

The role of spatial epidemiology to support public  
health policies : case studies applied to health  
promotion, noncommunicable and infectious diseases  
in the canton of Vaud, Switzerland

Présentée le 12 janvier 2024

Faculté de l'environnement naturel, architectural et construit  
Laboratoire de géochimie biologique  
Programme doctoral en génie civil et environnement

pour l'obtention du grade de Docteur ès Sciences

par

**Anaïs Laurence LADOY**

Acceptée sur proposition du jury

Prof. C. Grossiord, présidente du jury  
Dr S. Joost, Prof. I. Guessous, directeurs de thèse  
Prof. L. Waller, rapporteur  
Prof. S. Monod, rapporteuse  
Prof. M. Salathé, rapporteur





*La vérité n'a pas besoin d'être dite pour être manifestée,  
(...) on peut peut-être la recueillir plus sûrement,  
sans attendre les paroles et sans tenir même  
aucun compte d'elles, dans mille signes extérieurs,  
même dans certains phénomènes invisibles.*

— **Marcel Proust**

Le Côté de Guermantes (1920),  
À la Recherche du Temps Perdu





## ACKNOWLEDGEMENTS

Tout d'abord, j'aimerais remercier mes directeurs de thèse, Stéphane et Idris, qui m'ont fait confiance pour mener à bien ce doctorat. En particulier, merci à Stéphane de m'avoir laissé une grande autonomie et de m'avoir toujours soutenue, malgré mes doutes et mon flux de paroles / de pensées parfois difficile à suivre.

Merci aux membres de la DGS vaudoise qui ont cru au projet GEOSAN, permis son financement et/ou assuré son suivi, notamment Pierre-Yves Maillard, Stéphanie Monod, Karim Boubaker, Stéphanie Pin, Marjorie Audard, Pierre-Olivier Barman et Pierre Stadelmann.

Je mesure ma chance d'avoir pu travailler sur une thèse aussi interdisciplinaire et appliquée, ce qui a été rendu possible grâce à la collaboration de personnes inspirantes et à l'écoute qui m'ont fait découvrir le monde de la santé et ses dimensions sociales, politiques, médicales, et logistiques. Je pense notamment à Oriana et Delphine pour l'étude « Commune en santé », à Gilbert, Séverine, Damien, Claire et Yangji pour les études sur la COVID-19, à Pedro pour l'étude sur les facteurs de risque cardiovasculaires, et à Louis-Henri Delarageaz, Jean Raymond, Jérôme Mouton et Sonia Arnal pour la planification de la campagne de vaccination itinérante.

Merci également aux membres de mon jury de thèse, pour leurs précieux commentaires et le temps consacré à évaluer ce travail.

L'enseignement ayant été un aspect que j'ai particulièrement apprécié durant ce doctorat, je remercie François Golay de m'avoir fait confiance et donné autant de liberté dans la gestion du cours de SIG.

Le doctorat comprend son lot de remises en question et de solitude. C'était donc une chance d'avoir pu travailler (et râler) en compagnie de personnes aussi drôles et soutenantes que mes collègues du LASIG / GEOME: Estelle, Solange, Thibaud, Elia, Oliver, Annie, Gabriel, Marco et Manon. Les géographes de la santé étant peu nombreux à l'EPFL, j'étais heureuse de pouvoir compter sur les GIRAPHes de Genève, David et Juan, pour l'échange de codes et d'idées.

Parmi les piliers de mon bien-être mental durant cette dernière année, merci à Caroline, Marcel et Tigrou.

Un très grand merci à ma famille qui m'a apporté son soutien sans faille malgré la distance, et notamment à mon frère Kévin, et à mes parents, Pascale et Denis, pour m'avoir offert la

## Acknowledgements

---

possibilité de venir faire mes études en Suisse. Je pense aussi à mes grands-parents, Marcelle, André et Berthe, qui auraient été fiers de cet accomplissement.

Enfin, merci à ma famille de coeur que j'ai pu me construire au fil des années, et qui m'apporte une force et une bienveillance constante. Pour n'en citer qu'une petite partie, merci à Estelle, Félix, Auré, Egor, PA, Jean, Mathieu, Ben, à la team SIE - Lou, Charlotte, Noé, Lucas, Josquin, Loïc, Julien - à mes gars sûrs Valentin et Romain, et enfin à François et à Lou.

*Lausanne, December 13, 2023*

Anaïs



## ABSTRACT

In a context of escalating public health challenges, including the rise of chronic diseases, the impact of climate change, and the COVID-19 pandemic, certain populations bear a disproportionate burden. As a result, there is an urgent need to develop public health strategies that not only promote overall well-being but also mitigate these health inequities. Addressing these issues through a geographic lens is essential because health status is strongly influenced by the social determinants of health, i.e., the conditions under which individuals are born, grow, live, work, and age. Such spatial epidemiology studies could facilitate the prioritization of public health interventions, and the design of initiatives tailored to the characteristics of populations and their environments. In recent decades, the research field of spatial epidemiology has been boosted by the increasing availability of high-resolution spatial data and advances in computational techniques. However, academic findings are rarely translated into population health interventions.

This thesis aims to bridge this gap by exploring the potential of spatial epidemiology in supporting public health policies. To this end, the research was structured around case studies aligned with the challenges faced by the Public Health Department of the canton of Vaud.

First, indicators related to social determinants of health were developed and mapped at a fine spatial scale (hectare level) to address the challenges of a national health promotion program engaged with municipalities. These indicators were then associated with individual health data to investigate the influence of the physical and social environments on the spatial distribution of cardiovascular risk factors. By identifying a pronounced geographic pattern of hypertension, obesity, and diabetes in the adult population of the city of Lausanne, the results provided insights for prioritizing and adapting future prevention campaigns.

The role of spatial epidemiology in infectious disease surveillance was then explored in the context of the COVID-19 pandemic. Spatio-temporal approaches were applied to individual RT-PCR test data to identify emerging clusters of COVID-19 cases. Subsequent genomic analysis of these clusters demonstrated that incorporating geographic approaches could improve the effectiveness of current surveillance systems by guiding prioritization strategies for contact tracing and virus tracking. In the final case study, spatial approaches were used to design the

## Abstract

---

COVID-19 mobile vaccination campaign in the canton of Vaud, illustrating the translation of research into practice. This thesis demonstrates that fine-scale spatial epidemiology can inform strategic decision-making for various health challenges, and concludes with practical recommendations for adopting a geographic lens within public health departments.

**KEYWORDS:** spatial epidemiology, GIS, spatial statistics, public health, geographic information, health promotion, COVID-19, cardiovascular diseases, health-environment associations, applied research, population health, informed decision-making



## RÉSUMÉ

Face à l'augmentation des maladies chroniques, aux répercussions du changement climatique, et plus récemment, à la pandémie de COVID-19, certaines populations subissent un fardeau disproportionné. Il est donc urgent de développer des stratégies de santé publique qui viseront non seulement à favoriser le bien-être général de la population mais aussi à réduire les inégalités face à la santé. Aborder ces enjeux avec une perspective géographique s'avère essentiel étant donné que l'état de santé des populations est fortement influencé par les déterminants sociaux de la santé, c'est-à-dire les circonstances dans lesquelles les individus naissent, grandissent, vivent, travaillent, et vieillissent. De telles études en épidémiologie spatiale pourraient faciliter la priorisation des actions de santé publique, et le développement d'initiatives adaptées aux caractéristiques des populations et de leur environnement. Bien que le domaine de l'épidémiologie spatiale ait connu un essor considérable ces dernières décennies, propulsé par l'accès à des données spatiales haute résolution et par des avancées computationnelles, la conversion de ces découvertes académiques en actions concrètes de santé publique demeure limitée.

Cette thèse vise à combler ce déficit en explorant le potentiel de l'épidémiologie spatiale dans la conception des politiques de santé publique. À cette fin, la recherche s'est articulée autour d'études de cas alignées sur les défis rencontrés par Direction Générale de la Santé du canton de Vaud.

Premièrement, des indicateurs liés aux déterminants sociaux de la santé ont été développés et cartographiés à une échelle spatiale fine (niveau de l'hectare) pour répondre aux besoins d'un programme national de promotion de la santé engagé avec les municipalités. Ces indicateurs ont ensuite été associés à des données de santé individuelles pour étudier l'influence des environnements physiques et sociaux sur la distribution spatiale des facteurs de risque cardiovasculaires. En mettant en évidence une structure géographique marquée de l'hypertension, de l'obésité et du diabète au sein de la population adulte de la ville de Lausanne, les résultats de cette étude fournissent des orientations stratégiques pour prioriser et adapter les futures campagnes de prévention.

Le rôle de l'épidémiologie spatiale dans la surveillance des maladies infectieuses a ensuite été exploré dans le contexte de la pandémie de COVID-19. Des méthodes spatio-temporelles ont

## Résumé

---

été employées sur des données issues de tests RT-PCR individuels afin de détecter les foyers émergents de cas de COVID-19. Une analyse génomique de ces clusters a ensuite permis de montrer que l'intégration d'approches géographiques dans les systèmes de surveillance en vigueur pouvait optimiser les stratégies de priorisation pour la traçabilité des cas contacts et la surveillance des agents pathogènes. Dans la dernière étude de cas, des approches spatiales ont été utilisées pour planifier la campagne de vaccination mobile contre la COVID-19 dans le canton de Vaud, illustrant ainsi la conversion immédiate des résultats académiques en mesures de santé publique. Cette thèse démontre que l'épidémiologie spatiale à fine échelle a le potentiel de guider la prise de décision stratégique pour divers enjeux de santé, et conclut avec des recommandations pratiques pour l'adoption d'une perspective géographique au sein des départements de santé publique.

**MOTS-CLÉS :** épidémiologie spatiale, SIG, statistiques spatiales, santé publique, information géographique, promotion de la santé, COVID-19, maladies cardiovasculaires, associations santé-environnement, recherche appliquée, approche populationnelle, prise de décision éclairée



# CONTENTS

<b>Acknowledgements</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>List of figures</b>	<b>xi</b>
<b>List of tables</b>	<b>xv</b>
<b>List of abbreviations</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Health and place . . . . .	1
1.2 State of the art: spatial epidemiology approaches and challenges . . . . .	4
1.2.1 Data in spatial epidemiology . . . . .	5
1.2.2 Investigating spatial patterns . . . . .	6
1.2.3 Characterizing spatial clusters . . . . .	8
1.2.4 Deprivation indices and geodemographic classification . . . . .	9
1.3 Context: Swiss health system and situation in the canton of Vaud . . . . .	10
1.4 Thesis aim . . . . .	13
1.5 Methodology . . . . .	14
1.5.1 Collaboration with public health authorities . . . . .	14
1.5.2 Methodological choices . . . . .	15
1.6 Chapters outline . . . . .	17
<b>2 Supporting health promotion initiatives with geographic information systems</b>	<b>21</b>
2.1 Introduction . . . . .	21
2.2 Geodemographic classification of municipalities . . . . .	22
2.3 Mapping social determinants of health . . . . .	26
2.3.1 Hectare-level indicators . . . . .	26
2.3.2 Composite indicators . . . . .	29
2.4 Web application . . . . .	32
	vii

## Contents

---

2.5	Discussion . . . . .	33
<b>3</b>	<b>Spatial patterns and factors associated with cardiovascular risk factors</b>	<b>37</b>
3.1	Background . . . . .	38
3.2	Materials and methods . . . . .	39
3.2.1	Data . . . . .	39
3.2.2	Methods . . . . .	41
3.3	Results . . . . .	44
3.3.1	Intra-urban variation in cardiometabolic risk . . . . .	44
3.3.2	Impact of individual confounders on cardiometabolic risk factors . . . .	47
3.3.3	Associations with the physical and social environments . . . . .	47
3.4	Discussion . . . . .	53
3.4.1	Intra-urban variation in cardiometabolic risk factors . . . . .	53
3.4.2	Associations with the physical and social environments . . . . .	53
3.4.3	Implications for the development of public health policies . . . . .	55
3.4.4	Strengths and limitations . . . . .	56
3.5	Conclusion . . . . .	57
3.6	Feasibility of extending NCD spatial risk prediction to the canton level . . . .	58
<b>4</b>	<b>Detecting daily spatio-temporal COVID-19 case clusters</b>	<b>61</b>
4.1	Introduction . . . . .	62
4.2	Materials and methods . . . . .	65
4.2.1	Patients . . . . .	65
4.2.2	SARS-COV-2 RT-PCR . . . . .	65
4.2.3	Study area . . . . .	65
4.2.4	Spatio-temporal clusters . . . . .	66
4.2.5	Cluster evolution and diffusion zones . . . . .	66
4.3	Results . . . . .	67
4.3.1	Epidemic trajectories of positive cases . . . . .	67
4.3.2	Cluster detection and temporal dynamics . . . . .	70
4.3.3	Cluster composition . . . . .	70
4.3.4	Viral load in clusters . . . . .	70
4.3.5	Cluster size, duration and viral load . . . . .	72
4.3.6	Geographic distribution of the first epidemic wave . . . . .	73
4.4	Discussion . . . . .	76
4.4.1	New information on COVID-19 clusters . . . . .	76
4.4.2	Limitations . . . . .	79
4.4.3	Added value of the methods used . . . . .	80
4.5	Conclusion . . . . .	82

<b>5 Combining spatio-temporal and genomic analyses to improve COVID-19 surveillance system</b>	<b>83</b>
5.1 Introduction . . . . .	84
5.2 Materials and methods . . . . .	85
5.2.1 Study design . . . . .	85
5.2.2 SARS-CoV-2 genome sequencing . . . . .	86
5.2.3 Reads processing and quality control . . . . .	86
5.2.4 Genomic analyses . . . . .	88
5.2.5 Genomic and geographic visualization . . . . .	88
5.3 Results . . . . .	89
5.3.1 Description of selected spatiotemporal clusters . . . . .	89
5.3.2 Genetic similarity within spatiotemporal clusters . . . . .	89
5.3.3 Comparison of spatiotemporal clusters and genomic groups . . . . .	91
5.4 Discussion . . . . .	94
5.4.1 Congruence between two approaches in different contexts . . . . .	94
5.4.2 Combining genomic and spatiotemporal clustering approaches in infectious disease surveillance . . . . .	96
5.5 Conclusion . . . . .	97
<b>6 Prioritization of populations for mobile COVID-19 vaccination campaigns</b>	<b>99</b>
6.1 Introduction . . . . .	99
6.2 Targeting vulnerable or deprived populations (phase 1) . . . . .	101
6.3 Targeting areas with low vaccination rate (phase 2) . . . . .	105
6.4 Discussion . . . . .	107
<b>7 Synthesis and perspectives</b>	<b>111</b>
7.1 Summary of the contributions of spatial epidemiology to public health policy design . . . . .	111
7.2 Integration of geographic services within public health departments . . . . .	116
7.2.1 Expanding fine-scale health data acquisition . . . . .	117
7.2.2 Training and guidance for health practitioners . . . . .	118
7.2.3 Operationalization and automation of spatial epidemiology pipelines . . . . .	119
7.3 Maintaining the link with academic research . . . . .	119
7.4 General conclusion . . . . .	120
<b>Bibliography</b>	<b>123</b>
<b>Appendix</b>	<b>147</b>
A.1 Supplementary materials for Chapter 1 . . . . .	148
A.2 Supplementary materials for Chapter 2 . . . . .	151

## Contents

---

A.3	Supplementary materials for Chapter 3 . . . . .	157
A.4	Supplementary materials for Chapter 4 . . . . .	176
A.5	Supplementary materials for Chapter 5 . . . . .	181
A.6	Manuscript - Geographic footprints of life expectancy inequalities in the state of Geneva, Switzerland . . . . .	
A.7	Code availability . . . . .	

## Curriculum Vitae

## LIST OF FIGURES

1.1	Main determinants of health . . . . .	2
1.2	Pathways of residential environment contributions to health inequalities . . . .	3
1.3	Snow's map of cholera deaths (1855) . . . . .	4
1.4	Geographic context of case studies: Canton of Vaud, Switzerland . . . . .	11
1.5	Methodological approach . . . . .	16
2.1	Typology of municipalities according to 9 distinct profiles . . . . .	25
2.2	Composite indicators of vulnerability for the municipality of Aubonne . . . . .	32
2.3	Interface of the GEOSAN web application . . . . .	34
3.1	Log relative risk surface of hypertension, obesity, diabetes, and dyslipidemia . .	45
3.2	Local contextual determinants of hypertension . . . . .	49
3.3	Local contextual determinants of obesity . . . . .	51
3.4	Local contextual determinants of diabetes . . . . .	52
4.1	Evolution of cases and clusters through time . . . . .	68
4.2	Case cluster characteristics over time . . . . .	69
4.3	Number of significant ( $p \leq 0.05$ ) case clusters over time characterized according to the viral load of the cases documented in each cluster . . . . .	69
4.4	Characteristics (size, age and duration) of significant space-time COVID-19 clusters ( $p \leq 0.05$ ) over the study period, categorized according to the viral load of the cases involved. . . . .	71
4.5	Number of cases observed within significant space-time clusters ( $p \leq 0.05$ ) in function of the mean and maximal viral loads of the first three cases involved . .	72
4.6	Spatial distribution of case clusters and diffusion dynamics of transmission clusters for 6 key dates (March 11, March 15, March 19, March 24, March 27, April 4) during the first epidemic wave . . . . .	74
4.7	Diffusion zones identified by the MST-DBSCAN algorithm. . . . .	77

## List of Figures

---

5.1	Spatial distribution and characteristics of the 17 spatiotemporal clusters considered for genomic data analysis . . . . .	87
5.2	Genetic distance between and within spatiotemporal clusters . . . . .	90
5.3	Comparison between spatiotemporal clusters and genomic groups in phylogenetic trees . . . . .	92
5.4	Genetic relationship among genomic groups and singletons in minimum spanning tree . . . . .	93
5.5	Distribution of genomic groups within clusters . . . . .	93
5.6	Graphical representation of findings and conclusion. . . . .	97
6.1	Timeline of the vaccination campaign in the canton of Vaud . . . . .	101
6.2	Vaccination settings employed during the Vaud mobile campaign . . . . .	102
6.3	Optimal locations for the Phase 1 mobile vaccination campaign . . . . .	104
6.4	Spatial clusters of low and high vaccination rates identified using Getis-Ord Gi* statistics . . . . .	106
6.5	Optimal locations for the Phase 2 mobile vaccination campaign in Yverdon-les-Bains. . . . .	109
A.1.1	Population distribution in the canton of Vaud, Switzerland . . . . .	148
A.1.2	Flowchart of the geocoding algorithm developed to retrieve coordinates of individuals' home addresses . . . . .	150
A.2.1	Status of the "Commune en santé" label in the Canton of Vaud as of October 5, 2021. . . . .	151
A.2.2	Architecture of the GEOSAN web application . . . . .	152
A.2.3	Program of the workshop with public health practioners held on January 16, 2023 (in French). . . . .	156
A.3.1	Spatial variation of obesity risk using a Gaussian Kernel with incremental bandwidths . . . . .	157
A.3.2	Flow charts for the sample selection of the baseline (2003-2006), follow-up 1 (2009-2012), and follow-up 2 (2014-2017) participants . . . . .	159
A.3.3	Spatial distribution of contextual factors across the study area (Lausanne, Switzerland). . . . .	160
A.3.4	Delimitation of areas presenting a significant high-risk and low-risk for hypertension, obesity, diabetes, and dyslipidemia . . . . .	161
A.3.5	Comparisons of contextual factors among areas categorized as high-risk, low-risk and neutral risk for hypertension. . . . .	162
A.3.6	Comparisons of contextual factors among areas categorized as high-risk, low-risk and neutral risk for obesity. . . . .	163
A.3.7	Comparisons of contextual factors among areas categorized as high-risk, low-risk and neutral risk for diabetes . . . . .	164

A.3.8	Comparisons of contextual factors among areas categorized as high-risk, low-risk and neutral risk for dyslipidemia . . . . .	165
A.3.9	Spatial variation of hypertension risk during baseline (2003-2006), follow-up 1 (2009-2012), and follow-up 2 (2014-2017) periods . . . . .	166
A.3.10	Spatial variation of obesity risk during baseline (2003-2006), follow-up 1 (2009-2012), and follow-up 2 (2014-2017) periods . . . . .	166
A.3.11	Spatial variation of diabetes risk during baseline (2003-2006), follow-up 1 (2009-2012), and follow-up 2 (2014-2017) periods . . . . .	167
A.3.12	Spatial variation of dyslipidemia risk during baseline (2003-2006), follow-up 1 (2009-2012), and follow-up 2 (2014-2017) periods . . . . .	167
A.3.13	GWR results for adjusted hypertension . . . . .	169
A.3.14	GWR results for adjusted obesity . . . . .	170
A.3.15	GWR results for adjusted diabetes . . . . .	171
A.3.16	Maps of local condition number for GWR and MGWR models . . . . .	172
A.3.17	Spatial variation of hypertension risk among middle-aged and older adults in Lausanne. . . . .	173
A.3.18	Spatial variation of obesity risk among middle-aged and older adults in Lausanne	174
A.3.19	Spatial variation of diabetes risk among middle-aged and older adults in Lausanne	175
A.4.1	Definition of time windows with SaTScan and MST-DBSCAN . . . . .	176
A.4.2	Distribution of viral loads for individuals within significant clusters . . . . .	178
A.4.3	Distribution of viral loads for individuals outside significant clusters . . . . .	179
A.4.4	Boxplots of viral load per age class of individuals tested positive . . . . .	180
A.5.1	Selection of the 17 spatiotemporal clusters considered for genomic data analysis	182
A.5.2	Construction of minimum spanning tree from a network of SARS-CoV-2 genomes with their pairwise SNV distance . . . . .	183
A.5.3	Jaccard similarity index of 9 spatiotemporal clusters within the Lausanne region .	184
A.5.4	Spatial distribution of genomic groups within spatiotemporal clusters . . . . .	185





## LIST OF TABLES

2.1	Total number of measures adopted and distribution of the percentage across the 6 domains for each labeled municipality . . . . .	25
3.1	Logistic regression results for each CMRF . . . . .	47
4.1	Classification of the space-time clusters according to the viral load of the cases involved . . . . .	71
6.1	Potential barriers to COVID-19 vaccination . . . . .	100
6.2	Indicators and associated parameters used to construct the composite indicator of area-level priority. . . . .	103
A.1.1	List of public health practitioners interviewed with their respective fields of expertise and institutional affiliations . . . . .	149
A.2.1	Description of the 39 variables used in the typology of municipalities . . . . .	153
A.2.2	Description of SDOH indicators used to characterize inhabited hectares . . . . .	154
A.2.3	Description of SDOH indicators used to characterize inhabited hectares . . . . .	155
A.3.1	Comparative statistics of baseline, study sample, and excluded population . . . . .	158
A.3.2	Frequency distribution of outcome variables for the baseline, follow-up, and follow-up 2 samples. . . . .	159
A.3.3	Results of the ordinary least squares (OLS) regression model . . . . .	168
A.3.4	Model fit metrics for ordinary least squares, geographically weighted regression, and multiscale geographically weighted regression models . . . . .	168
A.3.5	Comparisons of the bandwidths used in the GWR and MGWR models. . . . .	173
A.4.1	Mean viral load of the first three cases within clusters . . . . .	179
A.4.2	Maximal viral load of the first three cases within clusters . . . . .	180
A.5.1	Characteristics of the individuals for whom we analyzed the SARS-CoV-2 genomic sequence. . . . .	181
A.5.2	Pangolin lineage and nucleotide sequence changes of 20 genomic groups.. . . .	186

## List of Tables

---

A.5.3 Advantages and disadvantages of spatiotemporal clustering and genomics in surveillance. . . . .	187
---	-----



## LIST OF ABBREVIATIONS

**AIC:** Akaike Information Criterion  
**CMRF:** Cardiometabolic Risk Factor  
**CVD:** Cardiovascular Disease  
**FOSS:** Free and Open-Source Software  
**GIS:** Geographic Information Systems  
**GWR:** Geographically Weighted Regression  
**KDE:** Kernel Density Estimation  
**LST:** Land Surface Temperature  
**MGWR:** Multiscale Geographically Weighted Regression  
**MST-DBSCAN:** Modified Space-Time DBSCAN  
**NCD:** Noncommunicable Disease  
**OLS:** Ordinary Least Squares  
**PCA:** Principal Component Analysis  
**SDOH:** Social Determinants of Health  
**SES:** Socioeconomic Status  
**SFSO:** Swiss Federal Statistical Office  
**SNC:** Swiss National Cohort  
**TTIQ:** Test-trace-isolate-quarantine  
**VIF:** Variance Inflation Factor



# 1 INTRODUCTION

*Vivre sainement est une question de volonté,  
affirmait un intervenant lors des débats du Conseil  
des Etats. La volonté peut certes beaucoup, mais  
elle doit également bénéficier de bonnes  
conditions pour pouvoir se développer.*

— **Thomas Mattig**

regarding the rejection of the  
Federal Law on Prevention and Health Promotion (LPrév),  
Le Temps newspaper (2012)

## 1.1 Health and place

The study of the relationship between geography and health can be traced back to ancient civilizations. Both Ayurvedic physicians in India during the 1st and 2nd millennia BC and Hippocrates in his treatise "Air, Water and Places," ~2,400 years ago, emphasized the importance of environmental factors in health (Picheral, 1982). In contemporary public health, the importance of "place" is integral to the Social Determinants of Health (SDOH), which define "the conditions in which people are born, grow, live, work, and age" that influence health status (M. Marmot, J. Allen, et al., 2012). The Dahlgren-Whitehead rainbow model, developed in 1991 and still widely used today, provides a hierarchical representation of these determinants and suggests different strata for targeted health interventions (Dahlgren and

M. Whitehead, 2021). As illustrated in Figure 1.1, the model asserts that health status is influenced by an interplay of factors. These range from non-modifiable determinants (e.g., age, gender, genetic predisposition) to modifiable aspects including lifestyle behaviors (e.g., smoking, hazardous drinking, diet, and physical activity), social and community networks (e.g., family, social cohesion), living and working conditions (e.g., housing, access to health services, education), and broader socioeconomic, cultural, and environmental conditions (e.g., taxation, air quality).

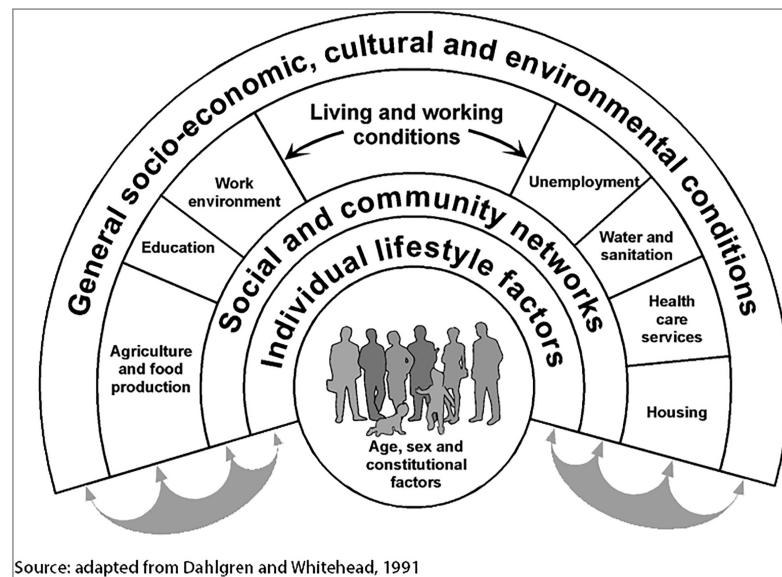


Figure 1.1 – **Main determinants of health.** Reprinted from Dahlgren and M. Whitehead (2021), with permission from Elsevier.

