References

1 van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. Ann Rheum Dis 2013; 72: 1747–1755.

2 Avouac J, Fransen J, Walker UA, et al. Preliminary criteria for the very early diagnosis of systemic sclerosis: results of a Delphi Consensus Study from EULAR Scleroderma Trials and Research Group. Ann Rheum Dis 2011; 70: 476–481.

3 Bellando-Randone S, Del Galdo F, Lepri G, et al. Progression of patients with Raynaud's phenomenon to systemic sclerosis: a five-year analysis of the European Scleroderma Trial

and Research group multicentre, longitudinal registry study for Very Early Diagnosis of Systemic Sclerosis (VEDOSS). Lancet Rheumatol 2021; 3: e834–e843.

4 Zhong L, Pope M, Shen Y, et al. Prevalence and incidence of systemic sclerosis: A systematic review and meta-analysis. Int J Rheum Dis 2019; 22: 2096–2107.

5 Bairkdar M, Rossides M, Westerlind H, et al. Incidence and prevalence of systemic sclerosis globally: a comprehensive systematic review and meta-analysis. Rheumatology 2021; 60: 3121–3133.

6 Rubio-Rivas M, Royo C, Simeón CP, et al. Mortality and survival in systemic sclerosis: systematic review and meta-analysis. Semin Arthritis Rheum 2014; 44: 208–219.

7 von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007; 370: 1453–1457.

8 Bernatsky S, Linehan T, Hanly JG. The accuracy of administrative data diagnoses of systemic autoimmune rheumatic diseases. J Rheumatol 2011; 38: 1612–1616.

9 Hilbe JM. Negative Binomial Regression. Cambridge University Press, 2011.

10 Quick H, Song G, Tabb LP. Evaluating the informativeness of the Besag-York- Mollié CAR model. Spat Spatiotemporal Epidemiol 2021; 37: 100420.

11 Butt SA, Jeppesen JL, Fuchs C, et al. Trends in incidence, mortality, and causes of death associated with systemic sclerosis in Denmark between 1995 and 2015: a nationwide cohort study. BMC Rheumatol 2018; 2: 36.

12 Bergamasco A, Hartmann N, Wallace L, et al. Epidemiology of systemic sclerosis and systemic sclerosis-associated interstitial lung disease. Clin Epidemiol 2019; 11: 257–273.

13 Dinse GE, Parks CG, Weinberg CR, et al. Increasing Prevalence of Antinuclear Antibodies in the United States. Arthritis Rheumatol 2022; 74: 2032–2041.

14 Conrad N, Misra S, Verbakel JY, et al. Incidence, prevalence, and co-occurrence of autoimmune disorders over time and by age, sex, and socioeconomic status: a population-based cohort study of 22 million individuals in the UK. Lancet. Epub ahead of print 5 May 2023. DOI: 10.1016/S0140-6736(23)00457-9.

15 Miller FW. The increasing prevalence of autoimmunity and autoimmune diseases: an urgent call to action for improved understanding, diagnosis, treatment, and prevention. Curr Opin Immunol 2023; 80: 102266.

16 Mayes MD, Lacey JV Jr, Beebe-Dimmer J, et al. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. Arthritis Rheum 2003; 48: 2246–2255.

17 Royle JG, Lanyon PC, Grainge MJ, et al. The incidence, prevalence, and survival of systemic sclerosis in the UK Clinical Practice Research Datalink. Clin Rheumatol 2018; 37: 2103–2111.

18 Gelber AC, Manno RL, Shah AA, et al. Race and association with disease manifestations and mortality in scleroderma: a 20-year experience at the Johns Hopkins Scleroderma Center and review of the literature. Medicine 2013; 92: 191– 205.

19 Morgan ND, Shah AA, Mayes MD, et al. Clinical and serological features of systemic sclerosis in a multicenter African American cohort: Analysis of the genome research in African American scleroderma patients clinical database. Medicine 2017; 96: e8980.

20 Pelkonen PM, Jalanko HJ, Lantto RK, et al. Incidence of systemic connective tissue diseases in children: a nationwide prospective study in Finland. J Rheumatol 1994; 21: 2143–2146.

21 Ciaffi J, Morabito MF, Ruscitti P, et al. Incidence, prevalence and mortality of systemic sclerosis in Italy: a nationwide population-based study using administrative health data. Rheumatol Int 2021; 41: 129–137.

22 Al-Dhaher FF, Pope JE, Ouimet JM. Determinants of morbidity and mortality of systemic sclerosis in Canada. Semin Arthritis Rheum 2010; 39: 269–277.

23 Steen VD, Medsger TA. Changes in causes of death in systemic sclerosis, 1972- 2002. Ann Rheum Dis 2007; 66: 940–944.

24 Knarborg M, Hyldgaard C, Bendstrup E, et al. Incidence, prevalence and regional distribution of systemic sclerosis and related interstitial lung Disease: A nationwide retrospective cohort study. Chron Respir Dis 2022; 19: 14799731221125559.

25 Yen EY, Singh DR, Singh RR. Trends in Systemic Sclerosis Mortality Over Forty- Eight Years, 1968-2015: A US Population-Based Study. Arthritis Care Res 2021; 73: 1502–1510.

26 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: 581– 590.

27 Hao Y, Hudson M, Baron M, et al. Early Mortality in a Multinational Systemic Sclerosis Inception Cohort. Arthritis Rheumatol 2017; 69: 1067–1077.

28 Ooi C, Solanki K, Lao C, et al. Mortality in the Waikato Hospital Systemic Sclerosis Cohort. Int J Rheum Dis 2018; 21: 253–260.

29 Jacobsen S, Halberg P, Ullman S. Mortality and causes of death of 344 Danish patients with systemic sclerosis (scleroderma). Br J Rheumatol 1998; 37: 750–755.

30 Muntyanu A, Ouchene L, Zhou S, et al. Geographical distribution of systemic sclerosis in Canada: An ecologic study based on the Canadian Scleroderma Research Group. J Am Acad Dermatol. Epub ahead of print 11 January 2022. DOI: 10.1016/j.jaad.2021.12.055.

31 Kassamali B, Kassamali AA, Muntyanu A, et al. Geographic distribution and environmental triggers of systemic sclerosis cases from 2 large academic tertiary centers in Massachusetts. J Am Acad Dermatol 2022; 86: 925–927.

32 Muntyanu A, Milan R, Rahme E, et al. Organic solvent exposure and systemic sclerosis: A retrospective cohort study based on the Canadian Scleroderma Research Group registry. J Am Acad Dermatol. Epub ahead of print 13 May 2023. DOI: 10.1016/j.jaad.2023.04.062.

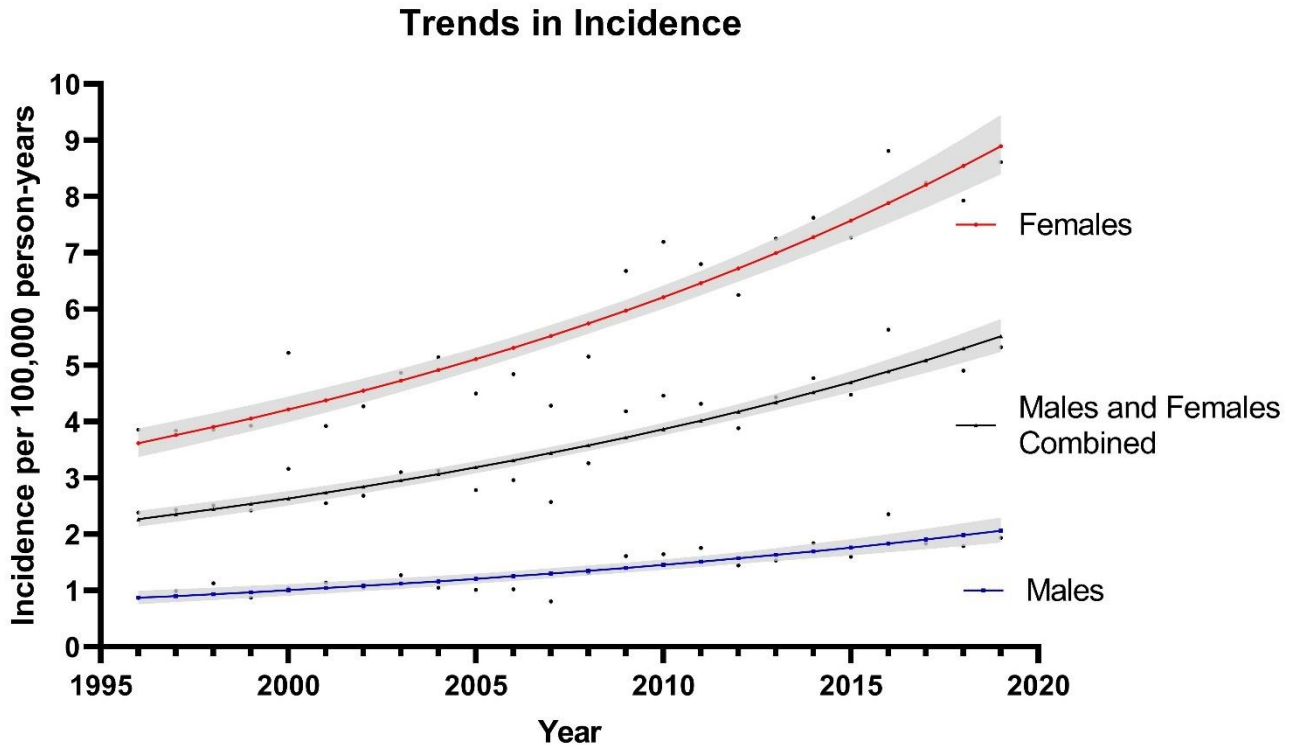


Figure 1. Age-standardized incidence rate (ASIR) over time (1996-2019) for females, males, and both. The grey shading depicts the 95% confidence intervals for the annual point estimates from the model. The points around the curve illustrate the true observed values for each year.

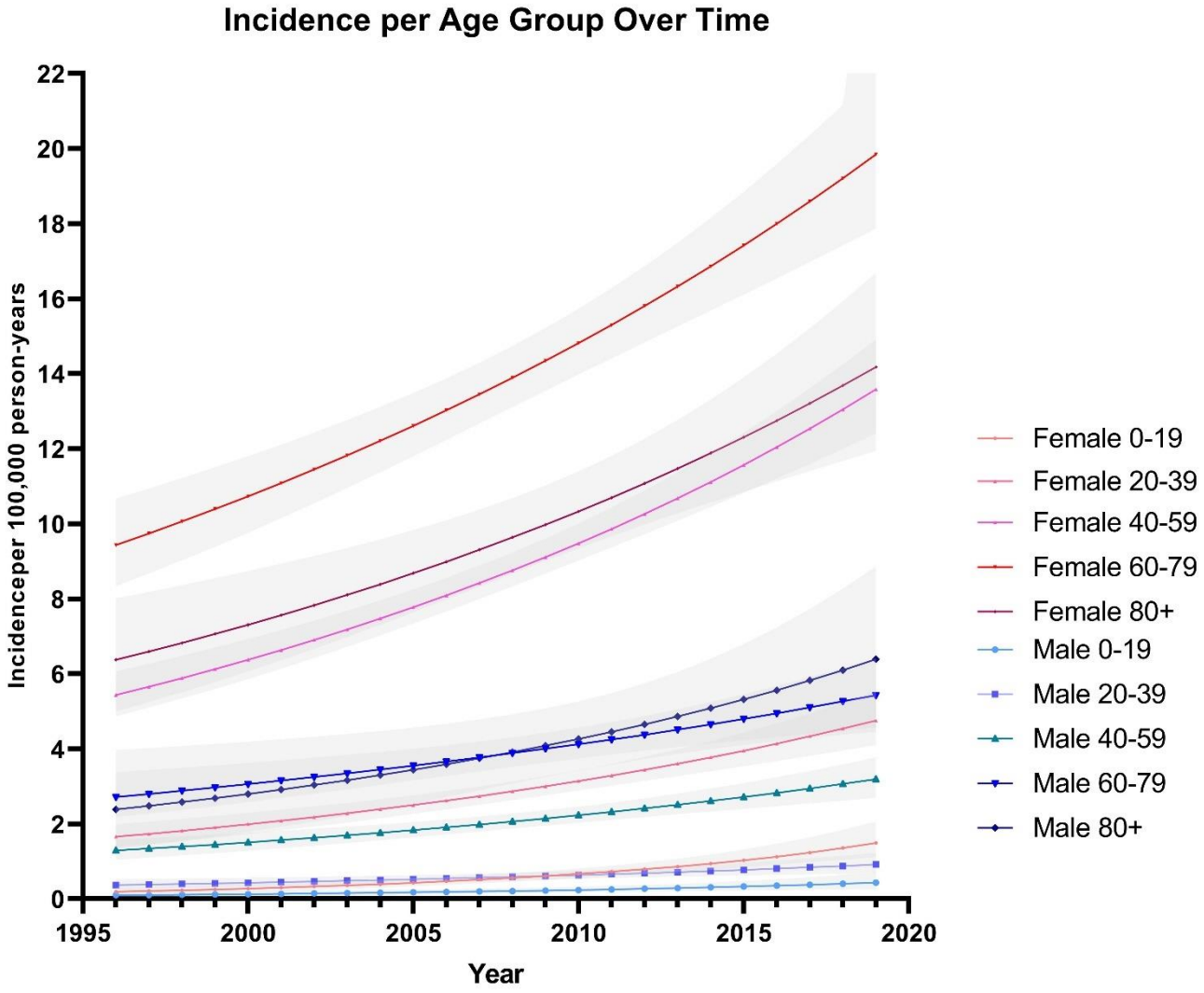


Figure 2. Crude incidence rate for females over time (1996-2019) per age group (0-19, 20-39, 40-59, 60-79, 80+) for males and females. Grey shading depicts the 95% confidence intervals for the annual point estimates from the model.

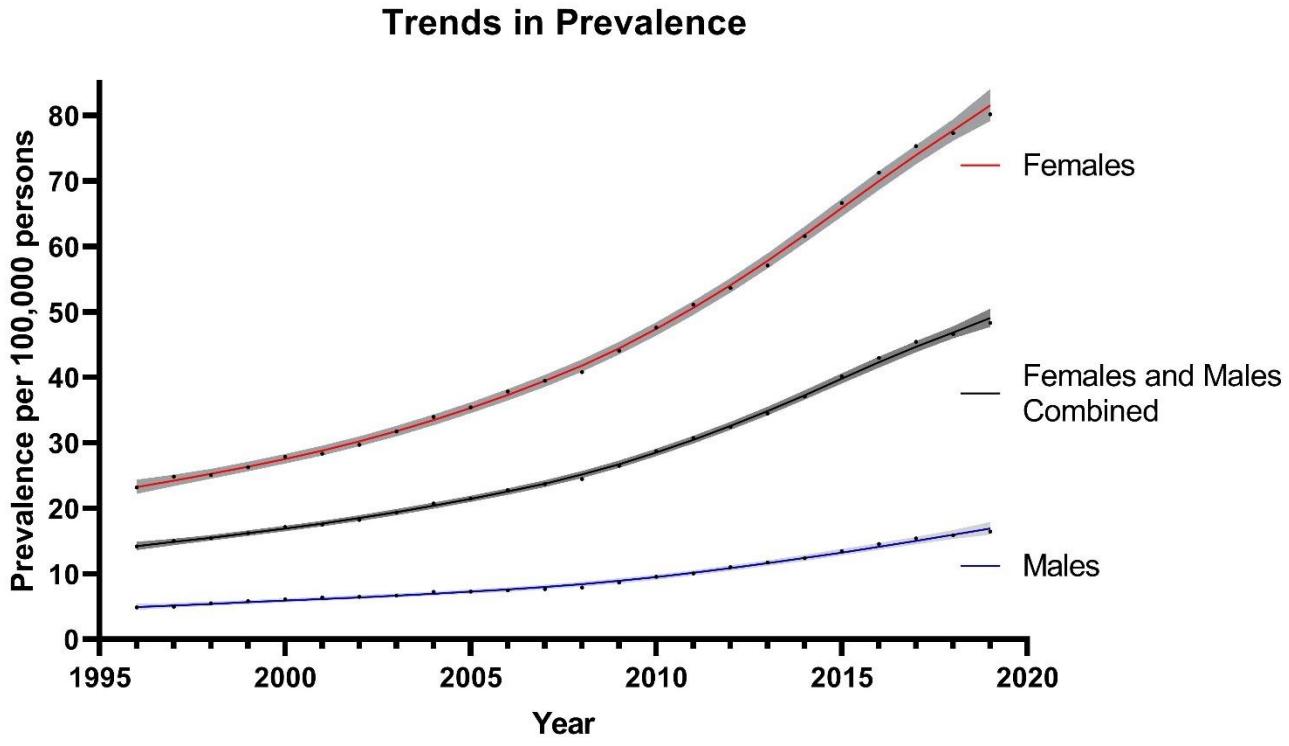


Figure 3. Prevalence rates over time (1996-2019) for females and males. The peak prevalence for females was 81.59/100,000 persons with a 5.6% (95% Confidence Interval, CI 5.36-5.88) average annual percent increase compared to 16.93/100,000 persons for males with a 5.5% (95% CI 5.04-5.97). average annual percent increase. Grey shading depicts the 95% CI for the annual point estimates from the model. The points around the curve illustrate the true observed values for each year.

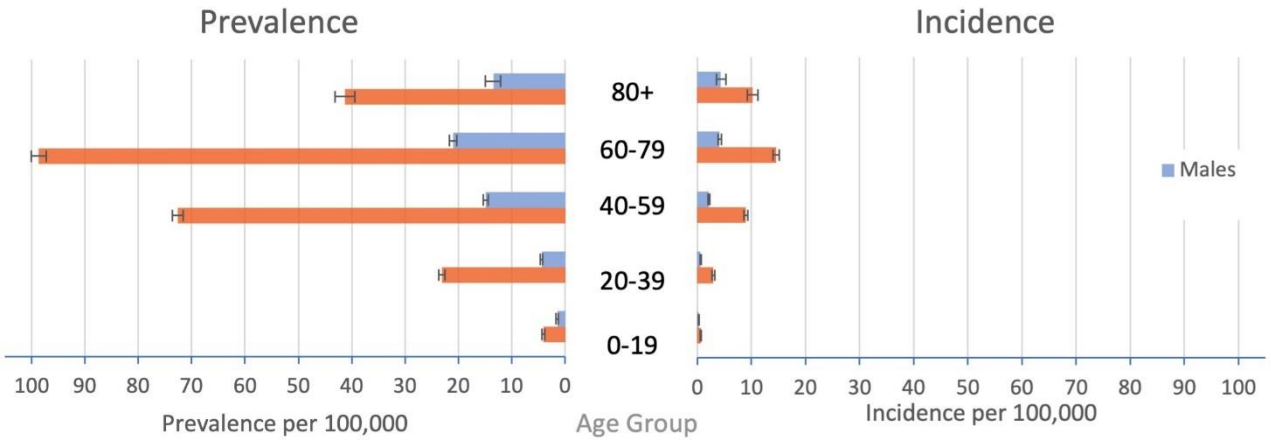


Figure 4. Average prevalence and incidence of systemic sclerosis per 100,000 persons and person-years, respectively, between 1996 and 2019 stratified by sex and age. The lines for each bar represent the 95% confidence intervals.

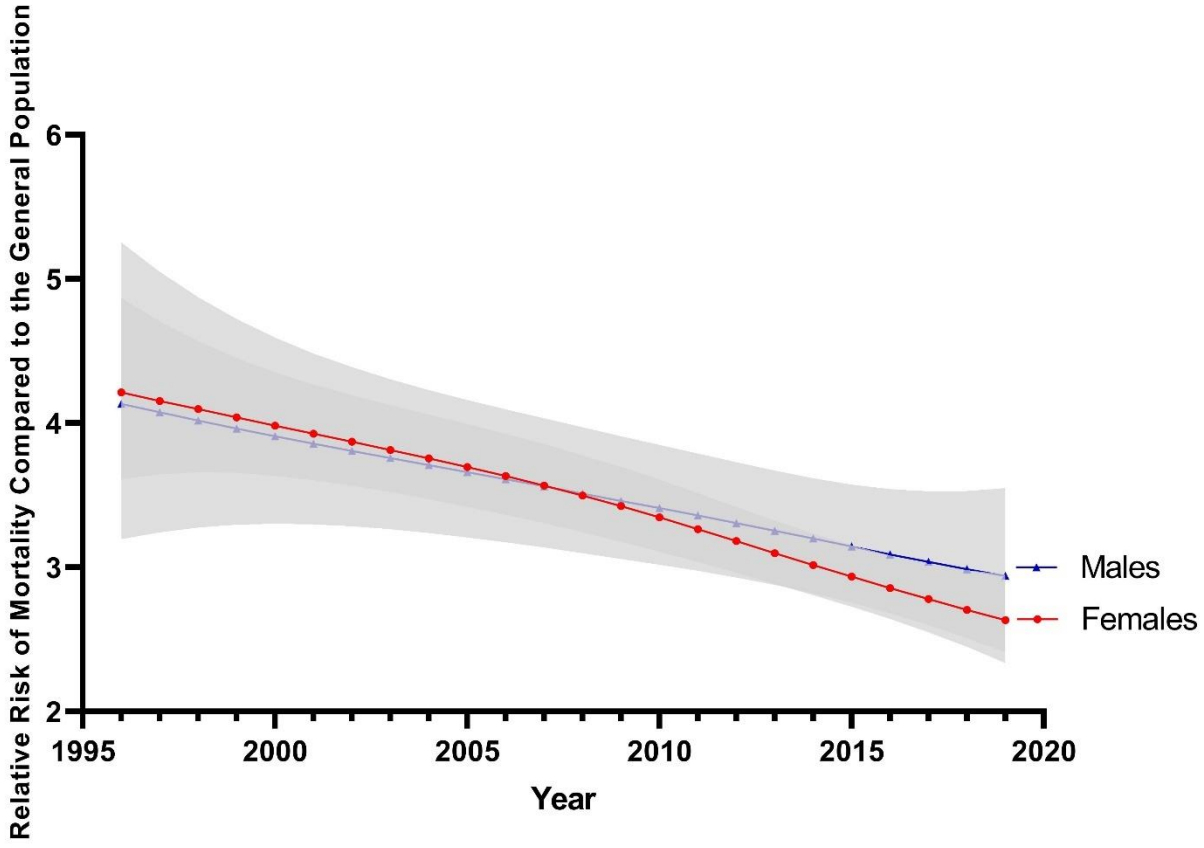


Figure 5. Trends in standardized mortality ratio (SMR), standardized by age and sex, over time (1996-2019) for females and males. Grey shading depicts the 95% confidence intervals.

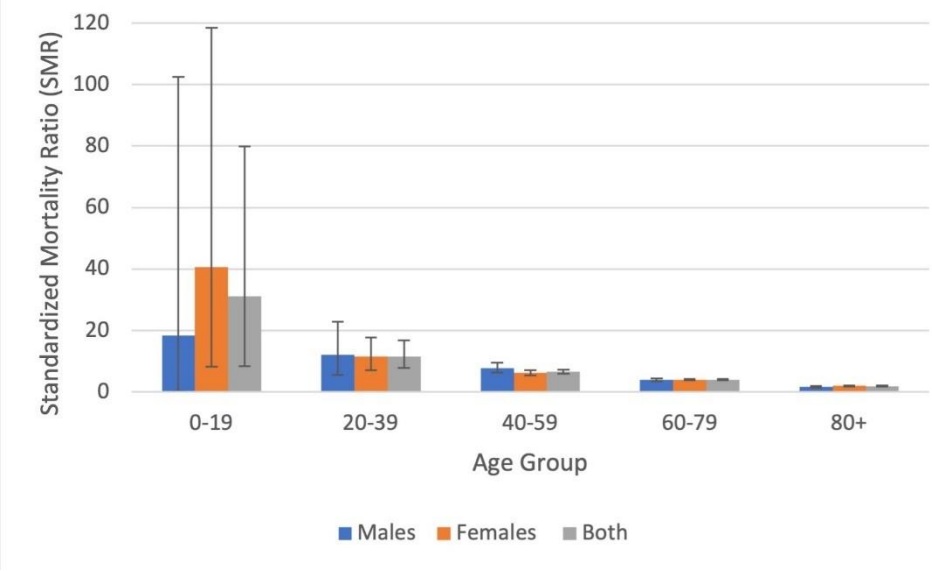


Figure 6. Standardized Mortality Ratio (SMR) over the study period per age group and sex.

The lines for each bar represent the 95% confidence interval.

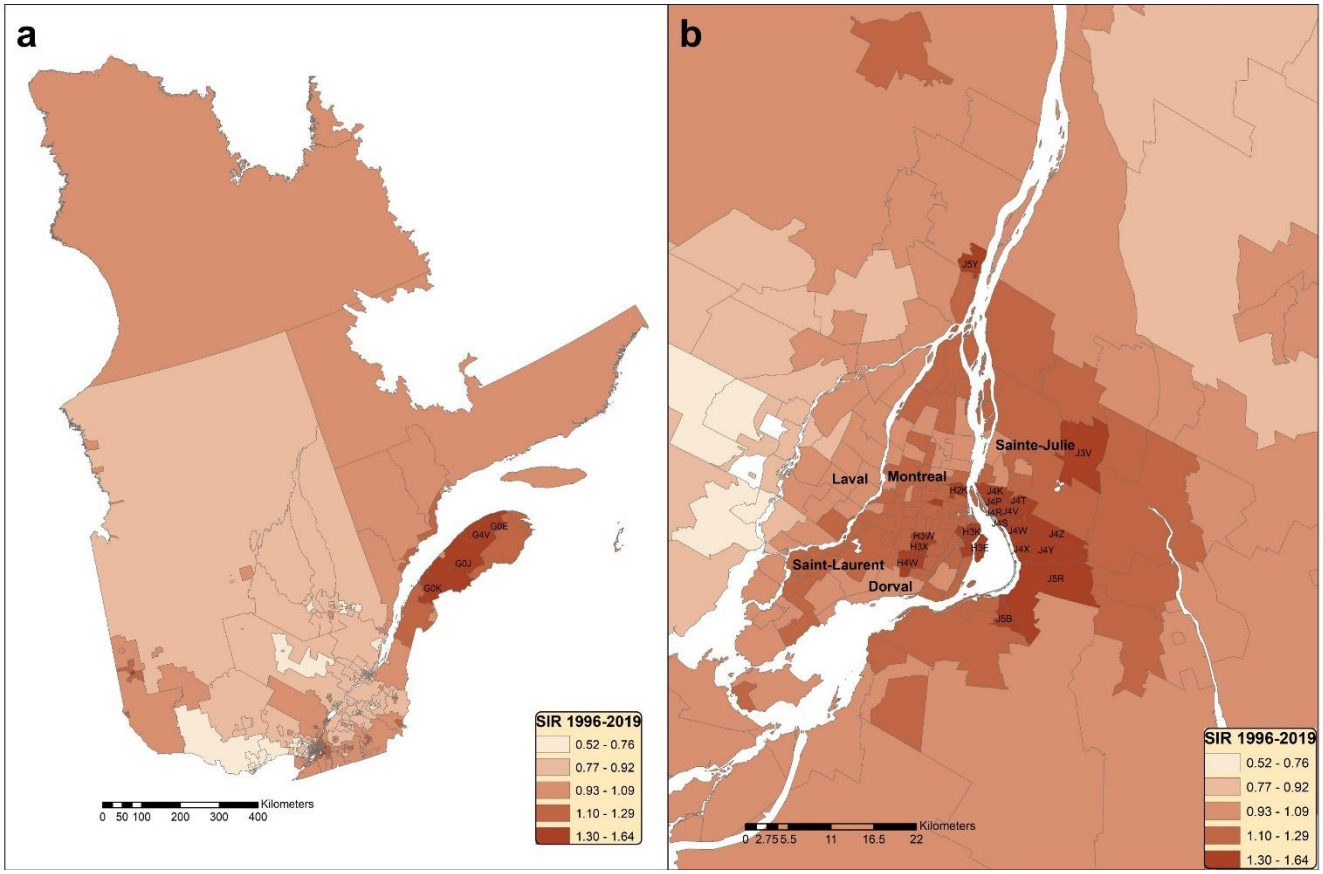


Figure 7. (a) Geographic Distribution of SIR in Quebec 1996-2019. (b) Geographic Distribution of SIR in the greater Montreal area over the study period (1996-2019).

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No.	Recommendation	Page No.	Relevant text from manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2	Population-based observational study as stated in the title and defined in the abstract.
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2	See abstract for key details of analyses and main findings.
Introduction				
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3	Paragraph 2 in introduction highlights existing studies and limitations in the literature; mainly that there are few epidemiological studies in North America and that age-specific and sex-specific data on SSc are needed.
Objectives	3	State specific objectives, including any prespecified hypotheses	3	Objectives listed in paragraph 3 of introduction describes that we aim to investigate the trends in the incidence, prevalence, and mortality of SSc in Quebec, Canada and perform detailed analyses stratified by sex, age groups and geographic regions.
Methods				
Study design	4	Present key elements of study design early in the paper	3	Study design is described in the second paragraph of the Methods.
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	3-4	The setting of Quebec, Canada and study timeline of 1989-2019 are stated in the second paragraph of the Methods. The databases for data collection were highlighted in "Data Sources" in the Methods. The only follow-up data collected was Standardized Mortality Ratios (SMR), which were estimated using the flexible negative binomial random walk model.
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	3-4	Cases were defined using a validated ICD-9 and ICD-10 case identification algorithm, as defined in the

				"Identification of Patients with Systemic Sclerosis" section in the Methods.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	3-4	Outcomes were defined as Incidence, Prevalence, Mortality, and Geographic Distribution, each with their own section in the Methods section. Diagnostic criteria for SSc were defined in "Identification of Patients with Systemic Sclerosis" in the Methods. Localized scleroderma cases were excluded to avoid misclassification of patients.
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	3-4	"Data Sources" in the Methods section defines sources of data. Diagnoses of SSc for incidence and prevalence were identified from Régie de L'Assurance Maladie du Québec (RAMQ). Hospitalization data was obtained from Maintenance et Exploitation des Données pour l'Étude de la Clientèle Hospitalière (MED-ÉCHO) and mortality statistics in the Provincial Vital Statistics (PVS) database. Each variable of interest (incidence, prevalence, mortality, and geographic distribution) under "Statistical Analysis" describes specific methods of assessment.
Bias	9	Describe any efforts to address potential sources of bias	4	Morphea (localized scleroderma) codes were excluded to avoid misclassification of SSc cases. Age- and sex-standardization was also performed to address the influence of these potential confounding variables.
Study size	10	Explain how the study size was arrived at	3-4	Methods of identifying incidence of SSc were described in the "Data Sources," "Identification of Patients with Systemic Sclerosis," and "Incidence" sections of Methods.

Continued on next page

Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	4	The Methods "Statistical Analysis" section describes that incidence groupings were made by sex, age, and Forward Sortation Area (FSA), prevalence by sex, and mortality by age and sex.
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	4	Described in the "Statistical Analysis" section of the methods.
		(*b*) Describe any methods used to examine subgroups and interactions	4	The "Statistical Analysis" section describes all analyses conducted (incidence, prevalence, mortality, and geographic distribution), including subgroups and which models were used for analysis. For example, Negative Binomial (NB) Generalized Linear Models were used for age standardized incidence (ASIR) analyses, NB random walk for prevalence and mortality.A Poisson Besag-York-Mollié regression model was used for spatial analysis.
		(*c*) Explain how missing data were addressed	n/a	n/a
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	n/a	n/a
		(*e*) Describe any sensitivity analyses	n/a	n/a
Results				
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	4	Results section identifies the number of individuals identified from Quebec populational databases (In total, 8,180 individuals received a new diagnosis of SSc from 1996 to 2019).
		(b) Give reasons for non-participation at each stage	n/a	n/a
		(c) Consider use of a flow diagram	n/a	n/a
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	4-6	Given this is populational data, we cannot describe individual risk factors or demographics beyond age, sex, and geographic distribution, which is described in the results.
		(b) Indicate number of participants with missing data for each variable of interest	n/a	n/a
Outcome data	15*	Report numbers of outcome events or summary measures	4-6	See results section.

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	4-6	All estimates are listed with 95% CI.
		(*b*) Report category boundaries when continuous variables were categorized	4-6	Age groups are defined whenever age-specific variables are reports. Sex was also categorized male vs female.
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	5	Crude mortality in the cohort was listed with a corresponding percentage in the "Mortality" section of the result.

Continued on next page

Supplementary Material Table of Contents

Statistical appendix

Negative Binomial models.

Count data are modelled using probability distributions that put mass on nonnegative integer values such as the Poisson and the Negative Binomial (NB) distributions. A Poisson linear trend model assumes $count_t$ (the count of incident cases at year t) follows a Poisson distribution with mean equal to $pop_t * \mu_t$ with $\log(\mu_t) = \alpha + \beta * t$. Here pop_t is the size of the population under study at year t, α is a global intercept and β is the trend coefficient. If we exponentiate β and subtract 1 we obtain the annual percent change in incidence rates.

In a Poisson linear trend model, μ_t (the mean or expected value of the rate ($count_t / pop_t$)) is also the variance of the rate at year t. In other words, the mean and variance of the rate are the same in a Poisson model. A Negative Binomial trend assumes the expected value of the rate at year is equal to μ_t and its variance is equal to $\mu_t + (1/k) * \mu_t^2$. The parameter k is estimated from the data and it allows the variance of the rate to be larger than its mean.

Flexible non-parametric random walk models

Let $count_t$ of incident cases at year t follow a Poisson distribution with mean equal to $pop_t *$

μ_t and let $\gamma_t = \log(\mu_t)$. A linear trend model assumes γ_t is a linear function of t (that is $\gamma_t = \alpha + \beta * t$).

When linearity of γ_t as a function of t does not hold and linear trend models do not provide a good fit to the data, flexible nonparametric models such as random walk models can be used.

A random walk model of order 1 ('rw1') by $\log(\mu_t) = \alpha + \gamma_t$ and $\gamma_t = \gamma (t-1) + \nu_t$; ν_t 's are independent randomly distributed variables with mean 0 and standard deviation σ_ν . In other words, a rw1 assumes the log of the expected value of the rate at year t is equal to the log of the expected rate at year t-1 plus a random term.

When random walk models of order 1 result in wiggly fitted curves, one can use random walks of order 2 ('rw2') that result in smoother fitted curves. When the true disease risk is believed to vary smoothly over time, rw2 models can be used. These models assume $\gamma_t = 2*\gamma_{t-1} - \gamma_{t-2} + \nu_t$, where the ν_t 's are independent randomly distributed variables with mean 0 and standard deviation σ_ν. In other words, a rw2 model assumes the log of the expected value of the rate at time t is a linear combination of the log of the expected value of the rate at times t-1 and t-2 plus a random term.

A negative binomial random walk is defined the same way as a Poisson random walk model except for the additional over-dispersion parameter that is included in the NB random walk model to allow for the variance to be estimated from that data and to be larger than the mean.

Standardised Incidence Ratios (SIR)

The expected number of cases in FSA_i and year t, assuming FSA_i in year t behaves the same way as the total population of Quebec from 2001 to 2019 (the standard population), is obtained as follows. First the expected number in each age group j is obtained by multiplying FSA_i population (in year j) in age group j by the rate observed in age group j in the standard population. Summing up the expected number of cases over age groups results in the expected number of cases

in FSA_i in year t.

The SIR for FSA_i in year t is as calculated as $SIR_{it} = count_{it}/E_{it}$ where $count_{it}$ is the observed number of cases in FSA_i in year t and E_{it} is the expected number of cases in FSA_i in year t assuming that for each age group j the risk in FSA_i in year t is the same the risk for age group j in the standard population. $SIR_{it} > 1$ ($SIR_{it} < 1$) means the risk in FSA_i at year t is higher (lower) than the risk in the standard population.

Spatial model

SIR was estimated using a Bernardinelli model[1] defined by $count_i \sim Poisson(E_i\eta_i)$ and $\log(\eta_i) = \alpha + u_i + v_i$, where $count_i$ is the observed number of cases, E_i is the expected number of cases, and η_i is the relative risk in FSA_i. The term α is a global intercept and the part u_i+v is modelled using a Besag, York and Mollié (BYM) model[2] which is a hierarchical spatial conditional correlation (CAR) model that assumes u_i is a weighted average of neighbouring u_i 's allowing for correlation between neighbouring FSA's. The v_i's are unstructured independently distributed spatial terms following a normal distribution with mean zero and standard deviation σ_v.

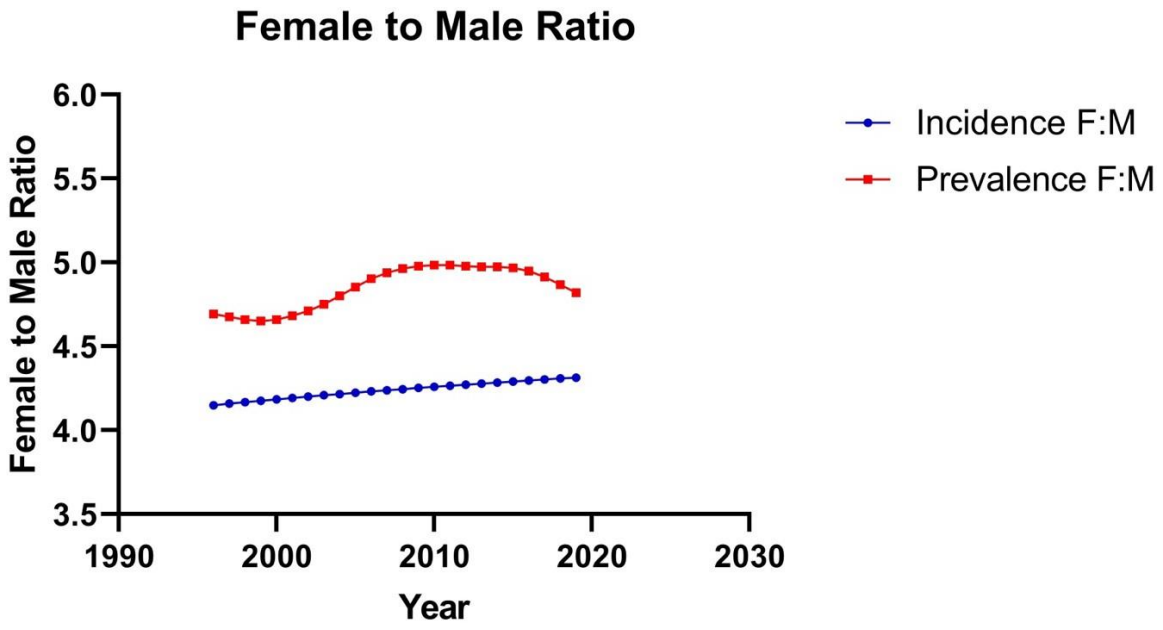
Trends in mortality risk

The expected number of deaths E_t in year t in the prevalent Systemic Sclerosis population assuming the SSC population and the general Quebec have the same annual risk of death is obtained as follows. First the expected number of deaths in each age group j is obtained by multiplying the SSC prevalent population in year t in age group j by the rate of death in the Quebec general population in year t in age group j. Summing up the expected number of deaths over age groups results in the expected number of death in the SSc prevalent population in year t. The Standardised Mortality Ratio (SMR) or relative risk of death in year t is defined as d_t/E_t. We used

a rw2 model for the trend in the relative risk of death in the SSC prevalent population that assumes $d_t \sim \text{Poisson}(E_t\theta_t)$ and $\log(\theta_t)=\alpha+u_t$; $u_t = 2*u_{t-1} - u_{t-2} + \nu_t$, ; the ν_t 's are independent randomly distributed variables with mean 0 and standard deviation σ_u

Statistical references

1. Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, Songini M. Bayesian analysis of space-time variation in disease risk. *Stat Med*. 1995;14(21-22):2433-43. doi: 10.1002/sim.4780142112.
2. Besag J, York J, Mollié A. Bayesian image restoration with applications in spatial statistics (with discussion). *Ann Inst Stat Math.* 1991;43:1-59. Available from: https://doi.org/10.1007/BF00116466.
3. Durbin J, Koopman SJ. Time series analysis by state space methods. *2nd ed. Oxford*: Oxford University Press; 2012.
4. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *J R Stat Soc Series B Stat Methodol*. 2009;71:319-92.
5. Blangiardo M, Cameletti M, Baio G, Rue H. Spatial and spatio-temporal models with R-INLA. *Spat Spatiotemporal Epidemiol*. 2013 Dec;7:39-55. doi: 10.1016/j.sste.2013.07.003. PMID: 24377114.
6. Rothman KJ, Boice JD. Epidemiologic analysis with a programmable calculator. Washington: U.S. Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health; 1972. Report No.: 79-1649.
7. Régie de l'assurance maladie Québec. La RAMQ en quelques chiffres. Régie de l'assurance maladie Québec. [n.d.]. Available from: https://www.ramq.gouv.qc.ca/en/node/89526 [Accessed 2024 Feb 4]



Supplementary Figure 1. Female to male ratio for Incidence and Prevalence between 1996-2019.

Supplemental Table 1. Quebec population counts by sex from 1996 to 2019.

Year	Females	Males	Total
1996	3678479	3568418	7246897
1997	3689236	3585375	7274611
1998	3702454	3593481	7295935
1999	3715436	3607814	7323250
2000	3730836	3626115	7356951
2001	3748718	3647738	7396456
2002	3768314	3673342	7441656
2003	3787821	3697932	7485753
2004	3810461	3725129	7535590
2005	3831147	3750329	7581476
2006	3854013	3777953	7631966
2007	3884061	3808855	7692916
2008	3918146	3843579	7761725
2009	3956348	3887035	7843383
2010	3997469	3931753	7929222
2011	4033102	3971988	8005090
2012	4057432	4003669	8061101
2013	4078756	4032124	8110880
2014	4096528	4053655	8150183
2015	4107383	4067889	8175272
2016	4130368	4095582	8225950

2017	4165840	4136223	8302063
2018	4211487	4190251	8401738
2019	4256672	4246811	8503483

Supplemental Table 2. Top 10 highest and lowest Forward Sortation Areas (FSAs) by Standardized Incidence Rate (SIR).

Top 10 FSAs		**Confidence Interval 95%**	**Lowest 10 FSAs**		**Confidence Interval 95%**
FSA	**SIR**		**FSA**	**SIR**	
G5M	1.64	(1.09-2.40)	**J9J**	0.61	(0.42-0.85)
G5N	1.57	(1.10-2.19)	**J8P**	0.60	(0.40-0.87)
H3X	1.56	(1.21-2.00)	**J9A**	0.58	(0.41-0.80)
J4X	1.56	(1.18-2.04)	**J8Z**	0.57	(0.39-0.81)
H3E	1.55	(1.19-2.01)	**J8V**	0.57	(0.40-0.77)
G5H	1.55	(1.00-2.33)	**J9B**	0.57	(0.40-0.77)
J4Y	1.54	(1.14-2.04)	**J8X**	0.55	(0.36-0.80)
J4W	1.53	(1.18-1.96)	**J8T**	0.53	(0.39-0.71)
G5L	1.53	(1.19-1.95)	**J8R**	0.53	(0.38-0.72)
J9X	1.46	(1.02-2.03)	**J8Y**	0.52	(0.34-0.76)

Chapter 4: Geographical distribution of SSc in Canada

Preamble to Manuscript 2

In this study, a national registry of SSc patients, the CSRG, was used. This database comprises of patients recruited from multiple centers across the country and represents approximately 10% of all prevalent cases. As with the previous study (Manuscript 1, above), an uneven geographic distribution was noted, and consequently the goal of this study was to assess whether air pollution and industrial density is associated with the geographic distribution of SSc and could help explain the regional variations observed. Notably, only ~30% of SSc patients reported an occupational exposure, indicating that larger scale exposures in a geographic vicinity may be contributing to disease development.