The model implicitly carries two important concepts:

*Population health.* Because some SDOH are common to groups of individuals (e.g., those living in the same geographic area), it is important to examine the distribution of risk factors across populations, rather than focusing solely on inter-individual differences. This population health approach was popularized by the work of Geoffrey Rose in the 1980s (G. A. Rose, Khaw, and M. G. Marmot, 2008). From a policy perspective, interventions that target all individuals in a population (i.e., population-level interventions), for example by creating health-promoting environments, have the potential to be more effective and less costly than individual approaches that focus on risk factor modification (Syme, 1986; Diez Roux, 2003).

*Health inequities.* Differences in SDOH can lead to health inequities, which M. Whitehead (1992) defines as unnecessary, unfair, and avoidable inequalities. These inequities often manifest along a social gradient, first enunciated by M. Marmot, Stansfeld, et al. (1991), indicating that individuals with lower socioeconomic status (SES) experience worse health

outcomes. Furthermore, these inequities are exacerbated by factors such as the degradation of the physical environment, including air pollution, noise, and other environmental stressors (Pearce et al., 2010). Health inequities exist globally and can be observed at different spatial scales in low-, middle- and high-income countries (M. Marmot, J. Allen, et al., 2012). Even in countries with high income and long life expectancy, such as Switzerland, health inequities persist and are increasing over time (Spiess and Schnyder-Walser, 2018). This underscores the need for urgent policy efforts to improve the health of the population while addressing and flattening health gradients.

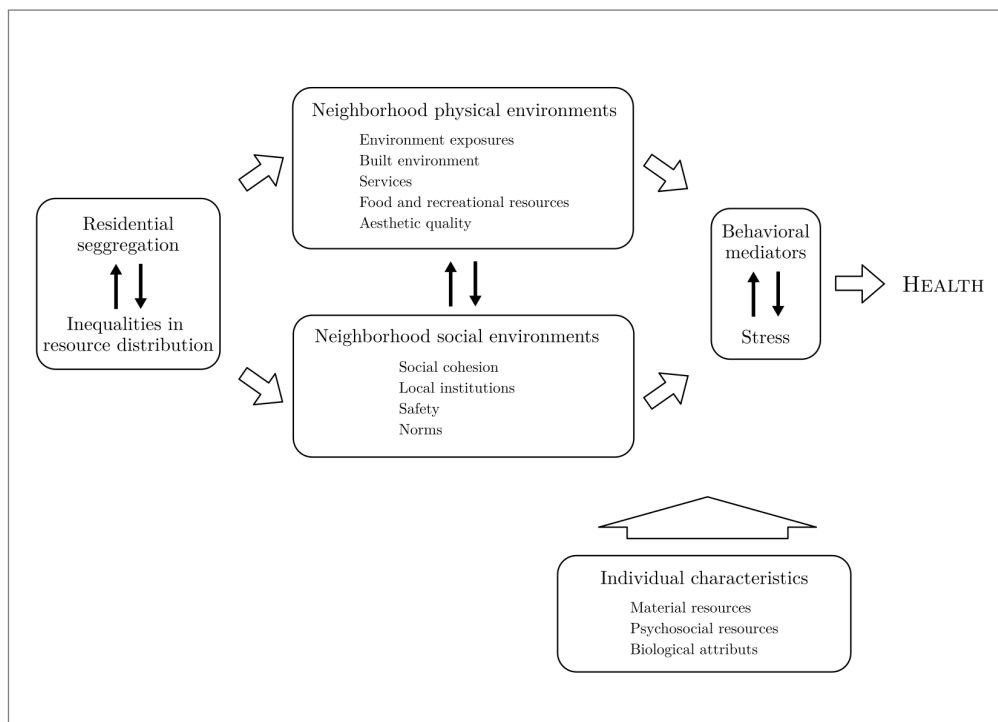


Figure 1.2 – **Pathways of residential environment contributions to health inequalities.** The figure is adapted from Diez Roux and Mair (2010).

Given the critical role of SDOH in shaping population health outcomes and the increasing recognition of health equity as a global priority, the geographic lens becomes paramount. In particular, a focus on residential environments emerges as highly relevant to addressing these challenges, as they possess both physical and social characteristics that can significantly affect the health of individuals and contribute to health inequities through the pathways illustrated in Figure 1.2 (Diez Roux and Mair, 2010). In addition, the neighborhood scale is policy relevant for the development of initiatives beyond the medical domain, such as the creation of health-promoting environments through urban planning (Lowe, Whitzman, and Giles-Corti, 2018).

The field of *spatial epidemiology* (also referred to as *geographical epidemiology*, or *health geography*) arises from this natural overlap between geography, which is concerned with spatial patterns and processes, and epidemiology, which focuses on the distribution of disease and health determinants (Mayer, 1983). By analyzing the spatial patterns and variations of health outcomes in relation to demographic, socioeconomic, and environmental factors, spatial epidemiology is playing a critical role in identifying and understanding health inequities, thereby facilitating the design of appropriate population-level interventions.

### 1.2 State of the art: spatial epidemiology approaches and challenges

Jon Snow's work on cholera (Snow, 1855) is considered one of the first applications of spatial epidemiology. Despite facing skepticism, Snow suspected that contaminated drinking water was the cause of the cholera outbreak in London. To test his hypothesis, he conducted door-to-door investigations of cholera deaths in the Golden Square area, a severely affected neighborhood of South London. Through these observations, Snow discovered a cluster of cases around the Broad Street water pump, which ultimately led to the closure of the contaminated pump, effectively curbing the epidemic. Figure 1.3 shows one of the famous maps Snow created to illustrate his findings, where the black rectangles correspond to cholera deaths, and where the dotted line encircles the area that was closer to the Broad Street pump, within walking distance, than any other street pump (Brody et al., 2000).

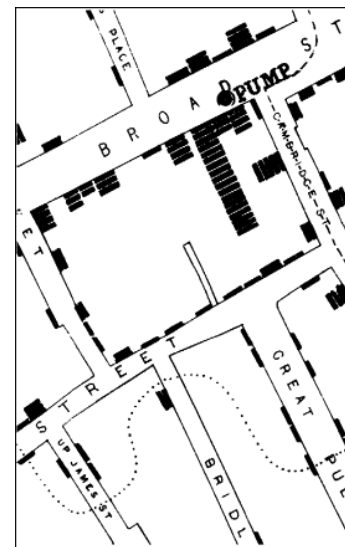


Figure 1.3 – **Snow's map of cholera deaths (1855).** Reprinted from Brody et al. (2000), with permission of Elsevier.

Since then, the development of spatial epidemiology has been somewhat discontinuous (Waller and Gotway, 2004). However, the increased availability of spatial data (also referred to as *geodata*) combined with the advances in computational technologies have fostered its growth since the 1990s. In particular, the improved access to geographic information systems (GIS) provided a powerful way to visualize, store, and explore spatial relationships, thereby effectively connecting individuals to their living environment (Nykiforuk and Flaman, 2011; Auchincloss et al., 2012).

The following sections aim to provide an overview of the characteristics of spatial data and the analytical methods used to incorporate the notion of place in health-related studies. The study of geographic variation in disease applies to both infectious diseases, in which a pathogen is



## 1.2. State of the art: spatial epidemiology approaches and challenges

---

transmitted from one infected individual to another, and noncommunicable diseases (NCDs), which arise from an interplay of biological, environmental, and lifestyle factors (Waller, 2021).

It is important to note that most of the approaches discussed below were not necessarily developed for public health, and applications can be found in a wide range of fields, including social sciences, environmental sciences, econometrics, or criminology.

### 1.2.1 Data in spatial epidemiology

Research in spatial epidemiology relies on the use of spatial data, which are characterized not only by a value (i.e., attribute) but also by a geographic location (i.e., a pair of geographic coordinates  $x,y$ ) (Rezaeian et al., 2007). Spatial data can pinpoint precise locations, such as street addresses (i.e., point data), or be associated with broader regions, such as postcodes or municipalities (i.e., areal data).

Two inherent properties of spatial data make the use of standard statistical methods complicated. The first, spatial dependence, refers to the tendency for spatial data to exhibit positive autocorrelation (Fischer and Getis, 2010), a concept captured by Tobler's First Law of Geography: "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). The second, spatial heterogeneity, implies that the relationships between variables vary across space, making it difficult to establish universal principles applicable to the entire Earth's surface (Anselin and S. J. Rey, 2010).

The granularity of spatial data introduces additional methodological considerations. In many studies, health data, such as mortality or disease incidence rates, are obtained from government databases and aggregated to a predefined administrative level to comply with confidentiality policies (Kirby, Delmelle, and Eberth, 2017). However, these administrative boundaries may not correspond to the natural scale of the phenomena being studied. Different levels of aggregation can thus lead to contrasting interpretations, a challenge known as the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1984).

In addition, the use of aggregated data to examine the relationship between health outcomes and contextual factors may expose researchers to the ecological fallacy, whereby relationships observed at the group level do not necessarily apply to individuals (Fotheringham, Brunson, and Charlton, 2002). Small-area studies serve as a valuable tool to mitigate ecological bias, but they also present their own unique challenges. This is particularly apparent when working with rare diseases, where the statistical instability introduced by the small-number problem can compromise the reliability of calculated incidence or prevalence rates (Waller and Gotway, 2004).

As investigations move to finer spatial scales, the accuracy of geocoding becomes increasingly important. Geocoding is the process of converting a given address into geographic coordinates,

typically achieved by matching it against external databases or through online geocoding services. In fine-scale studies, even slight positional errors can skew the results and potentially lead to misleading conclusions for decision makers (Kirby, Delmelle, and Eberth, 2017).

### 1.2.2 Investigating spatial patterns

Spatial dependence analysis, an integral aspect of spatial epidemiology, offers methods (1) to assess whether disease risk is randomly distributed across space and, conversely, (2) to identify high-risk areas that require immediate action (spatial clusters).

We generally distinguish between methods that evaluate global clustering, returning a unique value that summarizes the overall spatial pattern, and methods that evaluate local clustering, allowing the detection of specific spatial clusters. Knox (1989) defines a spatial cluster as "a geographically bounded group of occurrences of sufficient size and concentration to be unlikely to have occurred by chance".

Global clustering is assessed using spatial autocorrelation statistics, most notably Moran's I (Moran, 1950). These statistics capture the spatial dependence of data by highlighting how closely similar observations cluster. Positive spatial autocorrelation indicates high values near high values or low values near low values, while negative autocorrelation indicates high values near low values and vice versa (Fotheringham and Brunsdon, 1999). The degree of spatial autocorrelation is determined by comparing the matrix  $S$ , which represents the correlation between pairs of observations, with the matrix  $W$ , which represents their spatial proximity. The construction of the matrix  $W$ , also known as the spatial weights matrix, can take several forms, such as contiguity (a weight of 1 for contiguous observations, 0 otherwise), nearest neighbor (a weight of 1 if an observation is among the  $k$  nearest neighbors, 0 otherwise), or distance decay (decreasing weight with increasing distance). A high correspondence between the  $S$  and  $W$  matrices indicates strong spatial autocorrelation (Getis, 2008). To determine whether the observed spatial clustering is due to chance, a significance test uses Monte Carlo permutations to randomly reassign observations among spatial units.

In public health, the priority often lies in identifying spatial clusters of disease rather than summarizing the overall spatial pattern, thus, our primary focus will be on local clustering approaches. There are numerous analytical methods used to detect spatial clusters, and the choice of one over another will depend on various factors, such as the type of spatial data available (point or areal), the nature of the outcome (binary or continuous), or the inclusion of a temporal component.

Methods such as Getis-Ord  $G_i^*$  or  $G_i^*$ , developed by Getis and Ord (1992), or Local Moran's  $I$ , developed by Anselin (1995), are local versions of global spatial autocorrelation statistics. They classify each spatial unit into different types of spatial dependence. For example, Getis-

## 1.2. State of the art: spatial epidemiology approaches and challenges

---

Ord statistics identify "hot spots" and "cold spots" corresponding to clusters of high and low values, respectively. On the other hand, Local Moran statistics identify locations with high values surrounded by high values (High-High), locations with low values surrounded by low values (Low-Low), and spatial outliers, which are locations that exhibit negative spatial autocorrelation (High-Low or Low-High). Local autocorrelation statistics have primarily been used with aggregated health data (Hollands et al., 2013; R. Huang et al., 2015; Maroko, Nash, and Pavilonis, 2020). However, in the last decade, a growing number of studies have used these statistics with individual-level data to investigate the geographic patterns of Parkinson's disease (Fleury et al., 2021), life expectancy inequalities (Ladoy, Vallarta-Robledo, et al., 2021), Body Mass Index (BMI) (Joost, Duruz, et al., 2016), daytime sleepiness (Joost, Haba-Rubio, et al., 2018), or tobacco consumption (Vallarta-Robledo et al., 2022).

On the other hand, scan statistics detect spatial clusters by using a moving window over the study area and comparing the disease risk inside and outside the window. The method developed by Kulldorff (1997), widely used in the literature, calculates likelihood ratios for sets of concentric circles centered around each spatial unit. The algorithm returns the most unusual clusters with excess relative risk, along with their associated p-value. A notable advantage of this method is its independence from a predefined spatial weight matrix, since it evaluates different windows at each location. It is also compatible with both point and areal data and is well suited to varying population densities. The spatio-temporal extension, called space-time scan analysis (Kulldorff et al., 1998), serves as a tool for prospective cluster detection, making it suitable for public health monitoring. Space and space-time scan statistics have been used to investigate spatial patterns of lung cancer in Kentucky, United States (Christian et al., 2011) and childhood leukemia in Switzerland (Konstantinoudis et al., 2017). In the context of infectious diseases, they have been used to prioritize shigellosis case investigations in Chicago (R. C. Jones et al., 2006) and to assist decision makers during COVID-19 in resource allocation and case investigations (Desjardins, Hohl, and Delmelle, 2020; Greene et al., 2021).

Lastly, approaches derived from spatial point pattern theory are well suited to case-control datasets, which consist of two sets of points: the location of disease cases and the location of controls, representing the at-risk population. Central to this method is the calculation of the intensity function, i.e., a smoothed surface representing the expected density of events across the study area, generated using kernel density estimation (KDE) (Baddeley, Rubak, and R. Turner, 2016). Spatial variation in disease risk is estimated by comparing the intensity functions of cases and controls, and spatial clusters are identified by statistically significant local differences (Kelsall and Diggle, 1995b). Although less common in the literature due to the scarcity of case-control datasets, this approach has proven effective in several contexts, including a waterborne outbreak in Belgium (Braeye et al., 2015), the spatial pattern of autism spectrum disorders (Bakian et al., 2015), and the spatio-temporal distribution of COVID-19 cases in Berlin, Germany (Lambio et al., 2023).

### 1.2.3 Characterizing spatial clusters

Spatial clusters of health outcomes or determinants may indicate underlying etiological factors in the biological, physical, or social environments, or it may imply that individuals within these clusters share significant characteristics related to disease (MacMahon, Pugh, et al., 1970). Thus, investigation of health-environment associations is essential to identify underlying causes of observed patterns and to tailor population-level interventions.

The interpretation of such associations may be somewhat different for infectious and non-communicable diseases. For infectious diseases, identifying environmental factors can reveal sources of disease, such as the water pump in Snow's study (Snow, 1855), or elucidate factors that influence spread, such as mobility patterns, thereby helping to break transmission chains. Conversely, for NCDs, such analyses provide insights into potential drivers of health inequities.

Methods for explaining spatial clustering range from simple comparative statistics between high-risk, low-risk, and neutral areas to more statistical modeling approaches. While comparative statistics are helpful in suggesting potential drivers behind spatial patterns and guiding further investigation, regression analyses facilitate modeling, significance assessment, and prediction of health-environment relationships (MacMahon, Pugh, et al., 1970). The numerous existing approaches vary in computational complexity, ability to capture spatial heterogeneity, and to estimate parameter uncertainty (Auchincloss et al., 2012). Such approaches encompass multilevel models, spatial autoregressive models, and geographically weighted regressions.

Since the 1990s, the use of multilevel models has become a standard approach to study the effects of place on health. Operating within a hierarchical framework, these models consider individuals (micro-level) nested within neighborhoods (macro-level) as the unit of analysis (Diez-Roux, 2000). By simultaneously including individual and neighborhood-level predictors in regression models, these models can assess the contextual effect while accounting for individual confounders. They also examine spatial heterogeneity across neighborhoods and spatial homogeneity within neighborhoods (Owen, Harris, and K. Jones, 2016). For example, Diez-Roux, Link, and Northridge (2000) found an increased risk of BMI, hypertension, and sedentary behavior in lower-income neighborhoods within the United States, even after adjusting for individual-level income. Multilevel models have been criticized for the way they incorporate spatial information, treating neighborhoods as independent of each other and overstating the statistical significance of contextual variables in the presence of strong spatial dependence (Chaix, 2005).

On the other hand, simultaneous (SAR) or conditional autoregressive (CAR) models directly incorporate the notion of spatial dependence within the regression equation through a spatial weights matrix. These models extend the framework of ordinary least squares (OLS) to account for the spatial autocorrelation in residuals that often arises from spatial dependence. Such

## 1.2. State of the art: spatial epidemiology approaches and challenges

---

autocorrelation violates OLS assumptions and can bias parameter estimation and statistical inference (Yang, Shoff, and Noah, 2014). This phenomenon was illustrated by Chakraborty (2011) in his study of the effect of SES and ethnicity on the spatial distribution of cancer risk from traffic emissions in Tampa, Florida. These results highlighted how non-spatial regressions could misleadingly amplify the perceived influence of race and ethnicity on cancer risk.

While autoregressive models incorporate spatial dependence during the modeling phase, they assume consistent relationships between outcome and explanatory variables across the study area. This assumption can lead to biased estimates when confronted with spatial heterogeneity. In contrast, Geographically Weighted Regressions (GWR) (Fotheringham, Brunsdon, and Charlton, 2002) recognize and adjust for such heterogeneity by formulating an individual regression equation for each spatial unit. Within these equations, nearby features receive a higher weight, which is achieved using a spatial weight matrix (Fischer and Getis, 2010). The resulting coefficient estimates reveal the influencing factors relevant to each spatial unit (Auchincloss et al., 2012). For example, Edwards et al. (2010) showed that contextual factors associated with childhood obesity varied across neighborhoods in Leeds, UK. Their analysis allowed to derive specific policy implications tailored to each area, based on the covariates with the strongest associations with childhood obesity (e.g., access to supermarkets and leisure activities, availability of healthy options).

### 1.2.4 Deprivation indices and geodemographic classification

Having discussed the methods used to study and characterize spatial patterns of health, it is equally important to understand the spatial distribution of the underlying contextual determinants that influence these patterns. The construction of composite indicators derived from demographic, socioeconomic or environmental data is a common way of synthesizing some of the key determinants of health. Given the often limited availability of fine-scale health data, such indicators are also emerging as valuable proxies for assessing health inequities (Lalloué et al., 2013). While these proxy methods are not directly applicable to etiologic studies, they provide a critical resource for targeting preventive health interventions to areas populated by socially and economically deprived individuals (Rushton, 2003).

The concept of deprivation, first formulated by Townsend (1987), refers to "a state of observable and demonstrable disadvantage relative to the local community or the wider society or nation to which individual, family, or groups belong". The use of deprivation indices in health inequity research is justified by the well-established social gradient in health, where regions with lower income levels show higher mortality and morbidity rates compared to wealthier counterparts (Curtis, 2004; G. Rey et al., 2009). More recently, evidence has emerged of a similar gradient between SES and exposure to various environmental exposures such as air pollution, heat or noise (Pearce et al., 2010). Deprivation indices have been formulated at the national level to

quantify socioeconomic (Panczak et al., 2012; G. Rey et al., 2009) or environmental deprivation (Pearce et al., 2010; Ribeiro, Pina, and Mitchell, 2015; Richardson et al., 2010).

Such composite indicators provide a valuable means of ranking disadvantaged areas and thus offer an alternative approach to prioritizing interventions. However, public health practitioners may also seek to categorize neighborhoods that share similar characteristics to promote mutual learning. This categorization is achieved through geodemographic classification, a process that typically combines principal component analysis (PCA) with a clustering algorithm. Essentially, PCA reduces the dimensionality of the variables used to characterize areas, while the clustering algorithm groups these areas based on common profiles. For example, Y.-S. Li and Chuang (2009) used a classification approach to delineate different types of neighborhoods in Taiwan. They then integrated these groups into a multilevel model to examine associations with population health status. Richardson et al. (2010) concurrently developed a deprivation index (MEDIX) and a classification system for UK Census Area Statistics wards (MEDClass) to comprehensively assess multiple environmental deprivations within the UK population.

### **1.3 Context: Swiss health system and situation in the canton of Vaud**

Switzerland is a small country (8.8 million inhabitants in 2022, 41,284 km<sup>2</sup>) which operates under a decentralized administrative and political system comprising the federal government, 26 cantons, and 2136 municipalities, as of January 1, 2023. Notably, Switzerland's political landscape includes direct democracy, which allows citizens to participate directly in decision making through initiatives and referendums. The country's health system is known for its good healthcare quality and population satisfaction, but it also faces significant healthcare expenditures, which are among the highest in OECD countries (12.4% of GDP in 2016 and 5.3% of household consumption in 2015) (OCDE, 2017). With one of the highest life expectancies in the world (83.4 years in 2019), Switzerland also faces challenges related to population aging, including an increased prevalence of chronic diseases and multimorbidity (OBSAN, 2015).

The decentralized nature of the Swiss system makes healthcare financing and organization complex. The federal government's authority is limited to matters outlined in the federal constitution, which leads it to focus primarily on regulating health insurance, controlling pharmaceuticals and medical devices, and overseeing specific aspects of public health, such as infectious disease control and food safety. This leaves considerable autonomy to the cantons, which are responsible for providing health care, implementing health-related legislation, providing subsidies to low-income households, and coordinating prevention and health promotion activities. As a result, there is considerable heterogeneity among the cantons in the importance attached to health and in the allocation of public health responsibilities between cantons and municipalities (De Pietro et al., 2015; Monod et al., 2023).

### 1.3. Context: Swiss health system and situation in the canton of Vaud

Additionally, Switzerland places a strong emphasis on individual responsibility, as stated in Article 6 of the Federal Constitution: "All individuals shall take responsibility for themselves and shall, according to their abilities, contribute to achieving the tasks of the state and society". This focus on individual responsibility is also reflected in the country's healthcare vision, which is predominantly focused on medical sciences and prioritizes disease treatment rather than population-level approaches that address the SDOH (Monod et al., 2023).

These unique characteristics contribute to moderate regional health inequalities and a strong social gradient, with middle- and low-income populations experiencing lower life expectancy, poorer health status, and higher prevalence of chronic diseases (Abel, Hofmann, and Schori, 2013; Lutz, 2019). Switzerland also has one of the highest rates of people foregoing medical care among OCDE countries (20.9%, OCDE (2017)). The COVID-19 pandemic further strained the Swiss healthcare system and exacerbated pre-existing inequities (Monteverde and Eicher, 2023).

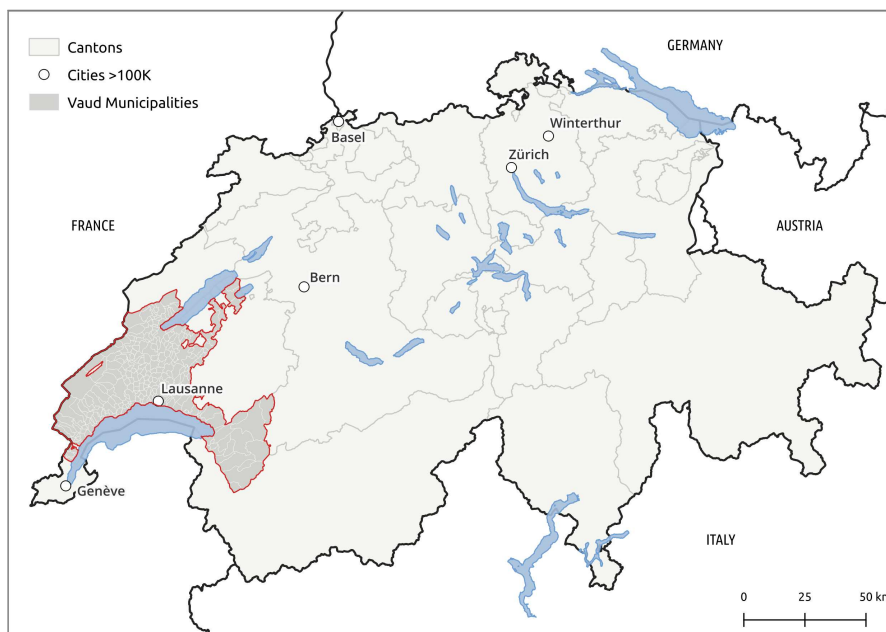


Figure 1.4 – Geographic context of case studies: Canton of Vaud, Switzerland.

The canton of Vaud is located in the French-speaking western part of Switzerland and covers an area of 3,212 km<sup>2</sup> with a population of 830,791 as of January 1, 2023 (see Figure 1.4 and supplementary Figure A.1.1). Its capital, Lausanne, is the fourth-largest city in Switzerland, with 141,418 inhabitants as of January 1, 2023. The canton faces public health challenges similar to those at the national level, including an increase in chronic diseases and an accentuation of social and cultural inequalities in health status and health behaviors (Pahud and Zufferey, 2019).

## Chapter 1. Introduction

---

In its 2018-2022 strategic report<sup>1</sup>, the cantonal health authority has attached great importance to the development of population health approaches that focus on the SDOH (DGS Vaud, 2018). Recognizing the limitations of a universal approach to addressing health inequities, i.e., where the same intervention is applied uniformly to the entire population, public health authorities have emphasized the importance of combining such approaches with targeted strategies. This approach, known as proportionate universalism, advocates implementing universal interventions while proportionately targeting the populations most in need (Marmot Review et al., 2010; Lutz, 2019).

However, the practical tailoring of these population health interventions remains elusive, with authorities citing a lack of comprehensive information on the health status and living environments of populations. In this context, the application of a geographic lens through spatial epidemiology approaches could prove to be instrumental. In addition, Switzerland has a wealth of fine-scale geospatial data at both the federal (e.g. geo.admin.ch) and cantonal (e.g. viageo.ch) levels that could be used to characterize residential environments.

Over the past two decades, there has been a remarkable growth in spatial health research, fueled by the increasing integration of "place" into medical databases. Beginning in January 1997, the Swiss Federal Statistical Office (SFSO) began recording the geographic location of all Swiss hospital discharges within designated "Medstat regions", which encompassed 605 regions across the country (Klauss et al., 2005). These regions (3,500 to 10,000 inhabitants each) were carefully delineated according to demographic, socioeconomic, and geographic factors, allowing spatial analyses of patient flows and health care utilization to guide resource allocation (Klauss et al., 2005; Matter-Walstra, Widmer, and Busato, 2006). This trend gained further momentum with the establishment of population-based cohorts in which participants' home addresses were documented, facilitating the conduct of fine-scale spatial epidemiology studies. For example, the Swiss National Cohort (SNC) (Bopp et al., 2009) has facilitated studies of the impact of the physical and social environment on various causes of death (Spoerri et al., 2010). Similarly, the Bus santé (Morabia et al., 1997) and Colaüs-PsyColaüs (Firmann et al., 2008) cohorts have collected georeferenced individual data to study the intra-urban spatial variation of cardiovascular outcomes and related determinants in the cities of Geneva and Lausanne, respectively.

However, there remains a significant gap between academic research and the translation of research findings into tangible population health interventions. This gap may be due to a variety of factors. On one hand, researchers sometimes design studies that lack reproducibility, alignment with policy needs, or the spatial and temporal scope required for intervention. On the other hand, practitioners may lack awareness of existing spatial approaches that

---

<sup>1</sup> *Rapport sur la politique de santé publique du canton de Vaud 2018-2022*, Direction Générale de la Santé (DGS) Vaud.



could be beneficial (Endacott et al., 2009) and might be reluctant to engage in academic collaboration, given the substantial time and resources required that could otherwise be allocated to interventions themselves (Hawe and Potvin, 2009).