In this study, we mapped the period prevalence of SSc cases and hot spots were observed in industrialized areas. Specifically, several clusters co-localized to Hamilton, Ontario and Montreal, Quebec. However, hot spots were also noted in rural areas which were far from recruitment centers. A significantly higher density of industries, as well as air pollution, was demonstrated in the FSAs with the highest period prevalence. We believe that this data is highly relevant, and future studies will focus on putative environmental triggers with biologic plausibility.

Initially, manuscript 2 was submitted as a full-length manuscript to the Journal of the American Academy of Dermatology (JAAD). After reviewing the manuscript, the editor of the journal proposed to publish the results as a research letter. Hence, additional information is summarized in the Supplementary Material as well as the general discussion (Chapter 8).

This study was published in the Journal of American Academy of Dermatology.

Muntyanu A, Ouchene L, Zhou S, Hudson M, Rezaeian M, LaChance A, Litvinov IV, Baron M, Netchiporouk E; Canadian Scleroderma Research Group. Geographical distribution of systemic sclerosis in Canada: An ecologic study based on the Canadian Scleroderma Research Group. *J Am Acad Dermatol. 2022 Nov;87(5):1095-1097. doi: 10.1016/j.jaad.2021.12.055. Epub 2022 Jan 11.*

We collaborated with Harvard University to conduct a similar study in Massachusetts, United States. This is further discussed in the general discussion chapter (Chapter 8)[106].

Geographical distribution of systemic sclerosis in Canada: An ecologic study based on the Canadian Scleroderma Research Group

Anastasiya Muntyanu, MD,[a] Lydia Ouchene, BSc,[b] Siriu Zhou, PhD,[c] Marie Hudson, MD, MPH,[c,d] Mohsen Rezaeian, PhD,[e] Avery LaChance, MD, MPH,[f] Ivan V. Litvinov, MD, PhD,[a] Murray Baron, MD,[d] Elena Netchiporouk, MD, MSc,[a] and The Canadian Scleroderma Research Group*

From the Division of Dermatology, Department of Medicine,[a] and the Faculty of Medicine, McGill University Health Centre, Montreal, Quebec, Canada[b]; the Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, Quebec, Canada[c]; the Division of Rheumatology, Department of Medicine, Jewish General Hospital, Montreal, Quebec, Canada[d]; the Department of Epidemiology and Biostatistics, Occupational Environment Research Center, Medical School, Rafsanjan University of Medical Sciences, Rafsanjan, Iran[e]; and the Department of Dermatology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts.[f]

*Investigators of the Canadian Scleroderma Research Group: M. Baron, Montreal, Quebec; M. Hudson, Montreal, Quebec; G. Gyger, Montreal, Quebec; J. Pope, London, Ontario; M. Larche, Hamilton, Ontario; N. Khalidi, Hamilton, Ontario; A. Masetto, Sherbrooke, Quebec; E. Sutton, Halifax, Nova Scotia; T.S. Rodriguez-Reyna, Mexico City, Mexico; N. Maltez, Ottawa, Ontario; C. Thorne, Newmarket, Ontario; P.R. Fortin, Quebec, Quebec; A. Ikic, Quebec, Quebec; D. Robinson, Winnipeg, Manitoba; N. Jones, Edmonton, Alberta; S. LeClercq, Calgary, Alberta; P. Docherty, Moncton, New Brunswick; D. Smith, Ottawa, Ontario; M. Fritzler, Mitogen Diagnostics, Cumming School of Medicine, Calgary, Alberta.

Funding sources: Supported by the Association des Médecins Specialistes Dermatologues du Quebec (AMSDQ) to Dr Muntyanu and Dr Netchiporouk. No funding bodies had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The Canadian Scleroderma Research Group (CSRG) is funded by the Canadian Institutes of Health Research (CIHR), Scleroderma Canada, the Scleroderma Society of Ontario, Sclerodermie Quebec, the Scleroderma Association of Saskatchewan, Scleroderma Manitoba, the Scleroderma Society of Nova Scotia, the Scleroderma Association of British Columbia, the Cure Scleroderma Foundation, the Canadian Blood and Marrow Transplant Group, and the Lady Davis Institute of Medical Research of the Jewish General Hospital, Montreal, Quebec. The CSRG has also received educational grants from Actelion pharmaceuticals and Mallinckrodt.

IRB approval status: Approved at the Jewish General Hospital and all participating sites.

Key words: Canada; environmental factors; epidemiology; geographic distribution; prevalence; scleroderma; systemic sclerosis.

Reprints not available from the authors.

Correspondence to: Elena Netchiporouk, MD, MSc,

Division of Dermatology, Department of Medicine,

McGill University Health Centre, 1650 Avenue Cedar, Montreal, Quebec H3G 1A4, Canada

E-mail: elena.netchiporouk@mail.mcgill.ca

Conflicts of interest

None disclosed.

To the Editor: Systemic sclerosis (SSc) is a life threatening systemic autoimmune disease thought to be induced by environmental triggers in genetically predisposed individuals. Occupational expo- sures to silica and organic solvents were reported in a minority of patients.[1] A highly variable (7-489 cases per million) prevalence, depending on geographic region and familial/geographic clustering (eg, Massachusetts and Libby, United States), suggests that additional investigation for environmental associations is required, in particular for air pollution and industrial emissions.[1,2] Given that the highest prevalence rates were reported in Native populations (commonly employed in the mining industry) and the rich mining/downstream processing economic portfolio in Canada,[3,4] we conducted an ecologic cross-sectional study analyzing geospatial SSc epidemiology and environmental associations using the Canadian Scleroderma Research Group registry. A total of 1505 patients, above 18 years old, between 2004 and 2019 were included. The period prevalence was calculated for each Forward Sortation Area (FSA; first 3 digits of a postal code) and mapped with ArcMap 10.7.1 software (Fig 1, *A* and Supplemental Figs S1 and S2). A Monte Carlo simulation was performed to predict the distribution of patients across the available FSAs, based on the well- demonstrated evidence that SSc is more prevalent in women above 45 years old, to obtain the expected SSc counts and 95% CIs per FSA (Fig 1, *B* and Supplemental Fig S3, available via Mendeley at https://doi.org/10.17632/dddx4fd6zb.1) and compare them to the observed rates.[5] Data on air pollution were obtained from The Canadian Urban Environmental Health Research Consortium and industrial/mining lands from Enhanced Point of Interest (Digital Mapping Technologies Inc). Analyses were conducted using MATLAB and GraphPad Prism. Additional details on methodology and supplemental tables/figures are available via Mendeley at

https://doi.org/10.17632/dddx4fd6zb.1.

The median age at SSc diagnosis was 47.1 years, and 86.3% of the patients were women. An uneven geographic distribution of cases was observed, with urban and rural clusters (Fig S1, available via Mendeley at https://doi.org/10.17632/dddx4fd6zb.1). Nineteen FSAs had significantly higher observed prevalence rates (2- to 5.5-fold higher) than expected (Supplemental Table I, available via Mendeley at https://doi.org/10.17632/dddx4fd6zb.1). There were no significant differences between the high-prevalence and average Canadian Scleroderma Research Group prevalence by demographic characteristics, clinical features, or occupational exposure to silica, organic solvents, or welding fumes. Interestingly, 6 of the 19 high-prevalence FSAs were in rural regions and often far from recruitment centers. Several urban FSAs were geographically clustered in Hamilton (Ontario) (Fig 1, *A*) and Montreal (Quebec) (Supplemental Fig S2).

In comparison to the 19 lowest-prevalence FSAs, the 19 highest-prevalence FSAs showed significantly higher densities for all industries combined (P=.0059) and for paper (P= .0364) and petroleum and coal product manufacturing (P =.0002) industries (Supplemental Table II, available via Mendeley at https://doi.org/10.17632/dddx4fd6zb.1). Similarly, air pollution ($PM_{2.5}$, O_3, NO_2, and/or SO_2) was statistically higher in 15 of the 19 high- prevalence FSAs, and the $PM_{2.5}$ levels were higher in only 8 of the 19 FSAs, favoring major urban centers (Supplemental Table III, available via Mendeley at https://doi.org/10.17632/dddx4fd6zb.1).

Our analysis revealed a nonuniform geographic distribution of SSc cases in Canada, including in highly industrialized, rural areas located far from recruitment centers. Among the 6 rural FSAs, previously identified areas of increased prevalence were confirmed - namely, Woodstock and Monteregie[3]. Among the 13 urban FSAs, 5 belonged to Hamilton (Ontario), the largest

manufacturer of steel (60% of all Canadian production), with significant benzene air pollution. A statistically higher density of industries, as well as air pollution, was demonstrated in the FSAs with the highest prevalence (Fig 1, *A*). We believe that this data is highly relevant, and future studies will focus on putative environmental triggers with biologic plausibility.

The authors would like to thank The Canadian Urban Environmental Health Research Consortium (CANUE) for the annual air pollution data utilized in this study— specifically, calculated ozone, $PM_{2.5}$, SO_2, and NO_2 metrics, all of which were indexed to Digital Mapping Technologies Inc. Spatial Inc. postal codes.

References

1. Ouchene L, Muntyanu A, Lavoue J, Baron M, Litvinov IV, Netchiporouk E. Toward understanding of environmental risk factors in systemic sclerosis. *J Cutan Med Surg*. 2021;25(2): 188-204.
2. Kassamali B, Kassamali AA, Muntyanu A, Netchiporouk E, Vleugels RA, LaChance A. Geographic distribution and environmental triggers of systemic sclerosis cases from 2 large academic tertiary centers in Massachusetts. J Am Acad Dermatol. 2021:S0190-9622(21):00608-3. https://doi.org/10. 1016/j.jaad.2021.03.055
3. Calderon LM, Pope JE. Scleroderma epidemiology update. Curr Opin Rheumatol. 2021;33(2):122-127.
4. Minerals Sector Employment. Government of Canada. Accessed January 8, 2021. https://www.nrcan.gc.ca/science-data/ science-research/earth-sciences/earth-sciences-resources/earth- sciences-federal-programs/minerals-sector-employment/16739

5. Pauling JD, McGrogan A, Snowball J, McHugh NJ. Epidemiology of systemic sclerosis in the UK: an analysis of the Clinical Practice Research Datalink. *Rheumatology (Oxford)*. 2021;60(6):2688-2696. https://doi.org/10.1016/j.jaad.2021.12.055

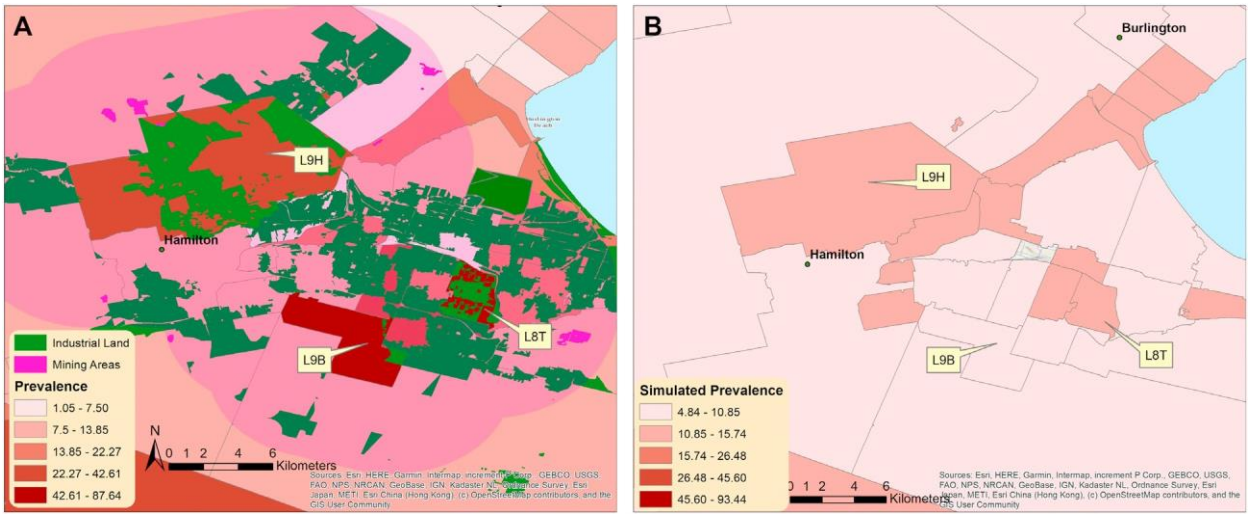


Fig 1. A, L8T, L9B, and L9H are 3 high-prevalence Forward Sortation Areas for systemic sclerosis that are found in Hamilton, Ontario. There is a strong industrial presence, mainly steel mining sites, within a 5-km area of each labeled Forward Sortation Area (outlined in pink). Green represents industrial sites, and fuchsia denotes mining sites. B. Expected prevalence of systemic sclerosis, from Monte Carlo simulation, at L8T, L9B, and L9H, based on age and sex alone.

Mendeley Supplementary Methods

Data Sources

CSRG is the largest registry of Canadian SSc patients and has comprehensive demographic, medical and imaging data allowing assessment of disease severity and clinical

course. The registry was approved by the ethics committee at the Jewish General Hospital (Montreal, Canada) and at each participating center and material transfer agreement was obtained. The patients were recruited from Canadian rheumatology centers. SSc diagnosis was verified by a rheumatologist and over 98% of the patients met the 2013 ACR/EULAR classification criteria for SSc. The characteristics of this SSc registry cohort are similar to other large SSc cohorts around the world.

Geospatial and Statistical Analyses

FSAs were available for each patient at recruitment into the registry. In Canada, postal codes are defined by six-character alphanumeric string, each accounting for about 19 households. Forward Sortation Areas (FSAs) use the first 3-digits of a postal code and correspond to a postal facility from which mail delivery originates and serves approximately 8,000 households (range 0 to >60,000). FSAs were categorized as urban or rural, defined by population density of more or less than 400 inhabitants/km, respectively, according to Statistics Canada.

In Canada 1621 separate FSAs were present in 2011 while only 665 contained SSc patients. Canadian census data for 2011 (midpoint year) was used as a denominator. Prevalence per FSA was used to map the cases with ArcMap 10.7.1 (Environmental Systems Research Institute [Esri], Redlands, California) as previously described. The base map by FSA was obtained from Statistics Canada. Increasing prevalence is demonstrated on the map with a gradient ranging from light pink to deep red in colour. Categories were created based on the Jenks Natural Breaks Classification.

Due to limited patient number, age-/sex- standardized rates could not be used for mapping. Population size, age, and sex distributions relative to the 2011 Census were obtained for each FSA

from Statistics Canada. A weight score for each FSA was calculated using the ratio of females >45-year-old in each FSA to females >45-year-old in all FSAs combined. Hence, FSA regions that are more populated and have a higher ratio of females >45-year-old should have more SSc cases based on the model and vice versa. The simulation was repeated 100,000 iterations**.** Similar simulations have been used in previous ecological studies. Observed CSRG period prevalence of SSc cases per FSA was compared to expected prevalence obtained from the simulation using a binomial probability test and $p<0.05$ was accepted as significant. Thus, high prevalence FSAs were defined as those with significantly higher prevalence than expected and 19 FSAs in our study met this definition. Analyses were conducted using MATLAB and Excel.

Identification of Exposures

The prevalence maps were linked with geographic data on air pollution for year 2011 for particulate matter 2.5 ($PM_{2.5}$), ozone (O_3), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2) for each FSA from The Canadian Urban Environmental Health Research Consortium (CANUE). Stability of concentration of air pollutants over time was assessed to ensure 2011 mid-point was appropriate. Pollutant concentrations in the 19 high prevalence FSAs were compared to the levels in the lowest 19 FSAs using Brown-Forsythe and Welch ANOVA tests with Games-Howell multiple comparisons test. Analyses were conducted with GraphPad Prism version 8.0.1.

ArcMap was used to assess the presence of industrial and mining lands, as well as a number of industrial facilities available from Digital Mapping Technologies Inc (DMTI) Enhanced Point of Interest (EPOI). Presence of facilities listed in Table S2, per FSA were investigated. Since other studies documented that most relevant exposures occur within a 5km radius from one's place of

residence, industries located within this distance were marked as potentially relevant. Proximity buffer in ArcMap was used to determine the number of manufacturing industries and mining sites within a 5km area of the FSA of interest. This was divided by the land area to determine industry density. Difference in industry density between the 19 FSAs of interest and the 19 lowest prevalence FSAs were analyzed using Welch's t-test with GraphPad Prism.

Patient demographics (age/sex), clinical disease severity (cutaneous subsets, Medsger Severity Score, and death), disease duration and reported occupational exposures were compared between areas of high prevalence *vs.* average CSRG prevalence to evaluate differences of these potential confounders as well as association with air pollution and industry density. This was performed using Welch's t-test for continuous variables and Fisher's test for categorical variables (GraphPad Prism version 8.0.1).

Results

Approximately a third of patients had dcSSc (Table S1). Two thirds of patients lived either in Quebec or Ontario, the 2 largest Canadian provinces. Provincial patient distribution was well represented except for British Columbia (2%), given no CSRG recruitment center was present after 2008. Patients resided across 665 different FSAs (Figure S1).

Based on the simulation model, more patients were expected to be identified in larger urban FSAs, near recruitment centers. However, aside from urban centers, many smaller remote communities were identified. Among the FSAs with the highest overall prevalence, were regions >200 km away from the recruitment centers such as Amherst (NS) (78.52 cases/100,000 people), Alma (QC) (49.49 cases/100,000 people), Flin Flon (MB) (34.93 cases/100,000 people), and

Miramichi (NB) (42.61 cases/100,000 people). To decrease the recruitment bias and false discovery, observed *vs*. expected prevalence (accounting for population demographics) for each FSAs were compared and only FSAs where at least 5 cases of SSc occurred were considered.

To assess whether the observed geographic distribution may be a result of patients with severe disease moving closer to treatment centers, demographic and disease features were compared between patients in the top 19 FSAs and the rest of CSRG patients. No statistically significant difference was observed for age, sex, SSc subtype, Medsger severity score and mortality. Given that a subset of SSc cases may be occupationally induced, patient self-reported data on occupational exposure to silica and aromatic compounds was assessed and no significant difference was found. Finally, patients had a similar disease duration across the 19 higher prevalence FSAs and the rest of the cohort, hence longer disease duration is unlikely to bias the obtained results.

Limitations

Our study has several limitations. Though representative of general SSc cohorts, CSRG data contains only ~10% of Canadian SSc cases. To overcome this limitation, simulation was used to simulate population-based distribution of SSc, taking into the account population age, sex, and size, to confirm the relevance of identified regions. The narrow confidence intervals of simulation analysis are the result of high number of simulations (100,000). The prevalence data could be a gross underestimate of real prevalence as this is not a populational study. The postal code information used to map SSc cases was derived from the day of patient recruitment into the registry and complete spaciotemporal history was unavailable. However, given that many identified FSAs were located far from recruitment centers, moving to be closer to a treatment center is unlikely.

We recognize that annual prevalence or incidence, stratified by sex, would have been of high interest, given the study population of 1505 patients, the period prevalence was chosen, and simulation took populational age and sex into account. Disease severity, duration, patient demographics (age and sex) and occupational exposures to pollutants (i.e. silica and organic solvents) were also compared between high prevalence FSAs and average CSRG data and no statistically significant difference was seen. Given that SSc is a rare disease, in future populational based studies, limitations based on low patient numbers may still exist, and smoothing techniques could be applied. This involves an empirical Bayes method, which produces a local risk estimate on the basis of the overall pattern of rates.

Mendeley Supplementary

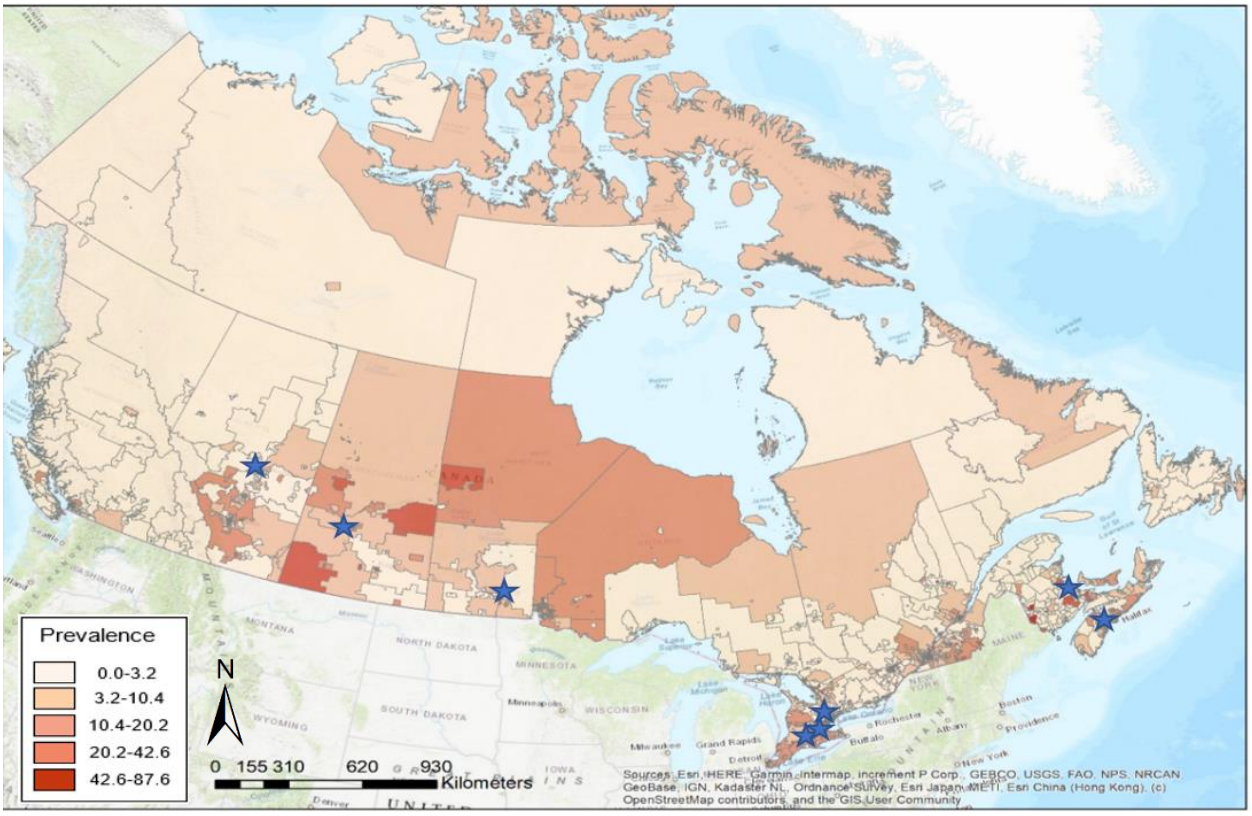


Figure S1. Geographic mapping of SSc patients in Canada (from 2004-2019) using the CSRG database.

Legend. While many CSRG patients live near urban recruitment centers, areas of higher Prevalence can be seen far away from recruitment centers, hospitals and specialists. Blue star represents CSRG patient recruitment centers outside of the province of Québec. CSRG, Canadian Scleroderma Research Group (Montréal, QC).

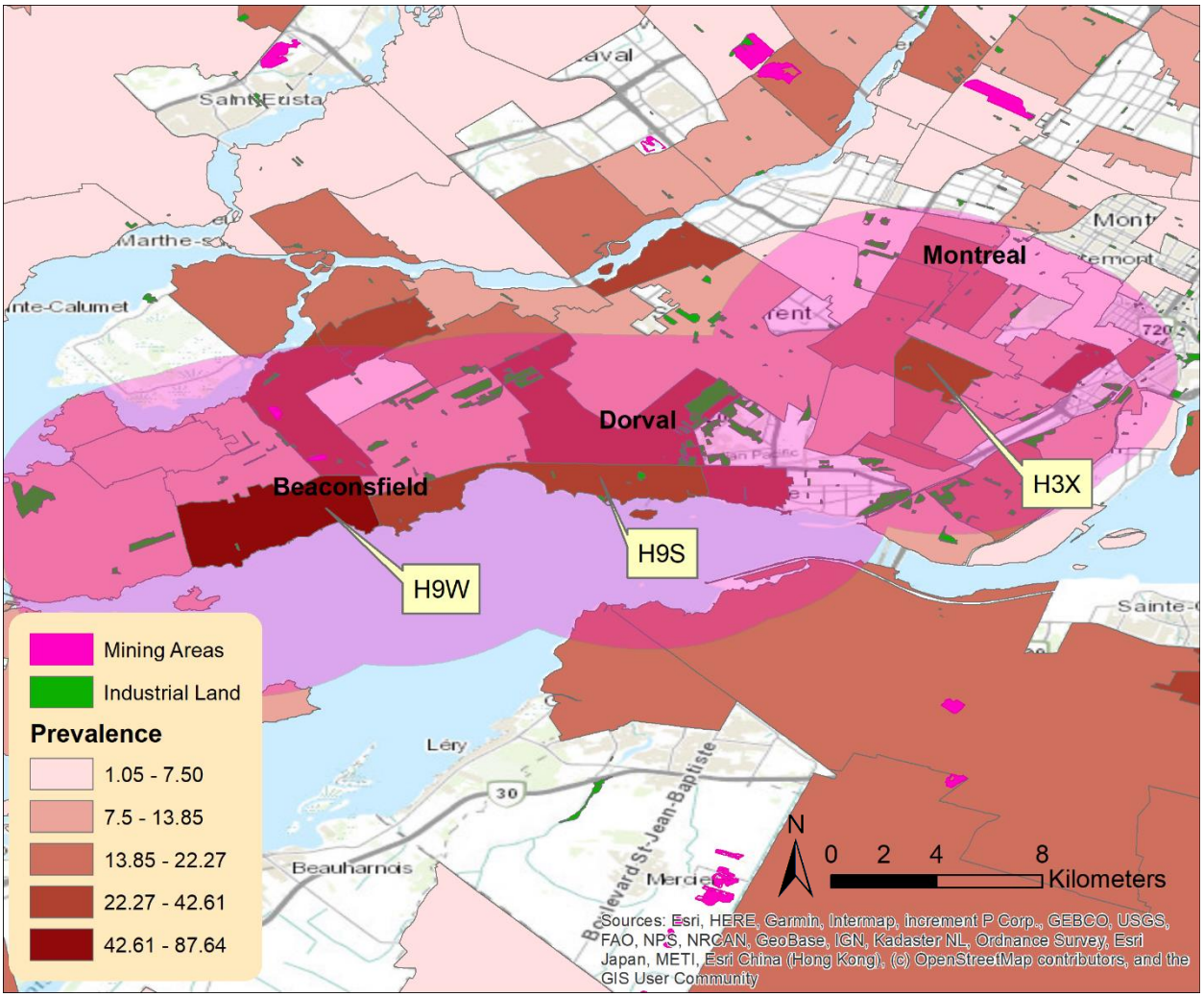


Figure S2. Montréal, Québec. Three high prevalence of SSc FSAs found near one another. Several mining and industrial areas are found within a 5 km radius of highlighted FSAs (outlined in pink). Green represents industrial sites and fuchsia color denotes mining sites.

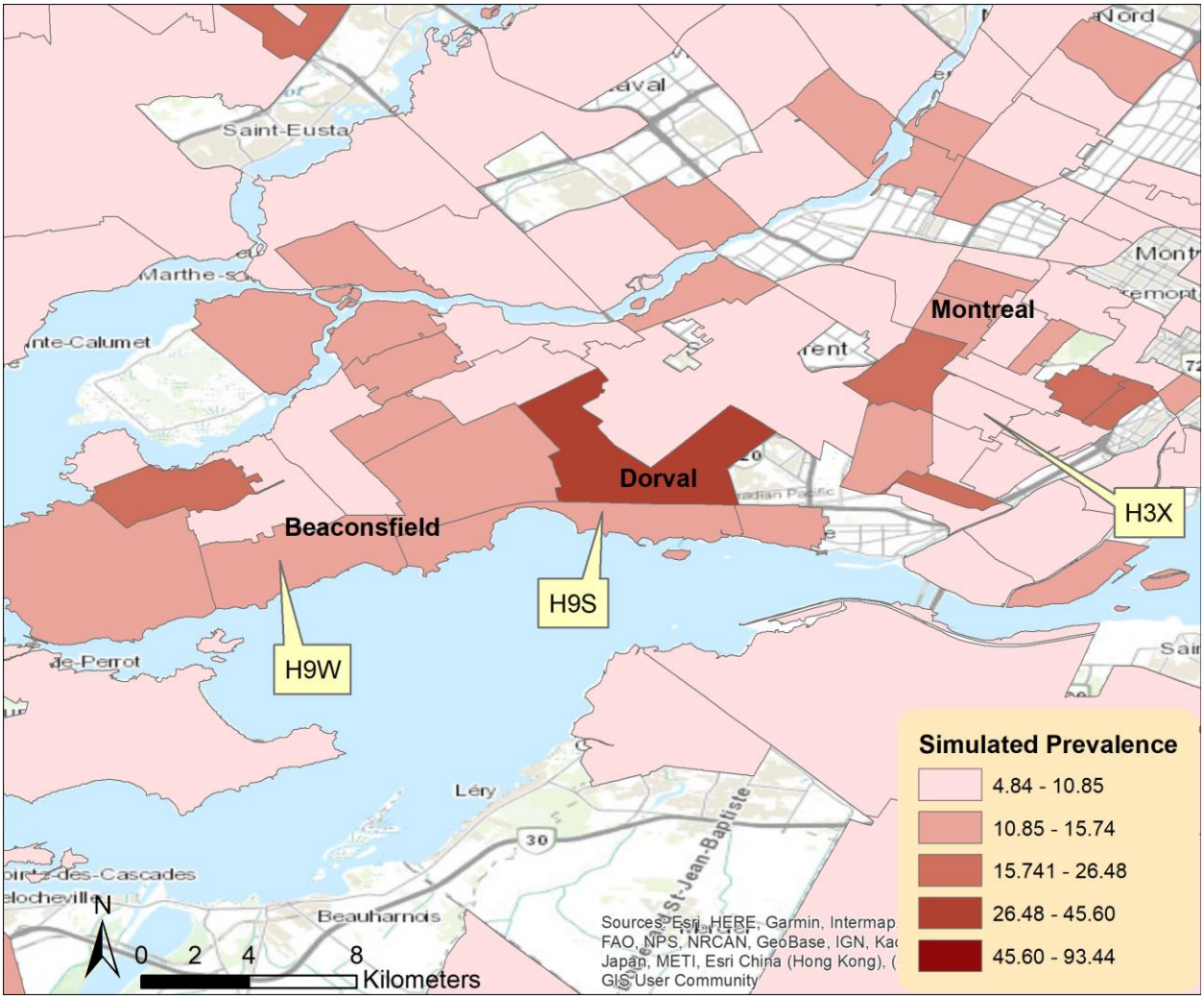


Figure S3. Expected prevalence of SSc, from Monte Carlo simulation, in Montreal, Quebec FSAs based on age and sex alone. Montreal, Quebec.

Table S1. Observed CSRG versus expected prevalence in top 19 significant high-prevalence FSAs with corresponding region name and rural/urban definition. The number of SSc cases at each FSA and the population size are also presented. R=rural defined as defined by population density of less than 400 habitants/km^2; U=urban; CI=confidence interval.

FSAs	CSRG prevalence	Predicted Prevalence (CI)	P value	Fold Change	U/R	Representative region (Province)	Number of Cases in CSRG	Population
E1N	42.61	15.02 (14.96-15.07)	0,03	2.84	R	Miramichi (NB)	5	11,733
H3S	21.04	7.32(7.30-7.35)	0.01	2.87	U	Côte-des-Neiges North - Montréal (QC)	7	33274
H3X	30.09	10.29 (10.25-10.32)	0,01	2.93	U	Hampstead/ Côte Saint-Luc- Montréal (QC)	7	23,261
H9S	31.82	12.10 (12.06-12.14)	0,02	2.63	U	Pointe-Claire/Dorval (QC)	7	21,999
H9W	46.14	11.37 (11.33-11.41)	0.0005	4.06	U	Beaconsfield (QC)	9	19,505
J0L	18.48	8.35 (8.34-8.38)	0,008	2.21	R	Montérégie (QC)	13	70341
J1N	24.24	8.53 (8.50-8.56)	0.01	2.84	R	Rock Forest - Sherbrooke (QC)	7	28879
L8T	52.20	11.71 (11.67-11.75)	0.0001	4.50	U	Hamilton (ON)	10	19,158
L8W	23.36	8.94 (8.91-8.97)	0.03	2.61	U	Hamilton (ON)	6	25686

L9A	28.68	10.33 (10.29-10.36)	0.01	2.78	U	Hamilton (ON)	7	24409
L9B	57.62	10,48 (10.44-10.52)	< 0.0001	5.50	U	Hamilton (ON)	12	20,827
L9H	28.49	11.19 (1.16-11.23)	0.01	2.55	U	Dundas - Hamilton (ON)	9	31593
N0N	19.60	9.74 (9.71-9.77)	0.05	2.01	R	Lambton county (ON)	8	40813
N4S	22.02	9.86 (9.83-9.89)	0.04	2.23	R	Woodstock (ON)	7	31793
R3R	31.91	10.88 (10.84-10.92)	0.007	2.93	U	Winnipeg (MB)	8	25,068
S0E	29.55	9.87 (9.84-9.90)	0.002	2.99	R	Eastern Saskatchewan (SK)	10	33,842
S7J	25.50	10.03 (10.00-10.07)	0.02	2.34	U	Saskatoon (SK)	8	31368
S7L	18.10	7.65 (7.63-7.68)	0.03	2.37	U	Saskatoon (SK)	7	38684
T3E	23.20	8.66 (8.63-8.69)	0.008	2.68	U	Calgary (AB)	9	38787

Legend: AB, Alberta; CI, 95% confidence interval; MB, Manitoba; NB, New Brunswick; ON, Ontario; QC, Quebec; SK, Saskatchewan; R, rural, U, urban. The representative region stated is corresponding to the official list from Canada Post. FSAs, Forward Sortation Areas; CSRG, Canadian Scleroderma Research Group.

Table S2. Comparison of industry density between top 19 and lowest 19 FSAs. Bold numbering indicates statistically significant values.

Industry	**19FSAs of interest**	**19 lowest prevalence FSA**	**p-value**
All Industries combined	0.452	0.363	**0.0059**
Paper Manufacturing	0.009658	0.003406	**0.0364**
Petroleum and Coal Products Manufacturing	0.1643	0.06408	**0.0002**
Chemical Manufacturing	0.01630	0.01571	0.507
Plastics and Rubber Products Manufacturing	0.05864	0.05176	0.158
Nonmetallic Mineral Product Manufacturing	0.05279	0.04629	0.084
Primary Metal Manufacturing	0.01363	0.01913	0.8200
Machinery Manufacturing	0.1722	0.1876	0.1948
Transportation Equipment Manufacturing	0.02770	0.02092	0.0648
Mining	0.01989	0.02707	0.0684

Table S3. Air pollution analysis for the top 19 significant high-prevalence FSAs compared to the lowest 19 FSAs represented in CSRG data. Pollution data obtained from CANUE for the year 2011. Concentration for O_3, NO_2, and SO_2 measured in parts per billions (ppb). NA indicates that only one measurement was available for that FSA and as a result statistical analyses could not be performed. Bold indicates statistically higher levels compared to the 19 lowest prevalence FSAs.

FSAs	**U/R**	**Representative region (Province)**	**Air pollution $PM_{2.5}$ - mg/m^3**	**p-value**	**Air Pollution O3 (ppb)**	**p-value**	**Air Pollution NO2 (ppb)**	**p-value**	**Air Pollution SO2 (ppb)**	**p-value**
E1N	R	Miramichi (NB)	3.808	<0.0001	23.50	<0.0001	4.254	<0.0001	0.144	<0.0001
H3S	U	Montreal (QC)	**8.767**	**<0.0001**	19.2	NA	**17.89**	**<0.0001**	0.54	NA
H3X	U	Montréal (QC)	**9.006**	**<0.0001**	19.2	NA	**17.71**	**<0.0001**	**0.530**	**<0.0001**
H9S	U	Pointe-Claire/Dorval (QC)	**8.135**	**<0.0001**	**25.16**	**<0.0001**	11.78	0.430	**0.695**	**<0.0001**
H9W	U	Beaconsfield (QC)	**8.285**	**<0.0001**	24.62	NA	9.466	<0.0001	0.248	0.6793
J0L	R	Montreal (QC)	5.514	0.0013	**26.61**	**0.0245**	4.610	<0.0001	**0.423**	**0.0076**
J1N	R	Sherbrooke (QC)	5.993	<0.0001	**29.14**	**<0.0001**	5.478	<0.0001	0.200	<0.0001
L9H	U	Hamilton (ON)	6.658	0.0188	**24.76**	**0.025**	8.423	<0.0001	0.200	<0.0001
L8T	U	Hamilton (ON)	**8.664**	**<0.0001**	23.22	NA	9.728	<0.0001	0.04	NA
L8W	U	Hamilton (ON)	7.028	0.122	23.23	<0.0001	10.74	<0.0001	0.04	NA
L9A	U	Hamilton (ON)	**9.039**	**<0.0001**		NA	10.96	<0.0001	0.04	NA

L9B	U	Hamilton (ON)	6.028	<0.0001	**25.31**	**0.0001**	9.587	<0.0001	NA	NA
N0N	R	Lambton County (ON)	5.667	0.0039	**31.17**	**<0.0001**	4.315	<0.0001	**1.164**	**<0.0001**
N4S	R	Woodstock (ON)	**8.564**	**<0.0001**	**29.05**	**<0.0001**	7.988	<0.0001	NA	NA
R3R	U	Winnipeg (MB)	6.578	<0.0001	22.01	<0.0001	8.392	<0.0001	0.110	<0.0001
T3E	U	Calgary (AB)	**9.088**	**<0.0001**	21.74	NA	**14.52**	**<0.0001**	0.126	<0.0001
S0E	R	Saskatoon (SK)	2.977	<0.0001	**27.33**	**<0.0001**	5.406	<0.0001	0.241	>0.9999
S7J	U	Saskatoon (SK)	6.354	<0.0001	23.91	NA	11.71	0.1369	0.076	<0.0001
S7L	U	Saskatoon (SK)	4.918	<0.0001	22.45	<0.0001	**12.94**	**<0.0001**	0.129	<0.0001
Average of 19 lowest prevalence FSAs			6.843		24.45		11.50		0.235	

Legend. AB, Alberta; CSRG, Canadian Scleroderma Research Group; FSAs, Forward Sortation Areas; MB, Manitoba; NB, New Brunswick; ON, Ontario; Ppb, parts per billion; QC, Quebec; SK, Saskatchewan; R, rural, U, urban. The representative region stated is corresponding to the official list from Canada Post. Bolded numbers represent statistically higher levels of air pollutants in a given FSA.

Chapter 5: Silica Exposure and SSc

Preamble to Manuscript 3

Given that the previous two studies demonstrated an uneven geographic distribution, we aimed to further explore the possible contribution of environmental factors (*e.g.* industrial density, air pollution, as above) and occupational exposures. In this study, we focused on silica, which has been closely linked to SSc development. However, limited data on the demographics of exposed patients, as well as the specific disease manifestations resulting from exposure is available. In this study, data was extracted from the CSRG national registry, between 2004-2019, consisting of 1,439 patients. The diagnosis of SSc was confirmed by a rheumatologist and all patients had documented occupational exposures, demographic information, and data on disease manifestations.

In the cohort, 6.6% of patients were exposed to silica (20% of male patients). They were younger at diagnosis, more likely to be male and have more severe disease features. Multivariate regression, controlled for multiple confounders, confirmed that silica exposure was associated with a younger age at diagnosis and worse disease severity and mortality. Hence, screening for silica exposure among higher risk individuals may be beneficial and these patients may benefit from closer monitoring for systemic disease.

This study was published in Frontiers in Medicine.

Muntyanu A, Milan R, Rahme E, LaChance A, Ouchene L, Cormier M, Litvinov IV, Hudson M, Baron M, Netchiporouk E; Canadian Scleroderma Research Group. Exposure to silica and

systemic sclerosis: A retrospective cohort study based on the Canadian Scleroderma Research Group. *Front Med (Lausanne). 2022 Sep 29;9:984907. doi: 10.3389/fmed.2022.984907.*

This study was featured on Consultant 360: Multidisciplinary Medical Information Network. https://www.consultant360.com/exclusive/exposure-silica-warrants-closer-monitoring-those-systemic-sclerosis

Exposure to silica and systemic sclerosis: A retrospective cohort study based on the Canadian Scleroderma Research Group

Anastasiya Muntyanu[1], Raymond Milan[2,3], Elham Rahme[3,4], Avery LaChance[5], Lydia Ouchene[1], Maxime Cormier[6], Ivan V. Litvinov[1,7,*], Marie Hudson[8,9], Murray Baron[9], Elena Netchiporouk[1] and the Canadian Scleroderma Research Group

[1] Division of Dermatology, Department of Medicine, McGill University Health Centre, Montreal, QC, Canada

[2] Division of Experimental Medicine, Department of Medicine, McGill University, Montreal, QC, Canada

[3] Centre for Outcomes Research and Evaluation, Research Institute of the McGill University Health Centre, Montreal, QC, Canada

[4] Division of Clinical Epidemiology, Department of Medicine, McGill University, Montreal, QC, Canada

[5] Department of Dermatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

[6] Division of Respiratory Medicine, Department of Medicine, McGill University Health Centre, Montreal, QC, Canada

[7] Division of Dermatology, University of Ottawa, Ottawa, ON, Canada,

[8] Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, QC, Canada,

[9] Division of Rheumatology, Department of Medicine, Jewish General Hospital, Montreal, QC, Canada

Abstract

Introduction: Systemic sclerosis (SSc) is thought to be induced by an environmental trigger in genetically predisposed individuals. This study assessed the demographic and clinical characteristics and disease severity of silica exposed SSc patients.

Methods: Data was obtained from the Canadian Scleroderma Research Group (CSRG) cohort, containing 1,439 patients (2004–2019). Univariate and multivariate logistic regression analyses were performed, to determine the phenotype and severity of silica-exposed SSc patients. Mortality was assessed using Cox Survival Regression and Kaplan-Meier analyses.

Results: Among 1,439 patients (86.7% females), 95 patients reported exposure to silica. Those exposed were younger, of male sex and with more severe disease. Sex differences were observed where male patients exposed to silica were more likely to be Caucasian and smokers whereas female patients were younger at SSc diagnosis compared to unexposed. Multivariate regression, controlled for multiple confounders, showed that silica exposure was associated with a younger age at diagnosis and worse disease severity and mortality.

Conclusion: Exposure to silica was reported in ~7% of CSRG cohort and ~20% of male patients and was associated with a worse prognosis in terms of age of diagnosis, organ involvement and mortality. Hence, screening for silica exposure among higher risk individuals may be beneficial and these patients may require closer monitoring for systemic disease.