As a result, the integration of geographic perspectives within public health agencies remains sporadic (Koschinsky, 2013), often limited to visualizing health indicators at the regional level or mapping the locations of health services. This observation was confirmed by a series of interviews conducted at the beginning of the thesis with various departments of the cantonal health authority (see details in Table A.1.1).

## 1.4 Thesis aim

Given the current state of knowledge in the spatial epidemiology literature, and the way public health policies are developed and applied in the canton of Vaud, we hypothesize that the use of spatial methods and data accessible to public authorities can enable the development of fine-scale geographic approaches; these approaches, in turn, have the potential to guide population health interventions according to proportionate universalism.

Building on our hypothesis, the aim of this thesis is to **develop spatial epidemiology approaches to support informed decision-making in the design of public health policies.**

To achieve this goal, the thesis is structured around three main objectives:

- **Characterize the population and its living environment** by collecting and analyzing social, economic, and environmental indicators, and make them accessible to public health practitioners;
- **Apply spatial epidemiology approaches to assist decision makers** in identifying areas of greatest need and local contextual determinants;
- **Enable the application of academic research findings by public health practitioners** through practical case studies.

The pursuit of these objectives raises several research questions:

- *Can social determinants of health be accurately captured by small area-level indicators?*
- *Is it feasible to conduct fine-scale spatial epidemiology studies from an applied public health perspective?*
- *How can fine-scale spatial epidemiology approaches guide the design of NCD prevention programs?*
- *How can fine-scale spatial epidemiology approaches enhance existing infectious disease surveillance systems?*
- *Can academic research findings be directly translated into public health actions?*

The spatial approaches developed should strike a balance between providing fine-scale insights for population-level interventions and being scalable to cover the entire territory, thus ensuring universal coverage of public health interventions and prevention initiatives. Furthermore, reproducibility and interdisciplinarity will be emphasized to enhance the relevance and applicability of the research. By achieving these goals, the thesis aims to make a significant contribution to informed public health decision-making and, ultimately, contribute to the reduction of health inequities.

## 1.5 Methodology

To address the research objectives, a comprehensive methodological framework was established, which was then adapted to the different case studies.

### 1.5.1 Collaboration with public health authorities

A first step was to establish collaborative relationships with key stakeholders in the cantonal health authorities. To facilitate this process, I dedicated one day per week to work at the Department of Public Health of the canton of Vaud (DGS). During this time, I presented the scope of my research, actively participated in strategic meetings, and conducted interviews with representatives from various sectors, including the public health and social services departments, as well as the University's Center for General Medicine and Public Health (Unisanté)<sup>2</sup>. This phase was important to align the research objectives of the thesis with the prevailing public health challenges, to assess the accessibility of health data sources, and to identify potential case studies for the application of spatial epidemiology approaches.

---

<sup>2</sup>Details of the interviews conducted at the cantonal health authority are provided in Appendix, Table A.1.1

### 1.5.2 Methodological choices

#### Spatial scale

The choice of an appropriate spatial scale was an important aspect of this research, as it should both inform population-level interventions at the community level and minimize biases related to the MAUP. Consequently, we chose to work at the level of inhabited hectares, i.e, a 100 x 100-meter area with resident populations. In Switzerland, each inhabited hectare is uniquely indexed by a *RELI* code and hosts a population ranging from 3 to 800 individuals. This is the smallest geographic unit for which census data are available from the Swiss Federal Statistical Office (SFSO).

#### Dataset acquisition

Two main types of secondary data sources were used in this research:

1. **Health data:** Obtained from non-spatial individual health databases, these data included information on the health status, potential risk factors (age, gender, lifestyle behaviors), and home addresses of a subpopulation of Vaud residents. It's important to note that these data were pseudonymized; all personally identifiable information was encrypted by the data provider prior to use in the research.
2. **Geographic contextual data:** These data were extracted from national or cantonal geodatabases and included demographic, social, socioeconomic, or environmental information.

#### Data preprocessing

Prior to analysis, two important preprocessing steps were applied to the datasets:

1. **Geocoding of health data:** The health datasets were initially non-spatial, necessitating a geocoding process. A home-made geocoding algorithm was developed to ensure high accuracy irrespective of the initial data source quality. For the detailed geocoding workflow, see the supplementary Figure A.1.2. Commercial geocoding services were not used to protect data privacy.
2. **Transformation of contextual data:** Geographic contextual data were sourced in diverse spatial formats, including point data for healthcare service locations, rasters for environmental exposures, and areal data for demographic statistics. To facilitate a coherent analysis, these datasets were usually standardized to the hectare level. This involved deriving meaningful indicators, such as the distance to the nearest health service, average air pollutant concentrations, or the proportion of the elderly population.

### Analytical approaches

Spatial epidemiology methods were used to achieve the research objectives, with three main strategies:

1. Analyzing the spatial pattern of health outcomes (Y) using spatial clustering methods to identify **high-risk areas**, i.e., spatial clusters.
2. Analyzing the spatial pattern of contextual variables (X) using composite indicators or geodemographic classifications to highlight **vulnerable areas**.
3. Investigating the associations between health outcomes (Y) and contextual variables (X) using spatial regression methods to guide the development of tailored and context-specific population health interventions.

The adoption of a particular strategy depended on the nature of the research question and the data available for each case study. A schematic representation of the methodological approach is provided in Figure 1.5.

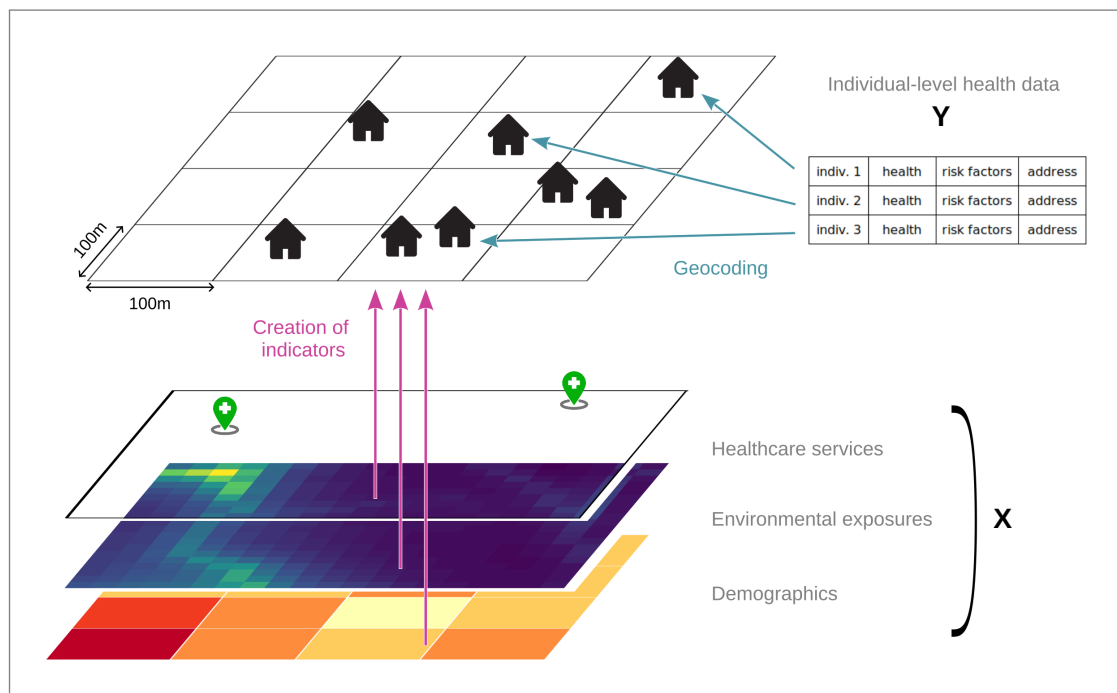


Figure 1.5 – Methodological approach.

### Data storage and protection

All research data were stored locally in a centralized Postgres/PostGIS database. To enhance scalability and future deployment, the database was organized into distinct compartments known as "schemas". Each schema acts as a structural framework designed for a specific case study, thereby organizing the data and their interconnections. This schema-based organization was conceived to facilitate the allocation of distinct access credentials for public health practitioners if the database has to be deployed on a remote server.

This research adhered to the Swiss data protection regulations, encompassing the Federal Law on Data Protection (Loi sur la Protection des Données, LPD) and the Federal Law on Human Research (Loi relative à la recherche sur l'être humain, LRH). To safeguard individual privacy, health data were anonymized through techniques such as aggregation or shifting of point locations. In addition, demographic statistics included a bias for inhabited hectares with less than 3 inhabitants.

### Reproducibility and data sharing

Certain methodological choices were made to improve reproducibility of the research work:

1. Contextual data were primarily sourced from open datasets, ensuring consistent updates and nationwide availability. This facilitates benchmarking with other cantons.
2. All processing codes are accessible online<sup>3</sup>, underscoring our commitment to transparency and reproducibility.
3. Only free and open-source software (FOSS) were employed, promoting financial independence and cross-platform interoperability.

These considerations have been made to encourage adoption of these spatial approaches within existing public health structures, and to facilitate collaboration with other researchers.

## 1.6 Chapters outline

The rest of the thesis is organized into five main chapters, including three scientific articles: two have been published and one has been submitted for peer-reviewed publication. In addition, two practical applications demonstrate the use of spatial approaches to guide cantonal health programs.

Starting with the challenges faced by a national program working with municipalities to

---

<sup>3</sup>The complete list of processing codes used in this research can be found in the Appendix A.7.

## Chapter 1. Introduction

---

develop health promotion measures, GIS approaches are used to develop and map fine-scale indicators that capture SDOH across the territory (Chapter 2). In Chapter 3, these contextual factors are linked to individual-level health data to better understand how characteristics of the residential environment influence the geographic variation in NCDs. The focus then shifts to infectious disease surveillance in Chapter 4, where spatio-temporal approaches are used to detect local clusters of COVID-19 cases. This is complemented with Chapter 5 that explores the synergies between spatio-temporal methods and genomics-based analyses to improve surveillance systems. As a continuation, Chapter 6 presents a unique application of a spatial approach to design a mobile vaccination campaign against COVID-19. Finally, Chapter 7 discusses the overall contributions of the thesis.

**Chapter 2** focuses on the collection, creation and mapping of SDOH using spatial approaches. This research was designed to meet the specific needs of the cantonal application of a national health program named "Commune en Santé" label. The program supports local health initiatives by helping municipalities to improve existing health promotion efforts and to develop health-promoting environments. Specifically, we conducted a geodemographic classification to explore how the characteristics of the municipalities influenced their participation in the labeling process. GIS techniques were used to generate a comprehensive set of indicators at the hectare level, covering different dimensions of SDOH. Composite indicators representing material deprivation, environmental burdens and distance to primary healthcare facilities were also developed. Collectively, these indicators were integrated into a web application (<https://geosan.epfl.ch>) and a GIS project, which allows for visualization, data querying, and map generation. This work serves as a foundation for the development of contextual factors that will be used in subsequent chapters.

In **Chapter 3**, the intra-urban variation of cardiometabolic risk factors (CMRFs) - namely, hypertension, obesity, diabetes, and dyslipidemia - was assessed within the city of Lausanne. The analysis was performed on 3,695 middle-aged and older adults participating in the ColaUS-PsyCola study, a population-based medical cohort conducted in Lausanne since 2003. We used intensity functions to predict the spatial risk of CMRFs and found persistent spatial clusters for obesity, diabetes, and hypertension. Geographically weighted regressions were then used to understand how contextual factors shape spatial variation in disease risk, even after controlling for important individual confounders. Notable local associations emerged with specific characteristics of the physical and social environments, particularly in relation to obesity. The chapter concludes with a discussion of the broader applicability of our findings to the entire territory.

**Chapter 4** addresses the topic of infectious diseases, which has become a prominent focus due to the emergence of COVID-19 shortly after the start of this thesis. The chapter presents a spatio-temporal analysis of COVID-19 cases during the first wave of the epidemic. In col-

laboration with the Institute of Microbiology of the University Hospital of Lausanne (CHUV), 33,651 SARS-CoV-2 RT-PCR tests were geocoded to patient's home addresses. We detected daily emerging clusters of COVID-19 across the canton of Vaud using space-time scan statistics and assessed how spatial clusters varied in terms of size, duration, and patient viral load. We also assessed the impact of the Swiss soft lockdown (March 19 - April 27, 2020) on the diffusion dynamics of transmission clusters detected using a modified space-time DBSCAN algorithm.

In **Chapter 5**, the synergies between genomic-based approaches and space-time scan statistics are explored to enhance public health surveillance systems for future epidemics. By sequencing the SARS-CoV-2 genomes of all COVID-19 cases within 17 representative geographic clusters identified in Chapter 4, we assessed the congruence between spatial and transmission clusters and identified two superspreading events potentially responsible for virus spread within the canton. Based on the results of the study, recommendations were made to optimize the efficiency of the cantonal infectious disease surveillance system.

**Chapter 6** presents a practical application of spatial approaches used in crisis management, specifically to guide the planning of the cantonal COVID-19 mobile vaccination campaign. This innovative strategy, launched by the cantonal health department in the spring of 2021, used mobile vaccination units to overcome potential barriers to vaccine access, such as geographic remoteness, restricted mobility, and socioeconomic deprivation. By creating a composite indicator of vulnerability and identifying clusters of low vaccination rates, we identified optimal locations for vaccination centers and subsequently advised the Head of the Vaud Civil Protection Office in campaign planning.

In **Chapter 7**, the discussion summarizes the research contributions, offers recommendations for integrating geographic approaches into public health departments, and outlines perspectives in the field of spatial epidemiology.





## 2 SUPPORTING HEALTH PROMOTION INITIATIVES WITH GEOGRAPHIC INFORMATION SYSTEMS

### 2.1 Introduction

Over the past two decades, compelling evidence has underscored the impact of social and environmental factors on health outcomes (M. Marmot and J. J. Allen, 2014), leading to the adoption of the concept of social determinants of health (SDOH). Effectively promoting population health and reducing health inequities requires multilevel interventions, addressing the health system, individual behaviors, as well as physical and social environments (Paskett et al., 2016). Such interventions require interdisciplinary cooperation across all government sectors (M. Marmot, J. Allen, et al., 2012; Amstutz and Villa, 2021).

In Switzerland, the lack of a federal legal framework for health promotion and prevention of NCDs hampers the coordination among the various stakeholders. Legislative efforts to fill this gap have failed, underscoring Switzerland's emphasis on individual responsibility (Mattig, 2013). This individualistic framing of NCD prevention predominantly tends to focus on the "big four" risk factors: diet, smoking, physical activity, and alcohol consumption (Macintyre, Ellaway, and Cummins, 2002). However, such an approach fails to address the underlying determinants that influence these behaviors, often referred to as the "causes of the causes" (M. Marmot and J. J. Allen, 2014).

The 2018-2022 strategic plan for public health policy in the canton of Vaud expresses the intention to strengthen the involvement of local authorities such as municipalities in health promotion (DGS Vaud, 2018). Due to their close relationship with the local population, the municipalities are well positioned to develop population health interventions by creating health-promoting environments. The national health promotion program "Commune en Santé", created in 2010 and implemented in 2015 in the canton of Vaud under the leadership of Unisanté, serves as an example in this regard. The program offers free advice to municipalities

## **Chapter 2. Supporting health promotion initiatives with geographic information systems**

---

with the aim of listing existing health promotion initiatives and advising on the introduction of new measures (Amstutz, Villa, et al., 2020). The labeling process evaluates measures in various areas related to municipal policies, leisure activities, family and solidarity, schools, public spaces, and occupational health. It includes validation by an external committee. As of October 1, 2021, only fifteen out of 302 municipalities in Vaud have obtained the label, with four others in the process (see Figure A.2.1 in Appendix). A recent external evaluation of the label revealed that small and rural villages were less committed than urban areas, possibly due to limited resources (Monin et al., 2020). This asymmetric response risks creating or exacerbating territorial health inequities.

In this context, the assets of spatial approaches become apparent. In particular, Geographic Information Systems (GIS) enable the integration and synthesis of heterogeneous data sources to characterize SDOH (Cromley and McLafferty, 2011). Furthermore, GIS allow for the generation of impactful visual representations that effectively convey key public health messages (Nykiyoruk and Flaman, 2011).

The research was structured around three main objectives. First, we assessed how the economic, environmental, and demographic characteristics of municipalities may influence their commitment to the "Commune en Santé" label. This assessment was informed by the development of a geodemographic classification for the municipalities in the canton of Vaud. Second, we used GIS techniques to construct fine-scale indicators designed to inform the labeling process and facilitate targeted health promotion efforts. Finally, we present how our findings were communicated to the relevant public health authorities. While the questions and methods were tailored to the Commune en Santé program, the resulting indicators and tools have broader applicability and may be useful for broader prevention and health promotion initiatives throughout the canton.

### **2.2 Geodemographic classification of municipalities**

In order to identify relevant variables for the geodemographic classification, we undertook a multi-step selection process. First, we gathered insights from previous typologies of Swiss municipalities conducted by the Swiss Federal Statistical Office (SFSO) (Zecha, Kohler, and Goebel, 2017) and Dumondel (1985), which provided information on salient differences between Swiss territories. Subsequently, empirical observations from the "Commune en Santé" team and findings from the label evaluation report (Monin et al., 2020) guided the inclusion of additional variables. Ultimately, our analysis included 39 variables encompassing information on the resident population, the level of urbanization, and the financial resources of the municipalities. A comprehensive list of these variables, along with their data sources, is available in Appendix, Table A.2.1.

## 2.2. Geodemographic classification of municipalities

---

To mitigate multicollinearity and enhance factor interpretability, a Principal Component Analysis (PCA) was performed on the selected variables. This technique transforms an original set of correlated variables into a new set of uncorrelated variables termed as principal components (PC). These components capture the most significant patterns and explain most of the variability in the data, thereby facilitating the delineation of distinct profiles. The analysis was conducted on standardized variables, with variables related to political orientations and income inequality given half the weight due to their relatively lower relevance. Based on the broken stick model test, we retained four principal components that explain 54.2% of the total variation.

The first principal component (1st PC - 23.3% of the variability) predominantly captures urban-rural and sociodemographic differences. It contrasts urbanized areas characterized by abundant public infrastructure, population growth, and socioeconomic disparities, with conservative rural localities that feature farmland and a high proportion of primary and secondary sector workers. The second principal component (2nd PC - 15% of the variability) reflects differences in SES and daily mobility patterns. It distinguishes areas with high income levels and a significant share of commuters from disadvantaged areas with an aging population that are far from urban centers. The third component (3rd PC - 9.5% of the variability) contrasts industrial peri-urban areas with isolated mountainous regions having financial resources but important socioeconomic disparities. The fourth component (4th PC - 6.4% of the variability) captures the age structure and geographical differences. It differentiates residential areas predominantly inhabited by retirees, from rural areas with a demographic skew towards children and the working-age population.

To group municipalities with similar profiles, we employed hierarchical agglomerative ascending classification, a clustering approach that does not require the number of groups to be predetermined (Kaufman and Rousseeuw, 2009). The algorithm, executed based on the positions of the municipalities across the four principal components, aims to maximize inter-group differences while minimizing intra-group variances. The optimal number of clusters was identified by maximizing the average silhouette coefficient, i.e., a metric that reflects the degree of similarity within groups. The geodemographic classification was performed in R using the *FactoMineR* and *stats* packages.

It ultimately led to the identification of nine distinct profiles (see Figure 2.1), each characterized by a unique combination of variables. These profiles are described below:

- PROFILE 1: AFFLUENT SUBURBAN CROWNS (9.6% of the population, 17.9% of the municipalities). Low-density residential zones near main urban centers with high SES, diverse nationalities, and elevated commuting rates.
- PROFILE 2: REMOTE RURAL REGIONS (4.1% of the population, 9.3% of the municipalities). Low- and mid-mountain rural areas remote from urban centers, characterized by limited

industrial and recreational opportunities and low residential attractiveness.

- PROFILE 3: COSMOPOLITAN EXPATRIATE HUBS (10.5% of the population, 6.3% of the municipalities). Urban and suburban areas characterized by significant proportions of allophones and European immigrants with high SES, but with notable income disparities.
- PROFILE 4: URBAN CENTERS AND AGGLOMERATIONS (42.7% of the population, 6.3% of the municipalities). Economically active urban cores populated by a young and working-age population, accompanied by lower levels of education compared to the rest of the canton.
- PROFILE 5: AGING AND CONSERVATIVE AGRICULTURAL COMMUNITIES (2.6% of the population, 18.2% of the municipalities). Small, agriculture-focused rural areas with older, conservative populations and low immigration rates.
- PROFILE 6: FAMILY-FOCUSED RURAL AND SUBURBAN AREAS (7.2% of the population, 23.5% of the municipalities). Inland low-density areas characterized by a high proportion of families, as indicated by a significant proportion of children and working-age individuals, and high commuting rates to nearby urban centers.
- PROFILE 7: INDUSTRIAL SUBURBS (8.1% of the population, 7.9% of the municipalities). Medium-density suburbs with industrial focus and high share of workers in the secondary sector.
- PROFILE 8: SOCIOECONOMICALLY DISADVANTAGED DISTRICTS (12.7% of the population, 6.3% of the municipalities). Areas that are heterogeneous in their level of urbanization, but all characterized by low SES, especially in educational attainment, accompanied by slightly above-average rates of immigration.
- PROFILE 9: REMOTE MOUNTAIN COMMUNITIES (2.6% of the population, 4.3% of the municipalities). Mountainous regions distant from urban centers, characterized by low income levels and a high proportion of retired individuals.

Table 2.1 shows summary statistics for the labeled municipalities as of October 2021, categorized by their respective profiles. It includes the number of measures adopted, and their distribution across the six main areas of the label: municipal policies, leisure activities, family and solidarity, school, occupational health, and public spaces.

Several notable observations can be made. First, Profile 4 is overrepresented with 8 of the 15 labeled municipalities (43.3%), which could support the conclusions of Monin et al. (2020). This observation aligns with the hypothesis that urban municipalities not only possess more abundant financial and human resources but also frequently have pre-existing health promotion measures implemented, which reduce the investment required to acquire the label. This is illustrated in the share of measures related to municipal policies, with the exception of Renens and Chavannes-près-Renens.

## 2.2. Geodemographic classification of municipalities

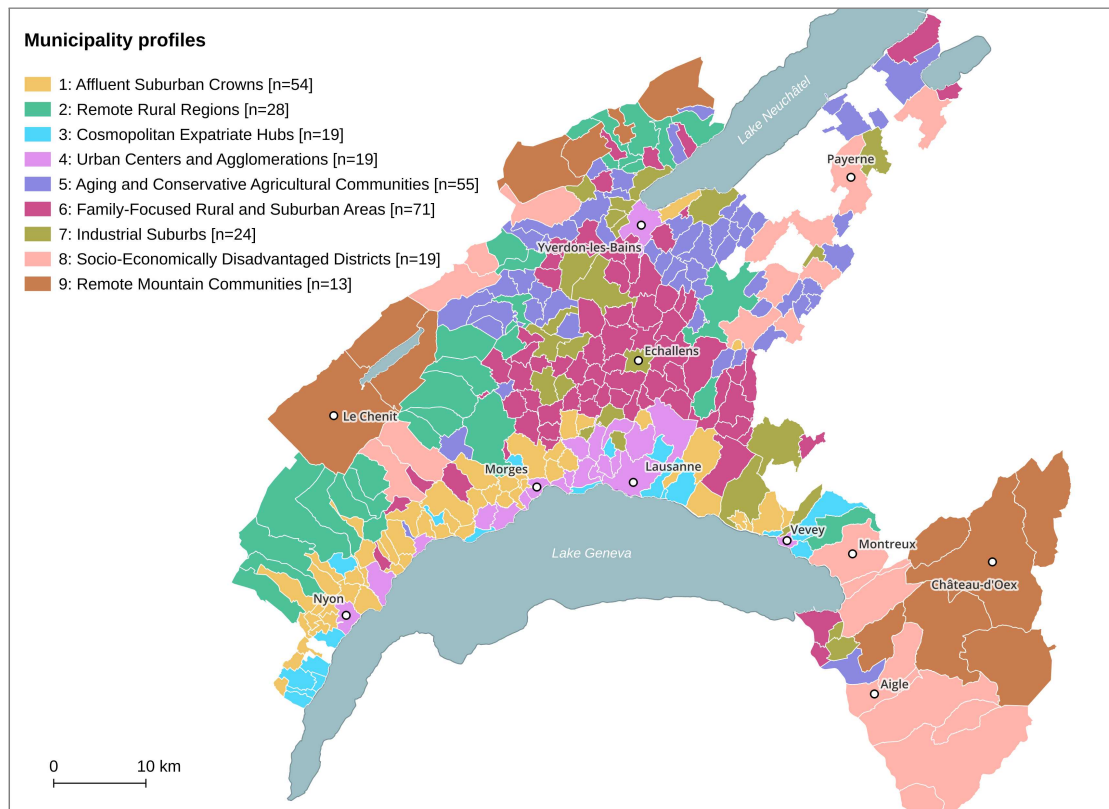


Figure 2.1 – Typology of municipalities according to 9 distinct profiles.

Municipality	Profile	Measures [#]	Municipal policy [%]	Leisure offers [%]	Family and solidarity [%]	School [%]	Occupational health [%]	Public spaces [%]
Le Vaud	1	19	21.05	15.79	21.05	10.53	5.26	26.32
Tévenon	2	19	15.79	21.05	21.05	10.53	15.79	15.79
Saint-Sulpice	3	28	25.00	10.71	14.29	25.00	10.71	14.29
Chavannes-près-Renens	4	28	21.43	14.29	28.57	10.71	10.71	14.29
Bussigny	4	32	31.25	9.38	15.63	9.38	12.50	21.88
Le Mont-sur-Lausanne	4	24	20.83	16.67	12.50	20.83	12.50	16.67
Ecublens	4	28	17.86	25.00	17.86	10.71	10.71	17.86
Gland	4	20	20.00	20.00	15.00	15.00	15.00	15.00
Renens	4	33	15.15	24.24	27.27	9.09	9.09	15.15
Nyon	4	68	26.47	19.12	20.59	8.82	7.35	17.65
Lausanne	4	79	32.91	17.72	15.19	13.92	7.59	12.66
Bercher	6	15	20.00	13.33	13.33	13.33	20.00	20.00
Montagny-près-d'Yverdon	7	19	31.58	15.79	10.53	10.53	10.53	21.05
Château d'Oex	9	25	16.00	8.00	24.00	12.00	12.00	28.00
Sainte-Croix	9	22	9.09	18.18	27.27	9.09	13.64	22.73

Table 2.1 – Total number of measures adopted and distribution of the percentage across the 6 domains for each labeled municipality.

Interestingly, Profile 9 (Remote mountain communities) shows a clear trend in the choice of health promotion domains, with a strong focus on measures related to family and solidarity and public spaces. This focus may reflect the development of self-help measures due to their geographical isolation, and an increased interest in nature conservation. Conversely, Profile 5 (Aging and conservative agricultural communities) and Profile 8 (Socioeconomically disadvantaged districts) do not have any municipalities that have been labeled or show interest in the label (Figure A.2.1). Further research, such as surveys of representative municipalities, could help to clarify the reasons behind this lack of interest.

Conclusive statements about other profiles (profiles 1, 2, 3, 6, 7) remain speculative because they are represented by only one labeled municipality. However, the dataset will expand as additional municipalities attain labeling, thereby allowing the emergence of potential trends.

### 2.3 Mapping social determinants of health

The geodemographic classification provided valuable insights for engaging communities that may have been left apart from the labeling process. However, while indicators at this spatial scale are informative for initial assessments, they may lack the requisite granularity for municipalities already engaged in the labeling process. As municipalities move forward in implementing health promotion measures, a more detailed and fine-scale understanding of the environment will be essential. This insight is crucial for various stages: territorial diagnosis, the planning of new health promotion measures, and the long-term monitoring and evaluation of interventions aimed at reducing health inequities.