Keywords

systemic sclerosis, silica, environmental triggers, occupation, mortality, gastrointestinal disease, interstitial lung disease, scleroderma

Introduction

Systemic sclerosis (SSc) is a chronic, fibrosing systemic autoimmune rheumatic disease (1). Most commonly affected organs are the skin, gastrointestinal (GI) tract, and lungs, which have the most contact to the outside world, and their involvement leads to significant morbidity and mortality (2). The prevalence of SSc in Canada in 2003 was estimated to be 74.4/100,000 females and 13.3/100,000 males (2). While SSc is more common in females, the prognosis has been consistently shown to be worse in male patients including more diffuse cutaneous SSc (dcSSc), more interstitial lung disease (ILD) and higher mortality (3, 4). It is not known whether this different SSc phenotype and prognosis seen in males is mediated by biological/hormonal influences or whether exposure/occupation related factors in males contribute to the process. As disease-modifying treatment options are limited, determining triggers and elucidating preventive strategies is of significant importance (3).

The pathogenesis of SSc, while not fully understood, is believed to be induced by environmental triggers in genetically predisposed hosts (3, 4). The nature of such triggers and factors accounting for disease severity/prognosis remain poorly understood. A recent review highlighted the environmental factors studied to date for association with SSc (5). The strongest evidence was observed for environmental or occupational exposure to silica and organic solvents. Specifically, a meta- analysis of cohort studies focusing on workers (often male) exposed to silica demonstrated an 18-fold increased incidence of SSc (5).

While there is evidence to suggest that occupational exposure to silica may be associated with an increased risk of SSc and a more severe phenotype, the proportion of SSc patients with history of occupational or other exposure to silica in North America remains to be clarified. It is

unclear whether certain patient characteristics should prompt assessment for prior silica exposure and whether these patients have a more severe SSc than unexposed SSc patients and may require a different clinical/screening approach for comorbidities and complications. Hence, using the Canadian Scleroderma Research Group (CSRG) (6), we aimed to assess the frequency of occupational exposure to silica among Canadian SSc patients, define the demographic and clinical characteristics of SSc patients exposed to silica, and study whether occupational exposure to silica confers a worse disease severity and mortality.

Materials and methods

Study population

The CSRG is the largest multi-center registry of Canadian SSc patients, extensively described elsewhere (7–10). Patients with a diagnosis of SSc followed in one of the 15 rheumatology centers (across Canada and Mexico), who accepted to participate in the registry, were prospectively recruited between 2004 and 2019. Detailed demographic, clinical, laboratory and imaging data were collected at study enrollment (first visit) and at subsequent visits (usually annually) thereafter, for up to 15 study visits. SSc diagnosis was verified by an experienced rheumatologist and over 98% of the patients met the 2013 ACR/EULAR classification criteria for SSc (11). Ethics approval for this study was obtained at the Jewish General Hospital, Montreal, Canada and at all participating CSRG study sites.

Design

This retrospective cohort study included SSc patients with ≥ 1 registry visit between January 2004 and September 2019. Patients were asked to complete a detailed questionnaire at

inception into the cohort which included yes/no questions regarding certain occupational exposures. Patients were categorized into silica exposed vs. silica unexposed based on their response. The following was indicated on the form "Please check the box if you have ever worked in environments that commonly involve the following substances or if you have ever been exposed to the medications or other exposures mentioned below. If you are not sure or would like to comment, please use the space provided." A patient was considered to be exposed to silica if they answered yes to any of the following exposures: silica dusts, hard rock mining, and/or coal mining. Patients also had an optional field to enter their current occupation title.

Socio-demographic and clinical characteristics

The following variables were extracted from the first patient's visit for all exposure groups: age, sex, ethnicity, smoking status (never or ever smoked), and disease duration (defined as the time between the onset of first non-Raynaud manifestation and recruitment date into the study). SSc subtype was reported as limited cutaneous (lcSSc) and dcSSc, defined as skin fibrosis involving the proximal limbs and/or trunk at any time (8). Severity of skin involvement was measured using the modified Rodnan skin score (mRSS) (score range 0–51) (8). Presence of abnormal nailfold capillaroscopy and history of finger necrosis/gangrene/amputation were recorded.

Presence and severity of internal organ involvement (cutaneous, cardiac, pulmonary, renal, and GI), SSc-specific and SSc-related antibodies and treatment were collected at baseline. Systemic organ involvement was defined as follows (ILD defined below). Pulmonary hypertension corresponds to a systolic pulmonary artery pressure of > 45 mmHg on right heart echocardiogram. SSc-specific renal involvement was defined as a history of renal crisis. Gastrointestinal (GI)

involvement was defined based on the median gastrointestinal (GI)-14 score (12). Other variables recorded by recruiting physicians at the first visit included presence of inflammatory arthritis, myositis, and history of cancer. Mean Medsger disease severity scale (DSS) (13), assessing the presence and severity of 9 individual organs, was also evaluated to corroborate results (14). Categories in this scale included a general domain, peripheral vascular, joint/tendon, muscle, GI, pulmonary, cardiac, renal, and the skin domain. Detailed definition and grading of Medsger DSS are explained elsewhere (14).

Antibody profiles including anticentromere (ACA), anti- topoisomerase 1 (ATA), anti-RNA polymerase III antibodies (anti-RNAP), anti-Ro52, and anti- nucleolus organizer region 90 (Nor90), anti-Ku, anti-Th/To, fibrillarin, anti-PM75, and anti- PM100 were detected by Euroline SSc profile LIA (Euroimmun GmbH, Luebeck, Germany) according to manufacturer's instructions. Antibody against anti-U1 ribonucleoproteins (U1RNP) was assessed by addressable laser bead immunoassays (ALBIA) (QUANTA PlexTM SLE8, INOVA Diagnostics, Inc.). All measurements were obtained from the initial registry visit. Antibodies were reported as negative or weak positive (considered absent) and moderate or strong positive (considered present) based on the accepted lab cut off point and numerical values were not available. Antibodies with nucleolar patterns were considered to be fibrillarin, anti-Th/To, Anti RNAP, PM75 and PM100 (15).

Medication history, including mycophenolate mofetil (MMF) or cyclophosphamide (CYC), was recorded by recruiting physicians as past use, current use, or never used. This was dichotomized for statistical analysis into exposed (past or current use) and never exposed.

Disease severity definition(s)

The following disease severity outcomes were considered: dcSSc phenotype, SSc-specific antibodies, younger age at SSc diagnosis, worse GI disease (GI-14 score) and higher risk and worse ILD. ILD was defined as present if a High Resolution Computerized Tomography (HRCT) of the lungs was interpreted by an experienced radiologist as showing ILD or chest x-ray findings of increased interstitial markings (not due to congestive heart failure) or fibrosis, and/or if a study physician reported findings indicative of ILD on physical examination based on a previously published decision rule (16). Patients with ILD were stratified into Forced Vital Capacity (FVC) of $\geq 70\%$ for mild disease and $< 70\%$ for moderate-to-severe based on their spirometry findings on the first visit.

Mortality

Mortality data was collected during annual visits using a standardized death case report form (17). Follow up was started at date of first registry visit and end of follow up was considered when mortality occurred. Patients were censored at the last available registry visit if they were lost to follow up and no mortality data was recorded.

Statistical analysis

Baseline patient characteristics were compared across the two groups (silica exposure vs. no silica exposure) using Chi- square or Fisher's exact test for categorical variables and ANOVA or Kruskal-Wallis test for continuous variables. Univariate and multivariate logistic regression was used to predict patients' characteristics associated with silica exposure, where silica exposure was considered as the outcome.

Additional univariate logistic regression models, where silica exposure was considered to be a predictor, were used for categorical variables and linear regression for the continuous variable (i.e., GI-14) to determine whether silica exposure was associated with SSc severity (as defined above). The multivariate model was adjusted for possible confounders.

Cox regression analysis for mortality was performed adjusting for age, sex, smoking, and disease duration. R Studio (version 1.4.1106) and SAS studio software was used to conduct all statistical analyses.

Subgroup analyses

As silica exposure is more common in males, separate analyses for silica exposure were performed by sex. Data on gender was not available.

Results

Patient characteristics

In total, 1,439 patients were included in this study, 86.7% were females with mean age at SSc diagnosis of 46.5 ± 13.7 years. Average disease duration at baseline was 9.83 ± 9.23 years. Ninety-five patients (6.6%) reported exposure to silica with female-male ratio of ~1:1 among exposed vs. 8:1 among unexposed patients (6.5:1 in the entire CSRG cohort). Specifically, 22.4% of CSRG males vs. 4.2% females were exposed to silica ($p < 0.0001$) (**Table 1**).

Baseline patient characteristics (**Table 1**) showed that SSc patients exposed to silica were significantly more likely to be younger at diagnosis (median age 44.9 vs. 47.2 years old; $p = 0.016$), males (45.3 vs. 11.1%; $p < 0.001$), have a dcSSc phenotype (51.6% vs. 35.3%; $p = 0.003$), more severe ILD (with higher proportion of low FVC (< 70%) 21.7% vs. 11.2%; $p = 0.016$;

higher Medsger score for lung disease 1.57 vs. 1.30; $p = 0.02$), worse skin fibrosis based on mRSS, 10 vs. 6; $p = 0.011$), and worse GI disease (median GI-14 score, 4 vs. 3; $p = 0.014$ and Medsger GI score 2.07 vs. 1.91; $p = 0.049$) compared to the non-exposed group. Furthermore, consistent with dcSSc phenotype, silica-exposed patients had higher ATA positivity (21.8% vs. 14.8%), lower ACA positivity (28.3% vs. 39.5%), higher prevalence of ILD (38.3% vs. 30.0%), and were more likely to be treated with cyclophosphamide (CYC) and/or MMF (14.9% vs. 8.3%), albeit statistical significance was not reached.

Results of the univariate logistic regression were similar. Silica-exposed patients were younger at diagnosis (OR 0.53; 95% CI: 0.33–0.84), males (OR 6.63; 95% CI: 4.27–10.28), and smokers (OR 1.69; 95% CI: 1.08–2.69) with worse disease phenotype, notably dcSSc (OR 1.96; 95% CI: 1.28–2.99), treatment with CYC and/or MMF (OR 1.95; 95%CI: 1.03– 3.45), lower ACA positivity (OR 0.60; 95% CI: 0.36–0.97), more severe ILD (FVC < 70% predicted, OR 2.20; 95% CI: 1.23–3.74 and higher mean lung Medsger severity score, OR 1.24; 95%CI: 1.03–1.48), worse skin fibrosis (mRSS, OR 1.03; 95%CI: 1.01–1.04), and worse GI disease (higher GI-14 and Medsger GI scores). All significant variables were considered in the multivariate model (**Table 2**).

Co-linearity assessment between the included variables in the multivariate model did not identify high collinearity ($r > 0.7$) (data not shown). The results showed that younger age (OR 0.42; 95%CI: 0.22–0.75), male sex (OR 7.87; 95%CI: 4.51–13.84), severe ILD (FVC < 70%) (OR 2.08; 95%CI: 1.00–4.27) and severe GI disease (GI-14) (OR 1.11; 95%CI: 1.01–1.21) were significant demographic and clinical characteristics of silica exposed patients.

Multivariate regression analyses stratified by sex and adjusted for confounders (all

significant variables identified in the univariate model), female patients exposed to silica were diagnosed younger (OR 0.40; 95%CI: 0.16–0.90) whereas male patients were more likely to be Caucasian (OR 12.06; 95% CI: 1.83–250.88), smokers (OR 4.70; 95% CI: 1.10–33.72), and had more severe ILD (OR 5.72; 95% CI: 1.51–24.27) (**Table 3**).

To assess whether exposure to silica may predict a worse disease prognosis, linear and logistic regression was performed (**Table 4**). SSc patients with reported silica exposure were significantly more likely to be diagnosed before age 50 (OR 0.53; 95% CI: (0.33–0.84) and have a worse disease. Notably, higher risk of dcSSc phenotype (OR 1.95; 95% CI: 1.28–2.99), more severe GI disease (β 0.85; 95% CI: 0.19–1.51), and a lower likelihood of ACA antibody positivity (OR 0.60; 95% CI: 0.37–0.98) were seen. A trend toward more ILD and more severe ILD was also observed, however this was not statistically significant. Multivariate model confirmed that SSc patients with silica exposure were significantly more likely to be diagnosed with both Raynaud's phenomenon (OR 0.48, 95% CI: 0.29–0.78) or SSc before 50 years of age (OR 0.47; 95% CI: 0.29–0.77) when adjusted for sex, ethnicity, smoking status and dcSSc disease phenotype. A strong trend for increased risk of ILD, severe ILD (OR 2.05; 95% CI: 0.96–5.36) and worse GI disease (β 0.67; 95% CI: -0.03 to 1.36) was observed when adjusting for multiple confounders including sex (**Table 4**). Similarly, an important trend toward dcSSc phenotype (OR 1.54; 95% CI: 0.99–2.42), higher prevalence of ATA antibodies and lower prevalence of ACA antibodies was seen after adjusting for age, sex, smoking, and ethnicity.

Mortality

Over the follow up period, 237 patients (of 1,439) were excluded for loss to follow up and/or missing data and 260 died (21.6%). Mortality rate of 71.4 (95% CI: 47.4–103.2) per 1,000

person-years (103.7 per 1,000 person-years in males and 50.5 per 1,000 person-years in females) was seen in silica exposed patients vs. 43.4 (95% CI: 38.0–49.4) (76.1 per 1,000 person- years in males and 39.9 per 1,000 persons-years in females) in the unexposed group (**Table 5**). Additionally, mortality in patients exposed to silica with disease duration of < 5 years was 86.1/1,000 person-years compared to 41.5/1,000 person years in the unexposed group. Unadjusted Kaplan Meier curve shows a significantly increased mortality rate in the silica-exposed group compared to the unexposed [Hazard Ratio (HR) 1.58, 95%CI: 1.07–2.35; $p = 0.0217$] (**Figure 1**). When the hazard ratio was adjusted for age, sex, smoking, and disease duration, a non-statistically significant trend for increased mortality was observed (HR 1.45, 95%CI: 0.96–2.19; $p = 0.0911$).

Discussion

Occupational exposure to silica has been strongly correlated with a higher risk of developing SSc and possibly confers a worse disease phenotype, however prior studies were limited by patient number and hence additional research were needed (5). Occupations with highest risk of exposure are reviewed elsewhere (5), but these include coremaker, bench molder, mineral-crushing machine operator, stone and gem cutter and finisher, concrete-mixer operator, and miners (5). Not surprisingly these occupations commonly employ male patients and males have been consistently shown to exhibit a worse SSc-related prognosis (3).

We showed that 6.6% of the CSRG participants reported exposure to silica and higher chance of reporting silica-exposure was associated with younger age (at diagnosis), male sex and more severe disease phenotype. These findings are aligned with previous reports where Marie et al. also showed that the 18 SSc patients exposed to silica vs. 82 patients not exposed to either silica nor

organic solvents in their study were more often males with severe disease (18). Our results are also aligned with a recent Australian SSc cohort study, which also found that patients exposed to silica were more likely to be male, smokers with worse disease features such as dcSSc, ILD, lower frequency of ACA (3). One study reported the dose dependent relationship between silica exposure and SSc although this could not be assessed in our cohort. The overall risk was found to increase with cumulative exposure from the time of entering the workforce for males [Incidence Rate Ratio (IRR) 1.07 (1.05– 1.09) per 50 mg/m^3 -years] and females [IRR 1.04 (0.99–1.10) per 50 mg/m^3 -years] (19).

We found that exposure to silica in SSc patients confers a two-fold increased risk of being diagnosed with Raynaud's phenomenon and SSc before age 50, despite adjusting for multiple confounders, including sex. Furthermore, silica exposure increased the risk of dcSSc and ATA positivity by almost 50% and increased the risk of severe ILD by twofold, with a confidence interval near statistical significance despite adjusting for multiple confounders. Previous literature supports the association between ILD, dcSSc phenotype and silica exposure (5, 20, 21). As expected, ACA positivity was lower in silica exposed patients as this antibody profile is typically associated with lcSSc phenotype and has been consistently shown to be protective against ILD and SSc related mortality in both lcSSc and dcSSc.

A robust trend toward worse GI disease was seen in patients exposed to silica in our study. While previous studies have shown the association between occupational silica exposure and gastric cancer, GI symptoms secondary to silica exposure in SSc patients have not yet been reported. Silica can come into contact with the GI tract as a result of ingestion following clearance

from the lungs (22, 23) and can lead to chronic local injury and inflammation in the GI tract (24). Thus, this area warrants further evaluation and assessment.

Sex is an important determinant of SSc prognosis. In our study, almost a quarter of male SSc patients reported silica exposure as opposed to only 4.2% of female SSc patients. This is also aligned with the Australian SSc cohort where 7.5% of SSc patients and 31.6% of male SSc patients reported silica exposure (3). These rates of silica exposure are much higher than rates expected in general population where ~1.1% of working Canadians may be exposed to silica in the workplace (25). The profile of SSc patients exposed to silica differed by sex. Male patients were more likely to be younger, Caucasian and smokers whereas female SSc patients exposed to silica were more likely to be younger at diagnosis compared to silica unexposed group. Despite adjusting for sex and multiple other covariates, we showed that silica exposure was associated with adverse outcomes.

Almost 35% of silica exposed and ~20% of unexposed SSc patients died over 14-year follow up with a mortality rate of 71.4 per 1,000 person-years in silica exposed vs. 43.4 in unexposed patients. As expected, mortality rate was higher for males and older patients. While significance was lost after adjustment, we believe the strong trend toward excess mortality needs further research.

Our study has several limitations. Missing data for individual variables usually ranged from 0 to 5% (**Table 2**). The exposure to silica was based on self-report and coded as Yes/No.

Timing, duration, and intensity of exposure were not available, although this is similar to other studies in the literature. Hence, it was not possible to evaluate dose dependent response and

development/severity of SSc. Additionally, exposure misclassification and recall bias are possible. However, we believe that both under- and overreporting of silica exposure are more likely to bias toward not finding any association. Few patients reported their occupation and industry type, which was used to verify likelihood of exposure to silica. However, different industry types and occupations may lead to low vs. high risk of silica exposure which could not be assessed in this study. While this is one of the largest studies assessing the association between silica exposure and SSc, the absolute number of patients with reported exposure to silica remains relatively low (95 patients, 6.6% in this registry). Finally, we did not have gender data and hence it remains to be confirmed whether the worse disease features seen in males with SSc are driven by biological factors (i.e., sex) or occupational/sociocultural exposures (i.e., gender).

Exposure to silica was seen in ~7% of the CSRG cohort and was more common patients younger at diagnosis, males, smokers and patients with a more severe disease. Differences by sex were observed where male patients exposed to silica were more often smokers and Caucasian vs. silica-exposed female patients were younger at SSc diagnosis. Furthermore, silica exposure was associated with worse SSc outcomes in these patients such as younger age at SSc diagnosis and a strong trend toward higher risk of dcSSc, ATA antibodies, more severe GI disease, ILD, and mortality. Hence, our results suggest that prior silica exposure among SSc patients in North America is common, particularly among males and younger females and these patients are at risk of worse outcomes. Patients with occupational silica exposure who present with new onset Raynaud's phenomenon should be thoroughly assessed for Very Early Diagnosis of SSc (VEDOSS) through physical examination (e.g., puffy fingers), nailfold capillaroscopy (using dermoscopy or videocapillaroscopy), and antibody testing (i.e., antinuclear and/or SSc-specific

antibodies). Earlier diagnosis of SSc could lead to counseling about discontinuation of silica exposure as well as earlier screening for systemic involvement and prompt treatment initiation. For patients diagnosed with SSc, ILD is considered to be an early complication often occurring in the first 3–5 years of disease onset. While there are no clear guidelines, experts usually suggest baseline ILD screening with high resolution CT scan and PFTs with DLCO and regular monitoring by spirometry for the first 3–5 years. SSc patients with silica exposure could benefit from regular follow up during the first 5 years and beyond. As any SSc patients, clinical signs or features of ILD should prompt an early specialist referral to minimize complications. Unfortunately, there is no specific treatment available for SSc associated with silica exposure aside from discontinuing exposure, smoking cessation and SSc management. Large prospective studies with detailed exposure/occupational questionnaires and job matrices are needed to further study the association between silica exposure and prognosis of patients with SSc.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: Canadian Scleroderma Research Group has the dataset. Requests to access these datasets should be directed to MB.

Ethics statement

The studies involving human participants were reviewed and approved by Jewish General Hospital, Montreal, Canada and at all participating CSRG study sites. The patients/participants provided their written informed consent to participate in this study.

Investigators of the Canadian Scleroderma Research Group

M. Baron, Montreal, Quebec; M. Hudson, Montreal, Quebec; G. Gyger, Montreal, Quebec; J. Pope, London, Ontario;

M. Larche, Hamilton, Ontario; N. Khalidi, Hamilton, Ontario;

A. Masetto, Sherbrooke, Quebec; E. Sutton, Halifax, Nova Scotia; T.S. Rodriguez Reyna, Mexico City, Mexico; N. Maltez, Ottawa, Ontario; C. Thorne, Newmarket, Ontario; P.R. Fortin, Quebec, Quebec; A. Ikic, Quebec, Quebec; D. Robinson, Winnipeg, Manitoba; N. Jones, Edmonton, Alberta; S. LeClercq, Calgary, Alberta; P. Docherty, Moncton, New Brunswick; D. Smith, Ottawa, Ontario; M. J. Fritzler, Department of Medicine, Cumming School of Medicine, Calgary, Alberta.

Author contributions

AM, RM, ER, MB, and EN designed the study. AM and LO collected and organized the data. AM prepared the manuscript. RM performed the statistical analyses with input from ER. AL, MC, IL, and MH provided feedback on the design, analyses, and manuscript. EN guided and supervised the study. All authors have read and approved the final version of the manuscript.

Funding

This research was supported by Association des Médecins Spécialistes Dermatologues du Québec and Canadian Dermatology Foundation grant to EN and AM.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent

those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Alba MA, Velasco C, Simeon CP, Fonollosa V, Trapiella L, Egurbide MV, et al. Early-versus late-onset systemic sclerosis: differences in clinical presentation and outcome in 1037 patients. *Medicine.* (2014) 93:73–81. doi: 10.1097/MD. 0000000000000018
2. Bernatsky S, Joseph L, Pineau CA, Belisle P, Hudson M, Clarke AE. Scleroderma prevalence: demographic variations in a population-based sample. *Arthritis Rheum.* (2009) 61:400–4. doi: 10.1002/art.24339
3. Patel S, Morrisroe K, Proudman S, Hansen D, Sahhar J, Sim MR, et al. Occupational silica exposure in an australian systemic sclerosis cohort. *Rheumatology.* (2020) 59:3900–5. doi: 10.1093/rheumatology/keaa446
4. Calderon LM, Pope JE. Scleroderma epidemiology update. *Curr Opin Rheumatol.* (2021) 33:122–7. doi: 10.1097/bor.0000000000000785
5. Ouchene L, Muntyanu A, Lavoue J, Baron M, Litvinov IV, Netchiporouk E. Toward understanding of environmental risk factors in systemic sclerosis [formula: see text]. *J Cutan Med Surg.* (2021) 25:188–204. doi: 10.1177/1203475420957950
6. Diab S, Dostrovsky N, Hudson M, Tatibouet S, Fritzler MJ, Baron M, et al. Systemic sclerosis sine scleroderma: a multicenter study of 1417 subjects. *J Rheumatol.* (2014) 41:2179–85. doi: 10.3899/jrheum.140236
7. Zheng B, Hudson M, Wang M, Baron M, Canadian Scleroderma Research Group.

Immunosuppressive treatment in diffuse cutaneous systemic sclerosis is associated with an improved composite response index (Criss). *Arthritis Res Ther.* (2020) 22:132. doi: 10.1186/s13075-020-02220-0

8. Nevskaya T, Zheng B, Baxter CA, Ramey DR, Pope JE, Baron M. Skin improvement is a surrogate for favourable changes in other organ systems in early diffuse cutaneous systemic sclerosis. *Rheumatology.* (2020) 59:1715–24. doi: 10.1093/rheumatology/kez529
9. Hoa S, Bernatsky S, Steele RJ, Baron M, Hudson M. Association between immunosuppressive therapy and course of mild interstitial lung disease in systemic sclerosis. *Rheumatology.* (2020) 59:1108–17. doi: 10.1093/rheumatology/kez407
10. Ferdowsi N, Huq M, Stevens W, Hudson M, Wang M, Tay T, et al. Development and validation of the scleroderma clinical trials consortium damage index (sctc-di): a novel instrument to quantify organ damage in systemic sclerosis. *Ann Rheum Dis.* (2019) 78:807–16. doi: 10.1136/annrheumdis-2018-214764
11. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 Classification criteria for systemic sclerosis: an American college of rheumatology/european league against rheumatism collaborative initiative. *Arthritis Rheumat.* (2013) 65:2737–47. doi: 10.1002/art.38098
12. Baron M, Hudson M, Steele R, Lo E. Validation of the ucla scleroderma clinica trial gastrointestinal tract instrument version 2.0 for systemic sclerosis. *J Rheumatol.* (2011) 38:1925–30. doi: 10.3899/jrheum.110060
13. Hedström AK, Åkerstedt T, Klareskog L, Alfredsson L. Relationship between shift work and the onset of rheumatoid arthritis. *RMD Open.* (2017) 3:e000475. doi:

10.1136/rmdopen-2017-000475

14. Medsger TA Jr., Bombardieri S, Czirjak L, Scorza R, Della Rossa A, Bencivelli W. Assessment of disease severity and prognosis. *Clin Exp Rheumatol.* (2003) 21(3 Suppl 29):S42–6.
15. Kayser C, Fritzler MJ. Autoantibodies in systemic sclerosis: unanswered questions. *Front Immunol.* (2015) 6:167. doi: 10.3389/fimmu.2015.00167
16. Steele R, Hudson M, Lo E, Baron M, Canadian Scleroderma Research Group. Clinical decision rule to predict the presence of interstitial lung disease in systemic sclerosis. *Arthritis Care Res.* (2012) 64:519–24. doi: 10.1002/acr.21 583
17. Richard N, Hudson M, Wang M, Gyger G, Proudman S, Stevens W, et al. Severe gastrointestinal disease in very early systemic sclerosis is associated with early mortality. *Rheumatology*. (2019) 58:636–44. doi: 10.1093/rheumatology/ key350
18. Marie I, Menard JF, Duval-Modeste AB, Joly P, Dominique S, Bravard P, et al. Association of occupational exposure with features of systemic sclerosis. *J Am Acad Dermatol.* (2015) 72:456–64. doi: 10.1016/j.jaad.2014. 11.027
19. Boudigaard SH, Schlunssen V, Vestergaard JM, Sondergaard K, Toren K, Peters S, et al. Occupational exposure to respirable crystalline silica and risk of autoimmune rheumatic diseases: a nationwide cohort study. *Int J Epidemiol.* (2021) 50:1213–26. doi: 10.1093/ije/dyaa287
20. De Decker E, Vanthuyne M, Blockmans D, Houssiau F, Lenaerts J, Westhovens R, et al. High prevalence of occupational exposure to solvents or silica in male systemic sclerosis patients: a belgian cohort analysis. *Clin Rheumatol.* (2018) 37:1977–82. doi:

10.1007/s10067-018-4045-y

21. Marie I, Gehanno JF, Bubenheim M, Duval-Modeste AB, Joly P, Dominique S, et al. Prospective study to evaluate the association between systemic sclerosis and occupational exposure and review of the literature. *Autoimmun Rev.* (2014) 13:151–6. doi: 10.1016/j.autrev.2013.10.002
22. Swaen GM, Meijers JM, Slangen JJ. Risk of gastric cancer in pneumoconiotic coal miners and the effect of respiratory impairment. *Occup Environ Med.* (1995) 52:606–10. doi: 10.1136/oem.52.9.606
23. Meyer MB, Luk GD, Sotelo JM, Cohen BH, Menkes HA. Hypothesis: the role of the lung in stomach carcinogenesis. *Am Rev Respir Dis.* (1980) 121:887–92. doi: 10.1164/arrd.1980.121.5.887
24. O'Neill C, Pan Q, Clarke G, Liu F, Hodges G, Ge M, et al. Silica fragments from millet bran in mucosa surrounding oesophageal tumours in patients in Northern China. *Lancet.* (1982) 1:1202–6. doi: 10.1016/s0140-6736(82)92335-2
25. Canada C. *Silica (Crystalline) Occupational Exposures: CAREX Canada*. Vancouver, BC: University of British Columbia (2022).

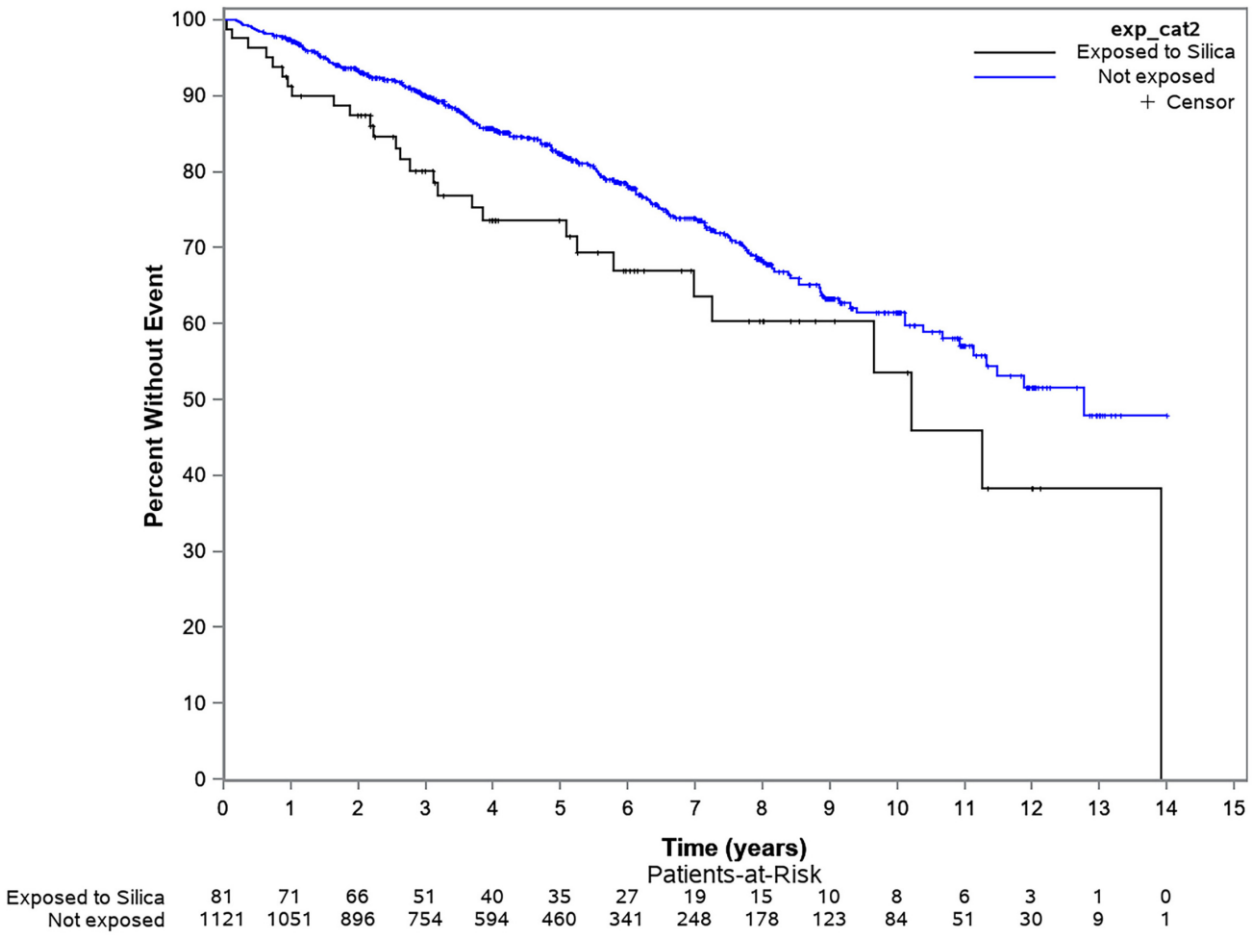


Figure 1. Kaplan-Meier curve evaluating mortality over time in silica exposed vs. unexposed groups

Table 1. Baseline patient characteristics.

Variables	Exposure to silica (*N* = 95)	No exposure to silica (*N* = 1,344)	*P*
Demographics			
Age ≥ 50 years, *N* (%)	25 (27.5)	546 (41.5)	**0.016**
Male sex, *N* (%)	43 (45.3)	149 (11.1)	<**0.001**
Caucasian, *N* (%)	88 (92.6)	1,204 (89.7)	0.626
Disease duration ≥ 5 years, *N* (%)	47 (51.6)	773 (58.9)	0.207
Smoking, *N* (%)	67 (70.5)	786 (58.7)	0.065
Clinical characteristics			
Diffuse disease, *N* (%)	48 (51.6)	470 (35.3)	**0.003**
Treatment with CYC or MMF, *N* (%)	14 (14.9)	110 (8.3)	0.082
ILD, *N* (%)	36 (38.3)	393 (30.0)	0.157
FVC < 70, *N* (%) (*n* = 1,244)	18 (21.7)	130 (11.2)	**0.016**
Pulmonary arterial hypertension, *N* (%) (*n* = 909)	8 (15.4)	140 (16.3)	0.208
Rodnan score, median (IQR)	10.00 [4.00, 17.50]	6.00 [2.00, 14.00]	**0.011**
Digital ulcer/pitting scars, *N* (%)	45 (47.9)	556 (41.8)	0.516
Necrosis/gangrene/amputation, *N* (%)	32 (34.0)	473 (35.6)	0.957
Nailfold capillaroscopy, *N* (%)	67 (71.3)	1,017 (76.5)	0.513
GI-14, Median (IQR) (*n* = 1,343)	4.00 [2.00, 7.00]	3.00 [1.00, 6.00]	**0.014**
Joint impairment, *N* (%)	23 (24.5)	377 (28.3)	0.721
History of renal crisis, *N* (%)	7 (7.4)	48 (3.6)	0.174
Cancer, *N* (%)	8 (8.5)	108 (8.1)	0.991
Antibody profile			
ACA antibody, *N* (%) (*n* = 1,244)	24 (28.3)	458 (39.5)	0.075
ATA antibody, *N* (%) (*n* = 1,244)	18 (21.8)	172 (14.8)	0.189
U1RNP, *N* (%) (*n* = 1,276)	6 (7.0)	63 (5.3)	0.758
Ro52, *N* (%) (*n* = 1,244)	22 (25.8)	309 (26.7)	0.664
Ku, *N* (%) (*n* = 1,244)	0 (0.0)	9 (0.7)	0.477
Nor90, *N* (%) (*n* = 1,244)	1 (1.1)	26 (2.2)	0.54
Nucleolar antibodies, *N* (%) (*n* = 1,243)	18 (21.4)	232 (20.0)	0.795
Medsger severity scores			
Medsger—general, mean (*SD*)	1.06 (1.37)	0.89 (1.18)	0.172

Medsger—peripheral vascular,Mean (*SD*) (*n* = 1,156)	1.88 (1.21)	1.63 (1.24)	0.097
Medsger—skin, mean (*SD*)	1.36 (0.77)	1.21 (0.70)	0.05
Medsger—joint/tendon,Mean (*SD*) (*n* = 1,095)	0.96 (1.32)	0.68 (1.18)	0.055
Medsger—muscle, mean (*SD*)	0.33 (0.90)	0.23 (0.72)	0.185
Medsger—GI tract, mean (*SD*)	2.07 (0.85)	1.91 (0.78)	**0.049**
Medsger—lung, mean (*SD*)	1.57 (1.21)	1.30 (1.11)	**0.02**
Medsger—heart, mean (*SD*)	0.56 (1.17)	0.46 (0.95)	0.354
Medsger—kidney, mean (*SD*) (*n* = 1,266)	0.16 (0.65)	0.11 (0.61)	0.473

IQR, interquartile range; SD, standard deviation; CYC, cyclophosphamide; MMF, mycophenolate mofetil; ACA, anticentromere antibody; ATA, anti-topoisomerase I antibody; U1RNP, anti-U1 Ribonucleoproteins antibody; ILD, interstitial lung disease; FVC, forced vital capacity; Unless a different denominator (n) is indicated, the missing number for remaining variables was < 5%.

Chi-square or fisher exact test for categorical variable. ANOVA or Kruskal-Wallis test for continuous variables. N provided where > 5% of data was missing.

Bold indicates statistically significant values.

Table 2. Univariate and multivariate logistic regression model for factors associated with exposure to silica among all study sample.

	Univariate logistic regression	**Multivariate logistic regression**
	Silica (N = 95)*	**Silica**
Variables	**OR (95% CI)**	**Adjusted OR (95% CI)**
Age ≥ 50 years	**0.53 (0.33–0.84)**	**0.42 (0.22–0.75)**
Male sex	**6.63 (4.27–10.28)**	**7.87 (4.51–13.84)**
Caucasian	1.15 (0.71–3.51)	-
Disease duration ≥ 5 years	0.75 (0.49–1.14)	-
Treatment with CYC or MMF	**1.95 (1.03–3.45)**	1.11 (0.45–2.48)
Smoking	**1.69 (1.08–2.69)**	1.05 (0.60–1.89)
Diffuse disease	**1.96 (1.28–2.99)**	1.49 (0.75–2.94)
Digital ulcer/ pitting scars	1.28 (0.84–1.94)	–
Necrosis/gangrene/ amputation	0.94 (0.60–1.44)	–
Nailfold capillaroscopy	0.76 (0.48–1.23)	–
Pulmonary arterial hypertension	0.93 (0.40–1.92)	–
Joint impairment	0.82 (0.49–1.31)	–
History of renal crisis	2.15 (0.87–4.60)	–
Cancer	1.05 (0.46–2.10)	–
ACA antibody	**0.60 (0.36–0.97)**	0.91 (0.48–1.66)
ATA antibody	1.54 (0.87–2.60)	–
U1RNP	1.36 (0.51–3.00)	–
Ro52	0.96 (0.57–1.56)	–
PDGFR	–	–
Ku	–	–
Nor90	0.52 (0.03–2.49)	–
Nucleolar antibodies	1.09 (0.62–1.83)	–
ILD	1.45 (0.93–2.22)	–
FVC < 70	**2.20 (1.23–3.74)**	**2.08 (1.00–4.27)**
GI-14	**1.08 (1.02–1.15)**	**1.11 (1.01–1.21)**
Rodnan score	**1.03 (1.01–1.04)**	1.00 (0.96–1.03)
Medsger—General	1.12 (0.95–1.31)	–

Medsger—Peripheral Vascular	1.18 (0.97–1.43)	–
Medsger—Skin	1.31 (0.99–1.72)	–
Medsger— Joint/Tendon	1.19 (0.99–1.41)	–
Medsger—Muscle	1.17 (0.90–1.47)	–
Medsger—GI tract	**1.33 (1.01–1.79)**	0.99 (0.69–1.43)
Medsger—Lung	**1.24 (1.03–1.48)**	1.08 (0.82–1.42)
Medsger—Heart	1.10 (0.89–1.33)	–
Medsger—Kidney	1.12 (0.77–1.48)	–

* Reference group no exposure to silica.

– Defines not available/not applicable.

Bold signifies significant values; Variable definitions as above.

Table 3. Univariate and multivariate logistic regression model for factors associated with exposure to silica among male and female patients.

	Univariate logistic regression		**Multivariate logistic regression**	
	Males	**Females**	**Males**	**Females**
	OR (95% CI)*	**OR (95% CI)***	**Adjusted OR (95% CI)**	**Adjusted OR (95% CI)**
Age ≥ 50 years Male sex	0.60 (0.29 201.21) –	**0.37 (0.17 200.72)** –	0.37 (0.12 201.03) –	**0.40 (0.16 200.90)** –
Caucasian	**3.94 (1.10 25.15)**	1.00 (0.43 202.93)	**12.06 (1.83 250.88)**	0.95 (0.37 202.97)
Disease duration ≥ 5 years	0.87 (0.43 201.74)	0.88 (0.50 201.59)		
Treatment with CYC or MMF	1.39 (0.51 203.48)	1.89 (0.76 204.07)	0.75 (0.16 203.08)	2.20 (0.66 206.32)
Smoking	**3.22 (1.19 11.26)**	0.89 (0.51 201.57)	**4.70 (1.10 33.72)**	0.68 (0.34 201.35)
Diffuse disease	0.87 (0.44 201.72)	**2.22 (1.25 203.93)**	1.57 (0.48 205.19)	1.50 (0.60 203.67)
Digital ulcer/pitting scars	1.03 (0.52 202.06)	1.04 (0.58 201.83)	–	–
Necrosis/gangrene/amputation	0.63 (0.28 201.32)	1.27 (0.71 202.23)	–	–
Nailfold capillaroscopy	1.00 (0.48 202.19)	0.71 (0.39 201.36)	–	–
Pulmonary arterial hypertension	1.54 (0.39 205.18)	0.79 (0.23 202.08)	–	–
Joint impairment	0.50 (0.19 201.16)	1.15 (0.61 202.08)	–	–
History of renal crisis	2.03 (0.60 206.25)	1.20 (0.19 204.06)	–	–
Cancer	0.93 (0.20 203.17)	1.22 (0.41 202.86)	–	–
ACA antibody	0.95 (0.37 202.22)	0.74 (0.39 201.37)	1.95 (0.59 206.43)	0.65 (0.28 201.45)
ATA antibody	1.63 (0.67 203.76)	1.24 (0.53 202.57)	–	–
U1RNP	2.29 (0.29 14.31)	1.59 (0.47 204.10)	–	–
Ro52	0.76 (0.30	1.19 (0.61	–	–

	201.75)	202.23)		
PDGFR	–	–	–	–
Ku	–	–	–	–
Nor90	0.80 (0.04 205.63)	–	–	–
Nucleolar antibodies	0.72 (0.29 201.67)	1.17 (0.54 202.31)	–	–
ILD	**2.20 (1.11 204.45)**	0.74 (0.37 201.40)	2.10 (0.80 205.70)	**0.25 (0.08 200.64)**
FVC < 70	**2.88 (1.14 207.12)**	1.75 (0.74 203.67)	**5.72 (1.51 24.27)**	1.75 (0.58 205.05)
GI_14	**1.13 (1.01 201.26)**	1.10 (1.02 201.20)	1.10 (0.93 201.31)	1.10 (0.98 201.23)
Rodnan score	0.99 (0.96 201.02)	**1.03 (1.00 201.05)**	0.98 (0.92 201.04)	1.01 (0.97 201.06)
Medsger—General	1.10 (0.85 201.4)	1.06 (0.93 201.60)	–	–
Medsger—Peripheral vascular	0.98 (0.72 201.34)	1.21 (0.93 201.60)	–	–
Medsger—Skin	0.80 (0.51 201.22)	1.40 (0.94 202.02)	–	–
Medsger—Joint/Tendon	0.97 (0.73 201.27)	1.19 (0.92 201.50)	–	–
Medsger—Muscle	1.65 (0.99 202.75)	1.10 (0.75 201.48)	–	–
Medsger—GI tract	1.16 (0.79 201.75)	**1.47 (1.01 202.21)**	0.77 (0.44 201.33)	1.39 (0.81 202.36)
Medsger—Lung	**1.37 (1.02 201.86)**	1.07 (0.83 201.37)	1.03 (0.65 201.63)	1.15 (0.79 201.65)
Medsger—Heart	1.09 (0.79 201.48)	1.01 (0.72 201.33)	–	–
Medsger—Kidney	1.02 (0.68 201.61)	0.89 (0.34 201.45)	–	–

*Reference group: no exposure to silica.
Bold indicates statistically significant values.

Table 4. Association between health outcomes and exposure to silica (yes vs. no)

	Unadjusted OR (95% CI)	**Adjusted OR (95% CI)**	**Unadjusted β (95% CI)**	**Adjusted β (95% CI)**
ILD (yes vs. no) (*N* = 1,371)[a]	1.44 (0.94–2.23)	1.21 (0.73–1.99)	–	–
Severe ILD* (vs. mild ILD) (*N* = 370)[b]	1.60 (0.74–3.50)	2.05 (0.96–5.36)	–	–
Diffuse disease (yes vs. no) (*N* = 1,397)[c]	**1.95 (1.28–2.99)**	1.54 (0.99–2.42)	–	–
GI-14 (*N* = 1,397)[d]	–	–	**0.85 (0.19–1.51)**	0.67 (-0.03–1.36)
ATA (yes vs. no) (*N* = 1,222)[e]	1.54 (0.89–2.66)	1.47 (0.82–2.64)	–	–
ACA (yes vs. no) (*N* = 1,222)[f]	**0.60 (0.37–0.98)**	0.76 (0.45–1.29)	–	–
Earlier age of onset of disease ≥ 50 years (*N* = 1,395)[g]	**0.53 (0.33–0.84)**	**0.47 (0.29–0.77)**	–	–
Age of onset of Raynauds ≥ 50 years *(N* = 1,397)[h]	**0.53 (0.33–0.85)**	**0.48 (0.29–0.78)**	–	–

[a] Model adjusted for age, sex, diffuse disease, immunosuppressive medication, disease duration, ethnicity, smoking and organic solvents.

[b] Model adjusted for age, sex, diffuse disease, immunosuppressive medication, disease duration, ethnicity, smoking and organic solvents.

[c] Model adjusted for age, sex, smoking, and ethnicity.

[d] Model adjusted for age, sex, disease duration, diffuse disease, immunosuppressive medication, smoking and organic solvents.

[e] Model adjusted for age, sex, smoking, and ethnicity.

[f] Model adjusted for age, sex, smoking, and ethnicity.

[g] Model adjusted for sex, smoking, ethnicity, and diffuse disease.

[h] Model adjusted for sex, smoking, ethnicity, and diffuse disease.

*Defined as presence of ILD and FVC < 70. ILD, interstitial lung disease; ACA, anticentromere antibody; ATA, anti-topoisomerase I antibody. Bold indicates statistically significant values.

Chapter 6: Organic Solvent Exposure and SSc

Preamble to Manuscript 4

Elaborating on occupational exposures contributing to SSc development and severity, in this study we aimed to evaluate the impact of another occupational exposure, organic solvents. Here we describe the clinical features and survival in SSc patients exposed to organic solvents compared to the unexposed group. The CSRG, a national registry, comprising of 1,439 patients was used for the analysis. While there are several studies and case reports in the literature documenting a 2-3-fold increased risk of SSc with exposure to organic solvents, limited data is available on the specific disease manifestations and demographics of exposed individuals.

Among the 1,439 patients, 20.2% reported an occupational exposure to organic solvents. Exposed patients were more likely to be males, smokers, and have features of more severe disease. Multivariate regression adjusted for multiple confounders confirmed the association with renal crisis and more severe gastrointestinal disease. Increasing trend towards mortality was also noted.

This was the first study in the literature documenting increased risk of renal and gastrointestinal involvement in patients exposed to organic solvents. This could be due to organic solvents depositing in internal organs leading to protein damage and triggering autoimmunity. This study highlights the importance of screening for detailed occupational exposures, as this can influence the organs at higher risk of involvement and hence should be prioritized for screening. Further studies in this realm could contribute to preventative counselling for patients and improving regulations in occupational settings to minimize exposure.

This study was published in the Journal of American Academy of Dermatology.

Muntyanu A, Milan R, Rahme E, Baron M, Netchiporouk E; Canadian Scleroderma Research Group. Organic solvent exposure and systemic sclerosis: A retrospective cohort study based on the Canadian Scleroderma Research Group registry. *J Am Acad Dermatol. 2024 Mar;90(3):605-607. doi: 10.1016/j.jaad.2023.04.062.*

This study was also featured in March 2024 Journal of American Academy of Dermatology in the section "Letter to the Editor: Chemical exposure and sclerosing disorders"[107].

Organic solvent exposure and systemic sclerosis: A retrospective cohort study based on the Canadian Scleroderma Research Group registry

Anastasiya Muntyanu, MD,[a] Raymond Milan, PhD,[b] Elham Rahme, PhD,[b,c] Murray Baron, MD,[d] Elena Netchiporouk, MD, MSc,[e] and the Canadian Scleroderma Research Group*

From the Division of Experimental Medicine, McGill University, Montreal, Quebec, Canada[a]; Centre for Outcomes Research and Evaluation, Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada[b]; Department of Medicine, Division of Clinical Epidemiology, McGill University, Montreal, Quebec, Canada[c]; Division of Rheumatology, Department of Medicine, Jewish General Hospital, Montreal, Quebec, Canada[d]; and Division of Dermatology, Department of Medicine, McGill University Health Centre, Montreal General Hospital, Montreal, Quebec, Canada.[e]

*Investigators of the Canadian Scleroderma Research Group: M. Baron, Montreal, Quebec; M. Hudson, Montreal, Quebec; G. Gyger, Montreal, Quebec; J. Pope, London, Ontario; M. Larche, Hamilton, Ontario; N. Khalidi, Hamilton, Ontario; A. Masetto, Sherbrooke, Quebec; E.Sutton, Halifax, Nova Scotia; T. S. Rodriguez Reyna, Mexico City, Mexico; N. Maltez, Ottawa, Ontario; C. Thorne, Newmarket, Ontario; P. R. Fortin, Quebec, Quebec; A. Ikic, Quebec, Quebec; D. Robinson, Winnipeg, Manitoba; N. Jones, Edmonton, Alberta; S. LeClercq, Calgary, Alberta; P. Docherty, Moncton, New Brunswick;

D. Smith, Ottawa, Ontario; M. J. Fritzler, Department of Medicine, Cumming School of Medicine, Calgary, Alberta; and E. Kaminska, Calgary, Alberta.

Funding sources: Association des Medecins Specialistes Dermatologues du Quebec and Canadian Dermatology Foundation grants to Drs Netchiporouk and Muntyanu.

IRB approval status: Ethics approval for this study was obtained at the Jewish General Hospital, Montreal, Quebec, Canada and at all participating CSRG study sites.

Patient consent: Not applicable.

Key words: disease severity; environmental triggers; mortality; occupation; organic solvents; systemic sclerosis.

Correspondence to: Elena Netchiporouk, MD, MSc, FRCPC, Division of Dermatology, Department of Medicine, McGill University Health Centre, Montreal General Hospital, 1650 Cedar Avenue, Montreal, Quebec, Canada H3G 1A4

E-mail: elena.netchiporouk@mcgill.ca

Conflicts of interest: None disclosed.

To the Editor: Systemic sclerosis (SSc) is a fibrosing systemic autoimmune rheumatic disease, hypothesized to be triggered by external factors in genetically predisposed individuals.[1,2] Although occupational exposure to silica is a well-established risk factor for SSc development/severity, data on organic solvents (OSs), and impact on SSc disease phenotype are scarce.[1,2]

OSs are carbon-containing liquids used to dissolve/disperse other materials.[3] They include aromatic (eg, benzene and toluene) and chlorinated solvents (eg, vinyl chloride and trichloroethylene) which have been associated with a higher risk of systemic autoimmune rheumatic diseases.[3] OSs are encountered in many occupations, including laboratory work, leather/tire/paint/glue industries, de- greasing, dry cleaning, and others.[2] Limited data suggest that OSs increase SSc risk up to 2- to 10-fold and are possibly associated with a worse phenotype.[2,3] Mechanistically, it has been proposed that chronic OSs exposure may lead to organ deposition triggering an immune reaction and microangiopathy, as well as modification of self-proteins generating autoantibody production such as seen in SSc.[2,3] In this study, we aimed to investigate the frequency of OSs exposure among patients with SSc and whether it confers a more severe disease phenotype.

Using data from the multicenter, Canadian cohort of patients with SSc (the Canadian Scleroderma Research Group), 2 comparison groups were created (OSs exposed and unexposed) similar to our prior work.[1] Patients with self-reported exposures to chlorinated (eg, vinyl chloride, trichloroethylene, and dry cleaning solutions) and aromatic OSs (eg, toluene, benzene, and paint thinners/removers) were classified as exposed. Multivariate logistic and linear regression models were performed to study OSs association with disease severity measures (from baseline visit) as previously described.[1] Cox regression mortality analysis was performed adjusting for age, sex,

smoking, and disease duration.

Of 1439 patients with SSc, majority meeting the 2013 American College of Rheumatology/European Alliance of Associations for Rheumatology classification criteria, 20.2% were exposed to OSs (Table I).

Although SSc is a female predominant disease, a reduced male/female ratio was seen among OSs-exposed patients (1:2 vs 1:10 in unexposed). Exposed patients were more likely to be men (29.7% vs 9.2%; $P<0.001$), smokers, and with features of severe disease (ie, severe gastrointestinal disease, more renal crisis, and less anticentromere antibodies).

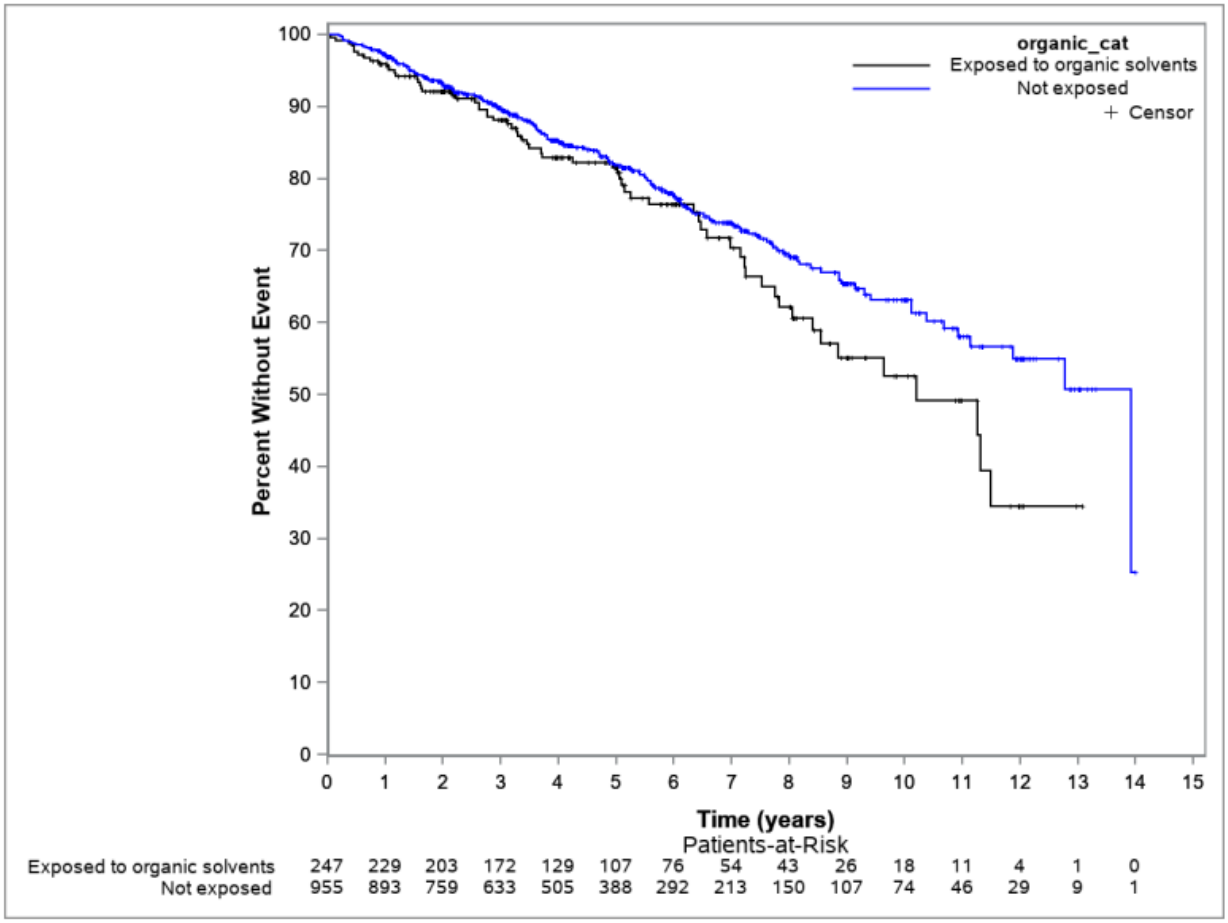
Multivariate regression confirmed the association between OSs exposure and a higher risk of renal crisis (odds ratio, 2.13; 95% CI, 1.15-3.93) and severe gastrointestinal disease (β 0.89; 95% CI, 0.47-1.31) despite adjusting for multiple confounders (Table II). A nonstatistically significant trend toward diffuse SSc, higher antitopoisomerase and lower anticentromere antibodies positivity was seen. Similarly, a trend toward increasing mortality rate was documented among exposed patients, especially men (Supplementary Table I and Supplementary Fig 1, available via Mendeley at https://data.mendeley. com/datasets/wf9j863h78/1). Our study's limitations include possible exposure misclassification due to household OSs exposure and/or recall bias (more likely to bias findings toward null) and the absence of data on the dose/duration of OSs exposure.

To our knowledge, this is the first study reporting a higher risk of renal crisis and gastrointestinal disease severity in patients with SSc exposed to OSs. However, previous research highlights that chronic OSs exposure may lead to internal organ deposition and subsequent organ and protein dam- age potentially leading to autoimmunity/which was reported in the lung, liver,

pancreas, and the kidney.[3] Additional studies have linked OSs exposure to severe progressive glomerulonephritis in men.[4,5] Hence, further research is needed.

References

1. Muntyanu A, Milan R, Rahme E, et al. Exposure to silica and systemic sclerosis: a retrospective cohort study based on the Canadian Scleroderma Research Group. Front Med (Lausanne). 2022;9:984907.Ouchene L, Muntyanu A, Lavoue J, Baron M, Litvinov IV, Netchiporouk E. Toward understanding of environmental risk factors in systemic sclerosis [formula: see text]. J Cutan Med Surg. 2021;25(2):188-204.
2. Barragan-Martinez C, Speck-Hernandez CA, Montoya-Ortiz G, Mantilla RD, Anaya JM, Rojas-Villarraga A. Organic solvents as risk factor for autoimmune diseases: a systematic review and meta-analysis. PLoS One. 2012;7(12):e51506.
3. Stengel B, C,en,ee S, Limasset J-C, et al. Organic solvent exposure may increase the risk of glomerular nephropathies with chronic renal failure. Int J Epidemiol. 1995;24(2):427-434.
4. Jacob S, H,ery M, Protois JC, Rossert J, Stengel B. Effect of organic solvent exposure on chronic kidney disease progression: the GN-PROGRESS cohort study. J Am Soc Nephrol. 2007; 18(1):274-281.



Mendeley Supplementary Figure 1. Kaplan-Meier curve evaluating mortality of systemic sclerosis (SSc) patients over time in organic solvent exposed vs. unexposed groups. Unadjusted Hazard Ratio (HR) (1.26; 95% CI 0.95-1.68, p=0.1033). Adjusted HR 1.14 (95% CI 0.85-1.53, p=0.3726) when adjusted for age, sex, disease duration, and smoking status.

Table 1. Baseline characteristics of patients with systemic sclerosis

Variables	Exposure to organic solvents (*n* = 290)	No exposure (*n* = 1149)	*P* Value[†]
Demographics			
Age ≥ 50 y, *N* (%)	104 (36.7)	467 (41.6)	.136
Male sex, *N* (%)	86 (29.7)	106 (9.2)	**<.001**
Caucasian, *N* (%)	262 (90.3)	1030 (89.7)	.839
Disease duration ≥ 5 y, *N* (%)	161 (56.9)	659 (58.8)	.846
Smoking, *N* (%)	197 (67.9)	656 (57.3)	**.003**
Clinical characteristics			
Diffuse cutaneous disease, *N* (%)	118 (41.1)	400 (35.1)	.163
Treatment with CYC or MMF, *N* (%)	32 (11.1)	92 (8.1)	.258
ILD, *N* (%)	90 (31.7)	339 (30.3)	.811
FVC < 70, *N* (%) (*n* = 1244)	32 (12.6)	116 (11.7)	.832
PAH, *N* (%) (*n* = 909)*	21 (13.0)	127 (17.0)	**.005**
Rodnan skin score, median (IQR)	7.00 [3.00, 14.75]	6.00 [2.00, 14.00]	.172
GI-14, median (IQR) (*n* = 1343)	4.00 [2.00, 7.00]	3.00 [1.00, 6.00]	<**.001**
Digital ulcer/pitting scars, *N* (%)	133 (46.2)	468 (41.2)	.249
Necrosis/gangrene/amputation, *N* (%)	95 (33.0)	410 (36.1)	.645
Nailfold capillaroscopy, *N* (%)	212 (73.1)	872 (75.9)	.386
Joint impairment, *N* (%)	83 (28.9)	317 (27.9)	.931
History of renal crisis, *N* (%)	20 (7.0)	35 (3.1)	**.009**
Cancer, *N* (%)	21 (7.3)	95 (8.3)	.702
Autoantibody profiles			
ACA antibody, *N* (%) (*n* = 1244)	84 (32.7)	398 (40.3)	**.037**
ATA antibody, *N* (%) (*n* = 1244)	47 (18.3)	143 (14.5)	.151
ARNAP, *N* (%) (*n* = 1244)	43 (17)	144 (15)	.39
U1RNP, *N* (%) (*n* = 1276)	13 (4.9)	56 (5.5)	.338
Ro52/TRIM21, *N* (%) (*n* = 1244)	54 (21.0)	277 (28.1)	**.034**
Ku, *N* (%) (*n* = 1244)	2 (0.8)	7 (0.7)	.478
NOR90, *N* (%) (*n* = 1244)	5 (1.9)	22 (2.2)	.463
Nucleolar antibodies, *N* (%) (*n* = 1243)	54 (21.1)	196 (19.9)	.52
Medsger severity scores mean (SD)			
General	0.97 (1.24)	0.88 (1.18)	.262
Peripheral vascular (*n* = 1156)	1.70 (1.23)	1.64 (1.24)	.475
Skin	1.25 (0.67)	1.22 (0.71)	.545
Joint/tendon (*n* = 1095)	0.72 (1.17)	0.70 (1.20)	.821
Muscle	0.24 (0.78)	0.23 (0.73)	.883

Gastrointestinal tract	2.02 (0.71)	1.90 (0.80)	**.019**
Lung	1.37 (1.12)	1.30 (1.12)	.348
Heart	0.46 (0.97)	0.47 (0.97)	.86
Kidney (*n* = 1266)	0.16 (0.68)	0.10 (0.59)	.164
Total (*n* = 832)	9.14 (4.34)	8.56 (3.93)	.085

Bold indicates significant values. Unless a different denominator (*n*) is indicated, the missing number for remaining variables was <5%. *ACA*, Anticentromere antibody; *ARNAP*, anti-RNA polymerase III antibody; *ATA*, anti-topoisomerase I antibody; *CYC*, cyclophosphamide; *FVC*, forced vital capacity; *GI-14*, gastrointestinal 14-question based score; *ILD*, interstitial lung disease; *IQR*, interquartile range; *MMF*, mycophenolate mofetil; *PAH*, pulmonary arterial hypertension; *U1RNP*, anti-U1 ribonucleoproteins antibody.

*PAH corresponds to a systolic pulmonary artery pressure of [45 mmHg on right heart echocardiogram.

[†]Chi-square/Fisher exact tests were used for categorical variables and ANOVA/Kruskal-Wallis test for continuous variables.

Table 2. Association between health outcomes and exposure to organic solvents (yes vs. no) among patients with systemic sclerosis

	N	**Unadjusted OR (95% CI)**	**Adjusted OR (95% CI)**
ILD (yes vs no)*	1371	1.07 (0.81-1.41)	0.92 (0.67-1.25)
Severe ILD*	370	0.91 (0.52-1.60)	0.82 (0.42-1.58)
Diffuse disease[†]	1397	1.29 (0.99-1.68)	1.10 (0.82-1.45)
ATA[†]	1222	1.31 (0.92-1.90)	1.27 (0.85-1.85)
ACA[†]	1222	**0.72 (0.54-0.96)**	0.82 (0.63-1.15)
ARNAP[†]	1222	1.18 (0.81-1.71)	1.08 (0.70-1.66)
Renal crisis[‡]	1398	**2.36 (1.34-4.16)**	**2.13 (1.15-3.93)**
PAH[§]	887	0.73 (0.45-1.21)	0.80 (0.47-1.34)
Age of systemic sclerosis onset (≥50 y) [†]	1395	0.81 (0.61-1.06)	0.77 (0.57-1.01)
GI-14[II]	1397	**Unadjusted β (95% CI) 0.95 (0.55-1.36)**	**Adjusted β (95% CI) 0.89 (0.47-1.31)**

Bold indicates statistically significant values.

ACA, Anticentromere antibody; *ARNAP*, anti-RNA polymerase III antibody; *ATA*, anti-topoisomerase I antibody; *GI-14*, gastrointestinal 14-question based score; *ILD*, interstitial lung disease; *OR*, odds ratio; *PAH*, pulmonary arterial hypertension.

*Model adjusted for age, sex, diffuse disease, disease duration, ethnicity, smoking, and silica.

[†]Model adjusted for age, sex, smoking, and ethnicity.

[‡]Model adjusted for age, sex, ethnicity, and diffuse disease.

[§]Model adjusted for age, sex, smoking, ethnicity, and disease duration.

[II]Model adjusted for age, sex, disease duration, diffuse disease, immunosuppressive medication, smoking, and silica.

Mendeley Supplementary Table 1. Mortality case count and incidence rate per 1,000 person years for systemic sclerosis (SSc). 'Exposed' corresponds to exposed to organic solvents.

	N of deaths	Person-year	IR per 1000 person-year (95% CI)
All patients (N=1202)			
Exposed (n=247)	64	1183	54 (42-69)
Not exposed (N=955)	196	4549	43 (37-49)
Male patients (N=162)			
Exposed (N=72)	30	298	101 (68-144)
Not exposed (N=90)	26	382	68 (44-99)
Female patients (n=1040)			
Exposed (N=175)	34	885	38 (27-54)
Not exposed (N=865)	170	4167	41 (35-47)
Age <50 years (N=712)			
Exposed (N=158)	33	784	42 (29-59)
Not exposed (N=554)	91	2733	33 (27-41)
Age ≥50 years (N=490)			
Exposed (N=89)	31	399	78 (53-110)
Not exposed (N=401)	105	1817	58 (47-70)
Disease duration < 5 years (N=493)			
Exposed (N=104)	23	466	49 (31-74)
Not exposed (N=389)	75	1730	43 (34-54)
Disease duration ≥ 5 years (N=709)			
Exposed (N=143)	41	718	57 (41-78)
Not exposed (N=566)	121	2818	43 (36-51)

Chapter 7: Tree-Based Machine Learning to Identify Predictors of Psoriasis Incidence at the Neighborhood Level

Preamble to Manuscript 5

From the previous studies discussed, it is evident that exogenous exposures can play an important role in triggering a disease which has a genetic predisposition. In this manuscript, we aimed to support these findings and methodology in another disease where the environment is also thought to play an important role: psoriasis.

Psoriasis is a chronic inflammatory skin condition affecting up to 5% of the population in North America. Given the associated symptoms and appearance, there is a significant impact on quality of life. Additionally, it is not a skin-limited condition, with an increased risk of developing numerous systemic comorbidities including other inflammatory diseases (*e.g.* psoriatic arthritis), metabolic disease, mental health comorbidities and malignancies. Studies have shown an increasing prevalence over time, however, estimates of incidence in Canada appear to be decreasing. Although there are safe and effective treatments available in terms of biologics and targeted therapies, these are quite costly to the patient and the healthcare system. While there is a strong genetic predisposition to psoriasis development, only ~30% of patients report a family history of psoriasis hence suggesting that potentially modifiable risk factors such as environmental, SES, behavioral and/or other exposures may be important. Therefore, with the documented increased morbidity, lower quality of life and the significant economic and health systems burden associated with psoriasis, an upstream approach to managing this condition is essential. Hence, psoriasis prevention is of high interest.

To better understand the epidemiology on a populational basis in Quebec, Canada, we used the provincial health administrative database to determine the incidence and prevalence of psoriasis as well as the geographic distribution. Then, using a comprehensive set of environmental and neighbourhood characteristics, a machine learning model was applied to determine which factors that could predict high psoriasis incidence.

In this study, similar to findings in the literature in Canada, we observed an increasing prevalence and a decreasing incidence over the study period. This is the first study to highlight highly variable psoriasis incidence rates on a regional level, suggesting that an individual's living environment can be implicated. For environmental factors, elevated UVR, temperature, vegetation and urbanization metrics showed a negative association with high psoriasis incidence, whereas nighttime light pollution showed a positive association. In the highest incidence neighbourhoods, mid-socioeconomic deprivation was observed. Studies of geographic distribution and contributing factors are important to reveal disparities, target preventative interventions and health resource allocation. Similar to our work on SSc above, and our previous work on mucocutaneous malignancies[108-111], helping identify risk factors can lead to improved patient education, advocacy for interventions to minimize these risk factors and inform resource allocation to high incidence areas.

This study was published in American Journal of Clinical Dermatology.

Muntyanu A, Milan R, Kaouache M, Ringuet J, Gulliver W, Pivneva I, Royer J, Leroux M, Chen K, Yu Q, Litvinov IV, Griffiths CEM, Ashcroft DM, Rahme E, Netchiporouk E. Tree-Based Machine Learning to Identify Predictors of Psoriasis Incidence at the Neighborhood Level: A

Populational Study from Quebec, Canada. *Am J Clin Dermatol. 2024 May;25(3):497-508. doi: 10.1007/s40257-024-00854-3.*

This study was featured in Analysis Group Press release. https://www.analysisgroup.com/news-and-events/press-releases/analysis-group-researchers-identify-neighborhood-characteristics-that-may-predict-psoriasis-in-quebec-canada/

Tree-Based Machine Learning to Identify Predictors of Psoriasis Incidence at the Neighborhood Level: A Populational Study from Quebec, Canada

Anastasiya Muntyanu[1,2], Raymond Milan[1], Mohammed Kaouache[3], Julien Ringuet[4], Wayne Gulliver[5], Irina Pivneva[6], Jimmy Royer[6], Max Leroux[6], Kathleen Chen[6], Qiuyan Yu[7], Ivan V. Litvinov[8], Christopher E. M. Griffiths[9], Darren M. Ashcroft[10], Elham Rahme[11], Elena Netchiporouk[8]

Elena Netchiporouk elena.netchiporouk@mcgill.ca

Wayne Gulliver WGulliver@newlabresearch.com

[1] Department of Experimental Medicine, McGill University, Montreal, Canada

[2] Division of Dermatology, University of Toronto, Toronto, Canada

[3] Centre for Outcomes Research and Evaluation, Research Institute of the McGill University Health Centre, Montreal, QC, Canada

[4] Centre de Recherche Dermatologique de Québec, Québec, Canada

[5] Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada

[6] Analysis Group, Montreal, Canada

[7] Ecological and Biological Sciences, Exponent Inc, Menlo Park, USA

[8] Division of Dermatology, Department of Medicine, McGill University Health Centre, Montreal General Hospital, Montreal, QC, Canada

[9] Centre for Dermatology Research, University of Manchester, Global Psoriasis Atlas, Manchester, UK

[10] School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Global Psoriasis Atlas, Manchester, UK

[11] Department of Medicine, Division of Clinical Epidemiology, McGill University, Montreal, QC, Canada

Abstract

Background. Psoriasis is a major global health burden affecting ~ 60 million people worldwide. Existing studies on psoriasis focused on individual-level health behaviors (e.g. diet, alcohol consumption, smoking, exercise) and characteristics as drivers of psoriasis risk. However, it is increasingly recognized that healthy behavior arises in the context of larger social, cultural, economic and environmental determinants of health. We aimed to identify the top risk factors that significantly impact the incidence of psoriasis at the neighborhood level using populational data from the province of Quebec (Canada) and advanced tree-based machine learning (ML) techniques.

Methods. Adult psoriasis patients were identified using International Classification of Disease (ICD)-9/10 codes from Quebec

(Canada) populational databases for years 1997–2015. Data on environmental and socioeconomic factors 1 year prior to psoriasis onset were obtained from the Canadian Urban Environment Health Consortium (CANUE) and Statistics Canada (StatCan) and were input as predictors into the gradient boosting ML. Model performance was evaluated using the area under the curve (AUC). Parsimonious models and partial dependence plots were determined to assess directionality of the relationship.

Results. The incidence of psoriasis varied geographically from 1.6 to 325.6/100,000 person-years

in Quebec. The parsimonious model (top 9 predictors) had an AUC of 0.77 to predict high psoriasis incidence. Amongst top predictors, ultraviolet (UV) radiation, maximum daily temperature, proportion of females, soil moisture, urbanization, and distance to expressways had a negative association with psoriasis incidence. Nighttime light brightness had a positive association, whereas social and material deprivation indices suggested a higher psoriasis incidence in the middle socioeconomic class neighborhoods.

Conclusion. This is the first study to highlight highly variable psoriasis incidence rates on a jurisdictional level and suggests that living environment, notably climate, vegetation, urbanization and neighborhood socioeconomic characteristics may have an association with psoriasis incidence.

Key Points

Higher ultraviolet radiation, temperature, vegetation and urbanization metrics were protective, whereas nighttime light pollution was associated with higher psoriasis incidence.

The highest incidence neighborhoods were characterized by mid-socioeconomic deprivation.

Studying neighborhood-level risk factors is important to reveal disparities in psoriasis risk across different geographic regions/population groups, forecast healthcare resource need and develop targeted populational interventions.

Introduction

Psoriasis is a chronic immune-mediated skin disease affecting ~60 million people worldwide with higher incidence in industrialized countries [1, 2]. In 2013, the World Health Organization recognized psoriasis as a major global health problem and a recent Global Burden of Disease (GBD) study ranked psoriasis as the second contributor to all skin-related disability-adjusted life-years, imposing a significant medical and economic burden on both the patient and society [1]. The identification of modifiable risk factors is crucial for development of effective prevention strategies and reduction of psoriasis burden worldwide.

Existing studies on psoriasis focused on individual-level health behaviors (e.g. diet, alcohol consumption, smoking, exercise), socioeconomic and demographic characteristics [3]. It is increasingly recognized that health behavior is not only driven by individual factors, but rather arises in the context of larger social, cultural, economic and environmental determinants of health [4]. Population health research in chronic diseases comorbid with psoriasis (e.g. diabetes mellitus, cardiovascular diseases) highlighted the importance of the neighborhood characteristics (the environment in which people live, work, and play) as a critical element to address population-level health differences [5–10], a cost- effective strategy to reduce chronic disease rates [11]. Unfortunately, little is known regarding neighborhood risk factors and the risk of developing psoriasis [12].

To address these research gaps, we aimed to identify a set of risk factors that significantly impact the incidence of psoriasis at the neighborhood level using populational data from the province of Quebec (Canada) and state-of-the-art advanced tree-based machine learning (ML) techniques.

Methods

Study Setting and Data Sources

This cross-sectional study, with yearly analysis, was reported in accordance with the STROBE checklist and was approved by the institutional ethics review board (#2021-6714).

The population of Quebec, the largest Canadian province by land area, was ~8 million in 2015 [13]. Quebec has a universal health care system, where all medical care is provided free of charge to its residents. Physician billing codes (based on the International Classification of Diseases [ICD]) are recorded in the *Régie de l'Assurance Maladie du Québec* (RAMQ) database and hospitalization data in the *Maintenance et Exploitation des Données pour l'Étude de la Clientèle Hospitalière* (MED-ECHO). RAMQ/MED-ECHO data for years 1997–2015 was used.

Data on > 400 neighborhood covariates including air/ noise/light pollution measures, greenness indicators, bluespaces, climate metrics, socioeconomic neighborhood characteristics (e.g. material/social deprivation, instability, marginalization) and built environment features (e.g. urbanization, proximity measures, active living environment [ALE]) were obtained from the Canadian Urban Environmental Health Research Consortium (CANUE) [14]. Proportion of females and median age for each forward sortation area (FSA, the first 3 digits of a postal code) were obtained from Statistics Canada (StatCan) census years (2001, 2006 and 2011) [15]. When needed, interpolation was used to account for missing data. All the variables included were continuous except for a binary variable for rural regions.

Case Definition

A previously defined algorithm was used where a diagnosis of psoriasis (ICD-9 and -10

codes -696.1, L40.x) was defined as the presence of either two or more billing codes by any physician in the physician billing claims database (RAMQ) within 2 years or one or more hospitalization with psoriasis as primary or secondary diagnosis (MED-ECHO database) [16, 17].

Inclusion Criteria

Adult psoriasis patients (≥ 20 years old) benefitting from the provincial drug insurance plan were included [18].

Data Analysis

Incidence

Years 1997–1998 were excluded to account for incidence and the year 2015 for case definition criteria. Annual incidence rate per FSA was determined per 100,000 people in the population. Annual population per FSA of individuals ≥ 20 years old with a provincial drug plan for corresponding years was used as the denominator. The FSA is the smallest possible geographic unit due to confidentiality rules and serves ~8000 households [19].

Geographic Mapping

ArcGIS 10.8 software was used to illustrate the geographic distribution of the average annual incidence rates as previously described [20–22].

Data Linkage

We generated a large neighborhood-level dataset by integrating information at the FSA level from RAMQ/MED- ECHO populational administrative databases for the years 1999–2014 and the CANUE database for corresponding years.

Machine Learning (ML) Model

The primary objective of the ML model was to use exposure data (neighborhood variables) to predict the likelihood of each FSA to belong to the top 10% incidence FSA (vs the other 90%). Exposure data from one year was used to predict incidence in the next year (e.g. neighborhood data in 2000 was used to predict psoriasis incidence in 2001) (Fig. S1, see electronic supplementary material [ESM]).

In the early stages of the project, several tree-based ML methods were tested and gradient boosting performed the best. Gradient boosting is a tree-based method that iteratively generates decision trees using randomly selected features and has been effectively used to predict health out-comes [11, 23–25]. Gradient boosting does not impose a limit on the number of predictors included in the model, allows a nonlinear relationship between predictors and out- come and can account for complex interactions between variables when building a prediction. Hyperparameters of the model (e.g. number of decision trees, tree depth, number of features considered at each split) were optimized through 5-fold cross-validation.

All available variables from CANUE for the study period were considered in order to decrease selection bias. Cor- relation networks were constructed to classify potential explanatory variables into clusters of highly correlated variables. Principal Component Analysis (PCA) was adapted to select the feature that contributed the most to the component (determined based on loading values) [26]. If, however, based on expert knowledge, a different variable within the cluster seemed more pertinent then it was selected and the PCA-selected one was dropped [26].

The final set of variables were used as inputs to the gradient boosting model. The model

was trained on 66% of the data, while 33% was held out for validation. Model performance was evaluated by the area under the curve (AUC) on the validation set. All analyses were conducted using scikit- learn (version 1.2.1) in Python (version 3.9.16). Additional details can be found in Supplementary Methods (see ESM). The output provided by the model included ranking of the variables in order of importance. Partial dependence plots were generated to illustrate graphically the effect and the directionality of each variable on the outcome while taking into account the influence of all other variables in the model. Parsimonious models were trained by adding features one at a time in order of importance, and model performance was assessed to determine a smaller set of most important predictors with similar predictive power to the full model (based on AUC). Results were reviewed during an expert consensus meeting to choose a model with a small number of clinically relevant variables and good accuracy.

Results

Study Population

A total of 43,663 patients with psoriasis were included **(**Table 1**)**. Of these, 22.9% were aged 20-44 years, 37.8% were 45–64 years, 21.8% were 65–74 years, and 17.5% were 75 years or older. Females comprised 50% of the cohort.

Incidence

A significant reduction in the crude annual incidence of psoriasis (slope = 0.264; R2 = 0.888; $p < 0.0001$) was recorded from 1999 (135.55/100,000 person-years) to 2014 (90.93/100,000 person-years) (Fig. S2, see ESM). A total of 413 FSAs in Quebec were included. Average annual incidence per FSA varied from 1.6 to 325.6 cases per 100,000 person-years.

A non-uniform geographic distribution of psoriasis incidence was observed during the study period **(**Figs 1A–C**)**, with higher incidence rates in northern and eastern Quebec. Variability in incidence of psoriasis per FSA was also seen within the Montreal metropolitan area. Lower incidence FSAs were generally clustered in the island of Montreal (Quebec's largest city).

Main Prediction Model

Of all the covariates considered, 48 neighborhood factors were included in the ML. Forty-six had some impact on predicting the probability of high psoriasis incidence (top 10%), where the higher ranking features were more important (Table 2**)**. These spanned domains of climate, socioeconomic neighborhood characteristics and built environment features. The final model had an AUC of 0.80.

Parsimonious Models

The model with the top four features including UV index (measured by average value of mean noon vitamin D index at sea level for the months of June, July and August), annual average of daily maximum temperature, social deprivation (measured by social factor score quintile) and proportion of females in the FSA had a good performance (AUC 0.71). However, the nine-variable parsimonious model had an AUC of 0.77 and was retained **(**Fig. 2**)**. The additional five features in this model included material deprivation (measured by the deprivation principal factor score of the Canadian marginalization index), soil moisture, nighttime light brightness, open urban form and distance to expressways.

Directionality Assessment

Partial dependence plots for the top variables are shown in Figs 3A–D and S3-18 (see ESM). Higher values of the top two predictors, the UV index and temperature, had a negative association (i.e. clinically protective) with psoriasis incidence. For temperature, there was a sharp decrease in the predicted probability of high psoriasis incidence in FSAs with mean daily maximum temperature >9 degrees Celsius. Similarly, higher soil moisture (a surrogate marker for vegetation) had a negative association with high psoriasis incidence (Fig. S3, see ESM).

Socioeconomic neighborhood characteristics such as social (ranked 3rd) and material (ranked 5th) deprivation were important predictors of psoriasis incidence. For both variables, the probability of high psoriasis incidence increased from the lowest to middle deprivation and decreased thereafter. Highest incidence FSAs were characterized by a higher proportion of middle social and material deprivation neighborhoods (Figs 3C, S19A–B [see ESM]).

In terms of demographic factors, the proportion of females (ranked 4th) had a negative association with high psoriasis incidence **(**Fig. 3D**)**. Ethnic concentration ranked 11th and hence was not included in the parsimonious model. Nevertheless, at very low levels of ethnic diversity, there was a positive relationship with psoriasis incidence but after a certain level, the trend was reversed (Fig. S5, see ESM).

The remaining variables in the top nine predictors of high psoriasis incidence included nighttime light brightness, open urban local climate zone and distance to expressways (Figs. S6–8, see ESM). Higher nighttime light was associated with higher psoriasis incidence. At low levels of urbanization (seen in the majority of FSAs), the probability of high psoriasis incidence

decreased as urbanization increased. At high levels of urbanization (minority of FSAs), psoriasis incidence did not appear to be affected by increased urbanization. Within a short distance of expressways, psoriasis incidence increased and decreased thereafter.

Additional factors of interest that were not included in the final parsimonious model (i.e. less important predictors) were air pollutants including particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2), and ozone (O_3). For $PM_{2.5}$ and NO_2, a positive relationship with psoriasis incidence was observed. However, for O_3, the trend was opposite (Figs S9–12, see ESM).

Discussion

In this study we document an uneven geographic distribution of psoriasis incidence on a province/country level. The geographic variability in psoriasis incidence observed in the province of Quebec (varying from 1.6 to 325.6/100,000 person-years) was comparable to the variation seen globally (30.3/100,000 to 321.0/100,000 person-years) [2]. Higher incidence FSAs were clustered together and seen in Northern Quebec where the climate is colder and UV irradiation is lower. Additionally, areas with higher incidence bordered Newfoundland and Labrador, a province that is reputed to have the highest Canadian prevalence of psoriasis [27]. Our findings support previous studies noting higher prevalence of psoriasis favoring drier/colder climate zones in Spain and Brazil [28, 29]. We also report a decline in incidence between 1999 and 2014, similar to previously published reports in other Canadian provinces and worldwide (e.g. United Kingdom, Russia, Italy and Taiwan) [2, 30–33].

Our main objective was to identify a set of predictors that can significantly impact the incidence of psoriasis at the neighborhood level using populational data from the province of

Quebec. Nine predictors (parsimonious model) were identified with a prediction accuracy of 0.77 (AUC) for the top 10% incidence FSAs during the study period. These spanned climate, socioeconomic/demographic factors and built environment characteristics.

The top two ranked predictors were UV index and average daily temperature, where higher values were negatively associated with psoriasis incidence (i.e. protective). Indeed, phototherapy is an efficacious therapy for psoriasis [34]. Similarly, the protective effect of ambient UV exposure against seasonal aggravations was demonstrated in previous studies [35, 36]. Hence, it is plausible that ambient UV exposure can reduce disease severity so that patients with milder disease may not seek medical care. As a result, a lower psoriasis incidence would be observed. Whether UV exposure may modulate the risk of developing psoriasis would require further research [37]. Globally, the effects of UV exposure on psoriasis are suggested through observations of lower incidence rates in countries closer to the equator [1, 2]. A sharp decrease in the probability of high psoriasis incidence was seen in FSAs with mean temperature >9 degrees Celsius. Warmer temperatures may favor more outdoor time and larger surface area of the skin being uncovered, leading to more UV exposure. More outdoor time may also be associated with active living (lower probability of sedentarism) [38]. A survey study of psoriasis patients found that cold temperatures led to worsening of their skin and joint disease [39].

Increased monthly minimum soil moisture (which has been shown to correlate with the Normalized Difference Vegetation Index [NDVI]) [40, 41] had a protective effect towards psoriasis incidence. Higher greenness is known to mitigate harmful exposures and promote a healthy lifestyle [42–46]. Additional greenness metrics identified among predictors (but not in the top nine) included higher annual maximum NDVI, which was similarly associated with a

protective effect. While this is the initial study evaluating the effect of NDVI on psoriasis, lower NDVI has been well correlated with risk of metabolic, cardiovascular and mental health conditions which are comorbidities often seen in patients with psoriasis [43, 47–50]. Green space planning in high-risk neighborhoods could be an important aspect to mitigate the burden of psoriasis, however, further studies are required on the topic.

Social deprivation, which is an indicator of proportion of those separated, divorced or widowed, living alone or in a single-parent family ranked third. Material deprivation which factors in residential instability, economic dependency, ethno-cultural composition and situational vulnerability ranked fifth. High-incidence FSAs were characterized by a higher proportion of middle social and material deprivation neighborhoods. Although not in the top nine ranking variables, additional proxy for socioeconomic status such as quintiles of instability (proportion of the population who moved in the last 5 years, proportion of individuals living alone, and the proportion of rented units compared with owned), material factor score (low income, low education and low employment to population ratio), and quintiles of dependency (proportion of individuals that do not have income from employment) showed a similar association. To our knowledge, only two prior studies looked at the socio-economic deprivation and psoriasis prevalence/severity and reported a positive association; however, higher incidence in middle socioeconomic neighborhoods is a novel trend reported in our study [51, 52]. Middle socioeconomic class individuals may have different lifestyle patterns, access to resources, and healthcare utilization compared with other socioeconomic groups. Considering the free universal health care in Canada, affordable and accessible education and multiple social assistance programs, our findings may not be generalizable to other jurisdictions. A thorough understanding

of the socioeconomic determinants of health and their mechanism in psoriasis is essential to reduce health disparities and must be studied further.