Consequently, the subsequent phase of the project was devoted to the development of a large set of hectare-level indicators that serve as proxies for various SDOH. In addition, three composite indicators were created to provide a synthetic measure of material deprivation, environmental burdens, and proximity to primary healthcare facilities.

#### 2.3.1 Hectare-level indicators

Variables were chosen for their health relevance, while the corresponding geodata were selected based on comprehensive coverage of all inhabited hectares within the study area. Preference was given to public repositories that are regularly updated. This section describes the selected variables, rationalizes their inclusion based on their health relevance, and explains the methodology used to construct hectare-level indicators.

##### *Environmental exposures*

Environmental risks are estimated to contribute for 24% of the global burden of disease (WHO, 2011). **Anthropogenic air pollution** is the leading environmental risk factor worldwide. In

particular, long-term and short-term exposures to particulate matter (PM<sub>10</sub> and PM<sub>2.5</sub>), carbon monoxide (CO), ozone (O<sub>3</sub>), nitrogen dioxide (NO<sub>2</sub>), and sulfur dioxide (SO<sub>2</sub>) have been linked to CVD, lung cancer, respiratory diseases, and neurological disorders (Manisalidis et al., 2020). In Switzerland, targeted policies to reduce air pollution in the Lausanne-Morges agglomeration have led to measurable health improvements, including an estimated 1-2% reduction in annual all-cause mortality (Castro, Künzli, and Götschi, 2017). In this study, we used the 2020 Pollumap air quality model (Heldstab, Schäppi, and Künzle, 2020), which has a 20x20m resolution, to calculate mean exposure levels of NO<sub>2</sub>, PM<sub>10</sub>, and PM<sub>2.5</sub> at the hectare level.

**Transportation-related noise** is another important environmental health determinant. Extensive research has established a causal link between transportation noise and adverse health outcomes, including annoyance, sleep disturbance, diabetes, and CVD (WHO, 2011; Vienneau, Eze, et al., 2019). In 2010, transportation noise resulted in the loss of 6,000 years of life (YLL) in Switzerland, with external health costs estimated to CHF 1,800 million nationwide (Vienneau, Perez, et al., 2015). Notably, Héritier et al. (2017) found that populations highly exposed to road and rail noise tend to have lower SES compared to the general population. We used the 2015 sonBASE database (FOEN, 2009) to compute a single measure of mean nighttime noise from road and railway sources at the hectare level. The sonBASE model computed noise immissions from railway and road traffic sources at a resolution of 10x10m.

The multi-faceted challenge of **urban heat islands** (UHIs) underscores the complex relationship between climate change, urban design and health. Urban areas, with their extensive concrete and asphalt surfaces, complex street and building geometries, and accumulation of anthropogenic heat, amplify the effects of heat waves. This phenomenon leads to an increased risk of heat-related mortality, which disproportionately affects vulnerable groups such as the elderly and those with pre-existing health conditions (Thommen, 2005; G. Huang, Zhou, and Cadenasso, 2011). Land surface temperature (LST) derived from remote sensing for a summer day is a common approach to approximate ambient air temperature (G. Huang, Zhou, and Cadenasso, 2011; Bosch et al., 2021). Using Landsat 8 satellite imagery, we calculated the mean LST for three summer days when the temperature exceeded 25°C and cloud cover was minimal (June 29, 2019, July 20, 2021, and August 21, 2021) (USGS, 2022). The mean LST was aggregated at the hectare level and converted to Celsius degrees. Recent advances by Bosch et al. (2021) integrate additional factors such as tree shading, urban vegetation evapotranspiration, and land surface albedo to provide a more accurate assessment of the UHI. While implementation of their approach was impractical for the broad scope of our study, their reusable computational workflow may prove valuable in developing heat wave prevention strategies for specific areas.

## **Chapter 2. Supporting health promotion initiatives with geographic information systems**

---

### *Accessibility*

Access to health services is a crucial health determinant, directly affecting an individual's ability to receive timely medical care. Beyond this, access to basic amenities such as food, recreation, schools, natural environments, and transportation also shapes the broader spectrum of SDOH. Inequities in accessing these amenities can arise from geographic, economic, or social factors, thereby disproportionately impacting populations with restricted mobility - such as the elderly or individuals with disabilities - as well as those lacking motorized transportation or facing financial limitations (OFS, 2018).

The measurement of spatial accessibility typically revolves around factors such as distance or travel time to points of interest. For our study, we used hectare-level indicators provided by the SFSO (OFS, 2018). These indicators quantify travel distances via road networks to various amenities, using 2018 data. We included several of these indicators that were relevant from a health promotion perspective. These include distances to compulsory schools, secondary and tertiary educational institutions, police services, grocery stores, forests, lakes, sports and fitness facilities, swimming pools and beaches, public transport stops, and medical offices. In addition, we have integrated an indicator for the overall quality of public transport (ARE, 2022).

We also used recent datasets from the cantonal health department, which provide the exact locations of health services, including emergency facilities, pharmacies, dental clinics, and ambulances (Canton de Vaud, 2023a). This allowed us to calculate a more realistic measure of accessibility using travel times and considering two modes of transportation, as most people walk about 10 minutes to get to a point of interest (M. Du, X. Zhang, and Mora, 2021). We developed a Python algorithm that works as follows: walkable and drivable road networks were extracted from OpenStreetMap (OSM) across the study area. Travel times from hectare centroids to the nearest amenity were computed using both networks. Specifically, the travel time took into account driving limits for road routes and a pedestrian speed of 3.6 km/h for walking routes. If the walking travel time was equal to or less than 10 minutes, it was retained; otherwise, the motorized travel time was used. The same methodology was used to calculate accessibility to playgrounds, which were obtained from the OSM repository (OSM contributors, 2023).

One notable aspect that is increasingly being explored in environmental health research is the presence of green and blue spaces within residential surroundings. These spaces can act as protective factors against CVD and mental health problems such as depression (Diez Roux and Mair, 2010), and their uneven distribution can potentially exacerbate health disparities (Schüle et al., 2019). Using data from the 2023 Vaud cadastral survey (Canton de Vaud, 2023b), we computed ratios of green and blue spaces within an 800-meter radial buffer around each hectare's centroid. While there is no consensus on the distance to be considered (Browning



and K. Lee, 2017), we chose a radial distance of 800 meters which is commonly used in health-related research (Schüle et al., 2019) and accommodates both urban and rural contexts.

Lastly, we calculated the number of traffic accidents involving pedestrians or cyclists within an 800m buffer around the hectare centroids, using a federal dataset on traffic accidents (OFS, 2022a). This indicator directly ties into the concept of walkability, reflecting the extent to which a neighborhood encourages walking and, consequently, physical activity (De Courrèges et al., 2021).

### *Demography*

Age plays a central role in shaping and interacting with several SDOH and often guides population health initiatives. The inclusion of indicators related to age distribution was therefore essential. We used standard age group categories commonly used in public health: children (0-15), young adults (15-24), adults (25-64), and the elderly (65+ and 80+). In addition, we provided information on the total resident population and the breakdown of age groups by sex.

### *Social and economic*

Several social and economic indicators that fall within the domain of SDOH were obtained from MicroGIS ([www.microgis.ch](http://www.microgis.ch)) at the hectare level for the year 2019. These indicators, compiled from federal and cantonal databases, covered various aspects such as housing conditions (proportion of people living alone and average household size), socioeconomic factors (median household income, population with compulsory education, unemployment rate), marital status (divorced or widowed people), cultural diversity (foreign population, non-French speaking population) and population with disabilities. Together, they serve as indicators of social vulnerability, which can be broadly understood as the susceptibility of individuals or communities to cope with, adapt to, or resist environmental and health hazards (Andrew and Keefe, 2014; R. Shrestha et al., 2016).

The comprehensive set of 65 hectare-level indicators, detailing the SDOH in residential settings, can be found in Appendix (Table A.2.2). This list is a valuable tool for the "Commune en Santé" team and other decision makers across the canton involved in public health and prevention.

### **2.3.2 Composite indicators**

In practical decision making, it is often more efficient to start with an overall view of a dimension, such as social vulnerability, than to examine a large number of separate indicators spatially. Composite indicators (i.e., a compilation of several indicators) have emerged as a valuable technique for encapsulating complex and multidimensional aspects of a given concept.

## Chapter 2. Supporting health promotion initiatives with geographic information systems

---

From the pool of hectare-level indicators outlined above (see Table A.2.2), we constructed three composite indicators designed to encapsulate distinct dimensions of deprivation or disadvantage: (1) material deprivation, (2) environmental burdens, and (3) distance to primary healthcare facilities.

### *Material deprivation*

The concept of material deprivation revolves around the lack of access to goods and conveniences associated with modern living, thus closely aligned with poverty (Pampalon et al., 2009). To capture this dimension, we considered the median annual household income, the percentage of the population aged 15 and over with compulsory education, and the unemployment rate for those aged 15 and over. Low income and unemployment restrict material resources, while limited educational attainment hinders access to employment and services. We have consciously omitted other social vulnerability dimensions calculated at the hectare level, such as individuals living alone, disability, or ethnicity, following the rationale that belonging to these groups does not inherently signify deprivation, despite potential vulnerability (Allik et al., 2020).

### *Environmental burdens*

For environmental risk exposures, we included variables such as exposure to NO<sub>2</sub> and PM<sub>2.5</sub>, nighttime road and railway noise, and LST. The exposure to PM<sub>10</sub> was not included because it exhibited particularly high correlation with PM<sub>2.5</sub> (0.97).

### *Distance to primary healthcare facilities*

Lastly, for the composite indicator reflecting distance to primary healthcare facilities, we incorporate travel times to emergency facilities, pharmacies, and distance to medical offices. These services offer general health advice and medical assistance.

### *Aggregation of indicators*

Two prominent methods exist for aggregating indicators into a singular measure: standardizing individual variables and combining them through a geometric or arithmetic mean with normative weights, or using PCA to reduce redundancy and assign data-driven weights (OECD, 2008; Lalloué et al., 2013). While PCA offers methodological advantages, its spatial and temporal replicability is limited (Allik et al., 2020). Therefore, we opted for the former approach. Similar to the approach of R. Shrestha et al. (2016) in Germany, we used arithmetic aggregation for material deprivation and geometric aggregation for environmental burdens. Arithmetic aggregation allows for higher compensability between indicators than geometric aggregation. We argue that high incomes could compensate for material deprivation resulting from high unemployment rates, while areas with minimal noise should not fully compensate for the effects of air pollution. Following this reasoning, we used geometric mean aggregation to formulate the distance to primary healthcare facilities indicator.

## 2.3. Mapping social determinants of health

---

Before the aggregation process, preliminary steps were undertaken. These included imputing missing values using the nearest non-null hectare value (affecting 1.5% of all inhabited hectares), and log-transforming variables with skewed distributions (e.g., PM2.5 concentration, distance to medical offices, income, education, and unemployment). To ensure comparability, we standardized the indicators included in the composite indicator. We applied z-score standardization to indicators reflecting material deprivation, as is commonly done in the literature (R. Shrestha et al., 2016; Allik et al., 2020). However, this approach was not appropriate for environmental burdens and accessibility to primary healthcare facilities, as it was preferable to have only positive values for geometric aggregations. As a result, we adopted a min-max normalization for these indicators, scaling them all within an identical range of 0 to 1.

In terms of weights, an equal weight basis (all weights set to 1) was adopted for the three composite indicators. This choice arose from the absence of definitive evidence regarding the relative contribution of individual indicators to material deprivation, environmental burdens, or accessibility to primary healthcare facilities. However, a slight downweighting was applied to NO<sub>2</sub> and PM<sub>2.5</sub> (each with a weight of 0.75) to avoid an imbalanced structure where air pollution dominates noise and heat, particularly given the relatively high correlation between NO<sub>2</sub> and PM<sub>2.5</sub> (0.8).

Finally, quantiles were assigned to each inhabited hectare of the canton for the three composite indicators: material deprivation, environmental burdens, and distance to primary healthcare facilities.

Taken together, these composite indicators serve as powerful decision-support tools, enabling rapid assessment of geographic areas vulnerable to adverse social and environmental conditions. They also highlight regions facing "triple jeopardy", where multiple environmental burdens, material deprivation, and lack of access to primary healthcare facilities intersect.

The correlation between the composite indicator of environmental burdens and material deprivation was 0.20, suggesting that populations with lower SES tend to live in environmentally deteriorated neighborhoods. However, this gradient is not apparent for distance to primary healthcare facilities (-0.07). The correlation between the distance to primary healthcare facilities and environmental burdens was -0.44, suggesting a moderate negative correlation. This indicates that in regions characterized by greater distances to primary healthcare facilities, often rural or mountainous areas, there tends to be lower levels of environmental hazards.

To illustrate the results with a practical example, Figure 2.2 shows the maps of the composite vulnerability indicators for the municipality of Aubonne (3,576 inhabitants in 2021). Selected for its unique geography, Aubonne has extensive forests and meadows in the north, a historic central built-up area, and a large industrial zone in the south. This distinct landscape is reflected in the composite indicator of environmental burdens (Figure 2.2 (B)), which shows

a north-south gradient, with southern inhabited hectares around industrial zones facing increased exposure to air pollution, nighttime noise, and heat islands.

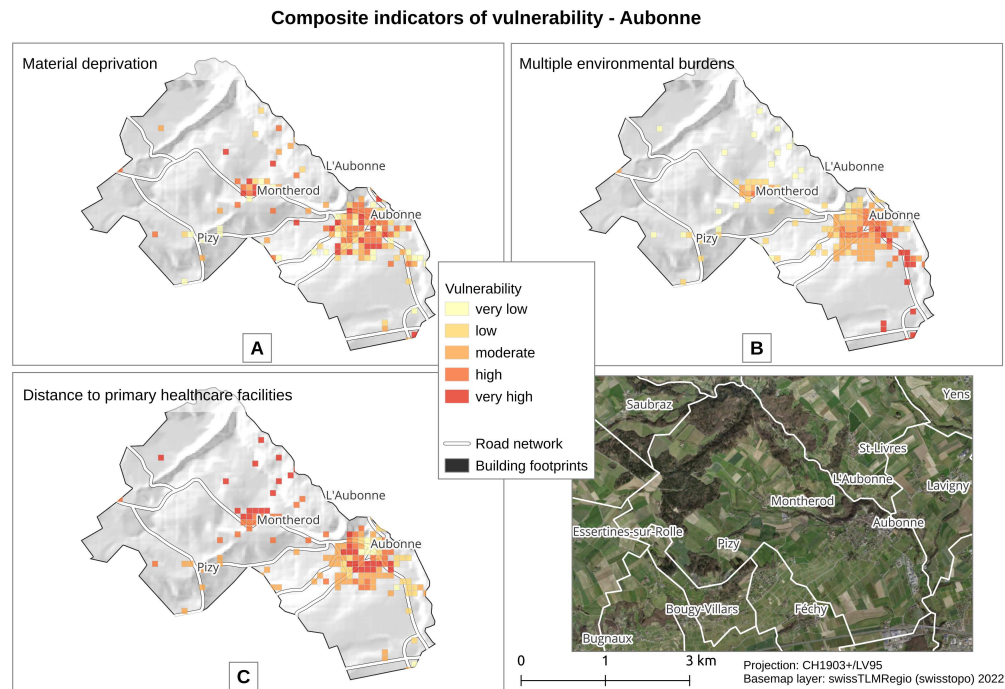


Figure 2.2 – **Composite indicators of vulnerability for the municipality of Aubonne:** (a) material deprivation, (b) multiple environmental burdens, and (c) distance to primary healthcare facilities. Vulnerability is assessed for each inhabited hectare in comparison to the rest of the canton of Vaud.

## 2.4 Web application

The practical utility of the developed indicators entirely depends on their accessibility and applicability by public health practitioners. To this end, we developed a web application as an integral part of this research project, using the Python Django open source framework (Django Software Foundation, 2019). This application is securely hosted on an EPFL server and provides access to both the Unisanté and the cantonal public health department. The architectural layout of the application is presented in Figure A.2.2.

With the aim of enhancing the decision-making process of "Commune en Santé", the web application provides answers to questions such as "How does this specific indicator vary within a municipality?" or "How does the municipality's performance compare to the rest of the canton for this particular indicator?"

Taking the previous example of the city of Aubonne, the user can query the night noise exposure for this specific locality. Figure 2.3 illustrates the output generated by the web application in response to such a query. Specifically, the displayed map highlights inhabited hectares within Aubonne and adjacent municipalities. These hectares are color-coded employing a sequential scheme, transitioning from green (low values) to red (high values). The color scheme makes it easy to quickly identify hectares where night noise has no significant biological effects (dark green, WHO Regional Office for Europe (2009)), may cause moderate health effects especially in vulnerable populations (light green, WHO Regional Office for Europe (2009)), may cause adverse health effects in all exposed populations (yellow, WHO Regional Office for Europe (2009)), may cause dangerous health effects including an increase in cardiovascular risk (orange, WHO Regional Office for Europe (2009)), are above the federal limit (red, OFEV (2023)). In the absence of internationally or federally recognized health-based guidelines, we used either quantile-based or equal-interval methods to set indicator thresholds. Hovering over individual hectares shows both their specific noise exposure levels and their total population. Additional metrics such as mean exposure for a municipality and statistical summaries (frequency distributions and boxplots) can be displayed. Health services locations and the typology of municipalities can also be overlaid on the map for better contextual understanding.

In parallel to the web application, a complementary QGIS project has been developed and distributed to all relevant stakeholders. This project incorporates all datasets elaborated upon in this study (see Tables A.2.1 and A.2.2) as well as additional layers relevant to daily operational needs. The additional layers include health services locations (Canton de Vaud, 2023a), playgrounds (OSM contributors, 2023), public transportation stops (ARE, 2022), administrative boundaries (swisstopo, 2023), and cadastral land use (Canton de Vaud, 2023b), among others.

To facilitate knowledge dissemination and foster the integration of geographical perspectives into current practices, an informative workshop was held on January 16th, 2023, at Unisanté<sup>1</sup>. The workshop presented the project and the web application to public health practitioners, equipping them with GIS training on the QGIS project to navigate geographic data layers, conduct spatial queries, and generate maps for diverse applications.

## 2.5 Discussion

This project provides a compelling example of how spatial approaches can be used to achieve the objectives of the national health promotion program "Commune en santé", in particular the objective of broadening the participation of municipalities in the labeling process.

The geodemographic classification of municipalities, which includes data on the demographic and socioeconomic characteristics of the resident population, municipal resources, and

---

<sup>1</sup>The program of the workshop is provided in Appendix, Figure A.2.3.

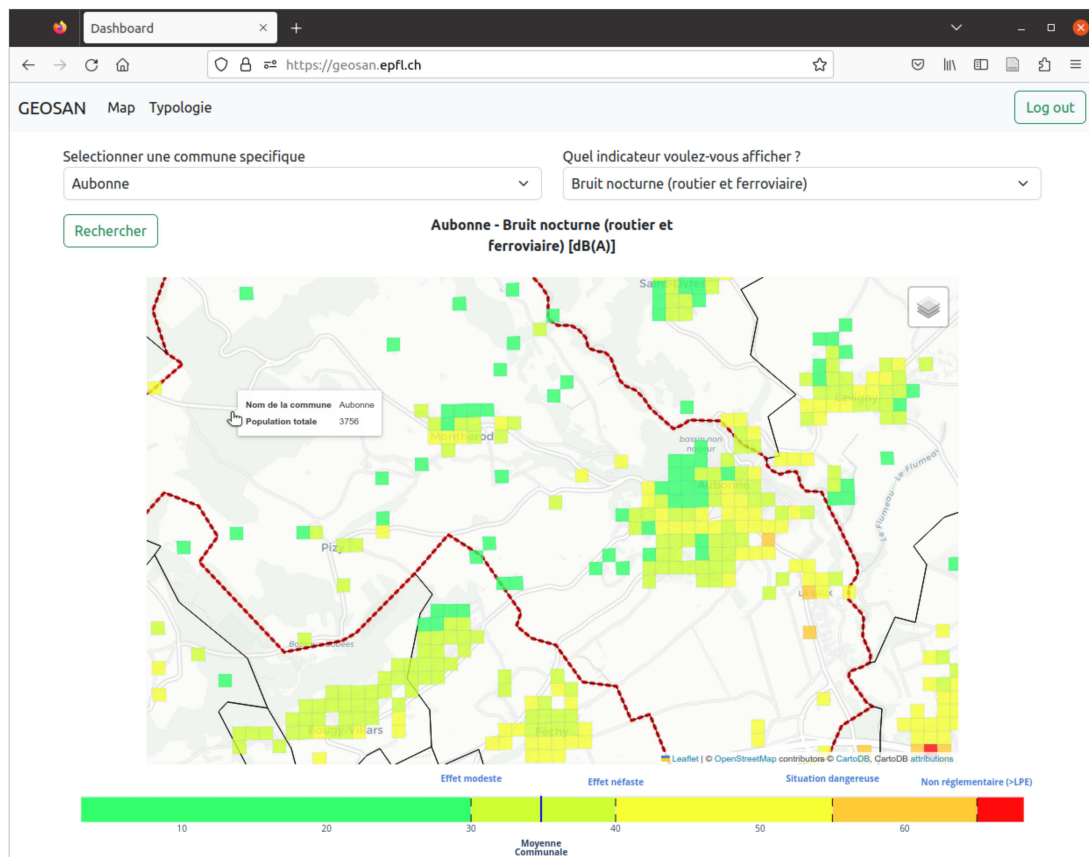


Figure 2.3 – **Interface of the GEOSAN web application.** The application is designed to display hectare-level indicators for a specific municipality along with its neighboring municipalities. Users can select the municipality of interest from the drop-down menu on the left, and select the indicator to be displayed from the drop-down menu on the right.

geographic features, has enriched the ongoing discussions following the external evaluation report of Monin et al. (2020). The creation of diverse SDOH indicators facilitates targeted health promotion measures, aligning with the principle of proportionate universalism and the specific population characteristics.

In this study, a comprehensive set of indicators was compiled using all fine-scale geospatial data relevant to the study area. The goal of building an extensive indicator base was motivated by recurring requests from public health stakeholders for the availability of health-related geographic data. However, this expansive scope comes with certain trade-offs, particularly in terms of indicator accuracy and temporal representation.

While the hectare scale may not always be optimal for interventions or for specific definitions

of environmental exposure, standardizing all indicators at this scale facilitates consistent analysis at small spatial scales and offers versatility for different research objectives. For example, when assessing health needs, one could aggregate groups of hectares, ensuring a minimum of 50 inhabitants to obtain consistent coefficients. In studies of the relationship between air pollution and health, the impact of surrounding hectares could be considered in addition to the average value of a specific hectare.

The formulation of composite indicators was intended to be flexible and adaptable to feedback from public health experts. These indicators sketch a preliminary outline of several dimensions of vulnerability, including material deprivation, environmental burdens, and distance to primary healthcare facilities. However, they are open to subsequent refinement. For example, an improvement to the composite indicator of primary healthcare accessibility might involve the inclusion of additional health services or an adjustment for service capacity relative to local population density. It would also be important to validate these indicators by comparing them to specific health outcomes to ensure their accuracy in reflecting population health variations across the entire territory (Allik et al., 2020). While such data were not available for this research, internal departments may have relevant datasets.

### *Limitations*

We need to address some comments about the origins and potential pitfalls of the data sources used to develop hectare-level indicators. Certain metrics, particularly those describing travel times to health facilities or the amount of green and blue space, are derived from cantonal databases. This origin introduces potential "edge effects" for hectares adjacent to neighboring cantons. Another limitation concerns the time lag between the generation of the data sources and their eventual publication. In particular, air pollution exposures were available for the year 2020; however, these were modeled on 2015 data and therefore do not take into account the potential impact of the pandemic (Heldstab, Schächli, and Künzle, 2020). This temporal misalignment is exacerbated at the hectare-level resolution. Therefore, when conducting territorial diagnosis with municipalities, it is important to prioritize broader spatial patterns over precise hectare values for indicators based on less recent information. The impact of recent urban planning projects should also be taken into account. Lastly, the social and socioeconomic indicators, acquired from the private entity MICROGIS, are fundamentally based on census data and thus accessible through cantonal statistical repositories like Statistiques Vaud. This situation underscores a broader issue: publicly available databases are not always indicative of the full range of data available to cantonal or federal institutions. In this context, the list of indicators presented here serves as a baseline, encouraging authorities to expand, refine, and more seamlessly integrate their internal datasets as deemed necessary.

### *Conclusion*

The typology of Vaud municipalities and the creation of 68 hectare-level indicators, revealed spatial inequalities in SDOH, both between and within municipalities. Our contribution to the "Commune en Santé" program can help cantonal health authorities in proactively recruiting municipalities into the label and supports the deployment of health promotion measures that are aligned with the target population and the principle of proportionate universalism. However, it is critical to note the quantitative nature of these indicators; for a more complete picture of SDOH, these quantitative metrics should be complemented by qualitative methods that include community engagement.

The time constraints of the thesis prevented project iterations, but future revisions are recommended based on feedback from the "Commune en santé" team regarding the application of the developed indicators. Nevertheless, the commitment to reproducibility in the development of fine-scale indicators, the creation of a web application for expanded outreach, and GIS-centered educational initiatives to inform about methodological considerations and potential biases collectively advocate for a successful collaboration between academics and public health practitioners. This work is a step forward in integrating the geographic perspective into the "Commune en Santé" initiative and the broader landscape of cantonal public health programs.

This project highlights the key role of local authorities, such as municipalities, in developing population health interventions, as they are at the forefront of creating health-promoting environments. The geodemographic classification allowed for the identification of certain barriers, including demographic and economic aspects, that may hinder their engagement. However, a significant barrier that could not be captured with spatial data is the broader policy context, which is also an SDOH. Discussions between municipalities and Unisanté have highlighted the lack of a legal framework that gives municipalities a mandate to act in health promotion. Therefore, it is crucial to pass legislation at the federal level that incentivizes and supports municipalities in implementing health promotion measures (Amstutz and Villa, 2021).

The geodata collected in this chapter have made it possible to calculate a comprehensive set of social determinants of health at the level of inhabited hectares. These indicators can then be linked to health outcomes, such as the prevalence of cardiometabolic risk factors, to elucidate how characteristics of the residential environment influence the spatial distribution of cardiovascular disease risk. Such analyses could, in turn, inform the design of population health interventions tailored to local contexts.



### 3 SPATIAL PATTERNS AND FACTORS ASSOCIATED WITH CARDIOVASCULAR RISK FACTORS

In Switzerland, noncommunicable diseases (NCDs) affect 2.2 million people and have an economic impact exceeding 50 billion francs per year (OBSAN, 2015). In addition, 20% of the population aged over 50 years have to cope with multiple chronic diseases, a phenomenon known as multimorbidity. In this context, cardiovascular disease (CVD) emerges as the primary cause of mortality and the third leading cause of hospitalization in the country (OFS, 2023a). To address the escalating burden of CVD, it is imperative to integrate individual-centered strategies, such as health education and early detection of risk factors, with population-level interventions aimed at reducing structural barriers to good health, like material deprivation and environmental exposures (Diez Roux, 2003; OFSP, 2019; Visseren et al., 2021). However, designing effective population health interventions requires a thorough understanding of the spatial distribution of CVD risk factors and of their contextual determinants.

In this chapter, we employed several methods to investigate the spatial pattern and local determinants of four major CVD risk factors - hypertension, obesity, diabetes, and dyslipidemia - collectively referred to as cardiometabolic risk factors (CMRFs). The study was conducted on a population-based cohort of middle-aged and older adults living in Lausanne, the largest city of the canton.

The work presented in this chapter is a preprint version of the article submitted to the *International Journal of Health Geographics*:

Ladoy, A., Marques-Vidal, P., Guessous, I., & Joost, S. (2023). *Identifying hot spots of cardiometabolic risk factors in a Swiss city: Impact of individual and environmental factors*. <https://doi.org/10.21203/rs.3.rs-3359714/v1>

As the first author of the article, I designed the study, processed and analyzed the data, and drafted the manuscript. The co-authors advised on the methods used and critically revised

the manuscript before submission.