While several demographic features were explored, including age and ethnic concentration, only the proportion of females in an FSA was among the top predictors. High proportion of females in an FSA showed a trend towards lower psoriasis incidence. These findings are aligned with the literature suggesting more severe and possibly more common disease in males [53].

In regard to other neighborhood characteristics, an increase in nighttime light brightness was associated with a higher incidence of psoriasis. Excessive or artificial light produced by human activities has been strongly associated with a disruption in the normal circadian rhythm and contributed to development of metabolic dysfunction, mood disorders and cancer [54]. Studies showed elevated cardiometabolic risk with elevated exposure levels and a 28% increase in risk of developing diabetes mellitus [55, 56]. While studies assessing light pollution in dermatology are scarce, especially for psoriasis, a Chinese study showed an association between artificial light at night in adolescence and development of atopic conditions [57].

Open urban, ranked seventh, represents the percent of pixels in a 1-km^2 neighborhood with high- to mid-rise buildings arranged in an open arrangement with scattered trees and plant cover as well as a moderate traffic flow [58]. Psoriasis incidence appeared to decrease with increased urbanization (i.e. increasing building density), while holding all other variables constant. A similar trend was noted for the dense urban variable (i.e. tall and medium-height buildings densely packed with little greenness and busy road traffic). Possibly this could be attributed to better access to educational/employment facilities, sports and food options as well as good active living

environment scores (i.e. walking distance to points of interest and efficient public transportation) favoring physical activity. Until now, there was no data in psoriasis looking at the association between incidence and different land-use forms. A handful of studies compared urban versus rural living and psoriasis prevalence with conflicting results [59, 60]. However, in other chronic conditions, previous studies showed that favorable active living characteristics were associated with a lower risk of metabolic and cardiovascular diseases [61].

This study provides evidence for the importance of neighborhood characteristics in regard to psoriasis risk. While there are several strengths to our study, our results must be interpreted within its context. Patients were identified based on ICD codes from populational databases using previously defined algorithms, however some misclassification cannot be ruled out. It is reassuring that the incidence rates and trends identified in our study were similar to those reported in the literature previously. While the model's overall performance was high (AUC 0.77 for the top nine variables taken together), the prediction impact of each feature taken separately on psoriasis incidence was beyond the scope of this work and will be a subject of future work. The tree-based ML algorithms have been shown to be biased for selecting categorical variables, but in our study, only one such variable (rural area) was included with all others being continuous. A sensitivity analysis was attempted by training two distinct models, one for rural and one for non-rural areas, however there were too few rural FSAs to build a well-performing model. Considering the population health focus, we cannot draw conclusions about a patient's individual risk factors. Given this study was based on the population in Quebec, it may not be generalizable to all populations and in the future we aim to confirm our findings in other populations.

Conclusion

To our knowledge, this is the first study of its kind to systematically and comprehensively assess neighborhood risk factors spanning domains of climate, pollution metrics, socioeconomic, sociodemographic and built environment characteristics in predicting psoriasis incidence using populational data. Using a state-of-the-art ML technique, we identified that a set of nine predictors had an AUC of 0.77 to predict high incidence areas of psoriasis in Quebec between 1999 and 2014. Unveiling differential exposures and vulnerabilities will help to advance our understanding of psoriasis pathogenesis and allow for populational interventions for psoriasis prevention similar to other chronic diseases [62].

References

1. Mehrmal S, Uppal P, Nedley N, Giesey RL, Delost GR. The global, regional, and national burden of psoriasis in 195 countries and territories, 1990 to 2017: a systematic analysis from the Global Burden of Disease Study 2017. J Am Acad Dermatol. 2021;84:46–52.
2. Parisi R, Iskandar IYK, Kontopantelis E, Augustin M, Griffiths CEM, Ashcroft DM, et al. National, regional, and worldwide epidemiology of psoriasis: systematic analysis and modelling study. BMJ. 2020;369: m1590.
3. Elmets CA, Leonardi CL, Davis DMR, Gelfand JM, Lichten J, Mehta NN, et al. Joint AAD-NPF guidelines of care for the management and treatment of psoriasis with awareness and attention to comorbidities. J Am Acad Dermatol. 2019;80:1073–113.
4. Biskup M, Macek P, Gozdz S, Terek-Derszniak M, Krol H, Kedzierawski P, et al. Two-year follow-up cohort study focused on gender-specific associations between socioeconomic status and body weight changes in overweight and obese middle-aged and older adults. BMJ

Open. 2021;11: e050127.

5. Renalds A, Smith TH, Hale PJ. A systematic review of built environment and health. Fam Community Health. 2010;33:68–78.
6. Chandrabose M, Rachele JN, Gunn L, Kavanagh A, Owen N, Turrell G, et al. Built environment and cardio-metabolic health: systematic review and meta-analysis of longitudinal studies. Obes Rev. 2019;20:41–54.
7. Durand CP, Andalib M, Dunton GF, Wolch J, Pentz MA. A systematic review of built environment factors related to physical activity and obesity risk: implications for smart growth urban planning. Obes Rev. 2011;12:e173–82.
8. Albrecht SS, Osypuk TL, Kandula NR, Gallo LC, Lê-Scherban F, Shrager S, et al. Change in waist circumference with longer time in the United States among Hispanic and Chinese immigrants: the modifying role of the neighborhood built environment. Ann Epidemiol. 2015;25:767-72.e2.
9. Bird C. The built environment and health. Perspect Public Health. 2012;132:105.
10. Prince SA, Lancione S, Lang JJ, Amankwah N, de Groh M, Jaramillo Garcia A, et al. Examining the state, quality and strength of the evidence in the research on built environments and physical activity among children and youth: an overview of reviews from high income countries. Health Place. 2022;76: 102828.
11. Hu L, Liu B, Ji J, Li Y. Tree-based machine learning to identify and understand major determinants for stroke at the neighborhood level. J Am Heart Assoc. 2020;9: e016745.
12. Neighborhood characteristics and the risk of psoriasis. A system- atic review. JAAD. International. 2023;13:100–1.

13. Government of Canada, Canada S. Population and Dwelling Count Highlight Tables, 2016 Census [Internet]. 2017 [cited 2023 Jun 2]. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Comprehensive.cfm
14. CANUE - The Canadian Urban Environmental Health Research Consortium [Internet]. [cited 2023 May 6]. https://www.canue data.ca/
15. Government of Canada, Canada S. Census Profile, 2016 Census - V5Y [Forward sortation area©] and Canada [Country] [Inter- net]. 2017 [cited 2023 May 8]. https://www12.statcan.gc.ca/ census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=FSA&Code1=V5Y&Geo2=PR&Code2=01&Data= Count&SearchText=V5Y&SearchType=Begins&SearchPR= 01&B1=All
16. Marrie RA, Patten SB, Tremlett H, Wolfson C, Leung S, Fisk JD. Increased incidence and prevalence of psoriasis in multiple sclerosis. Mult Scler Relat Disord. 2017;13:81–6.
17. Eder L, Widdifield J, Rosen CF, Alhusayen R, Cheng SY, Young J, et al. Identifying and characterizing psoriasis and psoriatic arthritis patients in ontario administrative data: a population-based study from 1991 to 2015. J Rheumatol. 2020;47:1644–51.
18. Milan R, LeLorier J, Brouillette M-J, Holbrook A, Litvinov IV, Rahme E. sex differences in the patterns of systemic agent use among patients with psoriasis: a retrospective cohort study in quebec. Canada Front Pharmacol. 2022;13: 810309.
19. More information on Postal code [Internet]. 2008 [cited 2022 Aug 12]. https://www12.statcan.gc.ca/census-recensement/2006/ref/ dict/geo035a-eng.cfm
20. Muntyanu A, Ghazawi FM, Nedjar H, Rahme E, Alakel A, Zubarev A, et al. Non-melanoma skin cancer distribution in the russian federation. Dermatology. 2021;237:1007–15.

21. Cattelan L, Ghazawi FM, Le M, Lagacé F, Rahme E, Zubarev A, et al. Geographic and socioeconomic disparity of gastric cancer patients in Canada. Curr Oncol. 2021;28:2052–64.
22. Muntyanu A, Savin E, Ghazawi FM, Alakel A, Zubarev A, Litvi- nov IV. Geographic Variations in Cutaneous Melanoma Distribu- tion in the Russian Federation [Internet]. Dermatology. 2020. p. 500–7. https://doi.org/10.1159/000507617
23. Shung DL, Au B, Taylor RA, Tay JK, Laursen SB, Stanley AJ, et al. Validation of a machine learning model that outperforms clinical risk scoring systems for upper gastrointestinal bleeding. Gastroenterology. 2020;158:160–7.
24. Eaton JE, Vesterhus M, McCauley BM, Atkinson EJ, Schlicht EM, Juran BD, et al. Primary sclerosing cholangitis risk esti- mate tool (PREsTo) predicts outcomes of the disease: a deriva- tion and validation study using machine learning. Hepatology. 2020;71:214–24.
25. Zhang Y, Wang S, Hermann A, Joly R, Pathak J. Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women. J Affect Disord. 2021;279:1–8.
26. Berman-Rosa M, Logan J, Ghazawi FM, Le M, Conte S, Netchi- porouk E, et al. Analysis of geographic and environmental fac- tors and their association with cutaneous melanoma incidence in Canada. Dermatology. 2022;238:1006–17.
27. Badaiki W, Penney M, Pyper E, Lester K, Skeard J, Shin J, et al. Real world studies of psoriasis and mental illness in newfoundland and labrador. J Cutan Med Surg. 2022;26:494–501.
28. Ferrándiz C, Bordas X, García-Patos V, Puig S, Pujol R, Smandía A. Prevalence of psoriasis in Spain (Epiderma Project: phase I). J Eur Acad Dermatol

Venereol. 2001;15:20–3.

29. Romiti R, Amone M, Menter A, Miot HA. Prevalence of psoriasis in Brazil - a geographical survey [Internet]. Int J Dermatol. 2017. p. e167–8. https://doi.org/10.1111/ijd.13604
30. Wei JC-C, Shi L-H, Huang J-Y, Wu X-F, Wu R, Chiou J-Y. Epi- demiology and Medication Pattern Change of Psoriatic Diseases in Taiwan from 2000 to 2013: A Nationwide, Population-based Cohort Study. J Rheumatol. 2018;45:385–92
31. Vena GA, Altomare G, Ayala F, Berardesca E, Calzavara-Pin- ton P, Chimenti S, et al. Incidence of psoriasis and association with comorbidities in Italy: a 5-year observational study from a national primary care database. Eur J Dermatol. 2010;20:593–8.
32. Eder L, Widdifield J, Rosen CF, Cook R, Lee K-A, Alhusayen R, et al. Trends in the prevalence and incidence of psoriasis and psoriatic arthritis in Ontario, Canada: a population-based study. Arthritis Care Res. 2019;71:1084–91.
33. Damiani G, Bragazzi NL, Karimkhani Aksut C, Wu D, Alicandro G, McGonagle D, et al. The global, regional, and national burden of psoriasis: results and insights from the global burden of disease 2019 study. Front Med. 2021;8: 743180.
34. Zhang P, Wu MX. A clinical review of phototherapy for psoriasis. Lasers Med Sci. 2018;33:173–80.
35. Pascoe VL, Kimball AB. Seasonal variation of acne and psoriasis: A 3-year study using the Physician Global Assessment severity scale. J Am Acad Dermatol. 2015;73:523–5.
36. Zheng X, Wang Q, Luo Y, Lu W, Jin L, Chen M, et al. Seasonal variation of psoriasis and its impact in the therapeutic manage- ment: a retrospective study on chinese patients. Clin Cosmet Investig Dermatol. 2021;14:459–65.

37. Karppinen TT, Ylianttila L, Kautiainen H, Reunala T, Snellman E. Empowering heliotherapy improves clinical outcome and qual- ity of life of psoriasis and atopic dermatitis patients. Acta Derm Venereol. 2015;95:579–82.
38. Beyer KMM, Szabo A, Hoormann K, Stolley M. Time spent out- doors, activity levels, and chronic disease among American adults. J Behav Med. 2018;41:494–503.
39. Balato N, Di Costanzo L, Patruno C, Patrì A, Ayala F. Effect of weather and environmental factors on the clinical course of pso- riasis. Occup Environ Med. 2013;70:600.
40. Sharma M, Bangotra P, Gautam AS, Gautam S. Sensitivity of normalized difference vegetation index (NDVI) to land surface temperature, soil moisture and precipitation over district Gau- tam Buddh Nagar, UP. India Stoch Environ Res Risk Assess. 2022;36:1779–89.
41. Adegoke JO, Carleton AM. Relations between Soil Moisture and Satellite Vegetation Indices in the U.S. Corn Belt. J Hydromete- orol. 2002;3:395–405.
42. Ke P, Xu M, Xu J, Yuan X, Ni W, Sun Y, et al. Association of residential greenness with the risk of metabolic syndrome in Chi- nese older adults: a longitudinal cohort study. J Endocrinol Invest [Internet]. 2022; https://doi.org/10.1007/s40618-022-01904-5
43. Feng S, Meng Q, Guo B, Guo Y, Chen G, Pan Y, et al. Joint expo- sure to air pollution, ambient temperature and residential green- ness and their association with metabolic syndrome (MetS): a large population-based study among Chinese adults. Environ Res. 2022;214: 113699.
44. Ponjoan A, Blanch J, Alves-Cabratosa L, Martí-Lluch R, Comas- Cufí M, Cirach M, et al. Impact of residential greenness on myo- cardial infarction in the population with diabetes: a sex-dependent association? Environ Res. 2022;205: 112449.

45. Khan JR, Sultana A, Islam MM, Biswas RK. A negative associa- tion between prevalence of diabetes and urban residential area greenness detected in nationwide assessment of urban Bangla- desh. Sci Rep. 2021;11:19513.
46. Doubleday A, Knott CJ, Hazlehurst MF, Bertoni AG, Kaufman JD, Hajat A. Neighborhood greenspace and risk of type 2 diabetes in a prospective cohort: the multi-ethncity study of atherosclero- sis. Environ Health. 2022;21:18.
47. Yang B-Y, Hu L-W, Jalaludin B, Knibbs LD, Markevych I, Hein- rich J, et al. Association between residential greenness, cardio- metabolic disorders, and cardiovascular disease among adults in China. JAMA Netw Open. 2020;3: e2017507.
48. Yeager R, Riggs DW, DeJarnett N, Tollerud DJ, Wilson J, Conklin DJ, et al. Association between residential greenness and cardio- vascular disease risk. J Am Heart Assoc. 2018;7: e009117.
49. Liu L, Yan LL, Lv Y, Zhang Y, Li T, Huang C, et al. Air pollu- tion, residential greenness, and metabolic dysfunction biomarkers: analyses in the Chinese Longitudinal Healthy Longevity Survey. BMC Public Health. 2022;22:885.
50. de Keijzer C, Basagaña X, Tonne C, Valentín A, Alonso J, Antó JM, et al. Long-term exposure to greenspace and metabolic syn- drome: a Whitehall II study. Environ Pollut. 2019;255: 113231.
51. Li X, Sundquist J, Hamano T, Sundquist K. Neighborhood Dep- rivation and Risks of Autoimmune Disorders: A National Cohort Study in Sweden. Int J Environ Res Public Health [Internet]. 2019;16. https://doi.org/10.3390/ijerph16203798
52. Löfvendahl S, Jöud A, Petersson IF, Theander E, Svensson Å, Carlsson KS. Income

disparities in healthcare use remain after controlling for healthcare need: evidence from Swedish regis- ter data on psoriasis and psoriatic arthritis. Eur J Health Econ. 2018;19:447–

53. Lagacé F, D'Aguanno K, Prosty C, Laverde-Saad A, Cattelan L, Ouchene L, et al. The Role of Sex and Gender in Dermatology - From Pathogenesis to Clinical Implications. J Cutan Med Surg. 2023;12034754231177582.
54. Bedrosian TA, Nelson RJ. Timing of light exposure affects mood and brain circuits. Transl Psychiatry. 2017;7: e1017.
55. Xu Y-X, Yu Y, Huang Y, Wan Y-H, Su P-Y, Tao F-B, et al. Expo- sure to bedroom light pollution and cardiometabolic risk: a cohort study from Chinese young adults. Environ Pollut. 2022;294: 118628.
56. Zheng R, Xin Z, Li M, Wang T, Xu M, Lu J, et al. Outdoor light at night in relation to glucose homoeostasis and diabetes in Chinese adults: a national and cross-sectional study of 98,658 participants from 162 study sites. Diabetologia. 2023;66:336–45.
57. Tang Z, Li S, Shen M, Xiao Y, Su J, Tao J, et al. Association of exposure to artificial light at night with atopic diseases: a cross-sectional study in college students. Int J Hyg Environ Health. 2022;241: 113932.
58. CANUE—The Canadian Urban Environmental Health Research Consortium [Internet]. [cited 2023 May 8]. https://www.canue data.ca/metadata.php
59. Nicolescu AC, Bucur Ștefana, Giurcăneanu C, Gheucă-Solovăstru L, Constantin T, Furtunescu F, et al. Prevalence and Character- istics of Psoriasis in Romania-First Study in Overall Population. J Pers Med [Internet]. 2021;11. https://doi.org/10.3390/jpm11 060523
60. Jankovic S, Raznatovic M, Marinkovic J, Jankovic J, Maksimovic

N. Risk factors for psoriasis: a case-control study. J Dermatol. 2009;36:328–34.

61. Mah SM, Sanmartin C, Riva M, Dasgupta K, Ross NA. Active living environments, physical activity and premature cardiometa- bolic mortality in Canada: a nationwide cohort study. BMJ Open. 2020;10: e035942.
62. Adams J, Mytton O, White M, Monsivais P. Why are some population interventions for diet and obesity more equitable and effective than others? The role of individual agency. PLoS Med. 2016;13: e1001990.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40257-024-00854-3.

Acknowledgements Datasets were accessed via the Canadian Urban Environmental Health Research Consortium (CANUE) Data Portal: https://www.canuedata.ca/. NDVI metrics, indexed to DMTI Spatial Inc. postal codes, were provided by CANUE (Canadian Urban Environ- mental Health Research Consortium). Tree Canopy metrics, indexed to DMTI Spatial Inc. postal codes, were provided by CANUE (Canadian Urban Environmental Health Research Consortium). Canadian Active Living Environments Index (Can-ALE), indexed to DMTI Spatial Inc. postal codes, were provided by CANUE (Canadian Urban Environ- mental Health Research Consortium). Dwellings needing major repair at the postal code level were obtained via the Canadian Urban Environmental Health Research Consortium (CANUE) Data Portal: https:// www.canuedata.ca/ GENUINE: Gentrification, Urban Interventions, and Equity data were provided by CANUE (Canadian Urban Environ- mental Health Research Consortium). Green roads metrics, indexed to DMTI postal codes, were provided by the Canadian Urban Environ-

mental Health Research Consortium (CANUE). DMSP-OLS metrics, indexed to DMTI Spatial Inc. postal codes, were provided by CANUE (Canadian Urban Environmental Health Research Consortium). Esti- mated noise metrics, indexed to DMTI Spatial Inc postal codes, were provided by the Canadian Urban Environmental Health Research Consortium (CANUE). Statistics Canada/CMHC proximity data, indexed to DMTI postal codes, were accessed via the Canadian Urban Environ- mental Health Research Consortium (CANUE) data portal (canuedata.ca). Sprawl score data were accessed via the Canadian Urban Environmental Health Research Consortium (CANUE) data portal (canuedata.ca). Multi-pollutant indices and input air quality data, indexed to DMTI Spatial Inc. postal codes, were provided by CANUE (Canadian Urban Environmental Health Research Consortium). PM2.5 metrics, indexed to DMTI Spatial Inc. postal codes, were provided by CANUE (Canadian Urban Environmental Health Research Consortium).

Declarations

Author contributions Conceptualization: AM, ER, EN. Methodology: AM, RM, MK, IP, JR, ML, KC, QY, ER, EN. Software: IP, JR, ML,

KC. Validation: IP, JR, ML, KC. Formal analysis: RM, MK, IP, JR, ML, KC. Resources (Data Sources): AM, RM, ER, EN. Data curation: AM, RM, MK, IP, JR, ML, KC, ER, EN. Writing—original draft preparation: AM. Writing—review and editing: AM, RM, MK, JR, WG, IP, JR, ML, KC, QY, IV, CG, DA, ER, EN. Visualization (graphs): AM. Supervision: IV, ER, EN. Project administration: EN. Funding acquisition: EN. All authors contributed to the article and approved the submitted version.

Funding No funding sources for this study.

Availability of data and materials (data transparency) Data from ISQ (Institut de la statistique

du Québec) is available following an extensive application and verification process of the study personnel. Due to confidentiality reasons, it cannot be shared outside of the approved research team. Data from CANUE (Canadian Urban Environmental Health Research Consortium) is available upon registering for a project and completing the according data use agreements. Data from Statistics Canada is publicly available online.

Code availability (software application or custom code) Not applicable.

Conflict of interest None to declare.

Ethics approval This study was approved by the institutional ethics review board (#2021-6714).

Consent to participate Not applicable, no individual patient data used in the manuscript.

Consent for publication Not applicable, no individual patient data used in the manuscript.

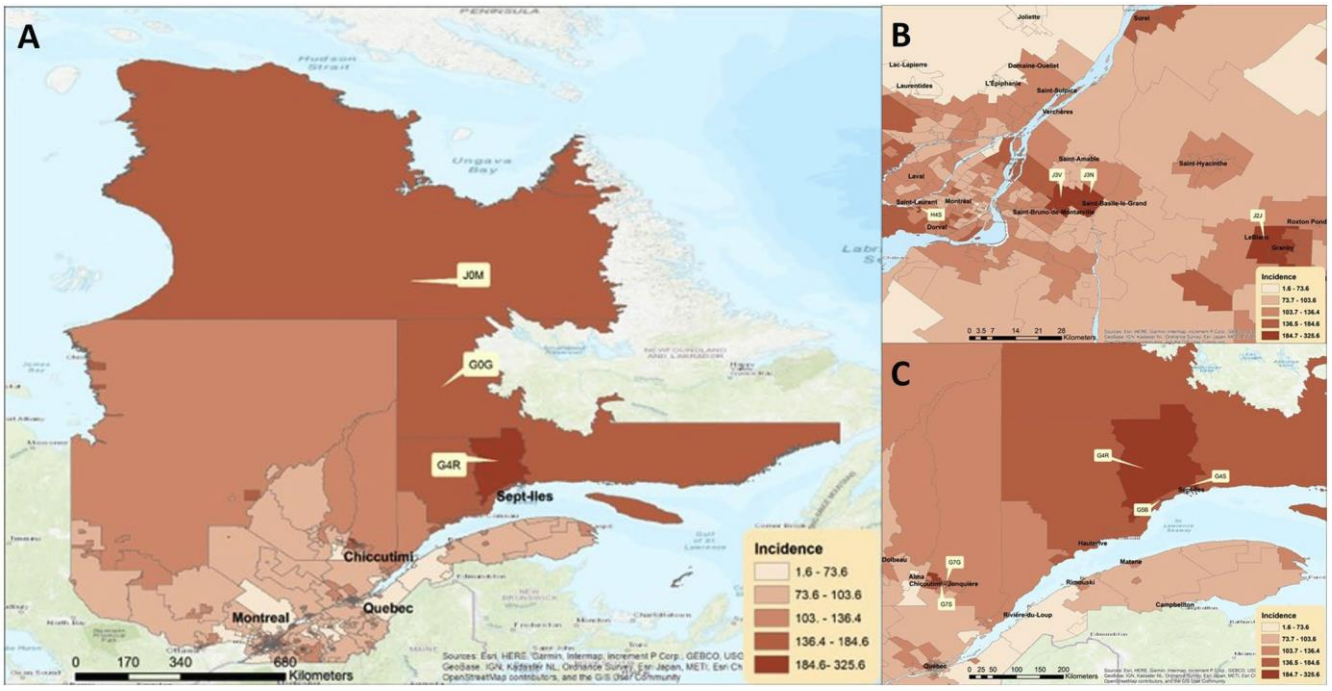


Fig. 1 A Geographic distribution of psoriasis incidence in Quebec over the study period. Zoom in to **B** Montreal Metropolitan and **C** Eastern Quebec. The figure illustrates the geographic variability in average psoriasis incidence rates over the study period.

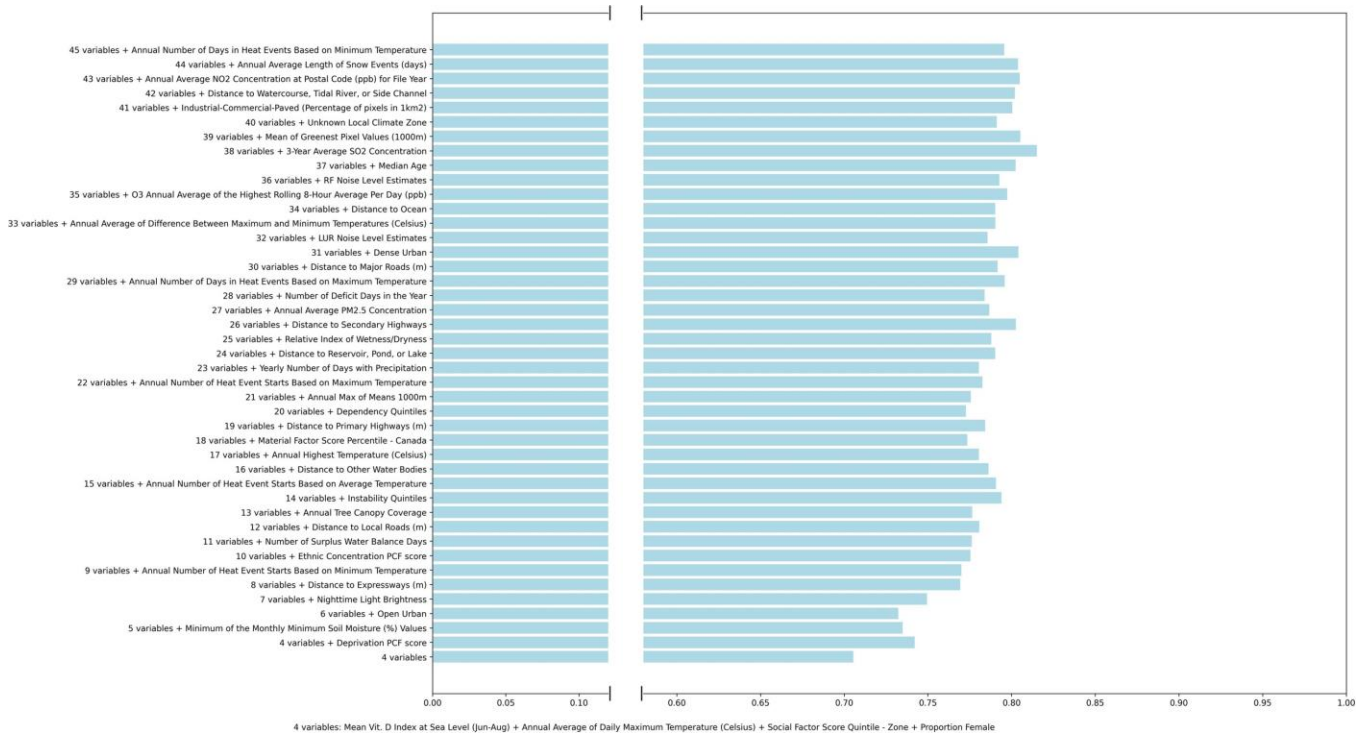


Fig. 2 Parsimonious plots demonstrating the change in the area under the curve with an increasing number of included variables. The four variables with the highest importance were (1) average value of mean noon vitamin D index at sea level for the months of June, July and August, (2) annual average of daily maximum temperature, (3) social factor score quintile within zone, (4) pro- portion female. *ALE* active living environments, *AUROC* area under the receiver operating curve, *LCZ* local climate zone, *LUR* land use regression, *mm* millimetres, *PCF* principal component factor, *PM* particulate matter, *ppb* parts per billion, *RF* random forest

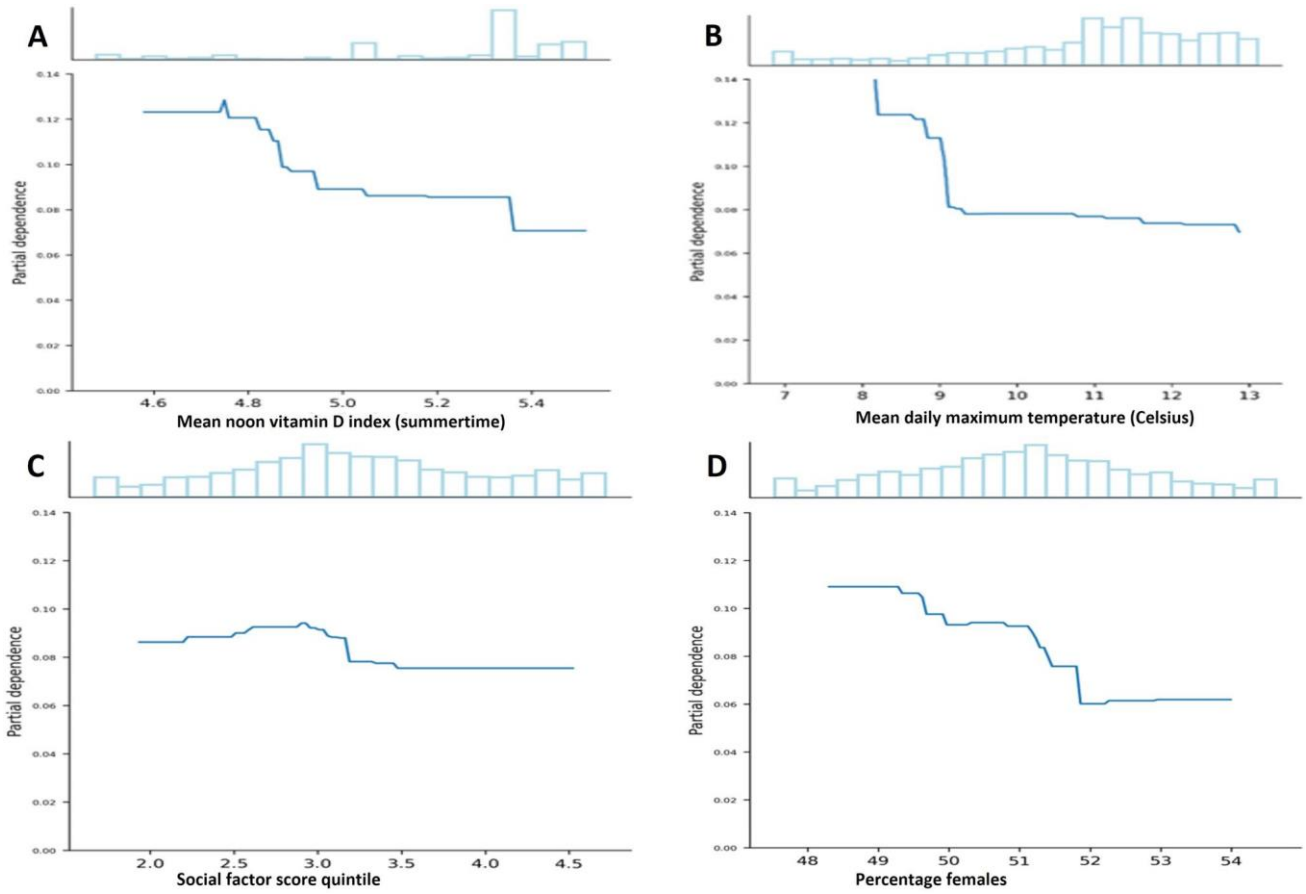


Fig. 3 Partial dependence plot showing directionality of the association between psoriasis incidence and the highest ranked features. (**A**) Higher Vitamin D level (proxy for the ultraviolet radiation index) has a protective effect against high psoriasis incidence (top 10%) at the neighborhood level. (**B**) Higher temperature is similarly associated with a lower psoriasis incidence. (**C**) Moving from the most socially marginalized neighborhoods (left on the x-axis) to mid marginalized (quintiles ~2–3.5), the probability of high psoriasis incidence increases and decreases thereafter (highest social quintiles). (**D**) Higher proportion of females is associated with lower predicted prob- ability of high psoriasis incidence at the neighborhood level

Table 1. Characteristics of patients with psoriasis included in the study

	Sex				Total
	F		M		
	N	%	*N*	%	*N*
Age group, *y*					
20–44	4952	22.7	5058	23.2	10,010
45–64	8069	37	8424	38.6	16,493
65–74	4616	21.2	4898	22.4	9514
>75	4178	19.2	3468	15.9	7646
Year of psoriasis diagnosis					
1999	1649	7.6	1643	7.5	3292
2000	1579	7.2	1549	7.1	3128
2001	1446	6.6	1464	6.7	2910
2002	1390	6.4	1416	6.5	2806
2003	1348	6.2	1365	6.2	2713
2004	1323	6.1	1404	6.4	2727
2005	1391	6.4	1299	5.9	2690
2006	1395	6.4	1297	5.9	2692
2007	1228	5.6	1171	5.4	2399
2008	1294	5.9	1332	6.1	2626
2009	1219	5.6	1288	5.9	2507
2010	1321	6.1	1332	6.1	2653
2011	1351	6.2	1343	6.1	2694
2012	1241	5.7	1330	6.1	2571
2013	1337	6.1	1265	5.8	2602
2014	1303	6	1350	6.2	2653
Total	21,815	100	21,848	100	43,663

Table 2. Ranking of all the environmental features analyzed in the Gradient Boosting Machine Learning Algorithm. The top nine highest ranking features are bolded

Average value of mean noon vitamin D index at sea level for the months of June, July and August	**1**	**0.0065**
Annual average of daily maximum temperature (Celsius)	**2**	**0.0063**
Social factor score quintile within zone	**3**	**0.0027**
Proportion female	**4**	**0.0026**
Principal component factor score—deprivation	**5**	**0.0025**
Minimum of the monthly minimum soil moisture (%) values	**6**	**0.0024**
Open urban (percentage of pixels in 1 km^2)	**7**	**0.0023**
Nighttime light brightness at postal code	**8**	**0.0022**
Distance to expressways (m)	**9**	**0.0021**
Annual number of heat event starts based on minimum temperature	10	0.0019
Principal component factor score—ethnic concentration	11	0.0018
Number of surplus days in the year	12	0.0018
Distance to local roads (m)	13	0.0017
Annual value at postal code	14	0.0017
Quintiles of instability	15	0.0016
Annual number of heat event starts based on average temperature	16	0.0015
Distance to other water bodies	17	0.0015
Annual highest temperature (Celsius)	18	0.0015
Material factor score percentile within Canada	19	0.0013
Distance to primary highways (m)	20	0.0013
Quintiles of dependency	21	0.0012
Annual max of means 1000 m	22	0.0011
Annual number of heat event starts based on maximum temperature	23	0.0011
Number of days in the year with snow on the ground	24	0.0011
Distance to reservoir, pond, or lake	25	0.0010
Relative index of wetness/dryness	26	0.0010
Distance to secondary highways	27	0.00098
Annual average $PM_{2.5}$ concentration (ug/m^3)	28	0.00086
Number of deficit days in the year	29	0.00079
Annual number of days in heat events based on maximum temperature	30	0.00072
Distance to major roads (m)	31	0.0007
Dense urban (percentage of pixels in 1 km^2)	32	0.00069
LUR noise level estimates	33	0.00061
Annual average of difference between maximum and minimum temperatures (Celsius)	34	0.00052

Distance to ocean	35	0.00049
O_3 Annual average of the highest rolling 8-hour average per day (ppb)	36	0.00048
RF noise level estimates	37	0.00046
Median age	38	0.00045
3-year annual average SO_2 concentration (ppb)	39	0.00039
Mean of greenest pixel values 1000 m	40	0.00029
Unknown LCZ (percentage of pixels in 1 km^2)	41	0.00023
Industrial-commercial-paved (percentage of pixels in 1 km^2)	42	0.00024
Distance to watercourse, tidal river, or side channel	43	0.00015
Annual average NO_2 concentration at postal code (ppb) for file year	44	0.000091
Annual average length of snow events (days)	45	0.000081
Annual number of days in heat events based on minimum temperature	46	6.96E-05

ALE active living environments, *LCZ* local climate zone, *LUR* land use regression, *mm* millimeters, *NO_2* nitrogen dioxide, *O_3* ozone, *PM* particulate matter, *ppb* parts per billion, *RF* random forest

Supplementary Methods

Inclusion Criteria

Only patients with public drug insurance coverage (44.3% of all Quebec citizens; 2015 data) were selected as all the data on medications received was available and could be used in the future to assess disease severity trends.

Exposures of Interest

Individual variables were available from CANUE per year for each 6-digit postal code [1-31] which accounts for 19 households on average [32]. To aggregate the 6-digit postal code level data to the FSA level, each postal code was weighted using its respective land area (sum of Local Delivery Unit (LDU) Areas) in the FSA (data available for years 2006-2014) from DMTI Spatial Inc. Postal code areas of 2006 were used to impute missing areas for years 2001 to 2005. Missing values were imputed using interpolation whenever possible (*i.e.* if other years had readings for the same postal code). When no other year was available, the national average for that year was used for imputation. Categorical variables (except for rural area) were excluded from the analysis due to the lack of an adapted method to aggregate categorical values at the FSA level.

Machine Learning Model

In early phases of the project development, we tested models using 1 and 5 years of environmental data, respectively, to predict psoriasis incidence in the next year, but the model with 1 year of prediction data performed better (higher Area Under the Curve [AUC]) and was retained. Similarly, several machine learning (ML) methods were tested and gradient tree boosting had the highest performance and hence, was used to develop the final model.

Gradient boosting is a tree-based method that iteratively generates decision trees using randomly selected features [33]. After the first decision tree is trained, a second tree is trained to predict the error term remaining from the predictions of the first tree, and the predictions of the two trees are subsequently added to update the prediction of the outcome. This process is repeated for a certain number of iterations.

Principal Component Analysis (PCA) was adapted to select a subset of variables with reduced collinearity. The PCA algorithm was applied to the full set of variables to extract the projected components. For each projected component among top components, the feature that contributed the most to the component (determined based on loading values) was retained.

References

1. Latifovic, R., Trishchenko, A., Chen, J., Park, W., Khlopenkov, K., Fernandes, R., Pouliot, D., Ungureanu, C., Luo, Y., Wang, S., Davidson, A., and Cihlar, J. 2005. Generating historical AVHRR 1-km baseline satellite data records over Canada suitable for climate change studies. Canadian Journal of Remote Sensing, vol. 31:324-34.
2. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment.
3. Sexton, J. O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K.M., Channan, S., DiMiceli, C., Townshend, J.R.G. (2013). Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS Vegetation Continuous Fields with lidar-based estimates of error. International Journal of Digital Earth, 130321031236007.

4. Ross, N., Wasfi, R., Herrmann, T., and Gleckner, W., 2018. Canadian Active Living Environments Database (Can-ALE) User Manual
5. Setton E, Redivo A. 2021. Building density metrics. Canadian Urban Environmental Health Research Consortium.\[2]. Microsoft. 2019. Canadian building footprints. Computer file: https://github.com/microsoft/CanadianBuildingFootprints.
6. Matheson et al. "Development of the Canadian Marginalization Index: a new tool for the study of inequality." Canadian Journal of Public Health, 2012;103(Suppl. 2):S12-S16.
7. Statistics Canada. 2023. (table). Census Profile. 2021 Census of Population. Statistics Canada Catalogue no. 98-316-X2021001. Ottawa. Released March 29, 2023.[2] Statistics Canada. 2019 (table). Census Profile. 2016 Census of Population. Statistics Canada Catalogue no. 98-316- X2016001. Ottawa. [3] CanMap Postal Code Suite v2016.3. [Computer file] Markham: DMTI Spatial Inc., 2016.
8. OpenStreetMap contributors. (2019) . Retrieved from GeoFabrik September 10, 2020 at https://download.geofabrik.de/north-america/canada.html.
9. Firth, C.L., Thierry, B., Fuller, D., Winters, M., Kestens, Y., 2020. GENUINE: Gentrification, Urban Interventions, and Equity Technical Document. Gentrification workgroup of INTErventions, Research, and Action in Citites Team. Centre de recherche du Centre hospitalier de l'Université de Montréal.
10. Setton E, Redivo A. Green roads metrics. (2020). Canadian Urban Environmental Health Research Consortium (CANUE). https//:canuedata.ca.
11. Pampalon, Robert, et al. "An Area-Based Material and Social Deprivation Index for Public Health in Québec and Canada." Canadian Journal of Public Health / Revue

Canadienne De Santé Publique, vol. 103, 2012, pp. S17–S22. JSTOR, JSTOR, www.jstor.org/stable/41995684.

12. CanMap Postal Code Suite v2015.3. [computer file] Markham: DMTI Spatial Inc., 2015.
13. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment.
14. Liu Y, Goudreau S, Oiamo T, Rainham D, Hatzopoulou M, Chen H, Davies H, Tremblay M, Johnson J, Bockstael A, Leroux T, Smargiassi A. Comparison of land use regression and random forests models on estimating noise levels in five Canadian cities. Environ Pollut. 2020 Jan;256:113367. doi: 10.1016/j.envpol.2019.113367. Epub 2019 Oct 10. PMID: 31662255.
15. Statistics Canada, Canadian Mortgage and Housing Corporation. (2020). Proximity Measures Database – Early release, April 17, 2020. [computer file].
16. Distance to Roads, prepared by the Canadian Urban Environmental Health Research Consortium. CanMap Postal Code Suite 2015, 2016, 2017, 2018 and 2019 [computer file] Markham: DMTI Spatial Inc. CanMap Road files 1996, 2001, 2006, 2011, 2016 and 2019 [computer file] Markham: DMTI Spatial Inc.
17. Distance to Water, prepared by the Canadian Urban Environmental Health Research Consortium CanMap Postal Code Suite 2015, 2016, 2017, 2018 and 2019 [computer file] Markham: DMTI Spatial Inc. CanMap Road files 1996, 2001, 2006, 2011, 2016 and 2019 [computer file] Markham: DMTI Spatial Inc.
18. OpenStreetMap contributors. (2019) . Retrieved from Planet OSM May 2021 at

https://planet.openstreetmap.org/

19. Luan, H., Fuller, D. 2020. "Urban Sprawl Index for Canada". https://doi.org/10.7910/DVN/YQX0SW. Harvard Dataverse.
20. van Donkelaar, A., R. V. Martin, R. J. D. Spurr and R. T. Burnett, High-resolution satellite- derived PM2.5 from optimal estimation and geographically weighted regression over North America, Environ. Sci. and Tech., doi 10.1021/acs.est.5b02076
21. van Donkelaar, A., R. V. Martin, et al. (2019). Regional Estimates of Chemical Composition of Fine Particulate Matter using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors. Environmental Science and Technology, 2019, doi:10.1021/acs.est.8b06392.
22. Hammer, M. S.; van Donkelaar, A.; Li, C.; Lyapustin, A.; Sayer, A. M.; Hsu, N. C.; Levy, R. C.; Garay, M. J.; Kalashnikova, O. V.; Kahn, R. A.; Brauer, M.; Apte, J. S.; Henze, D. K.; Zhang, L.; Zhang, Q.; Ford, B.; Pierce, J. R.; and Martin, R. V., Global Estimates and Long- Term Trends of Fine Particulate Matter Concentrations (1998-2018)., Environ. Sci. Technol, doi: 10.1021/acs.est.0c01764, 2020.
23. Aaron van Donkelaar, Melanie S. Hammer, Liam Bindle, Michael Brauer, Jeffery R. Brook, Michael J. Garay, N. Christina Hsu, Olga V. Kalashnikova, Ralph A. Kahn, Colin Lee, Robert C. Levy, Alexei Lyapustin, Andrew M. Sayer and Randall V. Martin (2021). Monthly Global Estimates of Fine Particulate Matter and Their Uncertainty Environmental Science and Technology, 2021, doi:10.1021/acs.est.1c05309
24. Hystad P, Setton E, Cervantes A, Poplawski K, Deschenes S, Brauer M, et al. 2011. Creating National Air Pollution Models for Population Exposure Assessment in Canada.