### 3.1 Background

Cardiovascular disease (CVD) is a major cause of death and disability in European countries, and there is growing concern about its increasing prevalence in the future (Kluge et al., 2020). The COVID-19 pandemic has further exacerbated the challenges faced by individuals with CVD. Not only do they face an increased risk of severe COVID-19, but they also experience collateral damage from pandemic response measures, including poor management of behavioral risk factors for noncommunicable disease during lockdown and disruptions in access to routine health services (Kluge et al., 2020; Maung and Marques-Vidal, 2023; Mennis, Matthews, and Huston, 2022).

Geographic variations in the prevalence of the most common CVD risk factors, such as diabetes, obesity, hypertension, and dyslipidemia (collectively referred to as cardiometabolic risk factors or CMRFs), have been observed at different spatial scales, and these variations have been associated with characteristics of the social and physical environment (Aguayo et al., 2020; Bentué-Martínez, Mimbreno, and Zúñiga-Antón, 2023; Toms et al., 2019). While obesity has been extensively studied compared to hypertension, diabetes, and dyslipidemia, all CMRFs consistently showed associations with the social environment, with higher risks found in low socioeconomic neighborhoods (Leal and Chaix, 2011). Obesity has also been associated with low neighborhood walkability, low greenness, and physical deterioration of the neighborhood (Ellaway, Macintyre, and Bonnefoy, 2005; Frank et al., 2022). Hypertension has been associated with traffic noise, air pollution, low walkability, and low greenness (Babisch et al., 1993; Brown et al., 2016; De Courrèges et al., 2021; Klompaker et al., 2019). Diabetes has been shown to be associated with road traffic noise, air pollution, low walkability, and low access to green spaces (Frank et al., 2022; Klompaker et al., 2019; Vienneau, Eze, et al., 2019; Y. Zhang et al., 2023). Dyslipidemia has been associated with low greenness and higher urbanization (Brown et al., 2016; De Groot et al., 2019).

These associations between CMRFs and contextual factors may explain the limited effectiveness of individual-centered prevention strategies, which often fail to address environmental and societal barriers to healthy lifestyles (Leal and Chaix, 2011; Hunter et al., 2009). In contrast, population-level interventions, which have been advocated for decades (G. Rose, 1992), focus on modifying the environment to promote lifestyle behaviors and minimize risk factors for the entire population (Visseren et al., 2021). The integration of both individual- and population-based strategies is therefore crucial to effectively reduce the overall prevalence of CVD and address health inequities (G. Rose, 1992; Visseren et al., 2021; Diez Roux, 2003).

Developing population-level interventions requires two key components: identifying high-risk

areas to prioritize interventions and understanding the specific characteristics of the social and physical environment on which to focus prevention efforts. However, achieving this at the local level can be challenging due to the limited availability of fine-scale health data, which are typically only available at the national or regional level (Aguayo et al., 2020; Toms et al., 2019). Furthermore, generalizing findings on health-environment associations is hampered by the heterogeneity of methods and indicators used, as well as by inherent contextual differences between regions and countries (Diez Roux and Mair, 2010; Paquet et al., 2016). Therefore, the aim of this study is to perform a spatial analysis using individual-level data from a large population-based cohort in Lausanne, Switzerland. The study aims to (1) assess the intra-urban spatial variation in cardiometabolic risk factors, (2) investigate the characteristics of the social and physical environment that may explain these variations while controlling for individual confounders, and (3) discuss the implications of these findings for the development of population-level interventions to address CMRFs.

## 3.2 Materials and methods

### 3.2.1 Data

#### Study population

Data were obtained from the CoLaus-PsyColaus population-based study (Firmann et al., 2008), which aims to assess the prevalence and identify the genetic, biological, and environmental determinants of cardiovascular disease in the city of Lausanne, Switzerland (41 km<sup>2</sup> and 140,202 inhabitants in 2021 (OFS, 2023b)). Participants were chosen to be representative of the population aged 35-75 years, following the selection procedure described in Firmann et al. (2008). The study was approved by the Institutional Ethics Committee of the University of Lausanne and complies with the 1964 Helsinki declaration. All participants provided written informed consent. A baseline recruitment phase took place between 2003 and 2006 (N=6733) and was followed by three follow-up studies carried out from 2009 to 2012 (N=5064), 2014 to 2017 (N=4881), and 2018 to 2021 (N=3751). In this study, we used the data from the second follow-up period (2014-2017) as the last follow-up was highly perturbed by the COVID-19 pandemic.

#### Health outcomes

The presence of cardiometabolic conditions (diabetes, hypertension, obesity, and dyslipidemia) was assessed by medical examination. Diabetes was defined as a fasting blood glucose level of  $\geq 7$  mmol/L and/or the use of antidiabetic treatment (Alberti and Zimmet, 1998). Hypertension was identified as a systolic blood pressure  $>140$  mmHg or a diastolic blood

pressure >90 mmHg and/or the use of antihypertensive drug treatment. Obesity was defined as a body mass index (BMI)  $\geq 30$  kg/m<sup>2</sup>, according to World Health Organization guidelines (WHO, 2000). Dyslipidemia was determined by a total cholesterol level >6.5 mmol/L and/or an LDL-cholesterol level >4.1 mmol/L, and/or the use of hypolipidemic medication (Abolhassani et al., 2017).

#### Individual factors

Demographic (age, sex, marital status, country of birth), socioeconomic (education, poverty), and behavioral risk factors (smoking status, alcohol consumption, physical inactivity) were assessed using validated questionnaires administered to the study participants. Home addresses provided by the participants at the time of the examination were geocoded by matching with the official Swiss address directory<sup>1</sup> (OFS, 2023c). Marital status was classified as married/cohabiting or living alone. Country of birth was categorized as "born in Switzerland" or "not born in Switzerland". Education level was divided into low (obligatory school or apprenticeship), medium (high school), or high (university degree). Poverty was considered for individuals who answered positively to "Are you experiencing real financial difficulties to meet needs?". Smoking status was classified as current or former/never smoker. Hazardous drinking was considered for participants consuming over 14 units per week (one unit=one glass of wine, one can of beer or one shot of spirit). Physical inactivity was defined as having a sedentary lifestyle, assessed using Physical Activity Frequency Questionnaires (PAFQ) (Verhoog et al., 2019).

#### Contextual variables

To assess the influence of the place of residence on CMRFs, we considered several indicators of the social and physical environments that have previously been associated with diabetes, hypertension, obesity, or dyslipidemia. Environmental indicators included greenness, street connectivity, nighttime traffic noise levels, and exposure to fine particulate matter (PM<sub>2.5</sub>) and nitrogen dioxide (NO<sub>2</sub>).

Greenness was quantified by calculating the proportion of green space within a 500-meter radius buffer around each hectare centroid using land cover data for the year 2015 (Source: Géodonnées Etat de Vaud). The choice of a 500-meter buffer was justified by its common use in assessing walkability in European urban areas (De Courrèges et al., 2021; Feuillet et al., 2015), corresponding to places accessible within a 5–10-minute walk. Street connectivity was measured by counting the number of intersections with three or more legs within the same buffer. The walkable street network from January 2016 was retrieved from OpenStreetMap

---

<sup>1</sup>Details of the geocoding algorithm used in this thesis are provided in Appendix, Figure A.1.2.

using the Overpass API, excluding private roads and nonrelevant intersections (parking aisles, service driveways, etc.).

Nighttime traffic noise exposure was obtained from the 2015 sonBASE dataset (10x10 m resolution) (OFEV, 2018), and the concentrations of PM<sub>2.5</sub> and NO<sub>2</sub> were extracted from the 2015 Pollumap immission model developed by MeteoTest (20x20 m resolution) (Heldstab, Schäppi, and Künzle, 2020). We then calculated the mean exposure in noise (dB) and air pollutants (µg/m<sup>3</sup>) in the 500-meter buffers around hectare centroids.

Social indicators included the proportion of the population with compulsory education, median neighborhood income, unemployment rate, and the proportion of non-Swiss citizenship. These indicators were obtained from MICROGIS for Swiss neighborhoods in 2015 (MICROGIS, 2023). All neighborhood factors were computed using Python and QGIS at the resolution of inhabited hectares (100x100 m cells) aligning with the scale used by the Swiss Federal Statistical Office for population statistics (SFSO, 2017). The assignment of contextual factors to ColaUS-PsyColaUS participants was achieved by spatial intersection.

### 3.2.2 Methods

#### Spatial relative risk of cardiometabolic risk factors

To assess the spatial variation in CMRFs across the study area, we compared the spatial density estimations between cases (i.e., individuals with the disease) and controls (i.e., non-cases) by modeling the individual's home locations as point patterns following a heterogeneous Poisson distribution (Lawson, 2006; Waller and Gotway, 2004). Random perturbations and edge correction techniques were applied to address potential issues with duplicate point locations and boundary effects (Davies, Marshall, and Hazelton, 2018).

The log relative risk surface, obtained by taking the ratio of the density functions of cases and controls, allowed the identification of high-risk and low-risk areas, indicating areas with significantly elevated or reduced probabilities of observing a case, respectively (Waller and Gotway, 2004). We also assessed the overall clustering of CMRFs across the study area using the method proposed by Kelsall and Diggle (1995b). The choice of the kernel form and, more importantly, the kernel bandwidth play a critical role in the estimation of spatial relative risk as it determines the degree of smoothing applied to the risk surface (see Figure A.3.1 in Appendix for bandwidth comparisons). A larger bandwidth might mask small-scale patterns and reduce local variations, whereas a bandwidth that is too small could introduce noise and overfitting. For this study, we selected a circular Gaussian kernel with a bandwidth of 200 meters for both cases and controls. This choice represents a compromise between the jointly optimal bandwidths proposed by Kelsall and Diggle (1995a) and Davies (2013).

### Chapter 3. Spatial patterns and factors associated with cardiovascular risk factors

---

Statistical inference of the relative risk surface involved testing the random labeling hypothesis through 999 Monte Carlo permutations. In each permutation, the labels (cases and controls) were randomly assigned to point locations. Tolerance contours were then derived from the resulting p-value surface, defining high-risk and low-risk areas at significance levels of 5% and 1%.

To ensure the robustness of the findings, the spatial patterns of CMRFs observed in the study sample (i.e., second follow-up of the ColaUS-PsyColaUS study) were compared with those observed at baseline and at the first follow-up. This comparison aimed to address concerns about the representativeness of the cohort and to identify consistent high-risk areas over a 10-year period.

As an initial exploration prior to the modeling phase, we analyzed differences in participant characteristics and contextual factors between areas with significantly elevated risk, reduced risk, or a neutral risk pattern (where the observed variation is not statistically significant) of CMRFs. Pairwise comparisons of study participants and inhabited hectares located in high-risk, low-risk, and neutral areas were performed using chi-square and Kruskal-Wallis tests for categorical and continuous variables, respectively. A Bonferroni correction was applied to p-values to account for multiple comparisons.

Analyses were performed in R using the *sparr* (Davies, Hazelton, and Marshall, 2011) and *smacpod* (French, 2022) packages.

#### Regression models

To explore the associations between CMRFs — hypertension, diabetes, obesity, and dyslipidemia — and their interactions with both physical and social environments, we employed a two-stage regression approach. This approach was designed to isolate the effects of contextual factors from known individual confounders, addressing the concern that observed health-environment relationships could be solely attributable to differences in individual characteristics (Leal and Chaix, 2011).

##### *Stage one - Logistic regression*

In the first stage, logistic regression was used to adjust the binary outcomes of CMRFs for key individual confounders: age, sex, poverty status (yes vs. no), education level (low, medium, high), and country of birth (Swiss-born vs. foreign-born). To address nonlinear relationships with conditions such as diabetes, obesity, and dyslipidemia, quadratic terms for age were incorporated. The Pearson residuals, representing the standardized differences between observed occurrences and model-estimated probabilities, captured the unexplained variation in CMRFs attributable to individual-level factors (Fleury et al., 2021).

#### *Stage two - Geographic Weighted Regressions*

The second stage involved investigating associations between contextual factors and CMRF variations adjusted for individual factors. This was achieved through Geographically Weighted Regression (GWR) analyses, using the Pearson residuals of the logistic regression as dependent variables.

Traditional regression models such as Ordinary Least Squares (OLS) assume spatial stationarity, that is constant relationships between outcomes and explanatory variables across space. However, this assumption may not hold true when dealing with spatial processes. GWR were developed to address spatial heterogeneity by fitting a regression equation at each observation location (Fotheringham, Brunsdon, and Charlton, 2002). It estimates model parameters based on nearby data points, weighting each point by its distance from the regression location, typically using a bisquare kernel. This approach contrasts with OLS, where all points in the study area are weighted equally. The multiscale extension of GWR, i.e., MGWR, further allows unique bandwidths for each explanatory variable. This recognizes that spatial processes may operate at different scales.

Initially, OLS models were calibrated to explore the global relationships between adjusted CMRFs and the contextual factors previously described: street connectivity, greenness, traffic noise, PM2.5 and NO2 exposure, median income, and unemployment rate. To identify potential multicollinearity issues, we evaluated the Variance Inflation Factor (VIF). Furthermore, the spatial autocorrelation of the residuals was assessed using the global Moran's I index, employing a Gaussian kernel with a bandwidth of 200 meters.

Subsequently, GWR and MGWR analyses were conducted. For these local models, kernel bandwidths were determined using a golden search method, with the objective of minimizing the Corrected Akaike Information Criterion (AICc). A bisquare adaptive kernel was selected to adjust for variations in point density. The models provided coefficient estimates, t-values, and local model diagnostics for each regression point. Consequently, we mapped both the coefficient estimates for each contextual factor and the locations where associations were statistically significant. These significant associations were identified based on the t-values obtained from the models, incorporating a correction for multiple comparisons<sup>2</sup> (T. Oshan et al., 2019).

Model comparisons relied on the AIC criterion, with a reduction in AIC values greater than 2 indicating substantial model improvement. The regression modeling steps were performed

---

<sup>2</sup>Significant observations are those where the absolute value of the t-value is larger than the critical t-value. For GWR, the critical t-value is derived from an adjusted alpha level, defined as  $\alpha = \frac{\xi}{ENP/p}$ , where  $\xi$  is the desired joint type I error rate (commonly 0.05), and the ratio  $\frac{ENP}{p}$  represents the effective number of tests. For MGWR, an adjusted alpha level for the  $j$ th model term is calculated using  $\alpha_j = \frac{\xi}{ENP_j}$ , with  $ENP_j$  being the effective number of parameters for the  $j$ th term.

using R (R Core Team, 2023) and the mgwr (T. Oshan et al., 2019) and PySAL (S. J. Rey and Anselin, 2007) libraries in Python.

### 3.3 Results

Out of the 4,881 participants included in the study during the second follow-up period (2014-2017), 61 individuals (1.25%) could not be geocoded, and 1,117 individuals (22.88%) were excluded because they resided outside the urban districts of Lausanne. Additional exclusions were made for participants with missing individual confounder data (N=4, 0.08%) and missing neighborhood characteristics (N=4, 0.08%). Thus, 3,695 participants (75.70%) were included in the final analysis (Figure A.3.2).

The mean age of the participants was  $64.1 \pm 10.5$  years, and 56.8% were females. The prevalence of CMRFs in the study sample was as follows: 48.2% were hypertensive, 17.7% were obese, 10.7% had diabetes, and 33.2% had dyslipidemia. Detailed information on outcome variables and participant characteristics, including demographic, socioeconomic, and behavioral factors, is presented in Tables A.3.1 and A.3.2. Table A.3.1 also provides comparative statistics between the study sample, the baseline population, and the population excluded from the study.

Regarding the characteristics of the physical and social environments, all variables displayed spatial variation across the city (see Figure A.3.3). Within the 500-meter radius buffer, the mean values of environmental characteristics were as follows: intersection density was  $1.0e-4$  (std:  $5.0e-5$ ), proportion of green spaces was 52.2% (std: 18.1), nighttime traffic noise was 41.6 dB (std: 3.67), PM2.5 concentration was  $10.5 \mu\text{g}/\text{m}^3$  (std: 0.32), and NO2 concentration was  $22.4 \mu\text{g}/\text{m}^3$  (std: 2.07). Additionally, the average median income per inhabited hectare was 41.2 kCHF/year (std: 19.2), (1 CHF=1.04 € or 1.14 US\$ as of 31st July 2023) the unemployment rate was 3.97% (std: 2.64), the proportion of the population with compulsory education was 20.2% (std: 12.2), and the proportion of the foreign population was 40.5% (std: 17.2).

#### 3.3.1 Intra-urban variation in cardiometabolic risk

Log relative risk surfaces of CMRFs are shown in Figure 3.1<sup>3</sup>, along with the delineation of areas presenting significantly elevated risk, at the 5% (dashed line) and 1% (solid line) significance levels. The delineation of areas with high-risk, low-risk and neutral-risk is shown in Figure A.3.4.

---

<sup>3</sup>For further geographic context on the spatial risk prediction of hypertension, obesity, and diabetes, consult Figures A.3.17, A.3.18, and A.3.19, respectively.



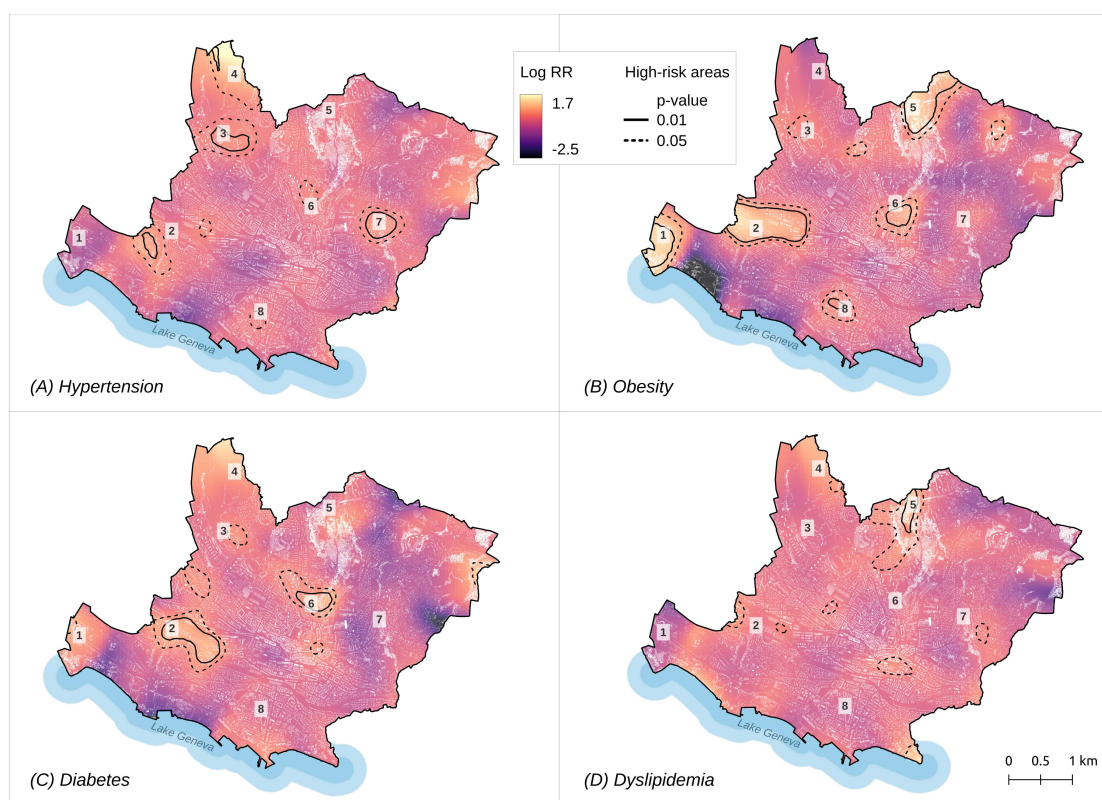


Figure 3.1 – **Log relative risk (Log RR) surface of CMRFs: (A) hypertension, (B) obesity, (C) diabetes, and (D) dyslipidemia.** Density estimations were performed using Gaussian kernels with a Bandwidth of 200 meters. High-risk areas were determined by tolerance contours (Tol. Intervals) based on 999 Monte Carlo permutations. Indicative landmarks are shown on the map to help interpret the results (#1-#8). The basemap layer consists of a Digital Heights Model (DHM) from 2012 (Géodonnées Etat de Vaud, 2012), and the maps were generated using QGIS (v. 3.22.16). The log relative risk surfaces for hypertension, obesity, and diabetes, along with geographic contextual elements to aid in intervention planning, are shown in Figures A.3.17, A.3.18, and A.3.19, respectively.

### Hypertension

Hypertension exhibited spatial variation across the city, with four large areas exceeding the 95% tolerance interval, in landmarks #2, #3, #4, and #7 (Figure 3.1 (A)). The test for global clustering was also significant, with a p-value of 0.04. High-risk areas were characterized by lower income and education levels than low-risk and neutral areas (Figure A.3.5). We also observed lower street connectivity and, surprisingly, higher greenness and lower environmental exposure (nighttime noise and air pollution) compared to low-risk areas, but the difference was not statistically significant compared to neutral areas. Study participants in high-risk areas were older ( $p=0.005$ ), had lower levels of education ( $p<0.001$ ), and were more likely to be foreigners ( $p=0.013$ ) and unemployed ( $p=0.029$ ).

### Obesity

Obesity showed a global clustering pattern ( $p=0.03$ ), and we identified five local clusters exceeding the 99% tolerance interval (Figure 3.1 (B)), in landmarks #1, #2, #5, #6, and #8. High-risk areas were associated with higher socioeconomic deprivation and a higher proportion of foreign population than low-risk and neutral areas (Figure A.3.6). While there were no significant variations in walkability, high-risk areas had statistically higher environmental exposures (nighttime noise and air pollution). In addition, study participants in high-risk areas were younger ( $p<0.001$ ) and more likely to be foreigners ( $p<0.001$ ), to live alone ( $p=0.002$ ), to have a low level of education ( $p<0.001$ ), and to experience financial difficulties ( $p<0.001$ ).

### Diabetes

Diabetes demonstrated spatial variation across the study area, with two local clusters detected at a significance level of 1% in landmarks #2 and #6 (Figure 3.1 (C)). There was no evidence of global clustering ( $p=0.61$ ). High-risk areas had lower walkability and higher environmental exposures compared to low-risk and neutral areas (Figure A.3.7). Inhabited hectares in high-risk areas were also characterized by a lower median income, a higher proportion of the population with compulsory education, and a higher proportion of foreign population. Study participants in high-risk areas were more likely to have lower levels of education ( $p<0.001$ ), higher unemployment rates ( $p=0.020$ ), and higher financial difficulties ( $p=0.004$ ).

### Dyslipidemia

Dyslipidemia exhibited a less pronounced spatial pattern than the other CMRFs, but one local cluster in the northeastern part of the city (landmark #5) was significant at the 1% significance level (Figure 3.1 (D)). There was no evidence of global clustering ( $p=0.37$ ). High-risk areas

were characterized by lower median income, higher unemployment rate, and higher nighttime traffic noise and NO<sub>2</sub> exposure compared to other risk areas (Supplementary Figure A.3.8). No significant differences were observed between study participants, except for educational level ( $p < 0.001$ ).

The spatial pattern of CMRFs was generally consistent with the patterns observed at baseline and the first follow-up study period (Figures A.3.9-A.3.11), highlighting the long-term persistence of high-risk areas despite individuals leaving the cohort.

### 3.3.2 Impact of individual confounders on cardiometabolic risk factors

The results of the logistic regression adjusting CMRFs for individual confounders are shown in Table 3.1. Together, age, sex, poverty, education, and country of birth explained 11%, 4%, 12%, and 6% of the variation in hypertension, obesity, diabetes, and dyslipidemia, respectively. Males were more likely than females to have hypertension, obesity, diabetes, and dyslipidemia. Individuals with lower education levels, older age, and those born outside Switzerland were also at higher risk for these CMRFs. Although individuals with financial difficulties had an increased risk of hypertension, obesity, and diabetes, the association with dyslipidemia was not statistically significant.

	Hypertension	Obesity	Diabetes	Dyslipidemia
Age (Linear term)	1.09 (1.08-1.09) ***	1.20 (1.07-1.34) **	1.43 (1.23-1.68) ***	1.26 (1.15-1.38) ***
Age (Quadratic term)	n/a	~1 **	~1 ***	~1 ***
Male (Ref. Female)	2.01 (1.73-2.33) ***	1.36 (1.14-1.62) ***	3.23 (2.57-4.06) ***	1.31 (1.13-1.52) ***
Financial difficulties (Ref. No)	1.26 (1.06-1.49) **	1.56 (1.29-1.90) ***	2.13 (1.68-2.69) ***	1.12 (0.95-1.33)
Education (Ref. Low)				
Medium	0.73 (0.61-0.87) ***	0.65 (0.52-0.80) ***	0.72 (0.55-0.94) *	0.66 (0.55-0.79) ***
High	0.58 (0.47-0.70) ***	0.41 (0.32-0.53) ***	0.45 (0.32-0.63) ***	0.53 (0.43-0.65) ***
Born in Switzerland (Ref. No)	0.84 (0.72-0.98) *	0.71 (0.59-0.85) ***	0.76 (0.60-0.96) *	0.81 (0.69-0.94) **
Pseudo R <sup>2</sup>	0.11	0.04	0.12	0.06
AIC	4303	3228	2178	4204

Table 3.1 – **Logistic regression results for each CMRF** Results are expressed as Odds Ratio and (95% confidence interval). Pseudo R<sup>2</sup> is based on McFadden's criteria, which is a commonly used goodness-of-fit measure for logistic regression models. Ref.: reference category. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

### 3.3.3 Associations with the physical and social environments

The results of the OLS model, shown in Supplementary A.3.3, show a significant negative association between median neighborhood income and adjusted CMRFs, specifically for hypertension, obesity, and diabetes. However, no significant associations were observed for adjusted dyslipidemia. The detection of weak but significant positive spatial autocorrelation

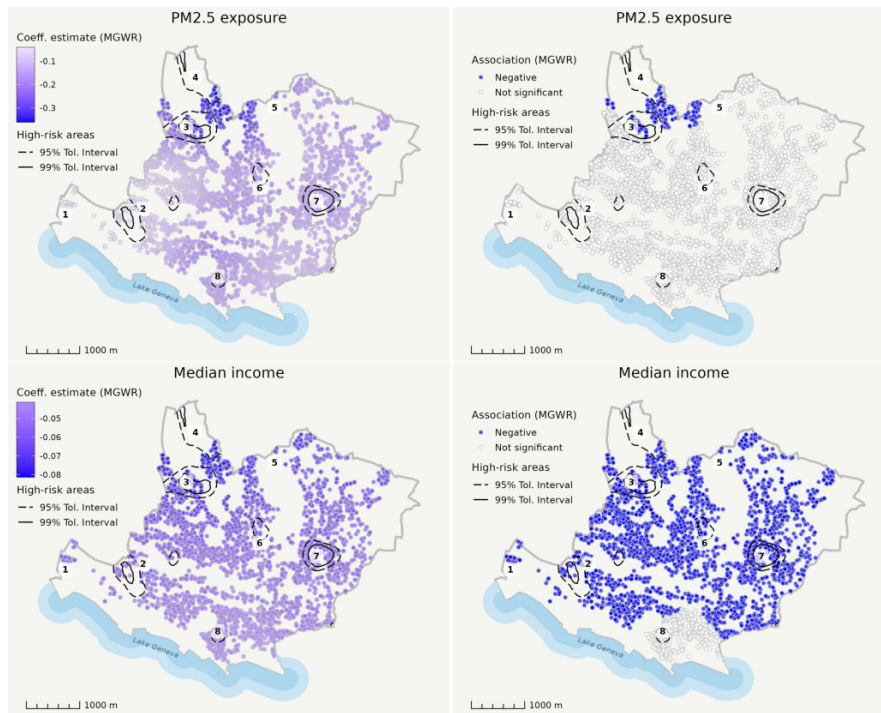
in the residuals, as indicated by Global Moran's I (ranging from 0.026 to 0.041), suggests that the OLS model is inadequate to fully capture the associations between adjusted CMRFs and neighborhood characteristics. This observation justifies the need for spatial analyses, leading to the application of both GWR and MGWR models. Note that the VIF values for each CMRF remained below the common threshold of 10 (O'Brien, 2007), suggesting minimal multicollinearity concerns. Accordingly, all variables were retained for the spatial regression analyses.