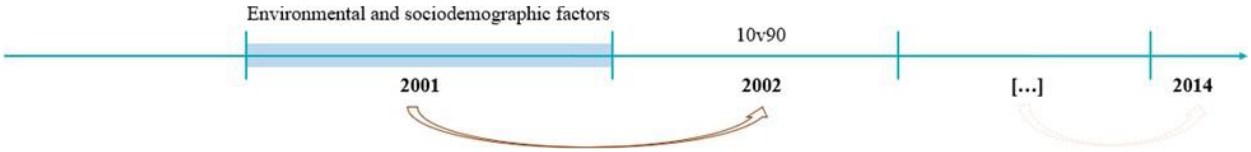
Environ. Health Perspect. 119:1123–1129; doi:10.1289/ehp.1002976.

25. Environment and Climate Change Canada, 2017. Air Quality Research Division, Toronto, Canada. Data files: CHRONOS_Ground-Level_O3_NA_2002.nc to CHRONOS_Ground- Level_O3_NA_2009.nc inclusive, generated July 2017.
26. Paul N, Henderson S. Canadian Optimized Statistical Smoke Model (CanOSSEM) daily PM2.5 estimates 2010 - 2019. Environmental Health Services, British Columbia Centre for Disease Control. 2021.
27. Environment and Climate Change Canada, 2017. Air Quality Research Division, Toronto, Canada. Data files: OMI_Ground-Level_SO2_NA_2005.nc to OMI_Ground-Level_SO2_NA_2015.nc inclusive, generated 2017-07-05.
28. Customized spatial climate data files prepared for the CanadianUrban Environmental Health Research Consortium by the Canadian Forest Service of Natural Resources Canada, October 2017.
29. Setton, E; Redivo A. (2022). Three-year warm-season maximum land surface temperature from LandSat 8. Developed for the Canadian Urban Environmental Health Research Consortium.
30. Bechtel, B.; Alexander, P.J.; Böhner, J.; Ching, J.; Conrad, O.; Feddema, J.; Mills, G.; See, L.; Stewart, I. Mapping Local Climate Zones for a Worldwide Database of the Form and Function of Cities. ISPRS Int. J. Geo-Inf. 2015, 4, 199-219.
31. Customized spatial climate data files prepared for the CanadianUrban Environmental Health Research Consortium by the Canadian Forest Service of Natural Resources Canada, October 2017.

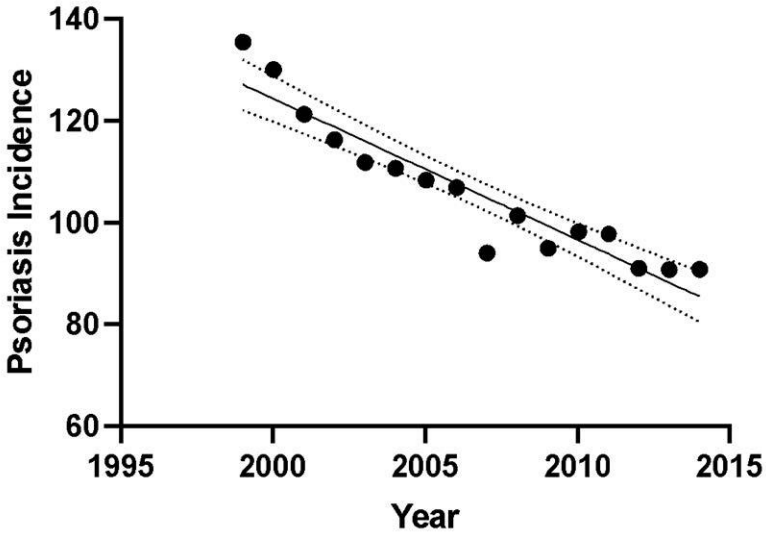
32. More information on Postal code. https://www12.statcan.gc.ca/census-recensement/2006/ref/dict/geo035a-eng.cfm.

33. https://jerryfriedman.su.domains/ftp/trebst.pdf

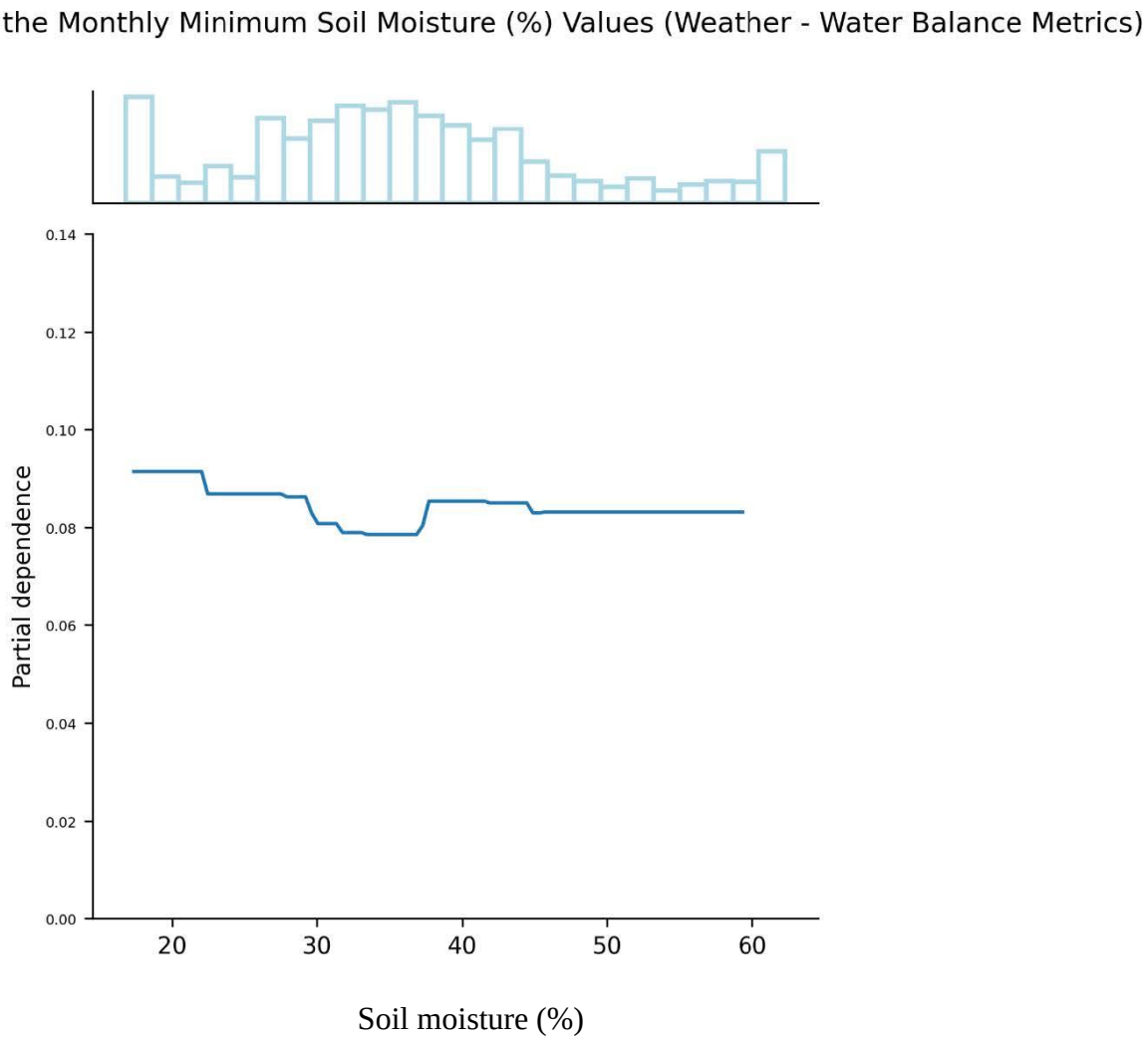
Supplementary Figure 1. Study design. Environmental and sociodemographic factors in the year prior were used to predict the psoriasis incidence in the year following



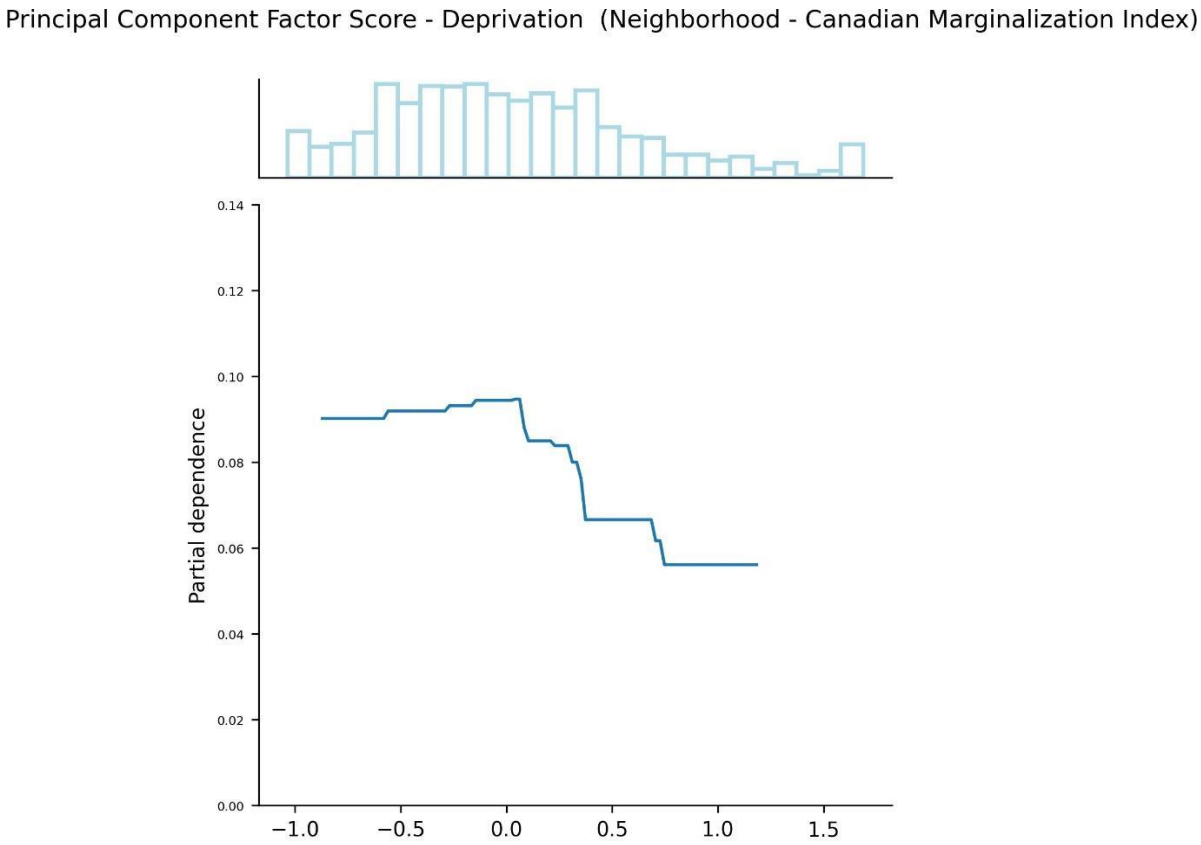
Supplementary Figure 2. Psoriasis incidence in Quebec during the study period



Supplementary Figure 3. Moving to the right on the x-axis corresponds to a higher soil moisture and moving up on the y-axis corresponds to increasing psoriasis incidence.

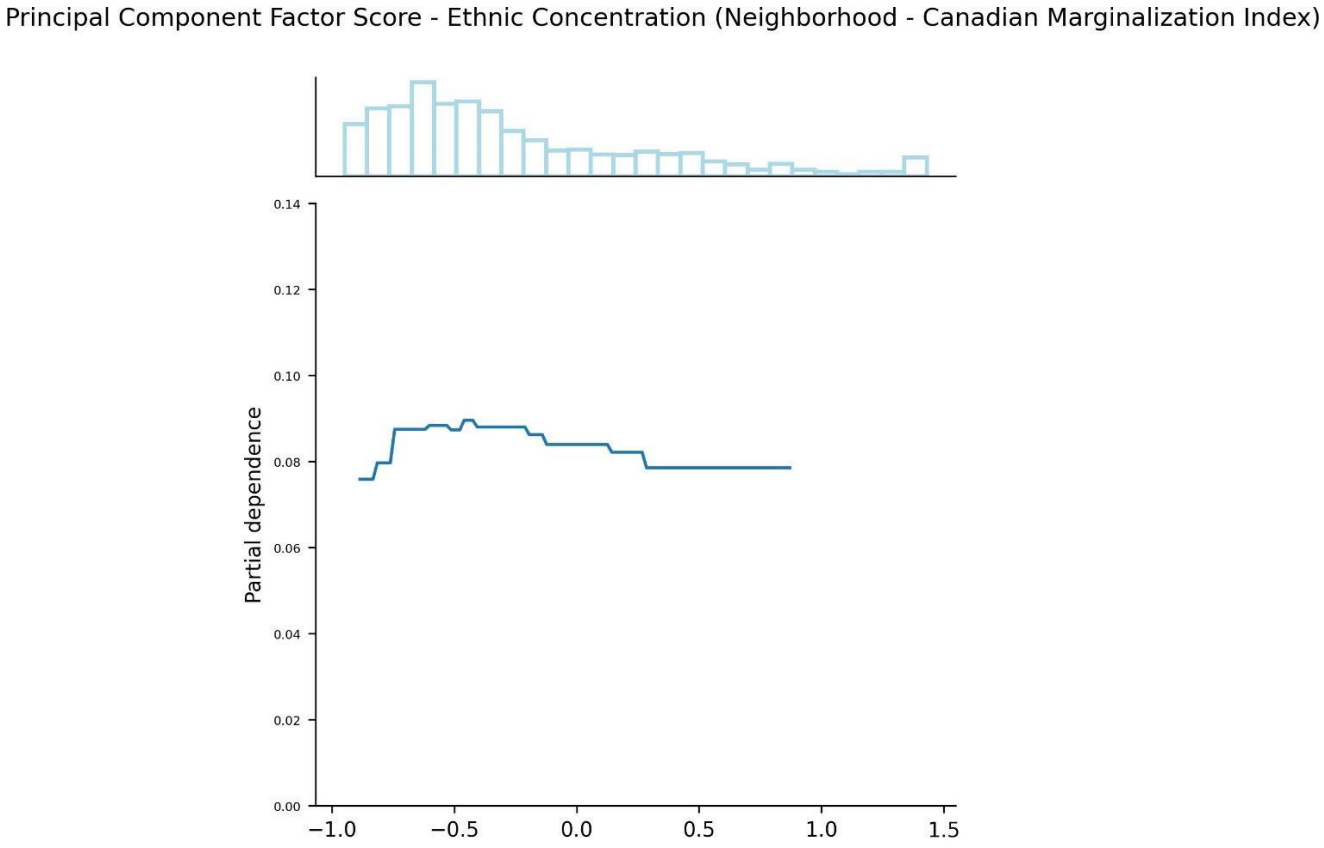


Supplementary Figure 4. Moving to the right on the x-axis corresponds to more deprivation in a neighbourhood and moving up on the y-axis corresponds to increasing psoriasis incidence.



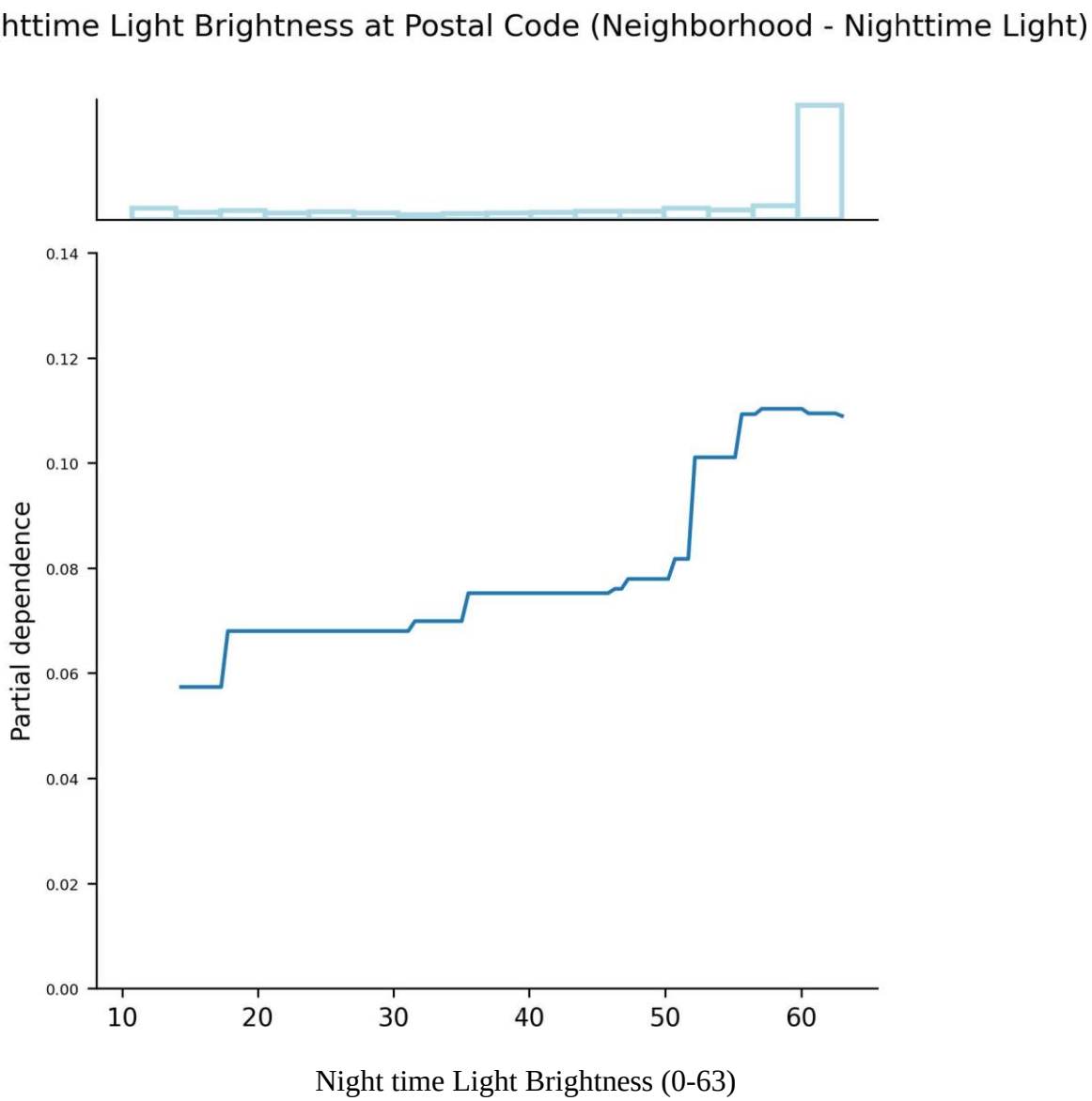
Principle Factor Score – Deprivation (-1 to 1)

Supplementary Figure 5. Moving to the right on the x-axis corresponds to a higher concentration of visible minorities and moving up on the y-axis corresponds to increasing psoriasis incidence.

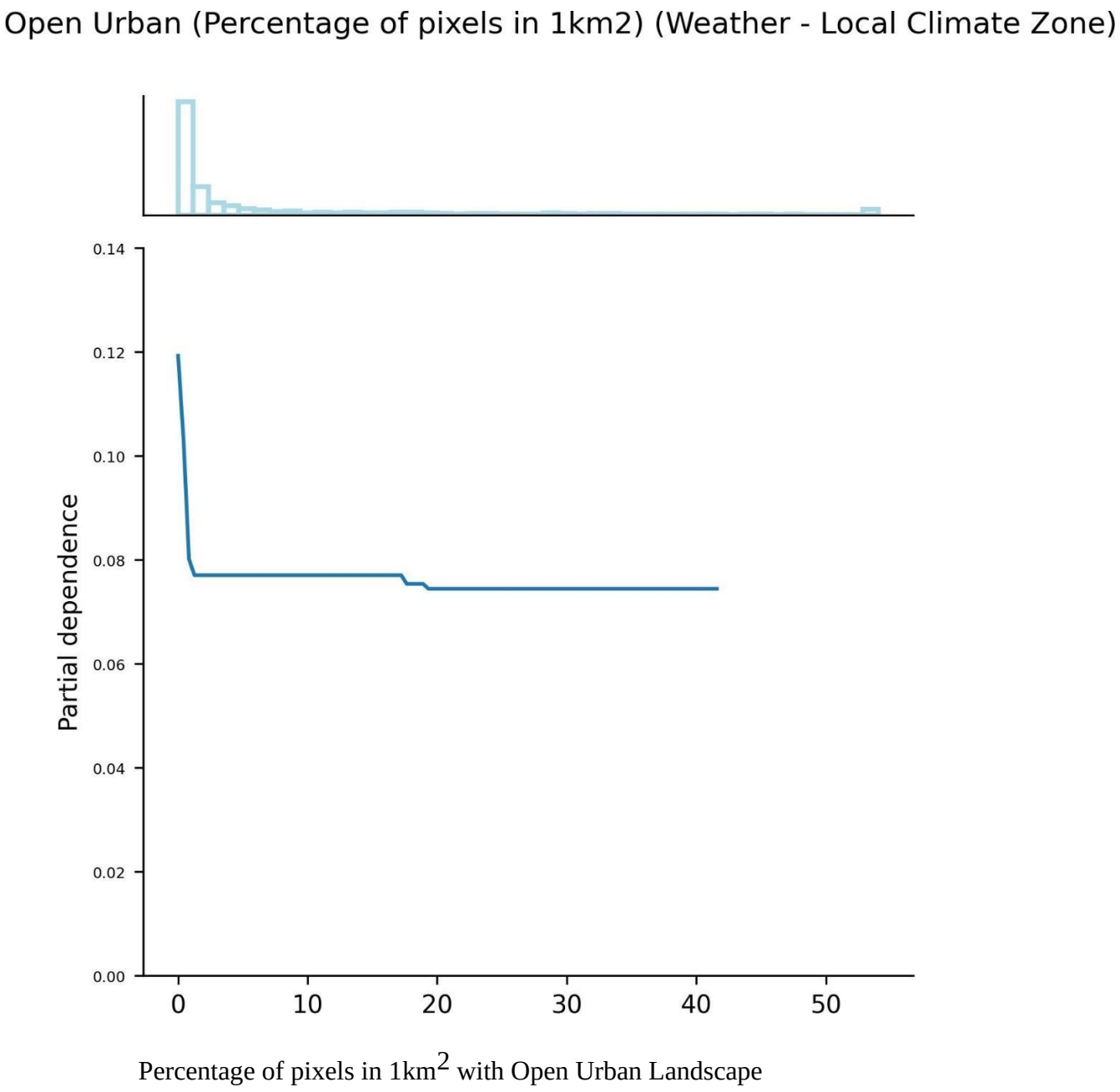


Principle Factor Score – Ethnic Concentration (-1 to 1)

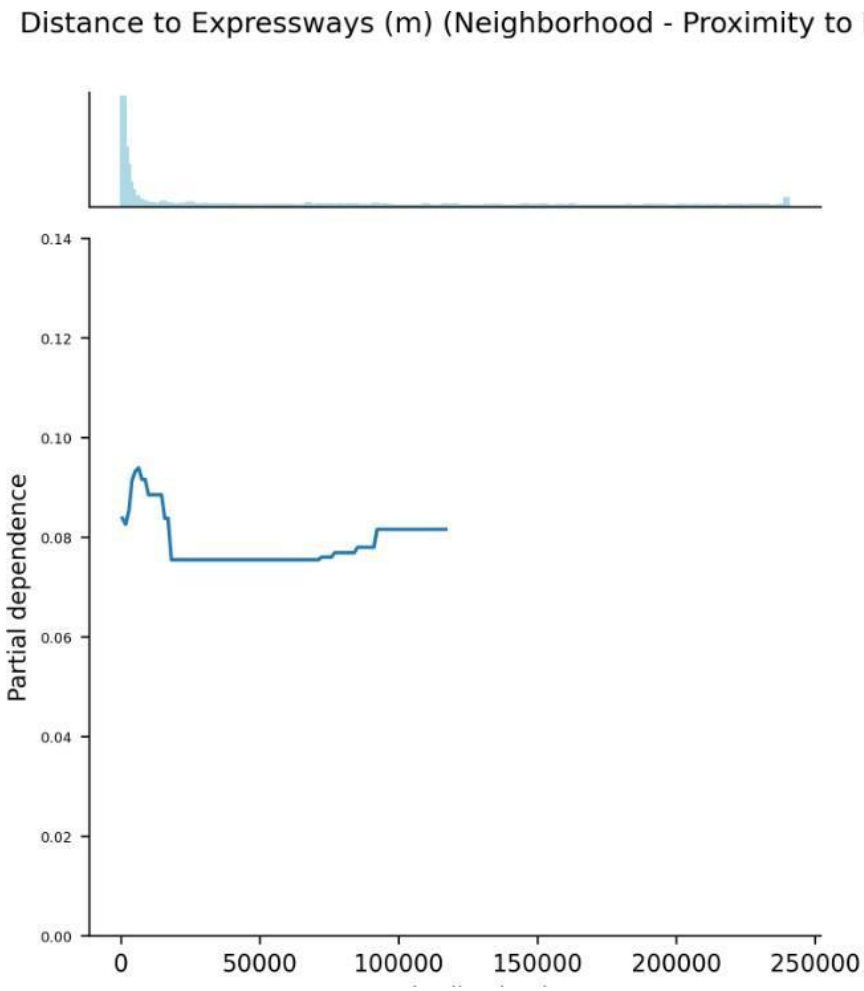
Supplementary Figure 6. Moving to the right on the x-axis corresponds to more nighttime light brightness exposure and moving up on the y-axis corresponds to increasing psoriasis incidence.



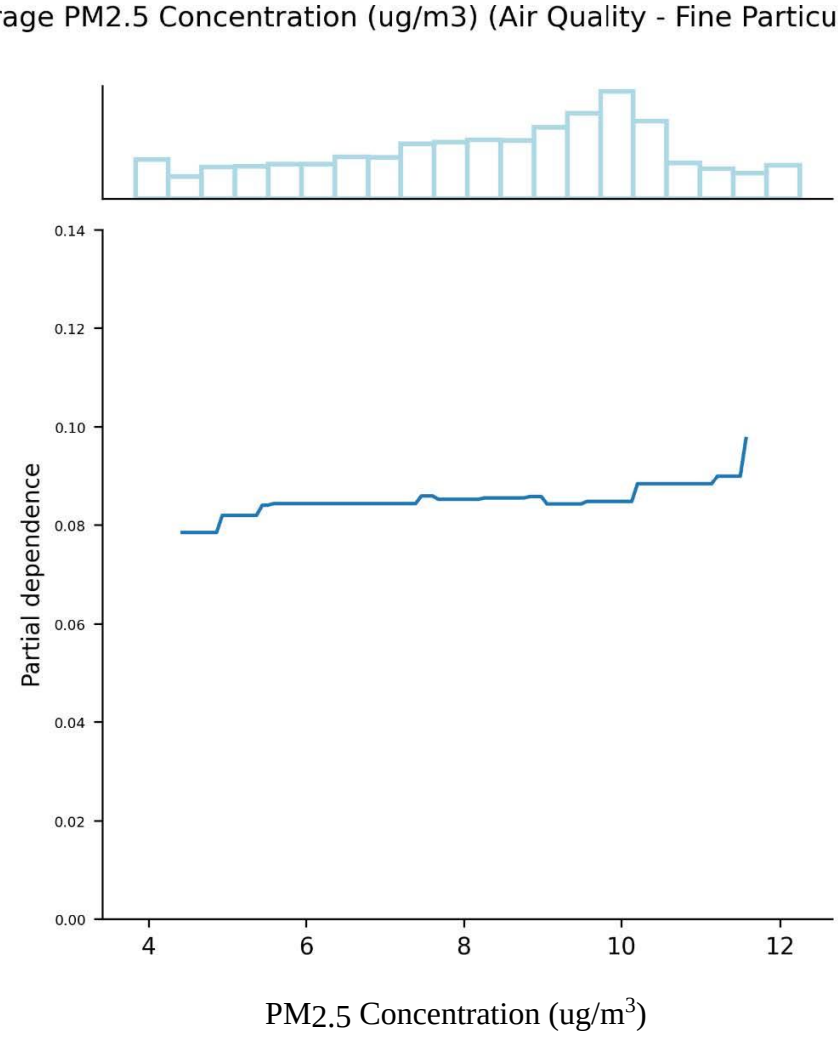
Supplementary Figure 7. Moving to the right on the x-axis corresponds to increased proportion of land covered by mid- and high-rise buildings in an open arrangement and moving up on the y-axis corresponds to increasing psoriasis incidence.



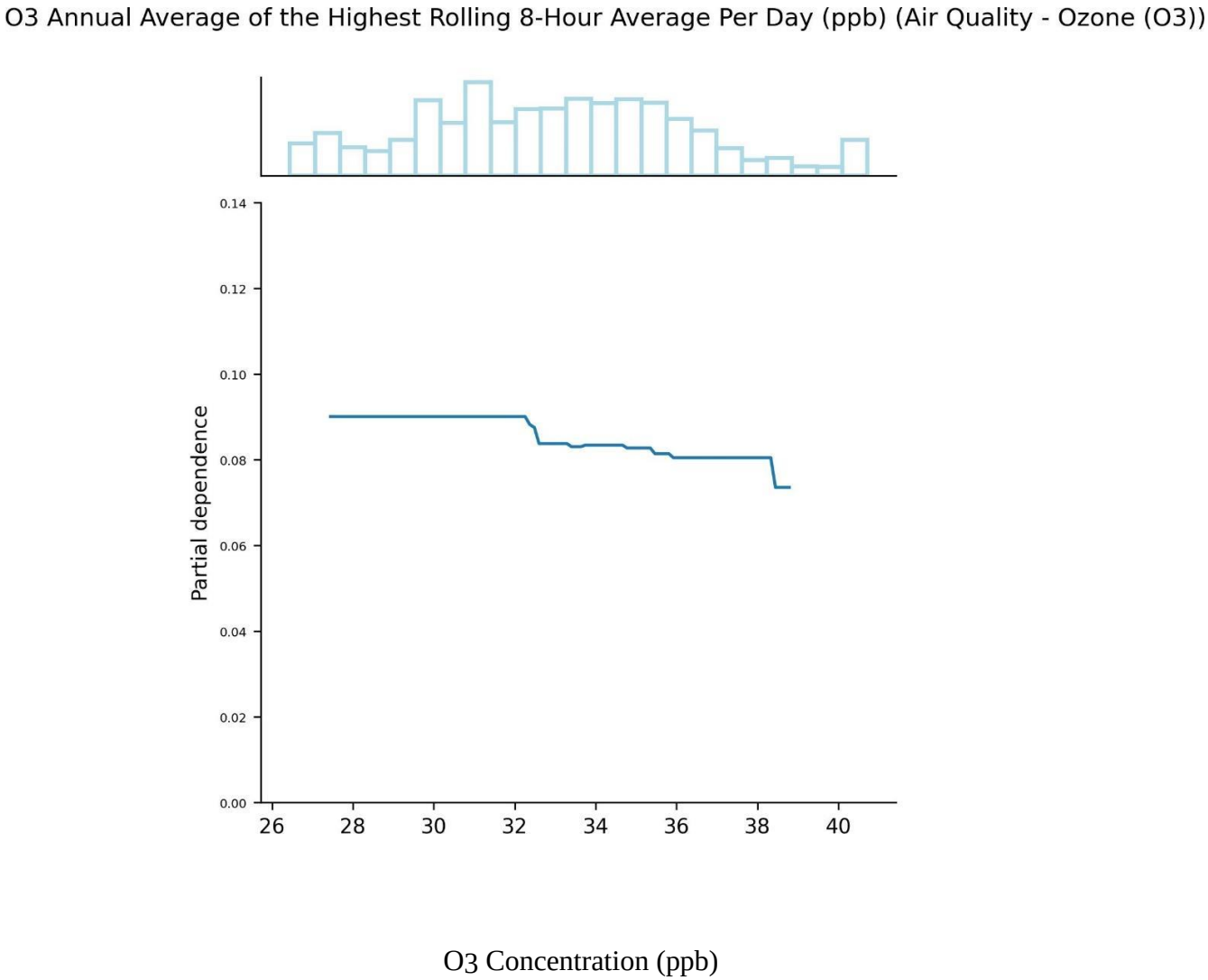
Supplementary Figure 8. Moving to the right on the x-axis corresponds to a longer distance to expressways and moving up on the y-axis corresponds to increasing psoriasis incidence.



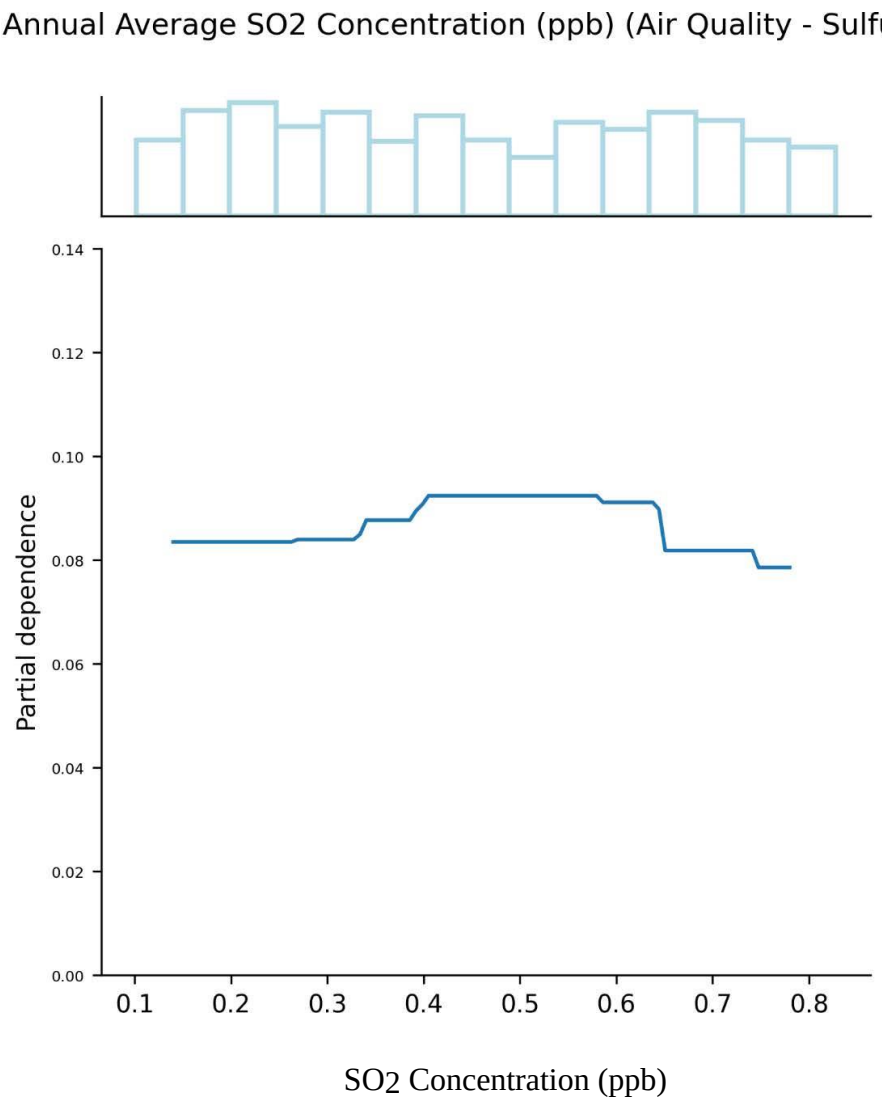
Supplementary Figure 9. Moving to the right on the x-axis corresponds to a higher concentration of $PM_{2.5}$ and moving up on the y-axis corresponds to increasing psoriasis incidence.



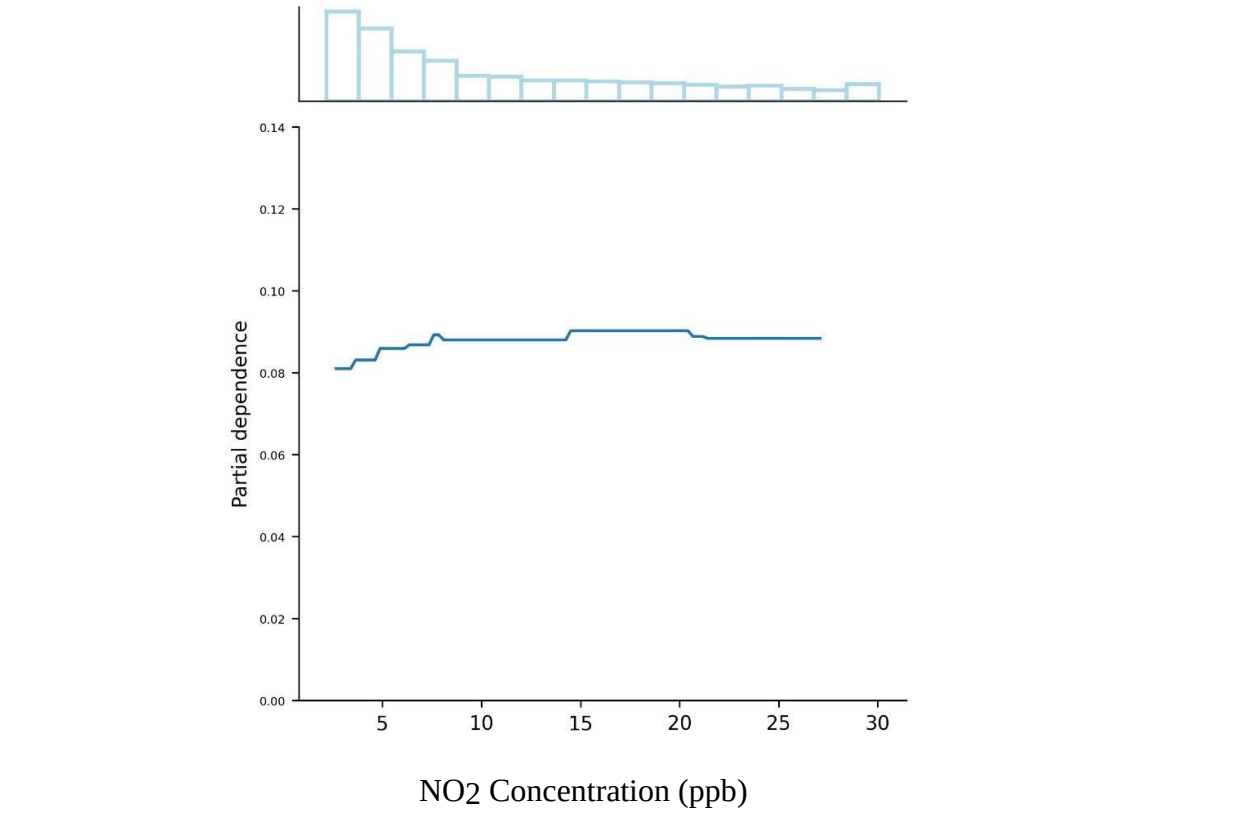
Supplementary Figure 10. Moving to the right on the x-axis corresponds to a higher concentration of O3 and moving up on the y-axis corresponds to increasing psoriasis incidence.



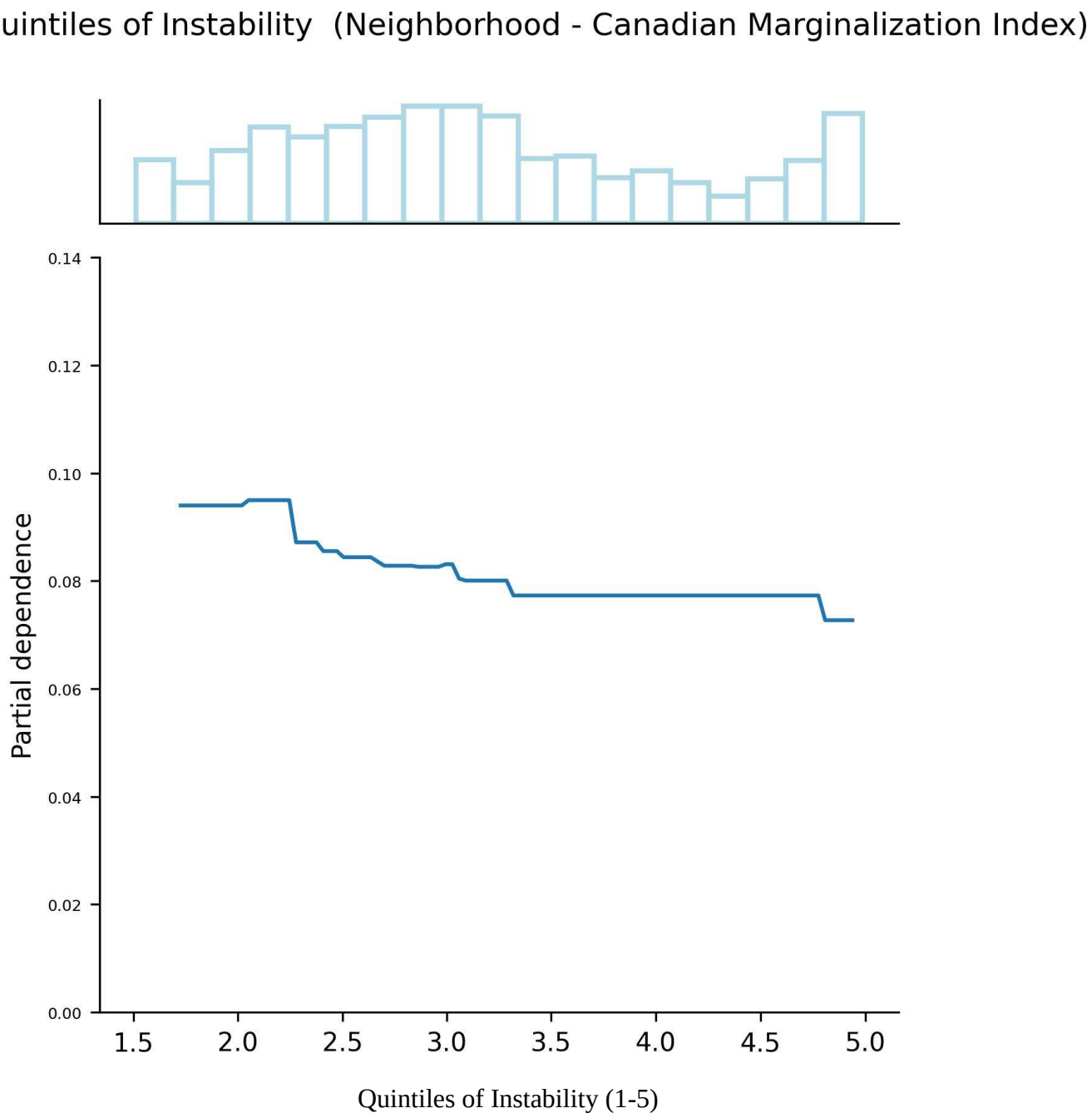
Supplementary Figure 11. Moving to the right on the x-axis corresponds to a higher concentration of SO_2 and moving up on the y-axis corresponds to increasing psoriasis incidence.