#### Hypertension

The MGWR model demonstrated superior performance in capturing the relationships between adjusted hypertension and contextual factors, as evidenced by its lower AIC (10,135) compared with the OLS (AIC=10,139) and GWR (AIC=10,141) models (Table A.3.4). In both spatial models, the local condition numbers remained below 15, indicating no multicollinearity issues (Figure A.3.16). The bandwidths used in the MGWR model were similar to those used in the GWR model, except for PM<sub>2.5</sub> exposure (1137 neighbors for MGWR bandwidth vs. 3568 for GWR, Table A.3.5). Contextual factors that significantly explained the spatial variation in adjusted hypertension are shown in Figure 3.2 for the MGWR model and in Figure A.3.13 for GWR. Neighborhood median income displayed a small negative association with hypertension, and this association was statistically significant throughout the city except in the southeastern region. In addition, we observed a negative local association between PM<sub>2.5</sub> exposure and adjusted hypertension, which was significant only in the MGWR model and overlapped the high-risk areas in landmarks #3 and #4.

#### Obesity

The MGWR model (AIC=9648) outperformed both the OLS (AIC=9676) and GWR (AIC=9654) models (Table A.3.4). The map of local condition numbers (Figure A.3.16) indicated that certain locations in the GWR model exceeded the common threshold of 30, suggesting potential multicollinearity problems in these areas. In addition, the different bandwidths used in the MGWR model (ranging from 1010 to 3409 neighbors, Table A.3.5) suggest that neighborhood characteristics influence obesity at different scales. For example, greenness, which was modeled using a bandwidth of 1010 neighbors, appeared to influence obesity in a more localized manner compared to neighborhood median income, which was modeled using a bandwidth of 2792 neighbors. This more regional influence of neighborhood income likely reflects broader socioeconomic patterns that influence lifestyle and health behaviors. The left panels of Figure 3.3 illustrate the spatial variation in the MGWR coefficient estimates. Neighborhood income and unemployment rate showed respectively negative and positive associations with obesity risk across the city, albeit with modest effects (coefficient estimates



**Figure 3.2 – Local contextual determinants of hypertension.** The figure shows the results of Multivariate Geographically Weighted Regression (MGWR) highlighting local associations between adjusted hypertension and characteristics of the physical and social environments. Maps in the left column show coefficient estimates reflecting the magnitude of associations with contextual factors, while the right column differentiates these associations by statistical significance: white dots for non-significant, blue for negative, and red for positive associations. Solid and dashed lines delineate high-risk areas for hypertension, as identified from the log relative risk surface. Indicative landmarks are shown on the maps to help interpret the results (#1-#8). Only determinants with statistically significant variation based on corrected t-values are shown. For privacy reasons, the Digital Heights Model (DHM) base map layer was intentionally excluded.

between -0.1 and 0.1). These spatial associations were regionally significant (right panels of Figure 3.3) and overlapped with high-risk areas identified in the log relative risk surface, specifically in landmarks #1, #2, #3, and #5 for median income, and #1, #2, #6, and #8 for unemployment. Both greenness and the proportion of compulsory education showed spatial heterogeneity across the city, with their associations changing sign depending on the area. In the south, a moderate local negative association between greenness and obesity risk was observed, partially overlapping with the high-risk area located in landmark #2. The proportion of compulsory education was positively associated with obesity risk in the eastern high-risk area (landmark #1). In contrast to the MGWR model, the GWR model (Figure A.3.14) did not reveal a significant association between education level and adjusted obesity but showed a significant positive association with greenness in the northern areas (landmark #5) and a negative association with street connectivity in landmark #1.

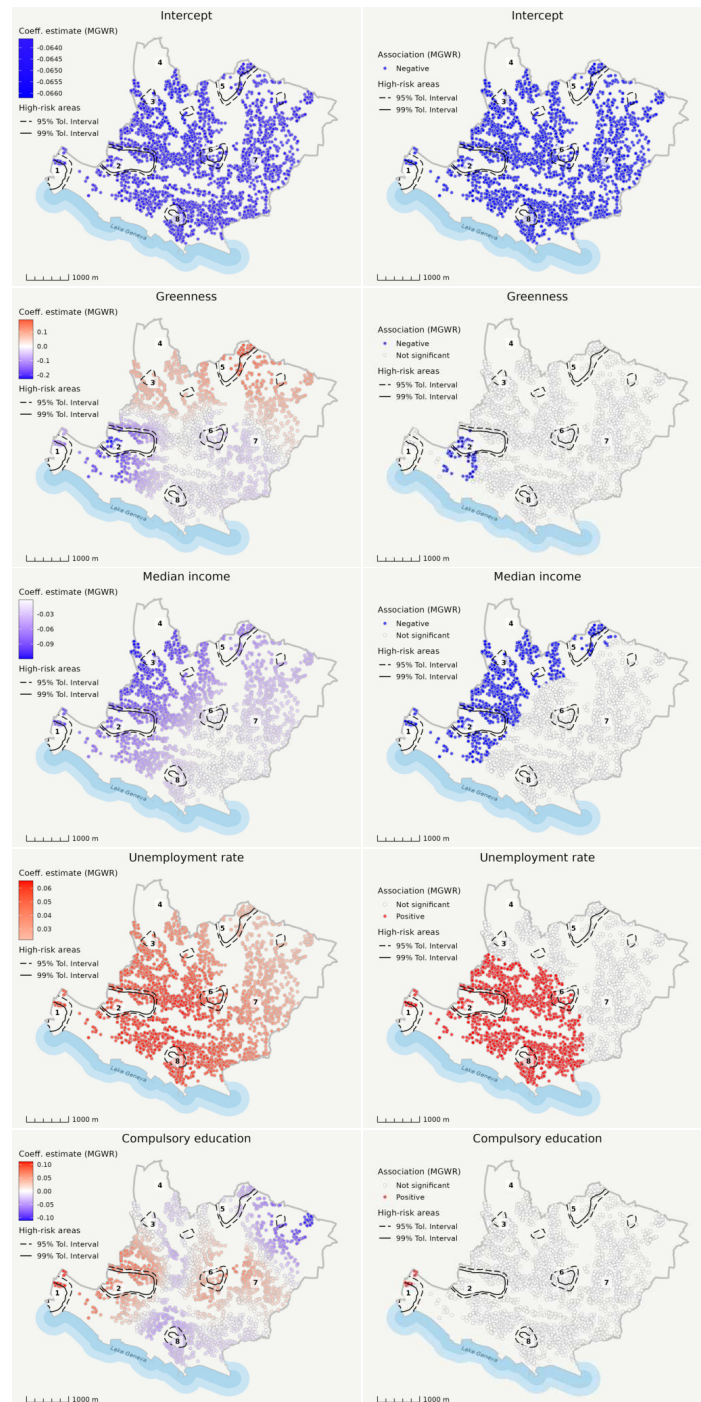
#### Diabetes

For adjusted diabetes, a modest increase was observed in the proportion of explained variance between the global and spatial models (GWR and MGWR), but model fit did not show improvement (Table A.3.4). The MGWR model (AIC=9742) slightly outperformed the GWR model (AIC=9745). Local multicollinearity was low across the study area (Figure A.3.16), and all explanatory variables operated at the global scale except for median income (659 neighbors for MGWR bandwidth vs. 3458 for GWR). Significant contextual associations with diabetes were exclusively related to social characteristics (Figure 3.4, Figure A.3.15). Both spatial models showed significant negative associations between neighborhood income and obesity in the northern areas. However, the coefficient estimate surface showed moderate spatially heterogeneous associations in the MGWR model and small negative associations in the GWR model, which can be explained by the different process scales considered in the two models. The proportion of the foreign population showed a slight negative association in both GWR and MGWR, with locally significant associations in the south of the city (landmark #8).

#### Dyslipidemia

As suggested by the global regression model, the GWR and MGWR models corroborated the negligible effect of the physical and social environments on adjusted dyslipidemia across our study area. This was highlighted by the lack of improvement in model performance and the absence of significant local associations for all contextual factors (Table A.3.4).

**Figure 3.3 – Local contextual determinants of obesity.** The figure shows the results of Multivariate Geographically Weighted Regression (MGWR) highlighting local associations between adjusted obesity and characteristics of the physical and social environments. Maps in the left column show coefficient estimates reflecting the magnitude of associations with contextual factors, while the right column differentiates these associations by statistical significance: white dots for non-significant, blue for negative, and red for positive associations. Solid and dashed lines delineate high-risk areas for obesity, as identified from the log relative risk surface. Indicative landmarks are shown on the maps to help interpret the results (#1-#8). Only determinants with statistically significant variation based on corrected t-values are shown. For privacy reasons, the Digital Heights Model (DHM) base map layer was intentionally excluded.





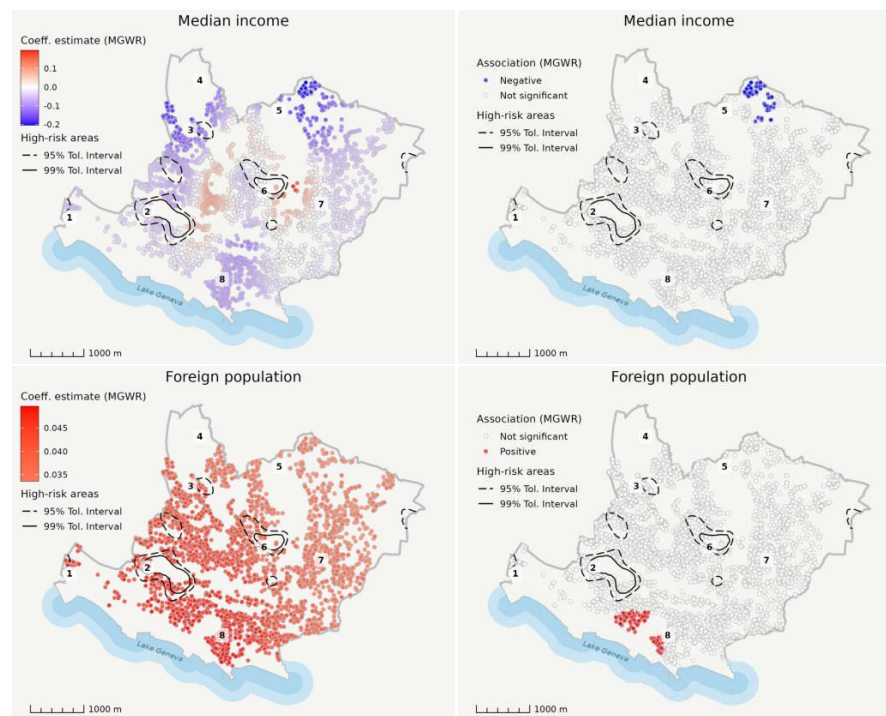


Figure 3.4 – **Local contextual determinants of diabetes.** The figure shows the results of Multi-variate Geographically Weighted Regression (MGWR) highlighting local associations between adjusted diabetes and characteristics of the physical and social environments. Maps in the left column show coefficient estimates reflecting the magnitude of associations with contextual factors, while the right column differentiates these associations by statistical significance: white dots for non-significant, blue for negative, and red for positive associations. Solid and dashed lines delineate high-risk areas for diabetes, as identified from the log relative risk surface. Indicative landmarks are shown on the maps to help interpret the results (#1-#8). Only determinants with statistically significant variation based on corrected t-values are shown. For privacy reasons, the Digital Heights Model (DHM) base map layer was intentionally excluded.



### 3.4 Discussion

Our study aimed to investigate the spatial variation in cardiometabolic risk factors (CMRFs) in an urban context and to explore the role of the physical and social environments in shaping these patterns while considering individual confounders. The results revealed significant spatial patterns in CMRFs as well as several local associations with the environment.

#### 3.4.1 Intra-urban variation in cardiometabolic risk factors

We observed substantial spatial variation in CMRFs across Lausanne, with obesity exhibiting the most pronounced pattern, followed by hypertension and diabetes. The log relative risk of these cardiometabolic outcomes exceeded the 99% tolerance interval in several areas, and we also observed a global clustering pattern for obesity and hypertension. In particular, our investigation revealed a higher risk of obesity in the western areas of the city (landmarks #1, #2, #3). This finding is consistent with previous research by Joost, Duruz, et al. (2016), who reported a striking westward gradient in BMI for the baseline and first follow-up of the same medical cohort, although the magnitude of the gradient was somewhat attenuated in our study.

Furthermore, our study identified overlapping high-risk areas associated with multiple cardiometabolic risk factors. Specifically, the neighborhood in landmark #2 had a concurrent elevated risk for obesity and diabetes, whereas the neighborhood in landmark #5 had an elevated risk for obesity and dyslipidemia. These findings suggest that targeted interventions should be prioritized in these areas, as they may indicate a higher risk of multimorbidity.

#### 3.4.2 Associations with the physical and social environments

The associations between cardiometabolic risk factors and characteristics of the physical and social environments were examined using geographically weighted regressions (GWR) on health outcomes previously adjusted for known individual confounders, including age, sex, education, country of birth, and poverty. For obesity, our analysis revealed significant local positive associations with unemployment rate and compulsory education, and negative associations with neighborhood income, even after controlling for individual socioeconomic status. These findings are in line with the literature, which consistently highlights the association between weight status and neighborhood socioeconomic position (Leal and Chaix, 2011). We also observed a significant negative local association between obesity and greenness in the southern areas, consistent with other studies (Ellaway, Macintyre, and Bonnefoy, 2005; Browning and K. Lee, 2017).

Hypertension displayed a negative global association with neighborhood income, which

was significant in the eastern and northern areas in our spatial models. This association between hypertension and socioeconomic deprivation is consistent with findings from other studies (Durfey et al., 2019; Wagner et al., 2016). However, it is interesting to note that while several variables were included in our study to characterize the socioeconomic position of the neighborhood, only neighborhood income emerged as a significant negative association with hypertension. This differs from the study by Chaix et al. (2010), where neighborhood education, not income, was identified as a significant factor. Surprisingly, we also found a local negative association between the concentration of PM<sub>2.5</sub> and adjusted hypertension, which might suggest potential mediating interactions with another environmental characteristic. It is worth noting that the neighborhoods in question (landmarks #3 and #4) are situated next to a small civil airport, and our variable for nighttime traffic noise did not include aircraft sources. This could provide a clue for further investigations, as associations between aircraft noise exposure and hypertension have been observed in other studies (Evrard et al., 2017; Petri et al., 2021).

Our study did not detect any local association between diabetes and the characteristics of the physical environment, in contrast with previous studies that reported significant associations between diabetes and walkability (Frank et al., 2022; Den Braver et al., 2018). These variations in findings between studies may be attributed to the use of different measures to assess greenness and street connectivity. However, we observed a significant negative association between adjusted diabetes and neighborhood income in the northern areas and a positive association with the proportion of the foreign population in the southern areas, which highlights the importance of addressing socioeconomic disparities in diabetes (Durfey et al., 2019; Kauhl et al., 2016).

Dyslipidemia showed no significant local associations with characteristics of the social and physical environments. Previous literature on the association between blood lipids and the built environment has often emphasized urban-rural differences, with higher blood lipid levels observed in urban areas (De Groot et al., 2019). Given that our study area was limited to an urban context, it is possible that the associations between dyslipidemia and the environment occur at a broader spatial scale, and thus we may not have sufficient spatial variation within the city to detect them.

Spatial associations between CMRFs and the contextual factors showed heterogeneity across the city. Some associations were consistent in direction but varied in magnitude, while others shifted in sign between areas. This underscores the value of geographically weighted regressions for capturing these intricate, local variations. Interestingly, similar to the findings of Kauhl et al. (2016) in northeast Germany, we observed a varying directional association between neighborhood income and diabetes across the city.

We recognized the importance of the multiscale extension of geographically weighted regres-

sion (MGWR), particularly for disentangling associations between obesity and the physical and social environments. The GWR model encountered problems with local multicollinearity and inadequately captured the multiple process scales among contextual factors. For diabetes and hypertension, where associations were predominantly global and no multicollinearity problems were identified, the multiscale extension was less critical. Nevertheless, while insights can still be gained from interpreting the GWR results, greater reliance should be placed on the results of the MGWR model due to its improved performance and better handling of multiscale effects (T. M. Oshan, Smith, and Fotheringham, 2020).

It is noteworthy that although our study revealed substantial differences in contextual factors between high-risk, low-risk, and neutral areas on the log relative risk surfaces for hypertension, diabetes, and obesity, these associations were not fully reflected in the GWR and MGWR models, except for obesity. This divergence could be explained by the fact that the spatial variation in diabetes and hypertension was largely attributable to individual-level characteristics rather than contextual factors.

#### 3.4.3 Implications for the development of public health policies

This study revealed a spatial structure of CMRFs across the city. Using GWR and MGWR models, our analysis revealed significant associations between neighborhood characteristics and local variations in obesity, hypertension, and diabetes. These findings highlight the critical role of place-based processes in generating health disparities related to CVD risk and argue for the design of interventions that are uniquely tailored to the distinctive population- and place-specific circumstances.

By pinpointing the regions where interventions are most needed and identifying potential local drivers of health determinants, our findings lay a foundation for Population Health Intervention Research (PHIR). PHIR focuses on investigating policy and program interventions, both within and outside the health sector, with the aim of reducing health problems and addressing health disparities (Hawe and Potvin, 2009).

In examining the specific local associations within areas of significantly elevated risk for CMRFs, we observed a consistent negative association with neighborhood median income in almost all high-risk areas for hypertension. For obesity, the large high-risk area in landmark #2 was characterized by a negative association with median income, a positive association with unemployment rate, and a negative association with greenness. In addition, the high-risk area for obesity in landmark #1 showed a negative association with neighborhood income and a positive association with the proportion of compulsory education. Although not statistically significant in the MGWR model, the GWR model also indicated a negative association with street connectivity, which should be considered in future policy planning, especially given the

proximity to surrounding highways.

It should be noted that the associations found with environmental or social factors do not imply the same type of interventions. Addressing social factors requires policy changes at a broader level, for example using empowerment measures or income support (Weber, 2020). The associations found with environmental characteristics could be implemented at a more local level through urban planning decisions. For example, improving street connectivity in a neighborhood by providing multiple route options and thus reducing travel distances could promote an environment for active mobility, which in turn could reduce the risk of obesity (F. Wang, M. Wen, and Y. Xu, 2013). Similarly, the development or enhancement of green spaces could promote physical activity and social interaction (A. Lee, Jordan, and Horsley, 2015). Future PHIR studies will be instrumental in determining the appropriateness and effectiveness of such interventions.

#### 3.4.4 Strengths and limitations

Our study has several strengths. First, we conducted our analysis using individual data from a population-based cohort, which provides a robust basis for assessing spatial variation in CMRFs. The use of health outcomes assessed during medical examinations, rather than relying on self-reported data, helps to mitigate common biases in epidemiological studies (Zellweger, Bopp, et al., 2014). Through the analysis of georeferenced home addresses of study participants, we mitigated inherent biases in ecological studies, such as the modifiable unit areal problem (Openshaw, 1984). Moreover, by adjusting for important individual-level confounders, such as age, sex, socioeconomic status, and country of birth, we reduced the likelihood that observed associations between neighborhood characteristics and health outcomes were solely attributable to unaccounted individual-level factors (Diez Roux, 2003).

Another strength is that we studied four major metabolic risk factors for CVD simultaneously, allowing comparisons of their respective geographic patterns. We also attempted to capture several dimensions of the social and physical environments that may be associated with CMRFs (Diez Roux, 2003; Diez Roux and Mair, 2010).

Our study benefits from the acquisition of environmental characteristics for the year 2015, ensuring temporal alignment with the study period, and we also observed residential stability for a significant proportion of study participants over a 10-year period (from baseline to study period). Despite these considerations, the cross-sectional nature of our study still limits our ability to fully account for participants' cumulative exposures over time.

Although the GWR and MGWR models identified local associations between obesity, hypertension, and diabetes and specific environmental characteristics, their explanatory power was limited. This is partly due to our use of health outcomes previously adjusted for individual

confounders (Pearson residuals of logistic regression), which explained 5% to 10% of the total variance of CMRFs. The contribution of contextual factors, approximately 2.5% for obesity and 1.5% for diabetes and hypertension, was modest compared to individual-level factors, consistent with literature (Diez Roux and Mair, 2010; Feuillet et al., 2015). In addition, modeling individual disease risk rather than area-level prevalence introduced noise and overfitting, which contributed to the observed lower R<sup>2</sup> values. Despite these limitations, our study may still fail to capture significant individual or contextual predictors of these CMRFs. Examining factors beyond the residential environment and including additional confounding variables may improve the explanatory power of future models and provide more comprehensive insights into the multifactorial determinants influencing CMRFs. From a policy perspective, this approach may also open the way to constructing a 'portrait' of populations susceptible to CMRFs and then using fine-scale area-level data to identify regions within the administrative territory with significant concentrations of such profiles. This proactive strategy could serve as a valuable tool for targeting preventive population health interventions.

Furthermore, it is important to consider the temporal context of our study, which reflects the state of the city in 2017. While the spatial health inequalities observed during the baseline period (2003-2006), the first follow-up (2009-2012), and the second follow-up (2014-2017) remained relatively stable, changes may have occurred since then. A third follow-up (2019-2022) was available, but the examination of participants was significantly affected by the COVID-19 pandemic, which may have introduced bias into the data. In addition, the departure of several participants from the cohort since baseline increases the possibility of deviations from the general population of Lausanne. However, it is noteworthy that we did not detect significant differences in population characteristics between baseline and the second follow-up, except for those related to aging (e.g., occupational activity, physical inactivity, proportion of females, and individuals living alone).

## 3.5 Conclusion

In conclusion, this study identified high-risk areas for CMRFs and elucidated several associations with specific physical and social environmental features independent of individual-level factors. These findings are important for prioritizing CVD prevention strategies and guiding future population health intervention research.

### 3.6 Feasibility of extending NCD spatial risk prediction to the canton level

Unlike the other studies in this thesis, which cover the entire canton of Vaud, the present case study focuses specifically on Lausanne, where we perform a spatial prediction of the risk of hypertension, obesity, diabetes, and dyslipidemia in middle-aged and older adults. This analysis elucidates associations between these CMRFs and several individual factors such as age, sex, education, income, and nationality, as well as with characteristics of the residential environment such as median income, unemployment rate, and greenness.

The question then arises as to whether the hectare-level indicators developed in Chapter 2 could be used to project the risk of developing one of the NCDs studied throughout the canton. Taking obesity as an example, we found significant associations with the participant's age, male gender, and the neighborhood unemployment rate. Could we therefore expect a higher risk category for obesity in the inhabited hectares of the canton where the proportion of men over 65 and the unemployment rate are high? Although tempting, the results of this study do not allow for such generalizations for several reasons. First, the observed associations accounted for no more than 15% of the overall variability in outcome, indicating limited explanatory power. Second, the spatial heterogeneity observed in the associations between health outcomes and environmental factors suggests that the magnitude and possibly the direction of these associations may vary across the canton's diverse geographic landscapes - including rural, mountainous, and urban areas - thus limiting the generalizability of the study's findings. Lastly, the sample studied was representative of the middle-aged and elderly population of the city of Lausanne, and any extrapolation to the broader, more heterogeneous cantonal landscape is likely to be biased.

In conclusion, although the findings of this study offer insights for guiding CVD prevention strategies within Lausanne<sup>4</sup>, extending the spatial prediction of CMRF risk to the broader canton would be inappropriate. Nevertheless, the methodological foundation established here serves as an instructive template for similar research efforts in other regions of the canton where NCD-related information is available, either from hospital discharges or local cohorts. One promising research avenue could involve leveraging pseudonymized Medical Statistics of Swiss Hospitals (MS) accessible to public health departments (Zellweger, Junker, and Bopp, 2019). These datasets include both the patient's place of residence and diagnoses coded according to the International Statistical Classification of Diseases and Related Health Problems (ICD-10) (OFS, 2022b). Alternatively, data from the emerging participatory cohort *Métasanté* (*Métasanté* | *Plateforme participative* 2023), currently deployed in two municipalities

---

<sup>4</sup>For additional context, consult Figures A.3.17, A.3.18, and A.3.19 in the Appendix, which present spatial risk predictions for hypertension, obesity, and diabetes, along with corresponding geographic contextual elements to facilitate intervention planning.

### **3.6. Feasibility of extending NCD spatial risk prediction to the canton level**

---

within the canton, represent another viable data source, as will be elaborated in the general discussion, section 7.2.

While the rise of NCDs was a primary health concern for the Department of Public Health to which I was affiliated for this thesis, the advent of COVID-19 in early 2020 drastically shifted these priorities. As a result, the scope of this thesis was adjusted to additionally examine the applicability of spatial epidemiology methods in addressing this emerging infectious disease threat.





## 4 DETECTING DAILY SPATIO-TEMPORAL COVID-19 CASE CLUSTERS

The declaration of COVID-19 as a pandemic by the World Health Organization on March 11, 2020 (World Health Organization, 2020), was shortly followed by the Swiss Federal Council's implementation of a soft lockdown on March 16 (Federal Council, 2020), an event that coincided with the date of the oral defense of my research plan (i.e., candidacy exam). In the face of this global health crisis, tracking the spatial spread of the virus became imperative to mitigate its impact, efficiently allocate resources and health services, and formulate evidence-based policies (Franch-Pardo et al., 2020). Consequently, scientists worldwide, particularly in spatial epidemiology, quickly redirected their research efforts to the COVID-19 emergency, leveraging spatial data science to manage the pandemic.

During the early stages of the pandemic, the scientific community benefited greatly from the development and maintenance of a near real-time GIS-based tracking web application by Johns Hopkins University (Dong, H. Du, and Gardner, 2020), which provided invaluable insights into the magnitude and geographic distribution of the pandemic. While several studies have focused on disease mapping, for example, by displaying case counts based on national or regional boundaries, the full potential of spatial epidemiological approaches has remained largely untapped (Ahasan and Hossain, 2021). Despite the demonstrated ability of spatial clustering methods to identify clusters of cases, deaths, and vulnerable populations that require immediate attention (Desjardins, Hohl, and Delmelle, 2020; De Ridder, Sandoval, Vuilleumier, Stringhini, et al., 2020), their use remained limited. This limitation has been attributed, in part, to a lack of access to fine-scale spatial and spatio-temporal datasets, which are critical for implementing complex spatial analyses (Franch-Pardo et al., 2021).

We have established a long-term partnership with the Institute of Microbiology at the University Hospital of Lausanne (CHUV) to address this research gap. Through this collaboration, we gained access to a unique dataset comprising all SARS-CoV-2 RT-PCR tests performed in the canton of Vaud during the first wave of the Swiss epidemic (January to June 2020). The dataset

provided valuable information, including patient demographics, home addresses, and viral loads associated with positive test results. This chapter presents the results of the fine-scale spatio-temporal analysis performed on this dataset with the aim of detecting COVID-19 case clusters, tracking their evolution across the state, and describing their characteristics. While the study was retrospective, we simulated a prospective daily frequency analysis to assess the potential opportunities and challenges associated with implementing this approach in Vaud for future epidemic waves.

This chapter represents a postprint version of the manuscript published in *Science of the Total Environment*:

Ladoy, A., Opota, O., Carron, P.-N., Guessous, I., Vuilleumier, S., Joost, S., & Greub, G. (2021). Size and duration of COVID-19 clusters go along with a high SARS-CoV-2 viral load: A spatio-temporal investigation in Vaud state, Switzerland. *Science of The Total Environment*, 787. <https://doi.org/10.1016/j.scitotenv.2021.147483>

As the first author of the publication, I designed the study, processed and analyzed the data, and co-wrote the manuscript.

### 4.1 Introduction

The novel coronavirus SARS-CoV-2 that causes the COVID-19 disease has impacted society at an unprecedented scale. The number of infected people increased rapidly around the globe, with over 141 million confirmed cases as of April 2021 and more than 3 million deaths (World Health Organization, 2021a; World Health Organization, 2021b). The rapid dissemination of the disease has challenged international experts and policymakers to implement strategies with regards to local viral spread, healthcare resources, economic and political factors (Nicola et al. (2020); see also the cross-country analysis of COVID-19 response: <https://analysis.covid19healthsystem.org/>). Contact tracing, lockdowns and quarantines have been implemented around the world in a bid to contain the virus spread, and has impacted over four billion people worldwide (Chu et al., 2020). These measures are aimed at protecting approximately 22% of the world's population at risk of severe COVID-19 complications (Clark et al., 2020), with important social and economic consequences (Chu et al., 2020; Ruktanonchai et al., 2020; Faber, Ghisletta, and Schmidheiny, 2020).

COVID-19 outbreaks occur via close contacts, which form clusters of positive cases. Critical challenges for containing the spread of the virus lie (i) in the early detection of clusters, which reflect active viral transmission (De Ridder, Sandoval, Vuilleumier, Stringhini, et al., 2020), and (ii) in the understanding of the spatial and temporal evolution of clusters (Hohl et al., 2020). Geospatial tools using the precise location of the place of residence of tested individuals

are highly effective to monitor an epidemic (Franch-Pardo et al., 2020; Keesara, Jonas, and Schulman, 2020). They allow for the implementation of targeted strategies to control the local spread of disease through space and time (Cromley, 2019).

Although widely used, there is no general agreement on the definition and concepts relating to clusters, outbreaks and hotspots, particularly given a spatial context. Yet, information available from public health departments around the world converge despite existing differences. The term “cluster” generally refers to a temporal aggregation and a spatial concentration of infection cases. Generally, COVID-19 clusters are defined as two or more test-confirmed cases – though this varies to three or more in France ([www.santepubliquefrance.fr](http://www.santepubliquefrance.fr)) and Switzerland, up to 10 or more in New Zealand ([www.health.govt.nz](http://www.health.govt.nz)) – between individuals associated with a specific non-residential setting with illness onset dates within 7 to 14 days. To further label clusters as an “outbreak”, one must also have either (i) identified direct exposure between at least two of the test-confirmed cases in that setting (for example under one meter face-to-face) during the infectious period of one of the cases, or (ii) if there is no sustained local community transmission, noticed the absence of an alternative source of infection outside the setting for the initially identified positive case (England, 2020). Clusters are also assimilated to the concept of “hotspot”, which is not clearly defined neither but often used in spatial epidemiology (Lessler et al., 2017). The World Health Organization (WHO) has defined a set of methods and procedures to identify epidemic hotspots for use in global surveillance of populations (World Health Organization, 2013). Additionally, infectious diseases studies have proposed methods to identify and characterize spatial clusters (Bejon et al., 2010; Bousema et al., 2012).

The identification of high prevalence areas for any phenomenon constitutes a specific research domain in spatial statistics. Point pattern analysis (Gatrell et al., 1996) and local spatial autocorrelation methods have previously been applied in the detection of disease clusters (Jacquez and Greiling, 2003). In the current COVID-19 pandemic, X. Zhang et al. (2020) used local Moran’s statistics to identify clusters in China at a large geographic scale using incident cases aggregated at the level of large administrative units. Among studies involving geospatial information reviewed by Franch-Pardo et al. (2020), few have characterized the spread of COVID-19 across space and time (e.g. Desjardins, Hohl, and Delmelle (2020)), and even fewer studies used spatial statistics to detect clusters at a local scale (De Ridder, Sandoval, Vuilleumier, Azman, et al., 2021; De Ridder, Sandoval, Vuilleumier, Stringhini, et al., 2020). More studies at local and regional scales that consider demographic characteristics of a population at risk are needed to provide timely information to enable accurate prevention and containment measures (Franch-Pardo et al., 2020). Indeed, the precise detection of spatial clusters, the description of their dynamics and evolution over time in a geographical context are key to inform decision-makers, to deploy smart testing overtime, and to provide targeted health and prevention interventions at a local scale (Kamel Boulos and Geraghty, 2020).

The persistence in time of clusters were shown to be associated with socio-economic deprivation (De Ridder, Sandoval, Vuilleumier, Azman, et al., 2021), but the size and duration of clusters are also likely to be due to “super-spreader” individuals or events (Danis et al., 2020). These super-spreader individuals or events are considered to greatly contribute to the transmission of an infectious disease (Lau et al., 2020). This process relates to the evidence for large variation in individual reproductive number, where some individuals contribute more than others to epidemics (Lloyd-Smith et al., 2005). Super-spreaders correspond to the small percentage of individuals (20%) within any population to control most (80%) transmission events (the 20/80 rule; see Stein (2011)). Super-spreaders are also present for the SARS-CoV-2 virus (Lau et al., 2020; Frieden and C. T. Lee, 2020); these individuals are more likely to be highly infectious, which is suggested to be related to high viral loads (Beldomenico, 2020). Notably, as recently shown in (Jacot, Greub, et al., 2020), the viral load of people infected with SARS-CoV-2 appears to be similar to what is observed for other respiratory viruses such as influenza B. It remains to be explored why SARS-CoV-2 exhibits such a high reproductive number ( $R_0$ ) of about 2 to 3.5 (Liu et al., 2020), and if the transmission pattern, cluster duration and size somehow correlate with viral load within a detailed spatio-temporal context.

Here, we characterize the spatial and temporal dynamics of the first wave of SARS-CoV-2 infections in the state of Vaud (western Switzerland) through the detection and location of clusters. Clusters are defined by location and time where individuals are expected to have been in contact (rather than by observed contacts between individual). For each cluster, we measure size, duration and composition (number and age of individuals, as well as their viral load). We use the results of the SARS-CoV-2 RT-PCR tests ( $n = 33,651$ ) performed by the Microbiology Laboratory of the Lausanne University Hospital (CHUV) between January 10 and June 30, 2020 (with a first positive case on March 2, 2020). The data used here include the results of RT-PCR tests, viral loads (copies/ml) for positive tests, individual age, and geographic location of residence for individuals tested. We used a spatial scan approach (Kulldorff, 1997; Moraga and Montes, 2011) to (i) detect spatio-temporal clusters of COVID-19 on a daily basis, (ii) disentangle the relationships between cluster size, duration and composition, and (iii) assess the importance of viral load in the evolution of clusters. We also implemented a Modified Space-Time DBSCAN (MST-DBSCAN) algorithm (Kuo, T.-H. Wen, and Sabel, 2018) to characterize the diffusion dynamics of transmission clusters. Finally, we discuss the effects of a soft lockdown used across Switzerland between March 19 and April 27, 2020, on the spread dynamics of the virus.

## **4.2 Materials and methods**

### **4.2.1 Patients**

Patients exhibiting symptoms compatible with COVID-19, including fever, cough, dyspnea, and loss of smell or taste, were tested using RT-PCR for the presence of the SARS-CoV-2 in their nasopharyngeal secretions, particularly for those considered as either vulnerable (e.g., immunosuppressed, obese, with chronic obstructive lung disease or age > 65 years) or likely to be exposed to vulnerable people (e.g. healthcare workers or those living with vulnerable persons). The studied population therefore predominantly includes vulnerable symptomatic individuals. Moreover, people who had been in contact with positive cases were also tested and included in the study, even if asymptomatic, to determine the necessity of a 10 day quarantine or isolation period. The precise residential address and age of the patient was collected at the time of sampling.

### **4.2.2 SARS-COV-2 RT-PCR**

Most RT-PCR were performed using the automated molecular platform implemented at the Institute of Microbiology (Lausanne University Hospital, CHUV). It uses the Magnapure automated RNA extraction method followed by PCR amplification on QuantStudio automated systems (Greub et al., 2016) with primers described by Corman et al. (2020), later slightly modified according to Pillonel et al. (2020) to further improve PCR sensitivity. From March 24, 2020, most RT-PCR were performed with the COBAS 6800 RT-PCR test, which exhibited similar performance than the home-brew automated approach (Opota et al., 2020). Some cases (n = 71) were tested using the GeneXpert approach to reduce processing time (Moraz et al., 2020). Viral load was calculated based on the “cycle threshold” (Ct) defined as the number of cycles required for the fluorescent signal to cross a given value threshold (Opota et al., 2020; Moraz et al., 2020).

### **4.2.3 Study area**

Data were collected in the south-western Swiss state of Vaud, north of Lake Geneva. Vaud has an area of 3212 km<sup>2</sup> (Fig. 4.6 (A)) with a population of 811,203 (end of 2019), giving an average density of 249 inhabitants/km<sup>2</sup>. Notably there are population density differences between the urban area of Lausanne-Morges on the shores of Lake Geneva ( 3000 inhabitants/km<sup>2</sup>) and rural areas towards the north ( 200 inhabitants/km<sup>2</sup>). An exception is the area of Yverdon-les-Bains, located directly south of the Lake of Neuchâtel with 2200 inhabitants/km<sup>2</sup>.

### 4.2.4 Spatio-temporal clusters

We used SaTScan software (version 9.6.1) to detect daily space-time clusters of individuals who were tested positive for SARS-COV-2 in the state of Vaud from March 2 to June 30, 2020, (no positive cases between January 10 and March 2, 2020). The algorithm developed by (Kulldorff, 1997) tests whether a disease is uniformly distributed among individuals over space and time. It uses a “moving cylinder”, with the base and height corresponding to the spatial and temporal components, respectively. Significance tests evaluate excess relative risk, i.e., more observed COVID-19 cases than expected within the moving cylinder relative to randomly distributed cases over space and time. We implemented this algorithm across a daily prospective surveillance analysis. We used a discrete Poisson model, where the number of events in the geographic area (total number of positive tests) is Poisson-distributed, according to a known underlying population at risk. Though typical SaTScan applications use the default value of 50% of the population at risk for the spatial size cluster’s radius, here we selected a radius covering a maximum of 0.5% of the total resident population (population at risk) in the state of Vaud ( $N = 811,203$  inhabitants; SFSO (2019)), as a smaller value emphasizes the discovery of small and homogeneous clusters (J. Chen et al., 2008). Tested individuals and the underlying population at risk were georeferenced at the centroids of a hectometric grid (SFSO, 2020) covering the entire study area. The minimum number of positive cases considered to constitute a cluster was set to three, and we restricted the temporal scanning window to a minimum of two days and a maximum of 14 days (see Figure A.4.1). The upper limit of 14 days accounts for the incubation period (generally 2 to 7 days) and infectious time (generally 7 to 10 days from symptom onset, as deduced from different culture-based and RT-PCR-based investigations; Jaafar et al. (2020); Jeong et al. (2020); Caruana et al. (2021); To et al. (2020)). The significance of the clusters was evaluated on the basis of 999 Monte-Carlo permutations that randomized both locations (Besag and Diggle, 1977) and times of the cases.

### 4.2.5 Cluster evolution and diffusion zones

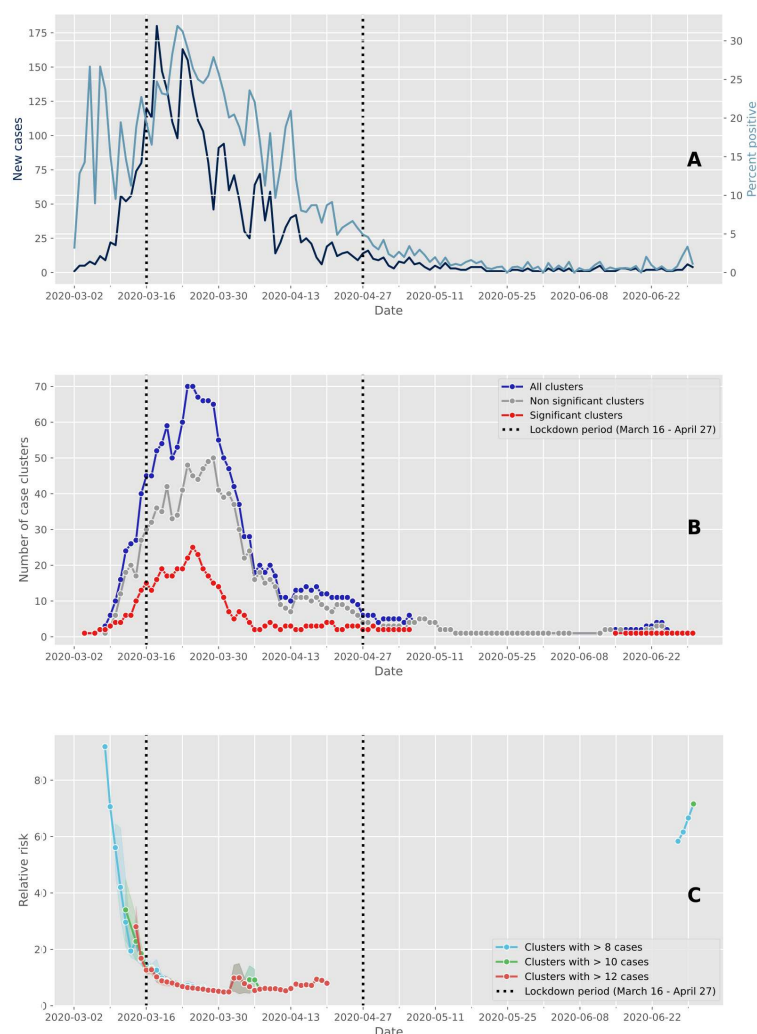
We used MST-DBSCAN (modified space–time density-based spatial clustering of application with noise; Kuo, T.-H. Wen, and Sabel (2018)) to characterize the diffusion dynamics of clusters. MST-DBSCAN is an algorithm used to detect, characterize, and visualize disease cluster evolution in geographic space and time. It geographically computes a kernel density that considers the effect of the incubation period of an infectious disease. It is based on DBSCAN (Ester, Kriegel, and X. Xu, 1996), a non-parametric density-based clustering algorithm that groups together objects (here, SARS-COV-2 positive cases) that are closely packed together (points with many nearby neighbors), marking points falling in low-density regions as outliers. The MST-DBSCAN identifies seven different cluster behaviors: a) emerge, b) grow, c) remain steady, d) merge, e) move, f) split or g) reduce.

We applied the MST-DBSCAN analysis to the 3317 COVID-19 positive cases identified (among 33,651 tested individuals), georeferenced at their precise residential address in the state of Vaud. Disease clusters were computed daily from March 4, 2020, to June 30, 2020. The maximum spatial radius considered was 1000 m, with a time window of 1 to 7 days to reflect the average infectious period after a positive test (see Figure A.4.1). A cluster was defined as a minimum of three positive cases. For all identified clusters, we established a typology of similar diffusion patterns in the geographical space. We associated clusters with postcode areas (557 units; MICROGIS (2019)) in the state of Vaud to use as spatial references. Then, we focused on three main behaviors to characterize the diffusion type through the postcode areas: a) Increase, if an area was covered by clusters whose evolution type was Emerge, Growth, or Merge; b) Keep, if an area was covered by clusters whose evolution type was Steady or Move; c) Decrease, if an area was covered by clusters whose evolution type was Reduction or Split. Postcodes with similar diffusion patterns over the entire study period were grouped using the Louvain method, a group detection algorithm that uses network analysis (Blondel et al., 2008). This approach synthesizes the spatio-temporal information and facilitates its visualization on a single map. Equations for the MST-DBSCAN approach are provided in supplementary materials (Note A.4).

## 4.3 Results

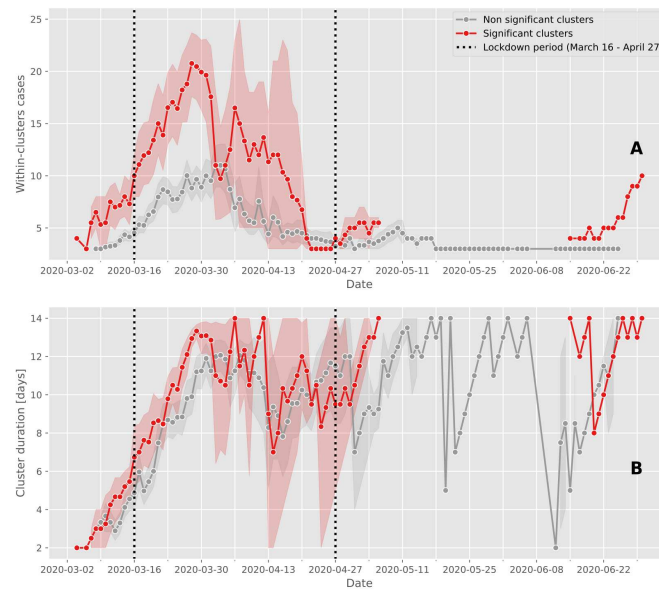
### 4.3.1 Epidemic trajectories of positive cases

A total of 33,651 individuals were tested over a period of 6 months (March 2 to June 30, 2020), of which 3317 (9.86%) were confirmed as positive using RT-PCR. Of these positive cases, 79% (2609/3317) occurred between March 9 and April 5, though this four week period corresponds to only 16% of the study period total duration (Figure 4.1 (A)). The peak of the first epidemic wave occurred on March 18, which was two days after the soft lockdown was implemented in Switzerland, that lasted from March 16 to April 27 (vertical dashed lines in Figure 4.1 (A)). Up to 180 individuals a day were documented as “positive” in our laboratory at the peak number of cases during the study period (4.1 (A), dark blue). The number of positive cases then decreased considerably from the first of May. The highest proportion of positive tests was observed four days after the peak of the epidemic wave, with a rate of positive tests reaching 32% (4.1 (A), light blue). The rate of positive cases was relatively high at the start of the epidemic when few individuals were tested. After this, the shape of the curve of the percentage of positive tests followed the one of the number of cases. This is likely because at the beginning of the epidemic, only individuals with symptoms and those at risk were tested, and it was only later that a much wider range of individuals were tested such that all symptomatic individuals and asymptomatic individuals in contact could access a test.

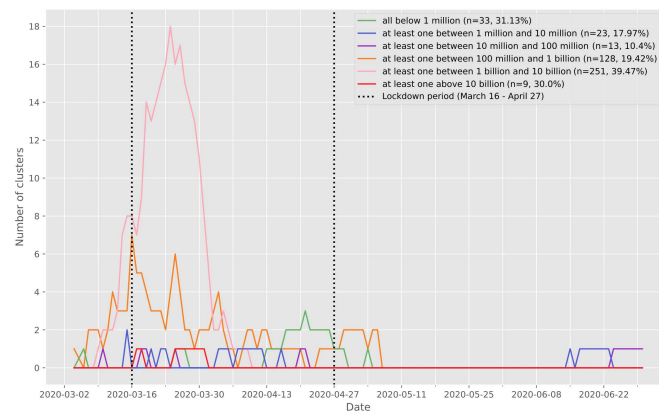


**Figure 4.1 – Evolution of cases and clusters through time.** The vertical dashed lines delimit the Swiss lockdown period (March 16 to April 27). (A) Epidemic trajectory of positive tests. The daily new confirmed cases are represented in dark blue and the percent of positive tests are represented in light blue. (B) Number of case clusters over time. The total number of clusters detected daily by space-time scan statistics are represented in dark blue, and the red and grey lines represent the proportion of significant clusters ( $p \leq 0.05$ ) and non-significant clusters ( $p > 0.05$ ), respectively. (C) Average relative risk of significant space-time clusters ( $p \leq 0.05$ ) over time according to within-cluster cases.





**Figure 4.2 – Case cluster characteristics over time.** The mean and confidence intervals of the number of cases within clusters (A) and the duration of clusters (B) are calculated for significant (red line) and non-significant clusters (grey line). Between May 6 and June 15, the prospective space-time scan statistic detected no significant clusters. The vertical dashed lines delimit the Swiss lockdown period (March 16 to April 27).



**Figure 4.3 – Number of significant ( $p \leq 0.05$ ) case clusters over time characterized according to the viral load of the cases documented in each cluster.** The vertical dashed lines delimit the Swiss lockdown period (March 16 to April 27). For each cluster, we extracted the positive test individuals intersecting the cluster both geographically and temporally, and we characterized the clusters according to the viral load of the individuals composing it. The clusters represented in green are composed only of individuals with a viral load of less than 1 million copies/ml. The clusters shown in blue are made up of at least one individual with a viral load between 1 million and 10 million copies/ml, and so on.

### 4.3.2 Cluster detection and temporal dynamics

We identified 1684 space-time clusters of more than two cases using residential address for individuals tested positive to SARS-CoV-2. Of these, 457 clusters were considered significant based on the within-proportion of positive cases compared to the total documented positive cases. Highest values of both significant and non-significant clusters were observed between March 9 and April 5 (Figure 4.1 (B)) and then the number of clusters decreased. The decrease in positive cases following the beginning of the soft lockdown (Figure 4.1 (A)) occurred approximately two weeks before the decrease in the number of clusters (Figure 4.1 (B)). The number of clusters displays a similar pattern through time but with a difference in amplitude. As shown in Figure 4.1 (C), the estimated relative risk for new clusters was greater before the soft lockdown and approximately 80 days after the end of the lockdown. The size of the clusters (i.e., number of cases within clusters) used to compute the relative risk did not strongly change the value of the relative risk during the core of the epidemic wave. However, cluster size affected relative risk when the number of positive cases was small, such as at the beginning and at the end of the epidemic wave.

### 4.3.3 Cluster composition

Significant space-time clusters generally involved a larger number of positive cases (maximum mean of 21 cases, where the largest cluster had 43 cases on March 25) compared to non-significant clusters (maximum mean of 11 positive cases; Figure 4.2 (A)). Notably, significant clusters with more than 15 positive cases were predominantly observed shortly after the soft lockdown was implemented from March 16 to April 27, with one exception on April 3 (Figure 4.2 (A)). Cluster durations – although limited to 14 days - increased over time from the start of the epidemic wave, showing little differences between significant and non-significant clusters. There was an absence of significant clusters from May 3 to June 16 (Figure 4.2 (B)).

### 4.3.4 Viral load in clusters

Clusters were defined by the presence of at least three positive cases within a limited geographic area, as documented in the SaTScan analysis. All clusters were then characterized according to the nasopharyngeal viral load of the cases for each cluster (Table 4.1). Five significant clusters were composed of three cases exhibiting a viral load below 10,000 copies/ml at time of testing, which is the same low load as found in non-significant clusters (Table A.4.1). However, significant clusters were more likely to be detected when viral loads were above 100 million copies/ml (Figure 4.3). Finally 18 significant clusters with at least one individual showing between 1 billion and 10 billion copies/ml were documented on March 24 (Figure 4.3, pink curve). There was a significant difference between the frequency distribution of

Case cluster viral load category	Total number of clusters	Number of significant clusters
All below 1 million	106	33 (31.13%)
At least one between 1 million and 10 million	128	23 (17.97%)
At least one between 10 million and 100 million	125	13 (10.40%)
At least one between 100 million and 1 billion	659	128 (19.42%)
At least one between 1 billion and 10 billion	636	251 (39.47%)
At least one above 10 billion	30	9 (30.00%)
<i>Total</i>	1684	457

Table 4.1 – **Classification of the space-time clusters according to the viral load of the cases involved.** Within-cluster cases were identified by matching both geographically and temporally positive test subjects, geocoded at the residential address, with space-time clusters. For example, a cluster was classified as "all below 1 million" if all individuals tested positive within the cluster during its active period had a viral load below 1 million copies/ml. For each cluster category, the total number of case clusters detected by prospective space-time scan statistics over the entire study period (March 2 to June 20) and the proportion of significant clusters ( $p \leq 0.05$ ) are reported.

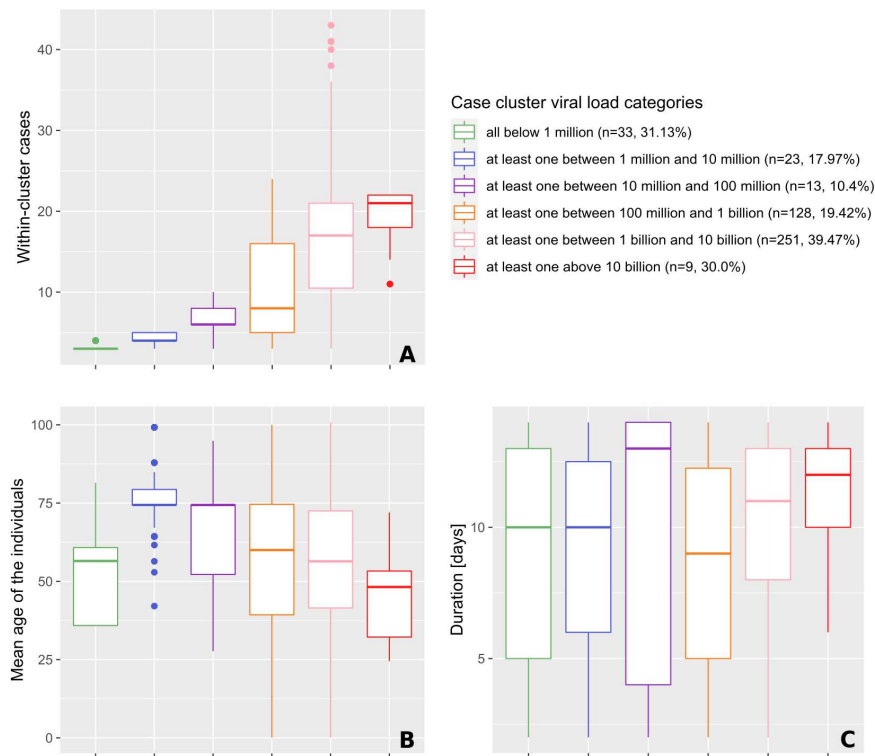


Figure 4.4 – **Characteristics (size, age and duration) of significant space-time clusters ( $p \leq 0.05$ ) over the study period, categorized according to the viral load of the cases involved.** The classification procedure is explained in more details in the legend of Table 4.1. Characteristics include the number of cases observed in clusters (A), the mean age of the positive tests individuals forming clusters (B), and the duration of clusters (C).

viral loads in significant clusters compared with non-significant clusters and outside clusters (Kolmogorov–Smirnov test, two-sample case,  $p < 0.001$ , see Figures A.4.3 and A.4.4).

The mean viral load of the first three cases was also studied, in order to gain insight of the possible relationship between nasopharyngeal viral load and contagiousness, indirectly measured by the documentation of subsequent clusters. For 20 significant clusters, all first three cases exhibited a viral load below 100,000 copies/ml, suggesting that individuals with fewer than 100,000 copies/ml may still be contagious (Table A.4.2). Moreover, the nasopharyngeal viral load of the first three cases was below 1 million copies for 40 significant clusters.

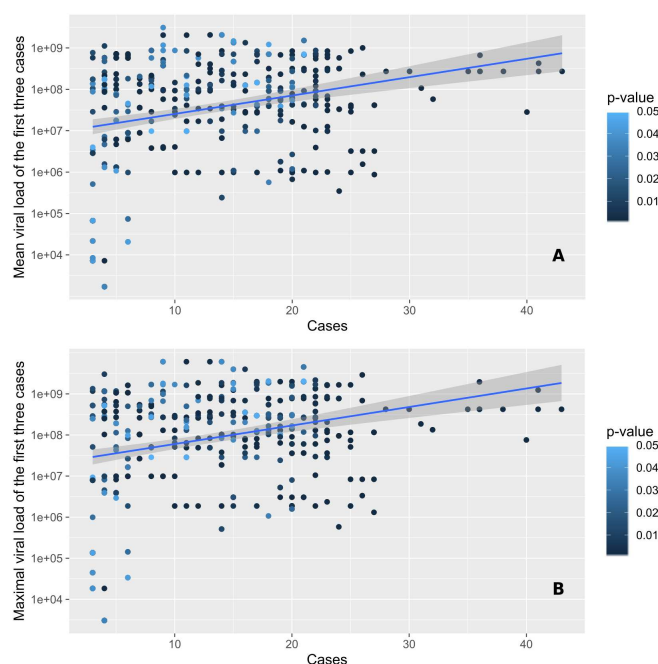


Figure 4.5 – **Number of cases observed within significant space-time clusters ( $p \leq 0.05$ ) in function of the mean (A) and maximal (B) viral loads of the first three cases involved.** Points are colored according to the significance level of the cluster, as assessed through 999 Monte Carlo random permutations.