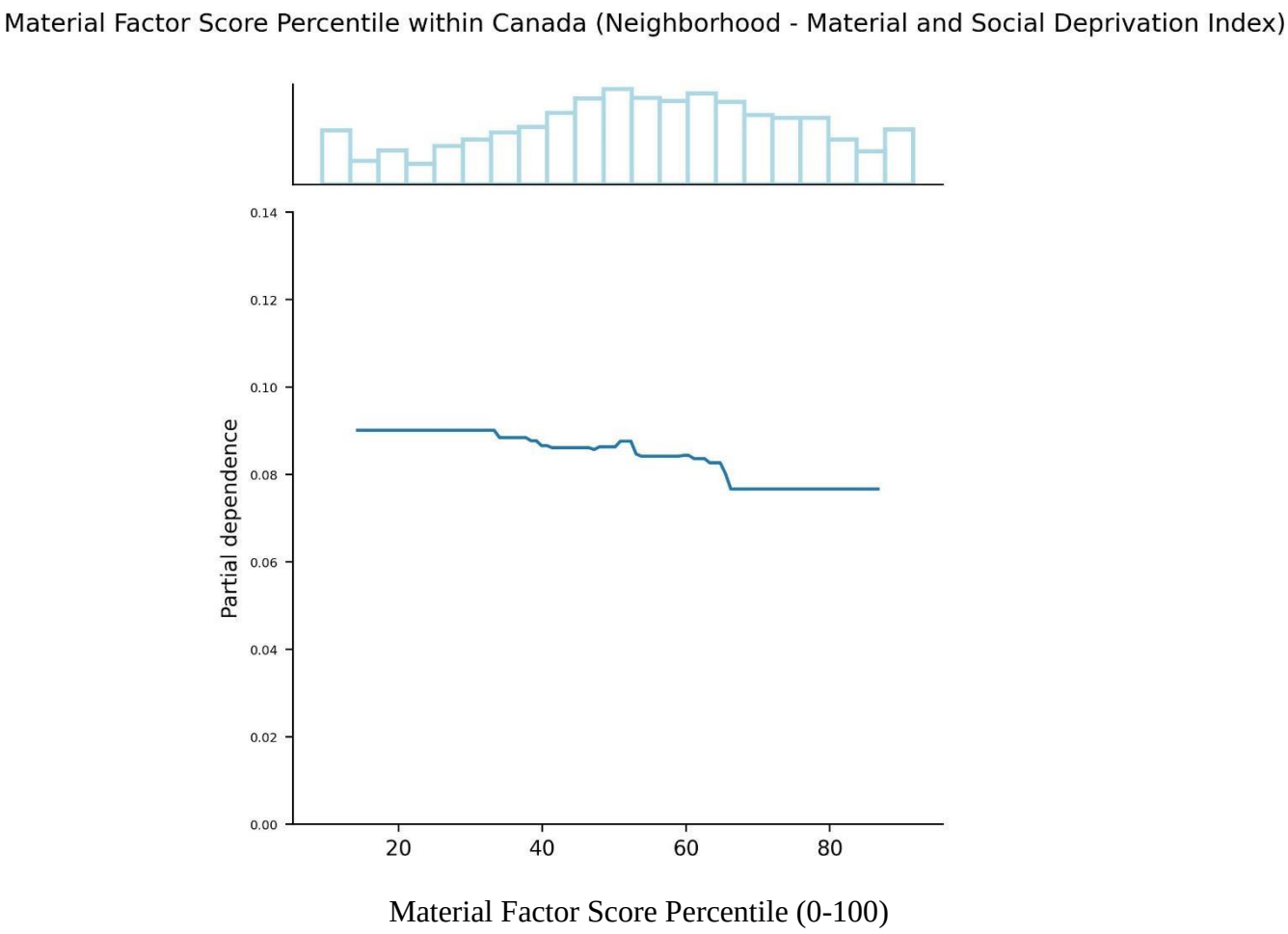
Supplementary Figure 12. Moving to the right on the x-axis corresponds to a higher concentration of NO_2 and moving up on the y-axis corresponds to increasing psoriasis incidence.



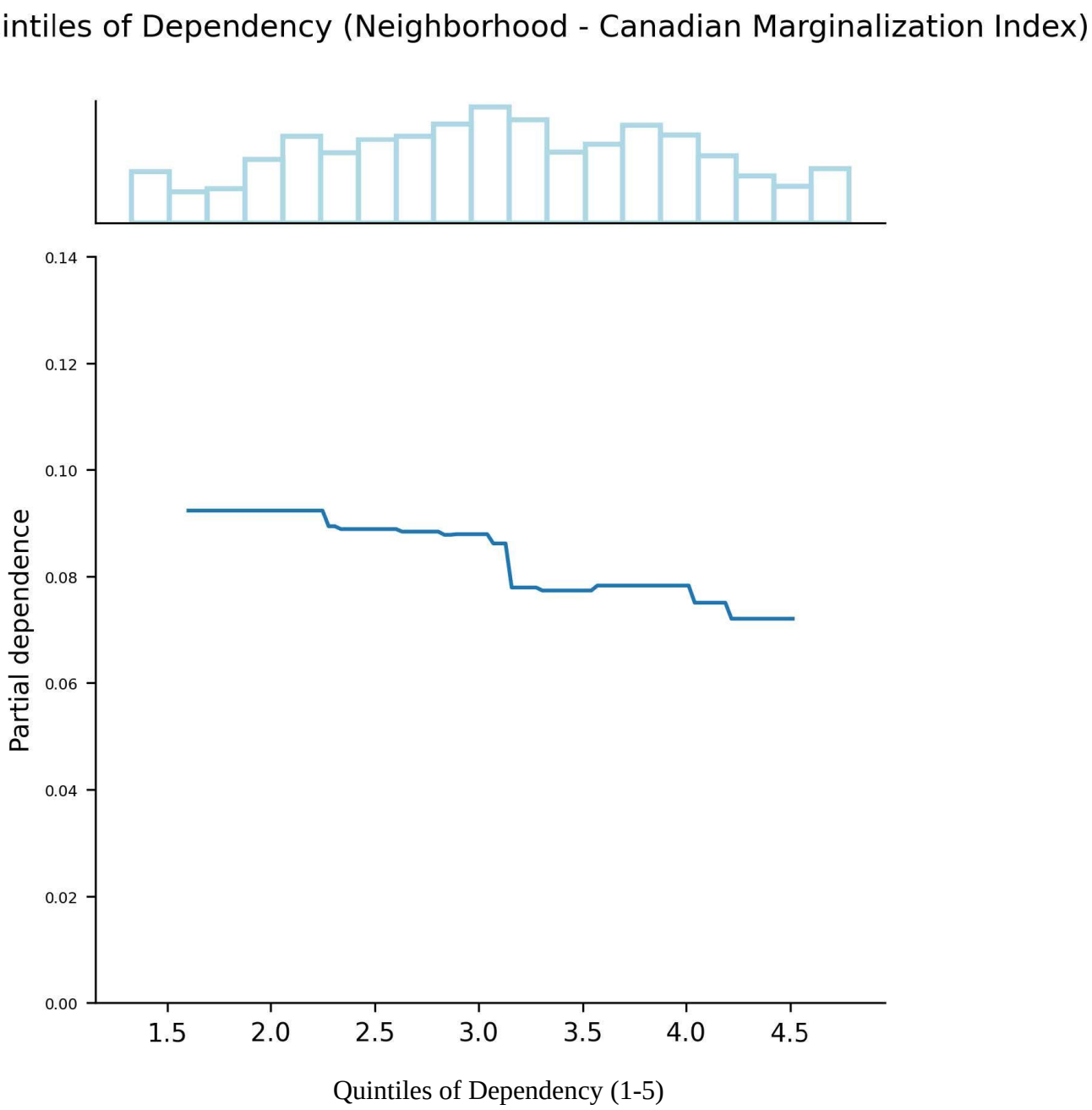
Supplementary Figure 13. Moving to the right on the x-axis corresponds to a higher level of instability in a neighbourhood and moving up on the y-axis corresponds to increasing psoriasis incidence.



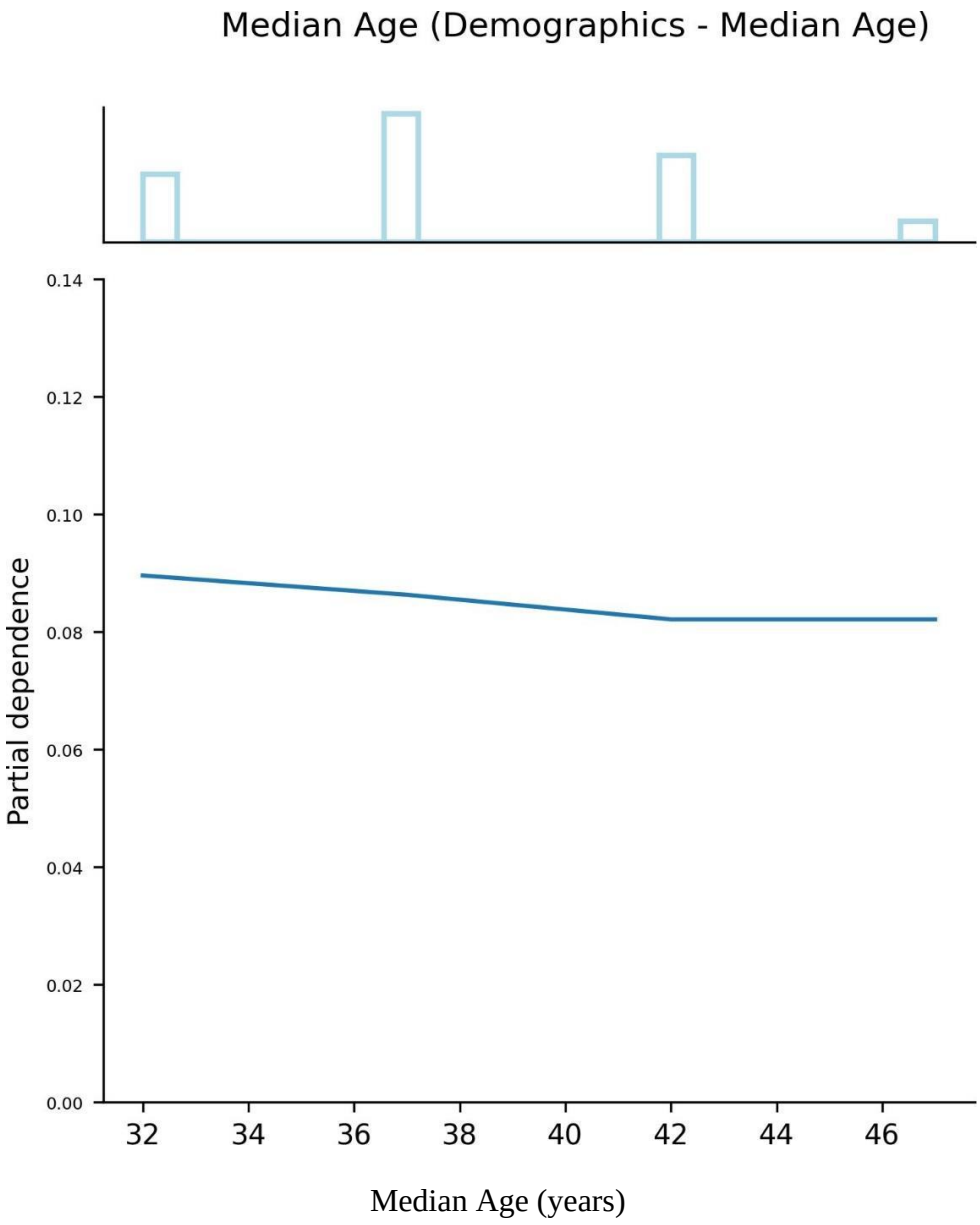
Supplementary Figure 14. Moving to the right on the x-axis corresponds to a higher level of material factor deprivation and moving up on the y-axis corresponds to increasing psoriasis incidence.



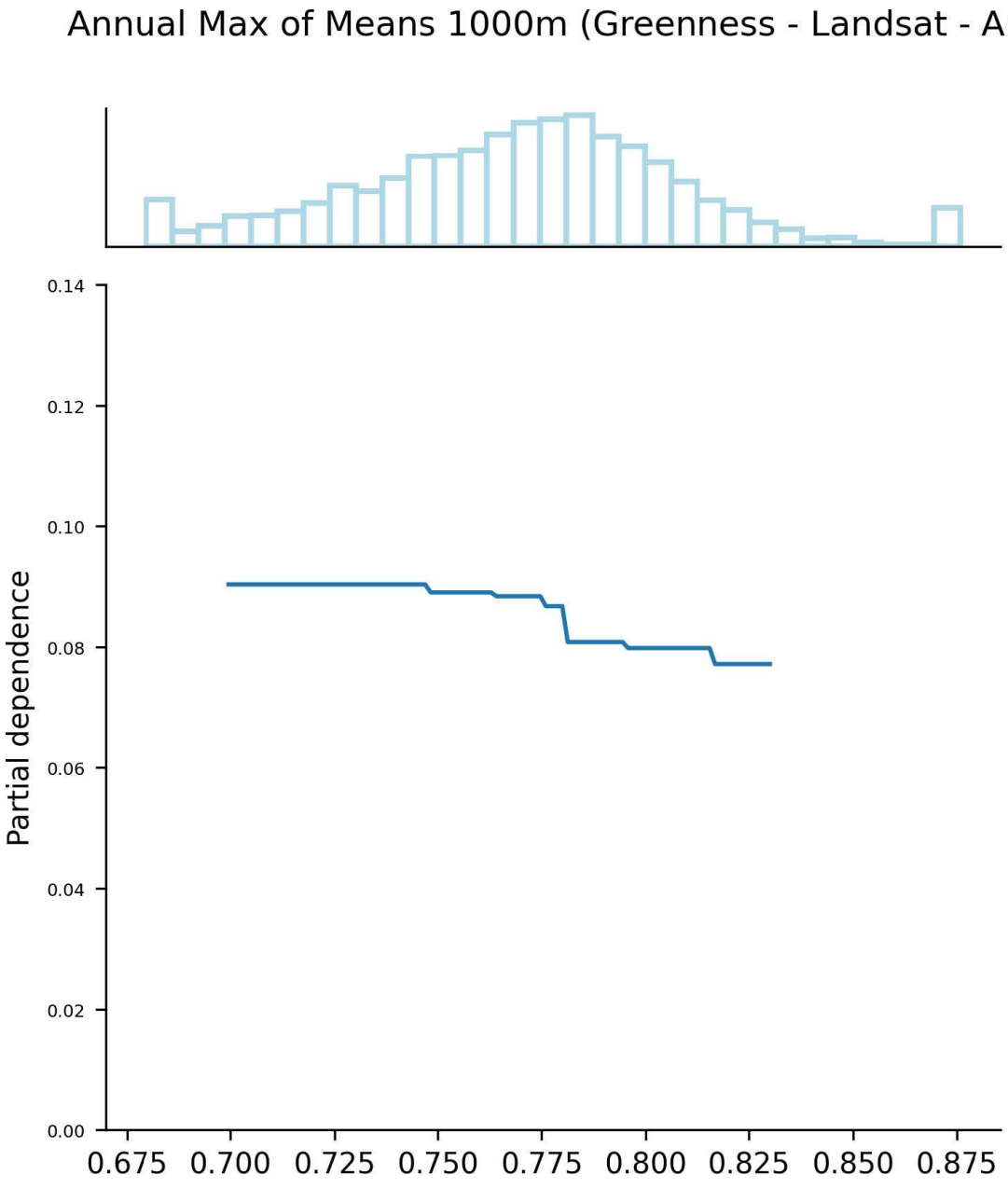
Supplementary Figure 15. Moving to the right on the x-axis corresponds to a higher level of dependency in a neighbourhood and moving up on the y-axis corresponds to increasing psoriasis incidence.



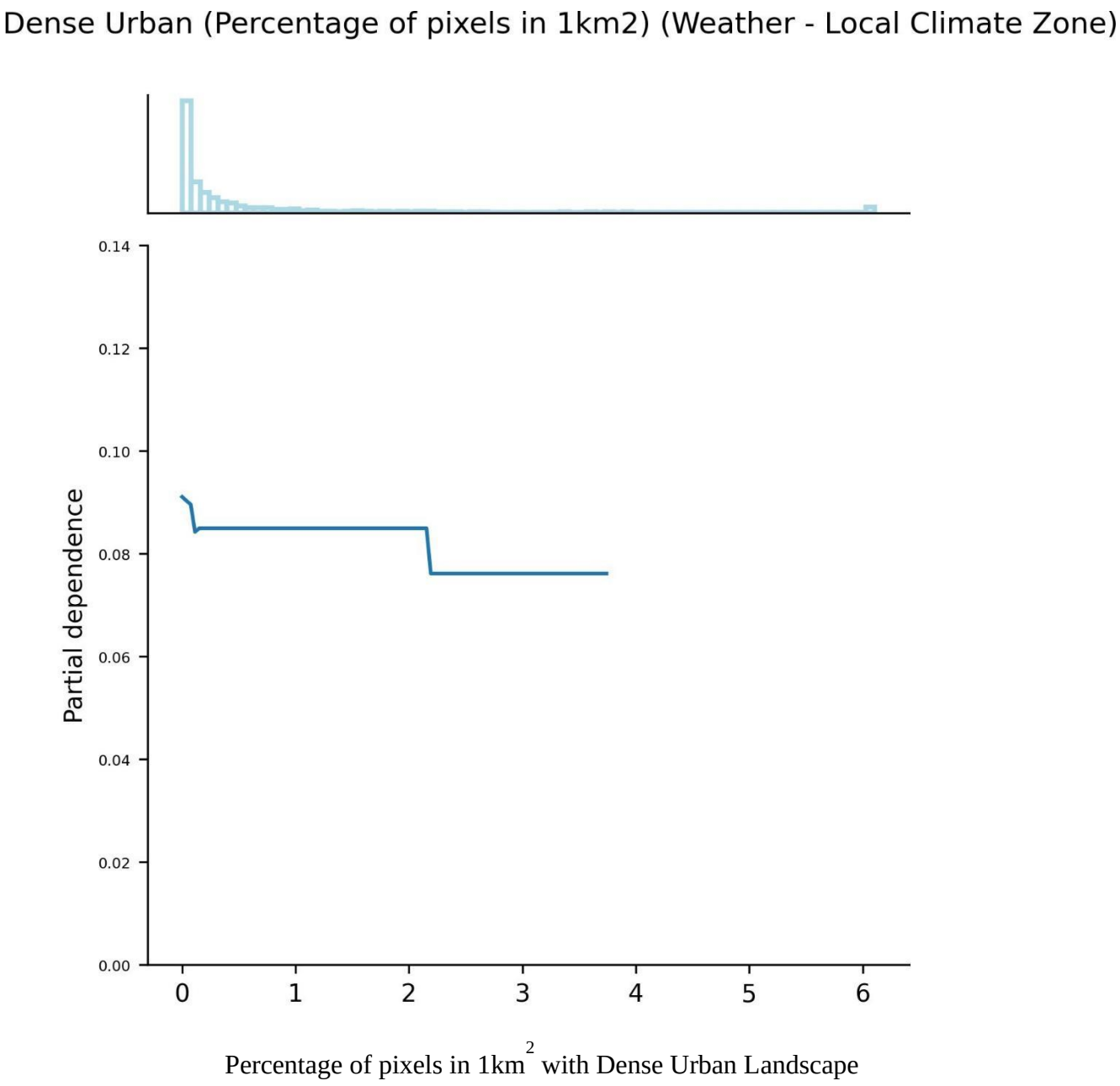
Supplementary Figure 16. Moving to the right on the x-axis corresponds to a higher median age at a Forward Sortation Area (FSA) and moving up on the y-axis corresponds to increasing psoriasis incidence.



Supplementary Figure 17. Moving to the right on the x-axis corresponds to a higher greenness level in a 1000m area and moving up on the y-axis corresponds to increasing psoriasis incidence.

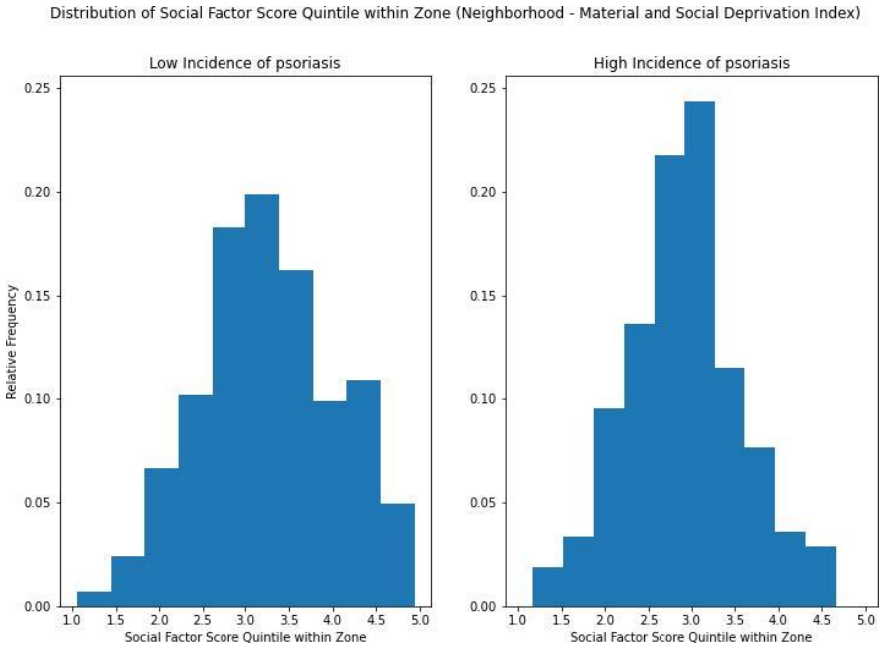


Supplementary Figure 18. Moving to the right on the x-axis corresponds to increased proportion of land covered by mid- and high-rise buildings in a dense arrangement and moving up on the y-axis corresponds to increasing psoriasis incidence.

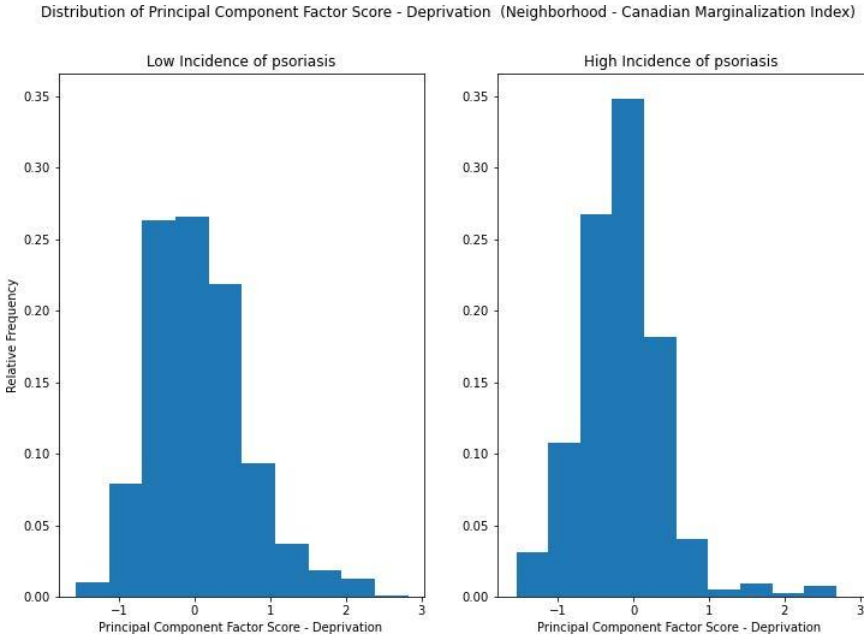


Supplementary Figure 19. **A** Distribution of the social factor score quintile between the high (top 10%) and low (other 90%) FSAs by psoriasis incidence. Higher values indicate lower socioeconomic status. **B** Distribution of principal component factor score for deprivation of the Canadian Marginalization Index between the high (top 10%) and low (other 90%) FSAs by psoriasis incidence. Higher values indicate more deprivation.

A.



B.


Distribution of Principal Component Factor Score - Deprivation (Neighborhood - Canadian Marginalization Index)
Low Incidence of psoriasis
High Incidence of psoriasis
Relative Frequency
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
−1
0
1
2
3
Principal Component Factor Score - Deprivation
Principal Component Factor Score - Deprivation

Chapter 8: Discussion

This thesis provides a comprehensive analysis of SSc and psoriasis in terms of epidemiology and assessment of contributing factors which helps address important gaps in the literature. Specifically, it aims to determine the incidence, prevalence, and mortality of SSc and psoriasis using populational data from Quebec, Canada. The CSRG registry was subsequently used to determine environmental and occupational exposures most notably, silica and organic solvents, and their impact on disease manifestations. Finally, a similar approach was used on another inflammatory condition, psoriasis, to determine updated epidemiology in Canada and identify contributing environmental factors.

Objective 1: Determine the incidence, prevalence, and mortality of SSc using populational data in Quebec, Canada

This study demonstrated increasing incidence and prevalence rates, overall and stratified by age and sex, of SSc between 1996-2019, while the mortality rates were decreasing indicating improved survival. An uneven geographic distribution of SIR was observed, suggesting that external factors can be at play leading to disease onset, outside of genetic predisposition. The year 2020 was excluded from the analysis, as the global pandemic was expected to decrease the number of diagnoses due to delay in accessing care[112-114].

To address this objective, a populational health administrative database was utilized to identify SSc cases between 1989-2019 using ICD-9/10 codes. This definition has previously been validated and was found to have a sensitivity of 80.5% and specificity of 94.9%[115]. Other studies in the literature using ICD codes to identify SSc patients, have found a high positive predictive

value (PPV) of 94% when compared to the accepted ACR/EULAR classification criteria[116]. Patients with localized scleroderma (*i.e.* morphea) were excluded based on the codes.

The cohort comprised of 8,180 SSc patients diagnosed between 1996-2019, where 80.3% were females and the average age at diagnosis was 57.3±16.3 years, comparable to the published literature[22]. The overall ASIR was 4.14/100,000 person-years (95% CI 4.05–4.24). This is in keeping with observed annual incidence rates from North America (1.4-5.6/100,000 person-years). Across several studies, North America has consistently showed elevated rates[117] and the ASIR in our study is greater compared to the global pooled incidence of 1.4/100,000 person-years. We observed an increasing trend in ASIR over the entire study period, with an AAPC of ~4%. This is similar to other autoimmune diseases. A study of 22 million patients across 19 autoimmune conditions showed that one in 10 patients has an autoimmune condition and that trends are increasing over time[34].

There is limited data available regarding SSc stratified by age group given the typically small sample sizes in most studies. In our study, the highest incidence rate was noted in the 60-79-year-old group for females and in the >80-year-old group for males. A similar study in the United States in 2003 showed the highest incidence in the 65–74-year-old group for females and the 75-84-year-old group for males[118]. A study in the United Kingdom comprised of 1,327 incident cases of SSc showed the highest incidence in the 55–69-year-old group[119].

The highest AAPC (~10%) in incidence was seen in children. However, children represented only 2.3% of all incident SSc cases in our cohort, with an incidence rate of 0.43/100,000 person years. This is comparable, although slightly higher, to the rates in the

literature (0.03-0.29/100,000 person-years)[120,121]. This rapid rise elucidated some concerns and the reason for this is not known, hence further investigation is required.

The average prevalence in our study was 28.96/100,000 persons (95% CI 28.72–29.20), with a 4.8-fold higher prevalence in females compared to males. The global estimates vary greatly between 3.1 to 144.5/100,000 individuals, with pooled estimates of 17.6/100,000 persons[117]. Similarly to incidence, consistently elevated rates in North America are seen indicating the occurrence is likely truly higher than in the rest of the world. This further highlights the urgent need to explore the reasons for this and identify preventative strategies. The increasing prevalence over time is expected as with increasing diagnosis (*i.e.* incidence) and increased survival there are more cases present in the population. The highest prevalence was observed in the 60–79-year-old group for females (as with incidence) and males. A study in the United Kingdom showed the highest prevalence in the 70–84-year-old group followed by the 55–69-year-old group[119].

SIRs varied geographically between 0.52 to 1.64. Spatial analysis confirmed an uneven geographic distribution of SSc in Quebec over time, with eight Forward Sortation Areas (FSAs) with ASIRs >5-times higher compared to the Quebec average. There appeared to be higher incidence in the northern and eastern part of Quebec. The uneven distribution of SSc in Canada was confirmed in our second study using the CSRG registry (discussed in Manuscript 2). Otherwise, limited data on geographic distribution is available in North America. Studies from Europe have also shown an uneven distribution with high prevalence in Italy, Spain and Sweden compared to France, Netherlands, and Norway[117].

The SMR in this study decreased from 4.18 (95% CI 3.64–4.76) in 1996 to 2.69 (95% CI 2.42–2.98) in 2019. Although the downward trend is reassuring, this was still higher compared to the general population. SMR decreased faster for females than males and after 2009, males consistently had higher mortality compared to females. Similarly, a study in the United States over a 48-year period showed a steady decline in mortality starting in 2001, where previously there was a slight increase[122]. In Canada, a study evaluating SMR found that the inception cohort had an SMR of 5.1, compared to 3.8 in the prevalent cohort[123]. As in our study, males also had higher SMR in both cohorts (8.6 for inception and 5.9 for prevalent) compared to females (4.4 for inception and 3.4 for prevalent). Other studies have also documented increased mortality in males, mainly due to delay in disease recognition and more severe disease phenotype[124,125]. However, there are also conflicting studies showing higher mortality among females[126]. Multiple factors can account for the overall decreasing mortality trend noted in our study, including improved recognition and diagnosis (*e.g.* classification criteria[42], nailfold capillaroscopy, autoantibody profiles), earlier diagnosis due to description of VEDOSS features in 2011[43], better treatment, including immunosuppressive agents, targeted therapies, and stem cell transplant[74], and mitigation of side effects from the treatment (*e.g.* reducing the use of high dose prednisone which can precipitate renal crisis)[118].

The highest SMR was observed among the 0–19-year-old group which could be attributed to several factors. Given only 2.3% of cases were in this age group, SMR can be inflated even with a small number of mortality cases. The increased mortality may also result from delayed diagnosis, given the rarity of this condition in children, or from underlying genetic factors that contribute to more severe disease phenotypes.

Objective 2: Using the CSRG national registry, determine the geographic distribution of SSc and evaluate the contribution of environmental triggers.

In the present study, using the CSRG patient registry, our analyses revealed an uneven geographic distribution of period prevalence of SSc in Canada, including both industrial/urban areas and rural areas located far from recruitment centers. High and low prevalence areas were compared in terms of industry density and the concentration of environmental pollutants, which could account for the regional variability of the SSc rates observed.

A total of 1,505 SSc patients were included in the registry, which is estimated to encompass 10% of all prevalent cases. Similar to other studies, the median age at SSc diagnosis was 47.1 (interquartile range 37.3-55.7) and 86.3% were females. Approximately a third of patients had dcSSc. Two thirds of patients in the cohort lived either in Quebec or Ontario, the two largest Canadian provinces. Provincial patient distribution was well represented except for British Columbia (2%), as there was no CSRG recruitment center there after 2008. Patients resided in 665 different FSAs across Canada.

Based on the simulation model considering the population size per FSA, age and sex distribution, more patients were expected to be identified in larger urban FSAs. Only 19 FSAs had a statistically significantly higher observed prevalence of SSc than expected based on the simulation and had at least 5 diagnosed patients with SSc. The overall prevalence in these 19 FSAs was 2-5.5-fold above the expected prevalence.

Thirteen urban FSAs were identified. They were located in the Montreal Metropolitan Region (H3S, H3X, H9S, H9W; QC), Hamilton/Dundas (L8T, L8W, L9A, L9B, L9H; ON), Winnipeg (R3R, MB), Saskatoon (S7J, S7L; SK) and Calgary (T3E, AB). Clusters of two or more high prevalence FSAs (without adjacent low prevalence FSA) were seen in Montreal and Hamilton. While all these cities are home to CSRG recruitment centers, the observed prevalence in these FSAs was 2-5.5-fold higher than predicted suggesting a non-random distribution.

Aside from urban centers, many smaller remote communities were also found to have high period prevalence, and this was greater than expected based on the simulation model. Interestingly, 6 out of 19 FSAs were in rural regions that are far from recruitment centers. They belonged to Miramichi (E1N, NB), Montérégie (J0L, QC), Rock Forest (J1N, QC), Lambton Country (N0N, ON), Woodstock (N4S, ON) and Eastern Saskatchewan (S0E, SK). We confirmed previously identified areas of increased prevalence, such as Woodstock (N4S,ON)[59] and Montérégie (J0L,QC)[127]. Montérégie is home to Kahnawake, a small Mohawk community of 7,000 people near Montreal, where 25 patients were documented to have SSc[127]. Several other rural FSAs (>200km from recruitment centers) had a high observed prevalence but did not reach statistical significance, likely due to low patient numbers. They included Flin Flon, Alma and Amherst.

To assess whether the observed geographic distribution may be a result of patients with severe disease moving closer to treatment centers, demographic and disease features were compared between patients in the top 19 FSAs and the rest of the CSRG registry patients. No statistically significant difference was observed for age, sex, SSc subtype, Medsger severity score[103] and mortality. Given that a subset of SSc cases may be occupationally induced, patient

self-reported data on occupational exposure to silica and organic solvents was assessed and no significant difference was found. Finally, patients had a similar disease duration across the 19 higher prevalence FSAs and the rest of the cohort, demonstrating that longer disease duration is unlikely to bias the obtained results.

Data on air pollution in the 19 high prevalence FSAs was obtained from CANUE and analyzed. Compared to the lowest 19 FSAs, air pollution ($PM_{2.5}$ and/or O_3 and/or NO_2 and/or SO_2) was statistically higher in 15 of 19 FSAs of interest, and the $PM_{2.5}$ levels were statistically higher in 8 of 19 FSAs, favoring major urban centers in Montreal, Hamilton and Calgary.

Canada is a leading country in mining, with key sectors including exploration, resource extraction, and downstream processing and manufacturing[128]. It is a top producer of many minerals/metals (*e.g.* gold, cobalt, nickel, platinum)[128]. Overall the minerals and metals industry has ~200 mines, ~6,500 sand/gravel/stone quarries[128], and 31 nonferrous smelters and refineries[129]. Therefore, Canada is particularly suitable to study geospatial SSc epidemiology and association with environmental factors.

Visual analysis of land use regions from DMTI-EPOI revealed important industrial/mining activities within 5 km of most rural FSAs and urban FSAs, including Hamilton and Montreal. Additional urban FSAs were located near international airports and/or a railyard (Montreal). In comparison to the 19 lowest prevalence FSAs, the 19 FSAs of interest showed significantly higher density for all industries combined ($p=0.0059$) as well as individually for paper manufacturing ($p=0.0364$) and petroleum and coal products manufacturing ($p=0.0002$) industries. For example, among rural FSAs, Miramichi was the site of the previous Heath Steele Mines, operational

1957-1999, and mainly focused on retrieving zinc, lead, silver, and copper. All patients from Miramichi were diagnosed with SSc prior to 1999. Lambton county (N0N, ON) is part of Canada's Chemical Valley and Sarnia-Lambton contains ≥60 refineries/chemical plants and >40% of all Canada's chemical industries. In 2011, it was listed by the WHO as the worst city for air quality in Canada[101]. Among urban FSAs, many high prevalence areas belonged to Hamilton, which is Canada's largest manufacturer of steel, responsible for 60% of all production in the country. Association between heavy metals and SSc development has been previously reported[130]. Aside from the heavy metal burden in this Hamilton region, Air Quality Report in Ontario in 2014 reported daily airborne benzene levels to be above the accepted upper limit [131].

In collaboration with the team at the Department of Dermatology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts we performed a study using similar methodology in the state of Massachusetts[106]. There were 2,196 SSc patients with a diagnosis confirmed by a rheumatologist or dermatologist identified from the patient data registry, based on medical records of Massachusetts General Hospital and Brigham and Women's Hospital between 1989-2019. The study found that presence of hazardous waste facilities (p=0.0039) and oil release or disposal sites (p=0.0203) were significantly associated with higher SSc prevalence. For chemical release sites no statistical significance was reached (p=0.3166). When all three exposures were combined, a significant increased risk was demonstrated (p=0.0002). Additionally, in areas of disease clustering, higher particulate pollution levels with spatial proximity to ash pollution and combustion facilities was seen. Investigations into the contribution of SES revealed that maps of low-income and minority communities had a spatial relationship with SSc clusters. Hence, in this study in Massachusetts, close vicinity to waste, chemical, and oil release sites, as well as air

pollution and industries with emission corresponded to clusters of SSc cases indicating potential impact on disease development.

Objective 3: Explore the impact that occupational exposure to a) silica and b) organic solvents played on clinical manifestations of SSc, disease severity, and mortality

a) Silica

Silica exposure has been shown to have up to 18-fold increased risk of developing SSc[22,132]. This study further explored the demographics and disease manifestation in exposed patients using a the large, CSRG registry.

Silica is a mineral component of the earth's crust and is a key constituent of soil, sand, and rocks[133]. It is frequently used as structural material because of the high melting point. It exists in two forms – amorphous and crystalline (*i.e.* quartz)[134]. Quartz is the most common form used in industries[135]. Exposure in an occupation often occurs through inhalation when products containing silica are disturbed through cutting, grinding, drilling, or chipping processes[136]. Occupations with high risk of exposure include mining, stone quarrying, sandblasting, concrete mixing, granite, and ceramic production[22]. As expected, these occupations are more commonly held by males. Estimates indicate that 429,000 Canadians are exposed to silica in an occupational setting, especially in construction, where 94% of these are males[134]. Outside of the occupational exposure, silica can be released into the air from natural, industrial, and agricultural activities, causing airborne exposure[22]. In addition to increased risk of developing SSc, silica exposure has also been linked to silicosis, a fibrotic lung disease, as well as cancer, particularly lung cancer[134]. Other associations include chronic obstructive pulmonary disease, an autoimmune disease such as rheumatoid arthritis, and pulmonary tuberculosis[134].

In our cohort, there were 1,439 patients included in the study where 86.7% were females. The average age at diagnosis was 46.5±13.7 years and an average disease duration was 9.83±9.23 years. Of these patients, ~7% reported occupational silica exposure (22.4% of CSRG males and 4.2% of females). Interestingly the male to female ratio was 1:1 among exposed patients, compared to the unexposed where the ratio was 8:1. As discussed above, the occupations where exposure typically occurs are mainly held by males. Patients that reported exposure were younger, more likely to be male, and have more severe disease (*e.g.* dcSSc, more severe ILD, worse skin fibrosis, and worse gastrointestinal disease). Males that were exposed were more likely to be Caucasian and smokers, whereas female patients were younger at SSc diagnosis, compared to unexposed individuals. These findings are similar to a study in Australia on SSc patients demonstrating that exposed patients are more likely to be male, smokers, and have worse disease features such as dcSSc, ILD, and lower frequency of ACA antibodies[137].

Multivariate regression, adjusted for multiple confounders, demonstrated that silica exposure was associated with a younger age at diagnosis and more severe disease with increased mortality.

Therefore, screening for silica exposure among SSc patients, especially those at higher risk, may be beneficial and these patients may require closer monitoring for systemic disease.

b) Organic solvents

Organic solvent exposure has been linked to a 2-10-fold increased risk of developing SSc[138-140]. This study further explored the demographics and disease manifestation in exposed patients using a large, CSRG registry.

Organic solvents describe a class of carbon containing compounds. The substances studied in terms of SSc risk include aromatic compounds (*e.g* benzene, toluene, xylene, and naphthalene) (odds ratio [OR] 8.17, 95% CI 2.29-36.5), white spirit (*e.g.* petroleum distillate) (OR 7.69, 95% CI 4.11-14.7), chlorinated solvents (OR 2.46, 95% CI 1.12-5.32), trichlorethylene (OR 2.29, 95% CI 1.04-5.22), ketone solvents (OR 4.20, 95% CI 2.19-8.06), and polyvinyl chloride (unknown risk)[22]. As they are generally used to dissolve other materials, they are commonly encountered in occupations such as laboratory work, leather/tire/paint/glue industries, degreasing, and dry cleaning[22]. In these occupations exposure or absorption can occur though contact with the skin or inhalation through the airway[22]. The proposed mechanisms for trichloroethylene leading to SSc include breakdown to epoxy compounds which can then bind to proteins and create autoantigens[141], as wells as solvent-induced[142] endothelial cell injury contributing to microvascular abnormalities and eventual organ fibrosis[143].

Of the 1,439 patients, 20.2% of patients reported an exposure to organic solvents. As discussed, females are more commonly affected by SSc, however males are more likely to report an occupational exposure compared to females. Consequently, the ratio of females to males in the exposed group was 2:1 compared to 10:1 in the non-exposed cohort. Previous studies have reported that the risk of SSc onset after solvent exposure is greater in males (OR 5.28 vs. 1.62 in females)[67]. Exposed patients were more likely to be males (29.7% vs. 9.2%), smokers, and have more severe disease (such as severe gastrointestinal disease, renal crisis, less ACA antibodies).

Interestingly, in this study, we found an increased risk of renal crisis (OR 2.13; 95% CI 1.15-3.93) in patients with organic solvent exposure history which has not been frequently reported in the literature. Another study demonstrated that males with glomerular nephropathies at

baseline had a 7-fold increased risk of progressing to chronic renal failure with high exposure to solvents[144]. The risk was found to increase with duration of exposure[145]. Hence, the microvascular effects of organic solvents could explain the increased risk of renal crises observed in our study.

We also observed increased severity of gastrointestinal involvement (β 0.89; 95% CI 0.47-1.31), although this has not been reported in other studies and requires further investigation. Several other studies in the literature have also reported more severe disease phenotype demonstrating increased risk of dcSSc (p=0.001), digital ulcers (p=0.01), ILD (p=0.02), myocardial dysfunction (p=0.04), and cancer p=0.003)[22,146,147].

Finally, a significantly lower rate of ACA positivity and a trend towards increasing ATA positivity was observed in exposed patients. This was expected, since ACA positivity is usually associated with localized phenotype, without significant renal involvement, while ATA is associated with dcSSc and significant internal organ involvement. While mortality was not statistically significant, a trend towards increased mortality was observed in the exposed group.

To our knowledge, this is the first description of increased risk of renal crisis and gastrointestinal disease severity associated with exposure to organic solvents in SSc patients. Further investigation of the association and subsequently implementation of workplace interventions to minimize exposure.

Objective 4: Evaluate the epidemiology of psoriasis using populational data in Quebec, Canada and comprehensively evaluate the environmental factors that could be contributing to high psoriasis incidence

There were 43,663 patients with psoriasis identified in Quebec between 1994-2014 who were covered by the provincial drug plan. An increasing prevalence was observed over time while the incidence decreased. Uneven geographic distribution of incidence across the province was noted. Machine learning algorithm identified 9 predictors in the parsimonious model with a prediction accuracy of 0.77 (Area under the Curve (AUC)). These spanned climate, SES/demographic factors and BE characteristics.

In this study, patients with psoriasis were identified based on ICD codes between 1994-2014 from the provincial populational health administrative database. Although the definition used was not validated using Quebec data, it has been previously validated in Ontario. Specifically, the following definition was used in the study: ≥ 1 diagnostic code in hospital records or ≥ 2 diagnostic codes (ICD-9: 696.1 and ICD-10: L40.x) in physician claims. The sensitivity of this algorithm was 52%, the specificity was 99% and the positive predictive value (PPV) was 62%[85]. Another population-based study using the Manitoba Health administrative database assessed the combined psoriasis and psoriatic arthritis ICD codes (ICD-9: 696.0, 696.1 and ICD-10: L40.x, M07.0, M07.2, M07.3) and found that ≥ 1 hospital visit, ≥ 1 physician claim or ≥ 1 treatment for psoriasis had a sensitivity of 72%, a specificity of 90% and a PPV of 25%[148]. In the same study, ≥ 1 hospital visit or ≥ 1 physician claim had a sensitivity of 44%, specificity of 97% and PPV of 44%[148]. One population-based cohort study conducted in the United States, using the Kaiser Permanente Northern California health database found that ≥ 1 physician claim by any

specialist for psoriasis (ICD-9 codes: 696.1) had a sensitivity of 100% and a PPV of 78%, while ≥ 1 claim for psoriasis by a dermatologist had a sensitivity of 91% and a PPV of 89%[149]. Finally, one study using the Swedish national health administrative database found that ≥ 1 claim for psoriasis by any specialist (ICD-10: L40.x) had a PPV of 81%[150].

We demonstrated increasing prevalence over the study period and decreasing incidence. This is similar to a study of global burden of psoriasis which showed a steady decrease in ASIR across each age group between 1990 to 2019[151]. Another review of 308 studies also noted a decreasing trend in incidence and an increasing trend in prevalence[152]. Increased prevalence can be explained by better disease recognition and decreased mortality, however, the reasons for a decreasing incidence are not clear and require further investigation. It could be in part due to local billing trends such as changes in ICD coding in Quebec, but a similar observation in other countries including the United Kingdom, Russia, Italy, and Taiwan suggest that other factors may be at play.

As expected, mapping of the average incidence rates per FSA revealed an uneven geographic distribution and varied between1.6 to 325.6/100,000 person-years, similar to variations reported across the world of 30.3-321.0/100,000 person-years[24]. Increased incidence was noted in northern Quebec, corresponding to regions with a cold climate and limited sun exposure. Other studies found that colder/drier regions of Brazil and Spain also had increased psoriasis prevalence[31,32]. Another high incidence cluster was found at the border with the province of Newfoundland and Labrador, which has the highest reported incidence of psoriasis in Canada[153]. However, there are also FSAs with increased incidence within the greater Montreal Area indicating that factors other than climate likely play a role.

Using conventional statistics, only a limited number of variables can be assessed. Furthermore, adjustment for multiple confounders needs to be considered when multiple assessments are being made as it is possible to have a positive result by chance alone[154]. Over >350 environmental/neighbourhood variables were available for years 1999-2014 per FSA. To remove the bias of manually selecting a few features of interest and bypass the limitations of conventional statistics, we implemented a machine learning model to evaluate which factors could predict high psoriasis incidence (in the top 10% of all FSAs) in the year following.

This model identified 46 features that had some impact on predicting the probability of high psoriasis incidence (top 10%). These belonged to the domains of climate factors, socioeconomic neighborhood characteristics and BE features. The resulting model had an AUC of 0.80. A parsimonious model was completed to identify the least number of features that could give the highest AUC. Through this approach, 9 features were identified giving an AUC of 0.77. The highest-ranking features included UVR, maximum daily temperature, proportion of females, soil moisture, urbanization, and distance to expressways, all of which had a negative association with psoriasis incidence. Nighttime light brightness had a positive association, whereas social and material deprivation indices suggested a higher psoriasis incidence in the middle SES neighborhoods.

Increased greenness in the neighbourhood has previously been shown to reduce harmful exposures and encourage a healthy lifestyle[36,37,91-93]. This measure has ben well studied in other metabolic, cardiovascular and mental health conditions, all of which are comorbid with psoriasis, and were found to have a strong correlation[37,155,156]. Hence, advocating for the addition of green

space in new residential areas and high-risk neighbourhoods could aid in addressing high psoriasis incidence.

Increased nighttime light brightness had a positive association with psoriasis in our study and has similarly been studied in metabolic and mental health conditions. One study from China showed that increased nighttime light brightness was associated with atopic dermatitis[157], another common inflammatory condition in dermatology with genetic predisposition and environmental influence.

Socioeconomic factors also played an important role, as social deprivation ranked 3rd and material deprivation ranked 5th amongst the features identified. The former considers social circumstances such as being separated, divorced or widowed, living alone or in a single-parent family. The latter considers residential instability, economic dependency, ethno-cultural composition and situational vulnerability. As opposed to the expected trend of increased psoriasis with increased deprivation, as seen in prior studies[158,159], we found that the middle classes were more likely to have high psoriasis incidence.

The results of this study highlight the importance of SES and neighbourhood characteristics in predicting high psoriasis incidence.

Limitations

Objective 1, 4

Using a populational administrative database can lead to several limitations such as missed or misclassified cases which could affect the incidence/prevalence/mortality estimates. However,

the definition used in the SSc study has been previously validated in Quebec and showed a high sensitivity and specificity. For psoriasis, the use of ICD codes for diagnosis was validated in Ontario and other studies around the world, but not in Quebec. ICD codes are obtained based on data filled out by the clinician at the patient's visit. Hence, there may be changing trends over time in coding practices which could lead to changes in case identification. However, it is reassuring that our data, with a large number of patients, parallels other studies in the literature. As this is a populational study, no individual level data was available to assess patient demographic features such as sex, ethnicity, age or disease manifestations/comorbidities. For SSc, a contribution to geographic differences observed could be due to a genetic founder effect in small communities, however, the data to evaluate this was not available.

Objective 2

In this study, the CSRG registry was used to identify patient cases. In this case, the diagnosis is confirmed by a rheumatologist and disease features are available per patient. The registry is estimated to account for only about 10% of all prevalent cases in the country and may be skewed to include patients near recruitment centers. However, a number of FSAs in our study were still located far from recruitment centers. To minimize the effect of this limitation, we used a simulation to predict the distribution of cases that would be expected based on population size, age and sex proportion per FSA, to confirm the relevance of identified regions. The specific FSA at patient recruitment into the registry was available, however, prior FSA data was not available and hence, it is possible that some spatiotemporal trends were missed.

Objective 3

For both the study on silica and organic solvent exposure among SSc patients, the CSRG registry was used. The occupational exposure is self-reported by the patients and the intensity and duration of the exposure could not be collected. There is also a risk of recall bias, although in higher risk occupations, it is likely that patients are informed of possible exposures and would remember upon questioning for the study. This approach also does not capture possible passive exposure from the environment (*e.g.* nearby mining operations) or household solvent exposures.

Strengths

Firstly, for objective 1 and 4, a populational database was used to identify a comprehensive set of patients over a two-decade period and allow detailed investigation of updated epidemiologic trends for both SSc and psoriasis. For SSc, the large cohort of patients identified allows stratification of incidence, prevalence, and mortality by age group and sex to further enhance our understanding which is limited in the literature currently.

In objective 2, a national registry, CSRG, was used with an ecological study design to study environmental risk factors. While this type of study design inherently has several limitations, it can provide critical information and lead to identification of novel causative factors as occurred with discovery of asbestos as the primary cause of mesothelioma[7]. Similar studies completed in the realm of cutaneous malignancies such as melanoma, non-melanoma skin cancer, and cutaneous T cell lymphoma have already translated to important public health initiatives[100,109,111,160].

In objective 4, this study was the first to comprehensively investigate the predictive environmental/neighbourhood factors contributing to high psoriasis incidence using populational

databases and sophisticated machine learning algorithms. The machine learning approach helps to overcome the limitations of conventional statistics and use all available variables/features to comprehensively study risk factors. We believe the findings of this study could translate to into targeted patient counselling and implementation of preventative strategies. Ultimately, this will benefit patients and their families and hopefully, in the future, decrease the global burden of psoriatic disease. Additionally, this approach can set a new standard to study the impact of the environment on the risk of chronic diseases in dermatology and beyond.

Hence, the goal of this thesis was to determine updated epidemiology and geographic distribution of SSc and psoriasis and evaluate environmental and occupational risk factors which can translate to patient education on disease prevention, advocacy for workplace/environmental interventions, and prioritization of resources to high-risk areas.

Chapter 9: Summary and Conclusions

This thesis provided an important contribution to understanding epidemiology of a rare disease such as SSc that is associated with high morbidity and mortality and of a common inflammatory skin disease -psoriasis. For both conditions the pathogenesis includes genetic predisposition and an environmental exposure which is not well characterized in most cases. This thesis highlights that exogenous factors encountered in the environment, neighbourhood, or occupation can contribute to disease development.

Findings from this thesis showed that there is an increasing trend in incidence and prevalence of SSc over the last two decades however, there is a declining mortality trend which is reassuring. There was uneven geographic distribution seen both at a national level and provincial

level (Quebec). Increased air pollution and high industrial density in FSAs correlated with areas of high SSc period prevalence. Occupational exposure to silica and/or organic solvents occurred in about 30% of the cohort. Exposed patients were more likely to be males, smokers, and have more severe disease phenotype.

For the study on psoriasis, while the prevalence was also increasing over the time period, the incidence was decreasing. Similarly, an uneven geographic distribution was observed. The top negative predictors from the machine learning model for high psoriasis incidence were found to be UVR, maximum daily temperature, proportion of females, soil moisture, urbanization, and distance to expressways. Nighttime light brightness was a positive predictor. Social and material deprivation indices suggested a higher psoriasis incidence in the middle SES neighborhoods.

Hence identifying contributing factors can help reduce the burden of disease, improved patient education, implement preventative strategies, and advocate for improved resource allocation/neighbourhood modifications in high-risk areas.

References

1. Bach JF. The hygiene hypothesis in autoimmunity: the role of pathogens and commensals. *Nat Rev Immunol* 2018; **18**(2): 105-20.
2. Dinse GE, Parks CG, Weinberg CR, et al. Increasing Prevalence of Antinuclear Antibodies in the United States. *Arthritis Rheumatol* 2022; **74**(12): 2032-41.
3. Ali N, Rosenblum MD. Regulatory T cells in skin. *Immunology* 2017; **152**(3): 372-81.
4. Sanchez Rodriguez R, Pauli ML, Neuhaus IM, et al. Memory regulatory T cells reside in human skin. *J Clin Invest* 2014; **124**(3): 1027-36.
5. In: Anaya JM, Shoenfeld Y, Rojas-Villarraga A, Levy RA, Cervera R, eds. Autoimmunity: From Bench to Bedside. Bogota (Colombia); 2013.
6. Ouchene L MA, Lavoue J, Baron M, Litvinov IV, Netchiporouk E. Toward understanding of environmental triggers of systemic sclerosis. *J Cutan Med Surg (In Press)* 2020.
7. McDonald JC. Epidemiology of malignant mesothelioma--an outline. *Ann Occup Hyg* 2010; **54**(8): 851-7.
8. Cutler JJ, Parker GS, Rosen S, Prenney B, Healey R, Caldwell GG. Childhood leukemia in Woburn, Massachusetts. *Public Health Rep* 1986; **101**(2): 201-5.
9. Jerjen R, Nikpour M, Krieg T, Denton CP, Saracino AM. Systemic sclerosis in adults. Part I: Clinical features and pathogenesis. *J Am Acad Dermatol* 2022; **87**(5): 937-54.
10. Takeshita J, Grewal S, Langan SM, et al. Psoriasis and comorbid diseases: Epidemiology. *J Am Acad Dermatol* 2017; **76**(3): 377-90.
11. Fang D, Chen B, Lescoat A, Khanna D, Mu R. Immune cell dysregulation as a mediator of fibrosis in systemic sclerosis. *Nature Reviews Rheumatology* 2022; **18**(12): 683-93.
12. Sieminska I, Pieniawska M, Grzywa TM. The Immunology of Psoriasis-Current Concepts in Pathogenesis. *Clin Rev Allergy Immunol* 2024; **66**(2): 164-91.
13. Alba MA, Velasco C, Simeon CP, et al. Early- versus late-onset systemic sclerosis: differences in clinical presentation and outcome in 1037 patients. *Medicine (Baltimore)* 2014; **93**(2): 73-81.
14. Joven BE, Almodovar R, Carmona L, Carreira PE. Survival, causes of death, and risk factors associated with mortality in Spanish systemic sclerosis patients: results from a single university hospital. *Semin Arthritis Rheum* 2010; **39**(4): 285-93.
15. Peytrignet S, Denton CP, Lunt M, et al. Disability, fatigue, pain and their associates in early diffuse cutaneous systemic sclerosis: the European Scleroderma Observational Study. *Rheumatology (Oxford)* 2018; **57**(2): 370-81.
16. Bernatsky S, Joseph L, Pineau CA, Belisle P, Hudson M, Clarke AE. Scleroderma prevalence: demographic variations in a population-based sample. *Arthritis Rheum* 2009; **61**(3): 400-4.
17. Matucci-Cerinic M, Bellando-Randone S, Lepri G, Bruni C, Guiducci S. Very early versus early disease: the evolving definition of the 'many faces' of systemic sclerosis. *Ann Rheum Dis* 2013; **72**(3): 319-21.
18. Gabrielli A, Avvedimento EV, Krieg T. Scleroderma. *N Engl J Med* 2009; **360**(19): 1989-2003.

19. Feghali-Bostwick C, Medsger TA, Jr., Wright TM. Analysis of systemic sclerosis in twins reveals low concordance for disease and high concordance for the presence of antinuclear antibodies. *Arthritis Rheum* 2003; **48**(7): 1956-63.
20. Ouchene L MA, Bernatsky S, Hudson M, Baron M, Netchiporouk E. Toward understanding of environmental triggers of systemic sclerosis. *JAMA Dermatology (submitted, under review)* 2020; (Under review).
21. Dolcino M, Pelosi A, Fiore PF, et al. Gene Profiling in Patients with Systemic Sclerosis Reveals the Presence of Oncogenic Gene Signatures. *Front Immunol* 2018; **9**: 449.
22. Ouchene L, Muntyanu A, Lavoue J, Baron M, Litvinov IV, Netchiporouk E. Toward Understanding of Environmental Risk Factors in Systemic Sclerosis. *J Cutan Med Surg* 2020: 1203475420957950.
23. Nakamura K, Jinnin M, Harada M, et al. Altered expression of CD63 and exosomes in scleroderma dermal fibroblasts. *Journal of dermatological science* 2016; **84**(1): 30-9.
24. Parisi R, Iskandar IYK, Kontopantelis E, et al. National, regional, and worldwide epidemiology of psoriasis: systematic analysis and modelling study. *BMJ* 2020; **369**: m1590.
25. Boehncke WH, Schon MP. Psoriasis. *Lancet* 2015; **386**(9997): 983-94.
26. Armstrong AW. Psoriasis. *JAMA Dermatology* 2017; **153**(9): 956-.
27. Salman A, Yucelten AD, Sarac E, Saricam MH, Perdahli-Fis N. Impact of psoriasis in the quality of life of children, adolescents and their families: a cross-sectional study. *An Bras Dermatol* 2018; **93**(6): 819-23.
28. Liu S, He M, Jiang J, et al. Triggers for the onset and recurrence of psoriasis: a review and update. *Cell Commun Signal* 2024; **22**(1): 108.
29. Feldman SR, Tian H, Wang X, Germino R. Health Care Utilization and Cost Associated with Biologic Treatment Patterns Among Patients with Moderate to Severe Psoriasis: Analyses from a Large U.S. Claims Database. *J Manag Care Spec Pharm* 2019; **25**(4): 479-88.
30. Mehrmal S, Uppal P, Nedley N, Giesey RL, Delost GR. The global, regional, and national burden of psoriasis in 195 countries and territories, 1990 to 2017: A systematic analysis from the Global Burden of Disease Study 2017. *J Am Acad Dermatol* 2021; **84**(1): 46-52.
31. Ferrandiz C, Carrascosa JM, Toro M. Prevalence of psoriasis in Spain in the age of biologics. *Actas Dermosifiliogr* 2014; **105**(5): 504-9.
32. Romiti R, Amone M, Menter A, Miot HA. Prevalence of psoriasis in Brazil - a geographical survey. *Int J Dermatol* 2017; **56**(8): e167-e8.
33. Karppinen TT, Ylianttila L, Kautiainen H, Reunala T, Snellman E. Empowering heliotherapy improves clinical outcome and quality of life of psoriasis and atopic dermatitis patients. *Acta Derm Venereol* 2015; **95**(5): 579-82.
34. Conrad N, Misra S, Verbakel JY, et al. Incidence, prevalence, and co-occurrence of autoimmune disorders over time and by age, sex, and socioeconomic status: a population-based cohort study of 22 million individuals in the UK. *Lancet* 2023; **401**(10391): 1878-90.
35. Miller FW. The increasing prevalence of autoimmunity and autoimmune diseases: an urgent call to action for improved understanding, diagnosis, treatment, and prevention. *Curr Opin Immunol* 2023; **80**: 102266.
36. Doubleday A, Knott CJ, Hazlehurst MF, Bertoni AG, Kaufman JD, Hajat A. Neighborhood greenspace and risk of type 2 diabetes in a prospective cohort: the Multi-Ethncity Study of Atherosclerosis. *Environ Health* 2022; **21**(1): 18.

37. Feng S, Meng Q, Guo B, et al. Joint exposure to air pollution, ambient temperature and residential greenness and their association with metabolic syndrome (MetS): A large population-based study among Chinese adults. *Environ Res* 2022; **214**(Pt 1): 113699.
38. Glass GE. Update: spatial aspects of epidemiology: the interface with medical geography. *Epidemiol Rev* 2000; **22**(1): 136-9.
39. Rezaeian M, Dunn G, St Leger S, Appleby L. The production and interpretation of disease mapsA methodological case-study. *Soc Psychiatry Psychiatr Epidemiol* 2004; **39**(12): 947-54.
40. Asano Y, Jinnin M, Kawaguchi Y, et al. Diagnostic criteria, severity classification and guidelines of systemic sclerosis. *J Dermatol* 2018; **45**(6): 633-91.
41. Calderon LM, Pope JE. Scleroderma epidemiology update. *Curr Opin Rheumatol* 2021; **33**(2): 122-7.
42. van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. *Arthritis Rheum* 2013; **65**(11): 2737-47.
43. Avouac J, Fransen J, Walker UA, et al. Preliminary criteria for the very early diagnosis of systemic sclerosis: results of a Delphi Consensus Study from EULAR Scleroderma Trials and Research Group. *Ann Rheum Dis* 2011; **70**(3): 476-81.
44. Domsic RT. Scleroderma: the role of serum autoantibodies in defining specific clinical phenotypes and organ system involvement. *Curr Opin Rheumatol* 2014; **26**(6): 646-52.
45. Valentini G, Iudici M, Walker UA, et al. The European Scleroderma Trials and Research group (EUSTAR) task force for the development of revised activity criteria for systemic sclerosis: derivation and validation of a preliminarily revised EUSTAR activity index. *Ann Rheum Dis* 2017; **76**(1): 270-6.
46. Nihtyanova SI, Sari A, Harvey JC, et al. Using Autoantibodies and Cutaneous Subset to Develop Outcome-Based Disease Classification in Systemic Sclerosis. *Arthritis Rheumatol* 2020; **72**(3): 465-76.
47. Ouchene L, Muntyanu A, Assayag D, et al. Skin disorders and interstitial lung disease: Part II-The spectrum of cutaneous diseases with lung disease association. *J Am Acad Dermatol* 2023; **88**(4): 767-82.
48. Kahan A, Coghlan G, McLaughlin V. Cardiac complications of systemic sclerosis. *Rheumatology* 2009; **48**(suppl_3): iii45-iii8.
49. Caron M, Hudson M, Baron M, Nessim S, Steele R, Canadian Scleroderma Research G. Longitudinal study of renal function in systemic sclerosis. *J Rheumatol* 2012; **39**(9): 1829-34.
50. Woodworth TG, Suliman YA, Li W, Furst DE, Clements P. Scleroderma renal crisis and renal involvement in systemic sclerosis. *Nat Rev Nephrol* 2018; **14**(2): 137.
51. Shreiner AB, Murray C, Denton C, Khanna D. Gastrointestinal Manifestations of Systemic Sclerosis. *J Scleroderma Relat Disord* 2016; **1**(3): 247-56.
52. Chifflot H, Fautrel B, Sordet C, Chatelus E, Sibilia J. Incidence and prevalence of systemic sclerosis: a systematic literature review. *Semin Arthritis Rheum* 2008; **37**(4): 223-35.
53. Zhong L, Pope M, Shen Y, Hernandez JJ, Wu L. Prevalence and incidence of systemic sclerosis: A systematic review and meta-analysis. *Int J Rheum Dis* 2019; **22**(12): 2096-107.
54. Robinson D, Jr., Eisenberg D, Nietert PJ, et al. Systemic sclerosis prevalence and comorbidities in the US, 2001-2002. *Curr Med Res Opin* 2008; **24**(4): 1157-66.

55. Silman AJ, Howard Y, Hicklin AJ, Black C. Geographical clustering of scleroderma in south and west London. *Br J Rheumatol* 1990; **29**(2): 93-6.
56. Valesini G, Litta A, Bonavita MS, et al. Geographical clustering of scleroderma in a rural area in the province of Rome. *Clin Exp Rheumatol* 1993; **11**(1): 41-7.
57. Diegel R, Black B, Pfau JC, McNew T, Noonan C, Flores R. Case series: rheumatological manifestations attributed to exposure to Libby Asbestiform Amphiboles. *J Toxicol Environ Health A* 2018; **81**(15): 734-47.
58. Arnett FC, Howard RF, Tan F, et al. Increased prevalence of systemic sclerosis in a Native American tribe in Oklahoma. Association with an Amerindian HLA haplotype. *Arthritis Rheum* 1996; **39**(8): 1362-70.
59. Thompson AE, Pope JE. Increased prevalence of scleroderma in southwestern Ontario: a cluster analysis. *J Rheumatol* 2002; **29**(9): 1867-73.
60. Jacobs L. Kahnawake Scleroderma Support Group. https://sclero.org/scleroderma/support/groups/canada/quebec/kahnawake/a-to-z.html.
61. Furst DE, Fernandes AW, Iorga SR, Greth W, Bancroft T. Annual medical costs and healthcare resource use in patients with systemic sclerosis in an insured population. *J Rheumatol* 2012; **39**(12): 2303-9.
62. Bernatsky S, Hudson M, Panopalis P, et al. The cost of systemic sclerosis. *Arthritis Rheum* 2009; **61**(1): 119-23.
63. Kuo CF, Luo SF, Yu KH, See LC, Zhang W, Doherty M. Familial risk of systemic sclerosis and co-aggregation of autoimmune diseases in affected families. *Arthritis Res Ther* 2016; **18**(1): 231.
64. Muntyanu A, Le M, Ridha Z, et al. Novel role of long non-coding RNAs in autoimmune cutaneous disease. *J Cell Commun Signal* 2022; **16**(4): 487-504.
65. De Martinis M, Ciccarelli F, Sirufo MM, Ginaldi L. An overview of environmental risk factors in systemic sclerosis. *Expert Rev Clin Immunol* 2016; **12**(4): 465-78.
66. World Health Organisation 2020. https://www.who.int/airpollution/news-and-events/how-air-pollution-is-destroying-our-health.
67. Zhao JH, Duan Y, Wang YJ, Huang XL, Yang GJ, Wang J. The Influence of Different Solvents on Systemic Sclerosis: An Updated Meta-analysis of 14 Case-Control Studies. *J Clin Rheumatol* 2016; **22**(5): 253-9.
68. Siemiatycki J, Lavoué J. Availability of a New Job-Exposure Matrix (CANJEM) for Epidemiologic and Occupational Medicine Purposes. *J Occup Environ Med* 2018; **60**(7): e324-e8.
69. Ndong Ba A, Verdin A, Cazier F, et al. Individual exposure level following indoor and outdoor air pollution exposure in Dakar (Senegal). *Environ Pollut* 2019; **248**: 397-407.
70. Pierce C, Fuhrman E, Xiong-Yang P, et al. Monitoring of airborne particulates near industrial silica sand mining and processing facilities. *Arch Environ Occup Health* 2019; **74**(4): 185-96.
71. Bernatsky S, Smargiassi A, Johnson M, et al. Fine particulate air pollution, nitrogen dioxide, and systemic autoimmune rheumatic disease in Calgary, Alberta. *Environ Res* 2015; **140**: 474-8.

72. Ghazawi FM, Alghazawi N, Le M, et al. Environmental and Other Extrinsic Risk Factors Contributing to the Pathogenesis of Cutaneous T Cell Lymphoma (CTCL). *Front Oncol* 2019; **9**: 300.
73. Ghazawi FM, Cyr J, Darwich R, et al. Cutaneous malignant melanoma incidence and mortality trends in Canada: A comprehensive population-based study. *J Am Acad Dermatol* 2019; **80**(2): 448-59.
74. Jerjen R, Nikpour M, Krieg T, Denton CP, Saracino AM. Systemic sclerosis in adults. Part II: management and therapeutics. *J Am Acad Dermatol* 2022; **87**(5): 957-78.
75. Spierings J, Chiu YH, Voortman M, van Laar JM. Autologous stem-cell transplantation in systemic sclerosis-associated interstitial lung disease: early action in selected patients rather than escalation therapy for all. *Ther Adv Musculoskelet Dis* 2021; **13**: 1759720x211035196.
76. Del Papa N, Pignataro F, Zaccara E, Maglione W, Minniti A. Autologous Hematopoietic Stem Cell Transplantation for Treatment of Systemic Sclerosis. *Frontiers in immunology* 2018; **9**: 2390-.
77. Griffiths CEM, Armstrong AW, Gudjonsson JE, Barker J. Psoriasis. *Lancet* 2021; **397**(10281): 1301-15.
78. Queiro R, Tejón P, Alonso S, Coto P. Age at disease onset: a key factor for understanding psoriatic disease. *Rheumatology (Oxford)* 2014; **53**(7): 1178-85.
79. Clinical Review Report: Guselkumab (Tremfya): (Janssen Inc.): Indication: For the treatment of adult patients with moderate-to-severe plaque psoriasis who are candidates for systemic therapy or phototherapy [Internet]. Ottawa (ON): Canadian Agency for Drugs and Technologies in Health; Appendix 5, Validity of Outcome Measures. 2018. https://www.ncbi.nlm.nih.gov/books/NBK534046/.
80. Pearl RL, Wan MT, Takeshita J, Gelfand JM. Stigmatizing attitudes toward persons with psoriasis among laypersons and medical students. *J Am Acad Dermatol* 2019; **80**(6): 1556-63.
81. Mazurkiewicz N, Krefta J, Lipowska M. Attitudes Towards Appearance and Body-Related Stigma Among Young Women With Obesity and Psoriasis. *Front Psychiatry* 2021; **12**: 788439.
82. Alpsoy E, Polat M, Yavuz IH, et al. Internalized Stigma in Pediatric Psoriasis: A Comparative Multicenter Study. *Ann Dermatol* 2020; **32**(3): 181-8.
83. Grover S, Mehra A, Dogra S, et al. Internalized Stigma and Psychiatric Morbidity among Patients with Psoriasis: A Study from North India. *Indian Dermatol Online J* 2021; **12**(1): 97-104.
84. Wan MT, Pearl RL, Chiesa Fuxench ZC, Takeshita J, Gelfand JM. Anticipated and Perceived Stigma Among Patients With Psoriasis. *J Psoriasis Psoriatic Arthritis* 2020; **5**(3): 93-9.
85. Eder L, Widdifield J, Rosen CF, et al. Trends in the Prevalence and Incidence of Psoriasis and Psoriatic Arthritis in Ontario, Canada: A Population-Based Study. *Arthritis Care Res (Hoboken)* 2019; **71**(8): 1084-91.
86. Petrella RJ, Gregory V, Luciani L, Barbeau M. Characteristics Of Chronic Plaque Psoriasis In Canada: A Retrospective Database Study. *Value in Health* 2014; **17**(3): A284.
87. Levy AR, Davie AM, Brazier NC, et al. Economic burden of moderate to severe plaque psoriasis in Canada. *Int J Dermatol* 2012; **51**(12): 1432-40.

88. Chan B, Hales B, Shear N, et al. Work-related lost productivity and its economic impact on Canadian patients with moderate to severe psoriasis. *J Cutan Med Surg* 2009; **13**(4): 192-7.
89. Huang YH, Kuo CF, Huang LH, Hsieh MY. Familial Aggregation of Psoriasis and Co-Aggregation of Autoimmune Diseases in Affected Families. *J Clin Med* 2019; **8**(1).
90. Luo OD, Ridha Z, Jfri A, et al. Neighborhood characteristics and the risk of psoriasis: A systematic review. *JAAD Int* 2023; **13**: 100-1.
91. Ponjoan A, Blanch J, Alves-Cabratosa L, et al. Impact of residential greenness on myocardial infarction in the population with diabetes: A sex-dependent association? *Environ Res* 2022; **205**: 112449.
92. Ke P, Xu M, Xu J, et al. Association of residential greenness with the risk of metabolic syndrome in Chinese older adults: a longitudinal cohort study. *J Endocrinol Invest* 2023; **46**(2): 327-35.
93. Khan JR, Sultana A, Islam MM, Biswas RK. A negative association between prevalence of diabetes and urban residential area greenness detected in nationwide assessment of urban Bangladesh. *Scientific Reports* 2021; **11**(1): 19513.
94. Zhao SS, Bellou E, Verstappen SMM, et al. Association between psoriatic disease and lifestyle factors and comorbidities: cross-sectional analysis and Mendelian randomization. *Rheumatology (Oxford)* 2023; **62**(3): 1272-85.
95. Debbaneh M, Millsop JW, Bhatia BK, Koo J, Liao W. Diet and psoriasis, part I: Impact of weight loss interventions. *J Am Acad Dermatol* 2014; **71**(1): 133-40.
96. Biskup M, Macek P, Gozdz S, et al. Two-year follow-up cohort study focused on gender-specific associations between socioeconomic status and body weight changes in overweight and obese middle-aged and older adults. *BMJ Open* 2021; **11**(7): e050127.
97. Renalds A, Smith TH, Hale PJ. A systematic review of built environment and health. *Fam Community Health* 2010; **33**(1): 68-78.
98. Canada S. More information on Postal code. 2006. https://www12.statcan.gc.ca/census-recensement/2006/ref/dict/geo035a-eng.cfm.
99. Canada S. Population Centre and Rural Area Classification 2016. 2017. https://www.statcan.gc.ca/eng/subjects/standard/pcrac/2016/introduction2020).
100. Lagace F, Ghazawi FM, Le M, et al. Penile Invasive Squamous Cell Carcinoma: Analysis of Incidence, Mortality Trends, and Geographic Distribution in Canada. *J Cutan Med Surg* 2020; **24**(2): 124-8.
101. Ghazawi FM, Le M, Cyr J, et al. Analysis of acute myeloid leukemia incidence and geographic distribution in Canada from 1992 to 2010 reveals disease clusters in Sarnia and other industrial US border cities in Ontario. *Cancer* 2019; **125**(11): 1886-97.
102. 2011 Census - Boundary files. 2011. https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2011-eng.cfm.
103. Medsger TA, Jr., Bombardieri S, Czirjak L, Scorza R, Della Rossa A, Bencivelli W. Assessment of disease severity and prognosis. *Clin Exp Rheumatol* 2003; **21**(3 Suppl 29): S42-6.
104. Enhanced Point of Interest. Markham: DMTI Spatial Inc.; 2013.
105. Kim G, Lee M-I, Lee S, Choi S-D, Kim S-J, Song C-K. Numerical Modeling for the Accidental Dispersion of Hazardous Air Pollutants in the Urban Metropolitan Area. 2020; **11**(5): 477.

106. Kassamali B, Kassamali AA, Muntyanu A, Netchiporouk E, Vleugels RA, LaChance A. Geographic distribution and environmental triggers of systemic sclerosis cases from 2 large academic tertiary centers in Massachusetts. *J Am Acad Dermatol* 2022; **86**(4): 925-7.
107. Elston DM. Letter from the Editor: Chemical exposure and sclerosing disorders. *Journal of the American Academy of Dermatology* 2024; **90**(3): 477-8.
108. Muntyanu A, Ghazawi FM, Pastukhova E, Zubarev A, Litvinov IV. National trends in incidence and geographic distribution of melanoma and keratinocyte carcinoma in the Russian Federation. *Front Med (Lausanne)* 2023; **10**: 1188038.
109. Muntyanu A, Nechaev V, Pastukhova E, et al. Risk factors and communities disproportionately affected by cervical cancer in the Russian Federation: A national population-based study. *Lancet Reg Health Eur* 2022; **20**: 100454.
110. Muntyanu A, Nechaev V, Pastukhova E, et al. Burden and geographic distribution of oral cavity and oropharyngeal cancers in the Russian Federation. *Front Oncol* 2023; **13**: 1197287.
111. Muntyanu A, Savin E, Ghazawi FM, Alakel A, Zubarev A, Litvinov IV. Geographic Variations in Cutaneous Melanoma Distribution in the Russian Federation. *Dermatology* 2020; **236**(6): 500-7.
112. Price S, Bailey S, Hamilton W, Jones D, Mounce L, Abel G. The effects of the first UK lockdown for the COVID-19 pandemic on primary-care-recorded cancer and type-2 diabetes mellitus records: A population-based quasi-experimental time series study. *Cancer Epidemiol* 2024; **91**: 102605.
113. National Cancer Institute: Surveillance E, and End Results Program. Impact of COVID on the April 2024 SEER Data Release. 2024. https://seer.cancer.gov/data/covid-impact.html.
114. Heer E, Ruan Y, Boyne DJ, et al. Impact of the COVID-19 pandemic on cancer diagnoses, stage and survival in Alberta. *CMAJ* 2023; **195**(23): E804-E12.
115. Bernatsky S, Linehan T, Hanly JG. The accuracy of administrative data diagnoses of systemic autoimmune rheumatic diseases. *J Rheumatol* 2011; **38**(8): 1612-6.
116. Butt SA, Jeppesen JL, Fuchs C, et al. Trends in incidence, mortality, and causes of death associated with systemic sclerosis in Denmark between 1995 and 2015: a nationwide cohort study. *BMC Rheumatol* 2018; **2**: 36.
117. Bairkdar M, Rossides M, Westerlind H, Hesselstrand R, Arkema EV, Holmqvist M. Incidence and prevalence of systemic sclerosis globally: a comprehensive systematic review and meta-analysis. *Rheumatology (Oxford)* 2021; **60**(7): 3121-33.
118. Mayes MD, Lacey JV, Jr., Beebe-Dimmer J, et al. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. *Arthritis Rheum* 2003; **48**(8): 2246-55.
119. 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.
120. Ciaffi J, Morabito MF, Ruscitti P, et al. Incidence, prevalence and mortality of systemic sclerosis in Italy: a nationwide population-based study using administrative health data. *Rheumatol Int* 2021; **41**(1): 129-37.
121. Pelkonen PM, Jalanko HJ, Lantto RK, et al. Incidence of systemic connective tissue diseases in children: a nationwide prospective study in Finland. *J Rheumatol* 1994; **21**(11): 2143-6.

122. Yen EY, Singh DR, Singh RR. Trends in Systemic Sclerosis Mortality Over Forty-Eight Years, 1968-2015: A US Population-Based Study. *Arthritis Care Res (Hoboken)* 2021; **73**(10): 1502-10.
123. Hao Y, Hudson M, Baron M, et al. Early Mortality in a Multinational Systemic Sclerosis Inception Cohort. *Arthritis Rheumatol* 2017; **69**(5): 1067-77.
124. Peoples C, Medsger TA, Jr., Lucas M, Rosario BL, Feghali-Bostwick CA. Gender differences in systemic sclerosis: relationship to clinical features, serologic status and outcomes. *J Scleroderma Relat Disord* 2016; **1**(2): 177-240.
125. Hughes M, Pauling JD, Armstrong-James L, Denton CP, Galdas P, Flurey C. Gender-related differences in systemic sclerosis. *Autoimmunity Reviews* 2020; **19**(4): 102494.
126. Li JX. Secular Trends in Systemic Sclerosis Mortality in the United States from 1981 to 2020. *Int J Environ Res Public Health* 2022; **19**(22).
127. Sclero.org. Kahnawake Scleroderma Support Group. https://sclero.org/scleroderma/support/groups/canada/quebec/kahnawake/a-to-z.html2020).
128. Canada Go. Minerals and the economy 2020. https://www.nrcan.gc.ca/our-natural-resources/minerals-mining/minerals-metals-facts/minerals-and-the-economy/205292020).
129. Plan TCMaM. The Canadian Minerals and Metals Plan. 2020. https://www.minescanada.ca/en2020).
130. Pedersen LM, Permin H. Rheumatic disease, heavy-metal pigments, and the Great Masters. *Lancet* 1988; **1**(8597): 1267-9.
131. Ministry of the Environment and Climate Change. Air quality in

Ontario: 2015 Report. Toronto, Ontario, Canada: Ministry of the

Environment and Climate Change, 2015.
132. Rubio-Rivas M, Moreno R, Corbella X. Occupational and environmental scleroderma. Systematic review and meta-analysis. *Clin Rheumatol* 2017; **36**(3): 569-82.
133. (IARC) IAfRoC. Monograph summary. 1997. https://publications.iarc.fr/86.
134. Canada C. Silica (Crystalline) Profile. https://www.carexcanada.ca/profile/silica_crystalline/.
135. Safety CCfOHa. Respirable Crystalline Silica: Breathe Easier. https://www.ccohs.ca/newsletters/hsreport/issues/2017/05/ezine.html.
136. (OSHA) OSaHA. Safety and health topics: Silica, crystalline. 2022. https://www.osha.gov/silica-crystalline.
137. Patel S, Morrisroe K, Proudman S, et al. Occupational silica exposure in an Australian systemic sclerosis cohort. *Rheumatology (Oxford)* 2020; **59**(12): 3900-5.
138. Barragan-Martinez C, Speck-Hernandez CA, Montoya-Ortiz G, Mantilla RD, Anaya JM, Rojas-Villarraga A. Organic solvents as risk factor for autoimmune diseases: a systematic review and meta-analysis. *PLoS One* 2012; **7**(12): e51506.
139. Kettaneh A, Al Moufti O, Tiev KP, et al. Occupational exposure to solvents and gender-related risk of systemic sclerosis: a metaanalysis of case-control studies. *J Rheumatol* 2007; **34**(1): 97-103.
140. Aryal BK, Khuder SA, Schaub EA. Meta-analysis of systemic sclerosis and exposure to solvents. *Am J Ind Med* 2001; **40**(3): 271-4.

141. Silman AJ, Hochberg MC. Occupational and environmental influences on scleroderma. *Rheumatic diseases clinics of North America* 1996; **22**(4): 737-49.
142. Haustein UF, Herrmann K. Environmental scleroderma. *Clinics in dermatology* 1994; **12**(3): 467-73.
143. Garabrant DH, Dumas C. Epidemiology of organic solvents and connective tissue disease. *Arthritis Res* 2000; **2**(1): 5-15.
144. Jacob S, Héry M, Protois JC, Rossert J, Stengel B. Effect of organic solvent exposure on chronic kidney disease progression: the GN-PROGRESS cohort study. *Journal of the American Society of Nephrology : JASN* 2007; **18**(1): 274-81.
145. STENGEL B, CÉNÉE S, LIMASSET J-C, et al. Organic Solvent Exposure May Increase the Risk of Glomerular Nephropathies with Chronic Renal Failure. *International Journal of Epidemiology* 1995; **24**(2): 427-34.
146. Marie I, Gehanno JF, Bubenheim M, et al. Prospective study to evaluate the association between systemic sclerosis and occupational exposure and review of the literature. *Autoimmun Rev* 2014; **13**(2): 151-6.
147. De Decker E, Vanthuyne M, Blockmans D, et al. High prevalence of occupational exposure to solvents or silica in male systemic sclerosis patients: a Belgian cohort analysis. *Clin Rheumatol* 2018; **37**(7): 1977-82.
148. Marrie RA, Patten SB, Tremlett H, Wolfson C, Leung S, Fisk JD. Increased incidence and prevalence of psoriasis in multiple sclerosis. *Mult Scler Relat Disord* 2017; **13**: 81-6.
149. Asgari MM, Wu JJ, Gelfand JM, et al. Validity of diagnostic codes and prevalence of psoriasis and psoriatic arthritis in a managed care population, 1996-2009. *Pharmacoepidemiol Drug Saf* 2013; **22**(8): 842-9.
150. Lofvendahl S, Theander E, Svensson A, Carlsson KS, Englund M, Petersson IF. Validity of diagnostic codes and prevalence of physician-diagnosed psoriasis and psoriatic arthritis in southern Sweden--a population-based register study. *PLoS One* 2014; **9**(5): e98024.
151. Zhang Y, Dong S, Ma Y, Mou Y. Burden of psoriasis in young adults worldwide from the global burden of disease study 2019. *Front Endocrinol (Lausanne)* 2024; **15**: 1308822.
152. Iskandar IYK, Parisi R, Griffiths CEM, Ashcroft DM. Systematic review examining changes over time and variation in the incidence and prevalence of psoriasis by age and gender. *Br J Dermatol* 2021; **184**(2): 243-58.
153. Badaiki W, Penney M, Pyper E, et al. Real World Studies of Psoriasis and Mental Illness in Newfoundland and Labrador. *J Cutan Med Surg* 2022; **26**(5): 494-501.
154. Pourhoseingholi MA, Baghestani AR, Vahedi M. How to control confounding effects by statistical analysis. *Gastroenterol Hepatol Bed Bench* 2012; **5**(2): 79-83.
155. Yang BY, Hu LW, Jalaludin B, et al. Association Between Residential Greenness, Cardiometabolic Disorders, and Cardiovascular Disease Among Adults in China. *JAMA Netw Open* 2020; **3**(9): e2017507.
156. Paoin K, Pharino C, Phosri A, et al. Association between greenness and cardiovascular risk factors: Results from a large cohort study in Thailand. *Environ Res* 2023; **220**: 115215.
157. Tang Z, Li S, Shen M, et al. Association of exposure to artificial light at night with atopic diseases: A cross-sectional study in college students. *Int J Hyg Environ Health* 2022; **241**: 113932.

158. Li X, Sundquist J, Hamano T, Sundquist K. Neighborhood Deprivation and Risks of Autoimmune Disorders: A National Cohort Study in Sweden. *Int J Environ Res Public Health* 2019; **16**(20).
159. Lofvendahl S, Joud A, Petersson IF, Theander E, Svensson A, Carlsson KS. Income disparities in healthcare use remain after controlling for healthcare need: evidence from Swedish register data on psoriasis and psoriatic arthritis. *Eur J Health Econ* 2018; **19**(3): 447-62.
160. Lagacé F, Noorah BN, Conte S, et al. Assessing Skin Cancer Risk Factors, Sun Safety Behaviors and Melanoma Concern in Atlantic Canada: A Comprehensive Survey Study. *Cancers (Basel)* 2023; **15**(15).