### 4.3.5 Cluster size, duration and viral load

Cluster size was positively associated with the presence of individuals with high viral loads (Figure 4.4 (A)). The highest viral loads measured were greater than 10 billion copies/ml and occurred in the largest clusters (median of 21 positive cases). This was used to identify super-spreading events. When comparing clusters harboring individuals with all viral loads below 1 million copies/ml with the ones where at least one case had a viral load above 1 million copies/ml, the cluster sizes were significantly different, with median cases per cluster

increasing from three to four ( $p < 0.001$ ; Figure 4.4 (A)). Similar relationships were observed when considering the mean and maximal values of viral loads of the first three positive cases (Figure 4.5 (A,B)).

Highest viral loads were found in clusters with individuals showing the lowest average age group considered in the investigation. Clusters composed of individuals in the highest average age group showed low to middle viral loads (Figure 4.4 (B)). The median age of individuals within a cluster was significantly higher when the cluster viral load was between 1 and 10 million copies/ml. The average age group then progressively decreased from 74 to 48 years, while viral load increased. Cluster duration was only significantly different between the orange category (100 million to 1 billion/ml) and the pink category (1 to 10 billion/ml), where clusters of the latter lasted half a day longer (mean of +0.46 days,  $p < 0.001$ ; see Figure 4.4 (C)). Clusters with individuals in the lowest average age group considered in the investigation and clusters with the highest viral loads (Figure 4.4 (B)) also constitute the largest clusters (Figure 4.4 (A)) and those that last the longest (Figure 4.4 (C)), respectively.

#### 4.3.6 Geographic distribution of the first epidemic wave

We chose six key dates to illustrate the evolution of the two cluster types during the first wave of the epidemic in the state of Vaud. Figure 4.6 shows the spatial distribution of space-time clusters (A–F) and compares it to information translating the diffusion dynamics of the clusters (A'–F'). A detailed description of the first SARS-CoV-2 epidemic wave in the of Vaud can be found in Box 4.3.6, illustrating the powerful and critical information that the approach offers. Cluster behaviors described in Box 4.3.6 were summarized with four diffusion zones shown in Figure 4.7 (A) and identified at the level of postal code areas using MST-DBSCAN (Figure 4.7 (B,C,D)). The grey diffusion zone corresponds to areas where no clusters emerged, while the green, orange and blue diffusion zones differ in the way clusters evolved over time. The green diffusion zones correspond to areas where the clusters immediately increased in size at the beginning of the epidemic wave (red line, Figure 4.7 (B)), but decreased drastically once the soft lockdown (vertical dash line) was implemented. We then observed a second peak associated with an important increase of clusters that reduced in size (red line, Figure 4.7 (B)). Both red and purple curves are bimodal and tended to decrease afterwards, with a few numbers of new small peaks that plateau forming a distribution with a long right tail. Conversely, orange and blue diffusion zones show a first peak of increasing clusters later, at about the time of the start of the soft lockdown (orange & blue areas, Figure 4.7 (C,D)). Both zones also show clusters that remained stable in size during the soft lockdown (blue line). Only the blue diffusion zone showed no further clusters after April 27, which corresponds with the end of the lockdown, and is the only zone that did not display a bimodal distribution of clusters. Note that no difference in viral load was documented among these different diffusion zones (Figure 4.7 (E)).

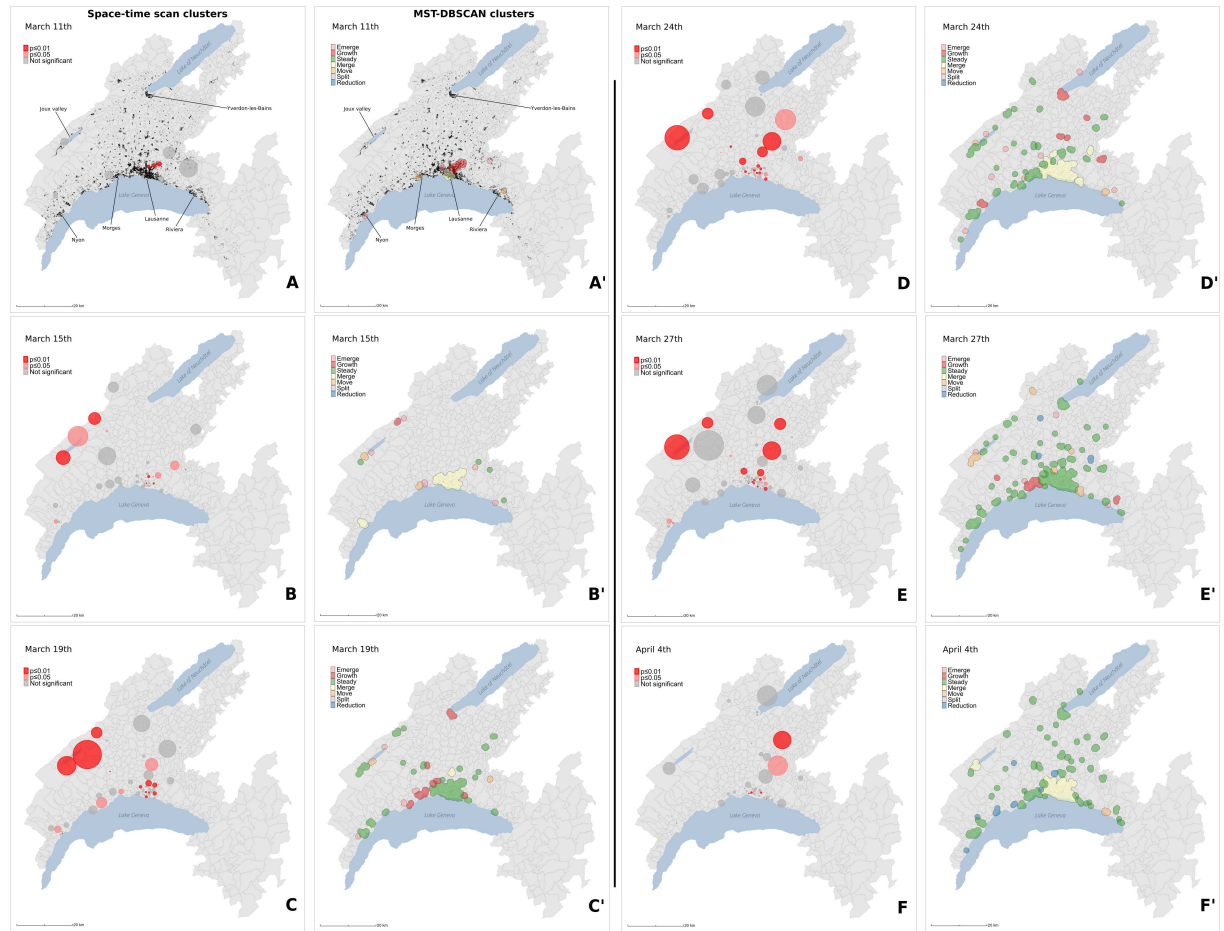


Figure 4.6 – **Spatial distribution of case clusters (A–F) and diffusion dynamics of transmission clusters (A’–F’)** for 6 key dates (March 11, March 15, March 19, March 24, March 27, April 4) during the first epidemic wave. Case clusters resulting from the prospective Poisson space-time scan statistics (A–F) are shaded according to their significance level: dark red for statistically significant clusters with  $\alpha = 0.01$ , light red for statistically significant clusters with  $\alpha = 0.05$ , and grey for non-significant clusters ( $p > 0.05$ ). Transmission clusters resulting from the MST-DBSCAN algorithm (A’–F’) are shaded according to their evolution type: emerge (pink), growth (red), steady (green), merge (yellow), move (orange), split (purple), and reduction (blue). Black points in (A) and (A’) represent the 33,651 individuals tested during the study period (January 10 to June 30).

**Box 4.3.6: The first SARS-CoV-2 epidemic wave in the state of Vaud, Switzerland.** On March 11, 2020, 7 days after the first detection of a positive case (Figure 4.6 (A)), we observe a phase of rapid growth and merging (see Figure 4.6 (A')), where a series of significant clusters appear directly north of Lausanne, the main city of the state. Interestingly, three out of the four clusters shown are located in wealthy areas. March 15 (Figure 4.6 (B,B')) is the day before the soft-lockdown. There are multiple clusters in the Lausanne area, among which a large fraction are significant (Figure 4.6 (B)). Figure 4.6 (B') shows that these clusters rapidly merged into a single "super cluster" deployed over the urban agglomeration. In the rural areas, active clusters emerge and grow north of the lake in the Joux valley, located in the Jura mountains, where population density is low. Four days later, on March 19 (Figure 4.6 (C,C')), the peak of the first wave is approaching (see Figure 4.1 (B)). The number of case clusters is high in the Lausanne area (Figure 4.6 (C)) but stabilizes. Similar behavior is observed towards the east along Lake Geneva; only one moving cluster is observed in the Riviera area (Figure 4.6 (C')), compared with new clusters appearing in the Morges area to the west. In the Joux valley, the activity remains important, and a cluster grows in Yverdon-les-Bains, south of the Lake of Neuchâtel. On March 24 (Figure 4.6 (D,D')), the peak of the first wave is reached (see Figure 4.1 (B)). New cases reactivate moving clusters in the center of Lausanne, while towards west the situation stabilizes and even reduces towards Geneva with no further significant cluster in the Nyon area. At the peak, a large significant cluster remains steady in the Jura, and several clusters grow in the remote, rural periphery north of the main urban area (Figure 4.6 (D,D')). In the north, close to the Lake of Neuchâtel, the clusters are not significant despite growing. On March 27 (Figure 4.6 (E,E')), all clusters start an important and rapid reduction phase (see Figure 4.1 (B)). The merged clusters of the Lausanne area split, and most clusters in the country-side remain steady. However, a significant cluster emerges in the west, in the Nyon area. On April 4 (Figure 4.6 (F,F')), the peak ends. The Joux valley cluster ends after 25 days and clusters in the state are either no longer significant, or are steady, split or reduce. There is one exception north of Lausanne with a single growing cluster located in a leisure area and is likely related to the presence of a school (with boarding).

### 4.4 Discussion

The discussion is divided into three major parts. The first highlights results that uncover new information on COVID-19 clusters, the second summarizes limitations of the interpretation of the results, and the third describes the added value of these methods for tackling epidemic problems and evaluating the effects of lockdown strategies.

#### 4.4.1 New information on COVID-19 clusters

##### **A temporal lag between documentation of positive cases and clusters burden**

Significant clusters were predominantly observed from March 15 to April 5 (red curve in Figure 4.1 (B)), while non-significant clusters in high population-density areas, such as Lausanne, were documented 4 to 5 days earlier and continuously occurred until mid-May (grey curve on Figure 4.1 (B), and Figure 4.6 (A)). We observed a time-shift between the decrease in the number of positive cases and the decrease in the number of clusters. This delay could be explained by the fact that most positive cases might have been at the origin of lasting clusters, i.e. clusters that last longer than 10 days from when their first positive cases are identified. Interestingly, the number of patients hospitalized at Lausanne University Hospital (CHUV) and the number of COVID-19-related deaths in the state of Vaud also followed the same epidemic curve, but with a two weeks delay (personal communication, G. Greub).

##### **Viral load is strongly informative about the presence and size of SARS-CoV-2 clusters**

Our results show that clusters at the peak of the SARS-CoV-2 epidemic wave were composed of individuals with a high viral load. Cluster size is positively associated with the presence of individuals with a high viral load in significant clusters, though in 40 clusters the first three cases exhibited a viral load below 1 million copies/ml, including 33 clusters with individuals that had a nasopharyngeal viral load below 1 million copies/ml. Moreover, as many as 20 clusters were composed of cases that initially had a viral load below 100,000 copies/ml, suggesting that subjects with fewer than 100,000 copies/ml may still have been contagious. The fact that significant clusters are composed of patients with viral loads as low as those found in non-significant clusters further supports the hypothesis that community transmission can occur with low levels of viral load. Nevertheless, this may also reflect a statistical bias as large clusters with more than 10 individuals are more likely to have at least one individual with a very high viral load.



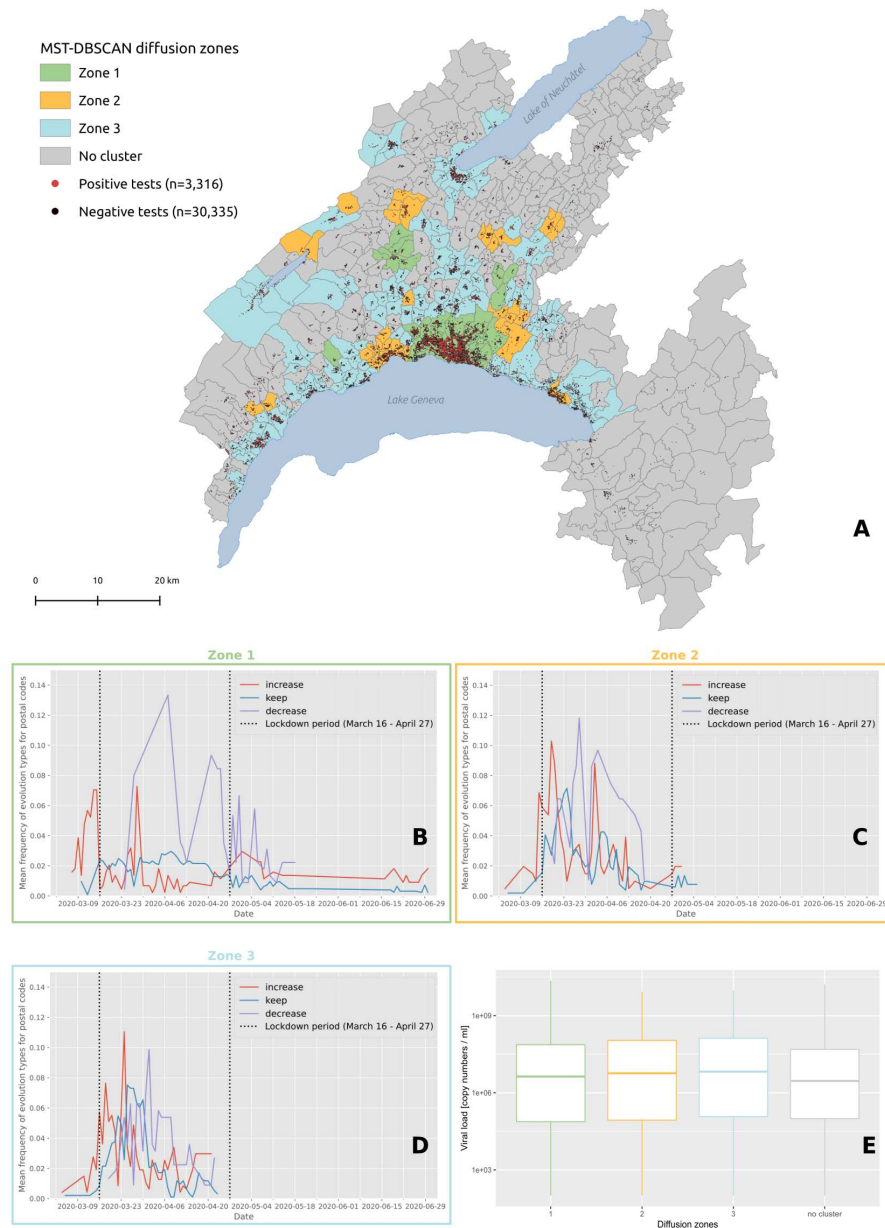


Figure 4.7 – **Diffusion zones identified by the MST-DBSCAN algorithm** (A). Postcode areas with the same color share similar diffusion patterns, and areas without any transmission clusters are represented in grey. The black dots on the map indicate the 33,651 individuals tested for COVID-19 in the state of Vaud between January 10 and June 30, 2020. Below the map are the frequencies of major evolution types over time for zone 1 (B), zone 2 (C) and zone 3 (D). The red line corresponds to the "increase" diffusion type whose area grows with time, the blue line corresponds to the "keep" diffusion type assigned to clusters whose area remains stable, and the purple line corresponds to the "decrease" diffusion type whose area becomes smaller. The vertical dashed lines delimit the Swiss lockdown period (March 16 to April 27). (E) Distribution of the viral load of test-confirmed cases living in each diffusion zone.

### **Advantage of RT-PCRs over antigen-based testing**

Given the relatively low sensitivity of antigen tests, we estimate that we would have missed or had delayed identification of approximately 24 clusters. Indeed, the 20 significant clusters with a viral load of the three first cases below 100,000 copies/ml would not have been detected with antigen tests, given that the best tests have a detection limit of about 100 to 200,000 copies/ml (Caruana et al., 2021). Moreover, the clusters with a case between 100,000 copies/ml and 1 million copies/ml would not have been detected in 5% of cases given an overall antigen sensitivity for such viral load of about 80% (Caruana et al., 2021). Thus, an antigen-based strategy would have missed about 5% (24/457) of the significant clusters. Therefore, for the second wave that began in October 2020 across western Switzerland, we advocated against the use of antigen tests for the vulnerable population and healthcare workers, as well as with non-vulnerable subjects during non-acute periods of SARS-CoV-2 infection (1 to 4 days of symptoms), despite encouraging results for antigen tests in subjects within the first four days of symptom onset (Schwob et al., 2020).

### **High viral load in large clusters within the youngest group age**

Within clusters, we found a clear negative relationship between age and level of viral load (Figure 4.4 (B)), and between cluster size and viral load (Figure 4.4 (A)). Indeed, while a high viral load was found in large clusters with the youngest group age considered in this study, low to intermediate viral load was measured in small clusters composed of older age groups. This suggests that large clusters were generated by active individuals from the working population, and super-spreader events may be at the origin of such large clusters. Surprisingly, when the level of viral load was analyzed across age groups, no relationship was found (Figure A.4.4), meaning that useful information emerges within clusters. Indeed, the characterization of the clusters provides a deeper analysis of the mechanisms behind the progression of an epidemic and the geographic analysis of clusters of cases might constitute a type of investigations to favor in the future.

### **Non-significant clusters also convey information on the progression of the epidemic**

Significant and non-significant clusters both show the same epidemiological trajectories. Indeed, they display similar patterns in terms of changes in size, differing only in amplitude. This suggests that the occurrence of clusters, even if non-significant, is a good estimator of the epidemic situation. Significant and non-significant clusters differ in terms of number of cases and measured viral load, but not in duration. This suggests that non-significant clusters (i) might correspond to transmission events unrelated to subjects with very high viral load, (ii) translate a lower impact on the population in terms of viral spread, and (iii) express a

transition towards or from a significant spatio-temporal configuration.

### 4.4.2 Limitations

#### **Tested population is not homogeneous through time**

During the course of the studied epidemic wave, recommendations for testing as requested by the authorities changed. Initially, only symptomatic patients at risk and health workers were tested. Then, from mid-March a wider proportion of the population was progressively tested, though younger individuals still were reluctant to get tested. This may have generated heterogeneity in our longitudinal investigation. Moreover, tests were likely performed at different stages of infection (early, late, etc.) and might not be representative of the correct window of infection. The day of the week may also have generated differences in the number of positive tests. Indeed the number of tests was often smaller during weekends, as some individuals preferred being tested only on the following Monday to avoid quarantining over the weekend.

#### **Positive cases might be missing**

Our estimate of the number of positive cases is not fully representative of the epidemic, particularly at the beginning when only symptomatic at-risk patients and health workers were tested. We might expect that close relatives of positive cases were also infected but not tested due to reagent shortage. However, this bias might have a limited impact on the assessment of cluster size, as any person in contact with a positive case (documented by the contact tracing team) was tested. An additional source of underestimation could be from the false negative RT-PCRs results due to imperfect nasopharyngeal sampling, though the clinical sensitivity of RT-PCR performed on nasopharyngeal samples in our laboratory is very sensitive, with 96 to 98% accuracy (Mueller et al., 2020; Schwob et al., 2020). Similarly, the rate of false positives in the same laboratory was estimated to be lower than 1/10,000 tests, due to full automation and bar-coding that was used to prevent human error and samples/tubes inversion (Greub et al., 2016). Asymptomatic patients might also contribute to the spread of the disease, though those individuals were not detected and thus could not be accounted for in this study unless they were among the contacts traced from positive cases. Finally, missing positive cases could be from people living or working at the state border, such that they were tested elsewhere. However, this seems to have a limited impact as over 80% of all samples tested for SARS-CoV-2 in 2020 were obtained from individuals living in the state of Vaud.

### Many space-time scan clusters

In this study, we adopted a posture to reproduce the daily monitoring of the epidemic. As a result, the analysis was repeated each day. In this configuration the number of clusters detected is unusually high compared to the total number of positive cases. Indeed, the same subset of positive cases can be responsible for several clusters if an area presents an excessive relative risk for several days in a row, as we consider a time window of 14 days. This is why the clusters' summary statistics for viral load, cluster's duration, and age of the subjects is smoothed.

#### 4.4.3 Added value of the methods used

##### Geographic clusters to characterize epidemics: a key tool for intervention

Despite the lack of a formal definition for clusters in a geographical context, the statistical approaches used here make implicit assumptions that – through different parameters – have a direct influence on cluster detection and interpretation. We used two complementary approaches that highlight different key aspects of disease clustering. Space-time scan statistics detect the geographical location of case clusters, assess their significance, and characterize their relative risk and duration. This prospective approach is particularly appropriate for the establishment of a daily surveillance system, as it identifies 'alive' clusters only, i.e. having an excess of relative risk on the day of analysis (Kulldorff, 2001). Unlike other detection methods, this approach searches for clusters without imposing the specification of their size and allows for the analysis of areas with heterogeneous population densities. Indeed, it identifies a cluster if the risk of disease within a space-time cylinder (radius = space, and height = time) is higher than outside the cluster. This information is key for public health authorities to target neighborhoods and calibrate protective or preventive measures to be deployed.

The MST-DBSCAN algorithm characterizes the diffusion dynamic of the epidemic. Here, the input parameters require a precise definition of the incubation period, the cluster transmission areas, and a minimum number of spatio-temporal neighbors required to form a cluster (Kuo, T.-H. Wen, and Sabel, 2018). The algorithm returns a typology of the evolution of transmission clusters, and identifies administrative units that are undergoing a similar diffusion process. Compared to space-time scan statistics, the MST-DBSCAN algorithm explicitly considers transmission relationships between cases but does not provide information about the cluster size (i.e., the within-number of cases) nor its statistical significance. The two approaches used together therefore allow for detailed monitoring of the disease's epidemic trajectory and populations at risk, and offer adequate tools for governments to both prioritize interventions on excess-risk locations and to develop adapted strategies to control cluster diffusion types.

### Maps reflect the chronology of the epidemic

The results displayed on static and animated maps reflect the chronology of the sanitary situation during the first wave of the epidemic. For instance, the major clusters in the Joux valley area can be clearly observed on different maps (Figure 4.6 (B,B',C,C')). Notably, these large clusters originate from a super-spreader event that took place at the end of February in a religious ceremony in Mulhouse, France. Many Swiss residents participated in this ceremony, and related clusters were then observed during the same period north of the Lausanne urban area and along the Jura mountains (e.g., in Morges and Nyon). Conversely, Lausanne was hit early-on by clusters, which is likely due to a first transmission event that occurred in Northern Italy.

Interestingly, the initial phase observed in the state of Vaud differs from what happened in Geneva, where the first clusters emerged in deprived neighborhoods eight days (March 5) after the first positive case (February 26) was detected (De Ridder, Sandoval, Vuilleumier, Azman, et al., 2021; De Ridder, Sandoval, Vuilleumier, Stringhini, et al., 2020). In Vaud, however, the initial cluster was directly detected the day of the first cases (March 4), with nine positive results in a wealthy neighborhood.

### Positive impact of soft lockdown

The soft lockdown was directly associated with a rapid reduction in the number of positive cases despite the increased rate of testing. This important reduction takes place in two clear phases in the main urban areas (see Figure 4.7 (B)), while the reduction of positive cases occurs as a succession of clusters increase and decrease in smaller urban centers and less dense areas (Figure 4.7 (D)). However, due to the time lag between the identification of positive individuals and the constitution of clusters, the cluster burden occurred directly after the implementation of the soft lockdown. Similarly, the largest clusters, the longest duration and the clusters with individuals showing large viral loads were observed just after the same time-lag. This time lag appears to be shorter in urban areas than in rural areas, likely reflecting the faster spread of the virus in large towns such as Lausanne, Morges, Nyon, Yverdon and along the Vaud Riviera. This faster spread is likely due to differences in social and cultural organization between rural and urban areas, including denser housing, which puts a higher risk of subsequent infection in families with lower socio-economic situations.

Our results highlight the efficacy of this soft lockdown strategy in controlling the epidemic and decreasing the number of positive cases. It also demonstrates the importance of acting quickly when the number of positive cases increases and not waiting for the settlement of clusters. Additionally, our results show that the relative risk remained very low throughout the lockdown period. Of note, the compliance of Swiss residents during the first soft lockdown is signaled

by the absence of any significant cluster from May 3 to June 16. Finally, it has not escaped our notice that it is already possible to observe the beginnings of the second wave from June 22, 2020 (Figure 4.2 (A)), which occurred exactly two weeks after a series of relaxations to the protective measures, including the authorization of public demonstrations of up to 300 people and the opening of nightclubs (June 6, 2020).

### 4.5 Conclusion

Our results highlight that cluster size is positively related to the presence of individuals with high viral loads, the latter being more commonly found in clusters harboring the youngest age group investigated in this study. This work also stresses the fact that cluster size and cluster duration are largely dependent on the viral load of a few number of individuals within a given cluster, underlying the impact of viral load on contagiousness.

Altogether, we provide robust data suggesting that transmission may occur despite source cases in a cluster presenting a viral load below 100,000 copies/ml. Such low viral load cases remain undetected by antigen testing, highlighting the importance of RT-PCRs assays in detecting cases and defining subsequent tracing strategies. This in-depth analysis suggests that even older at-risk individuals might get infected by SARS-CoV-2 despite active prevention, even if all cluster individuals exhibit a low viral load (below 100,000 copies/ml).

Finally, such a spatio-temporal characterization of clusters demonstrates the huge effect of the soft lockdown that took place in Switzerland from March 16 to April 27, 2020. These important results have been documented due to the contribution of the geospatial analysis of clusters.

This chapter showed that spatial epidemiology is effectively able to identify and characterize areas of high COVID-19 incidence. For seamless integration into existing surveillance frameworks, the role of geographic surveillance needs to be clearly delineated. Key questions include the extent to which geographically-defined clusters coincide with transmission clusters and how this type of surveillance can fill gaps in other methods, genomic investigations in particular.

## 5 COMBINING SPATIO-TEMPORAL AND GENOMIC ANALYSES TO IMPROVE COVID-19 SURVEILLANCE SYSTEM

In the previous chapter, we applied space-time scan and modified space-time DB-SCAN methods to identify COVID-19 case clusters and characterize cluster diffusion dynamics during the first epidemic wave in Switzerland.

Our results suggest that the use of prospective space-time scan statistics could aid decision making during epidemic periods. Notably, regions such as New York City (Greene et al., 2021) and Quebec (Lebel et al., 2021) have already incorporated these approaches into their pandemic response, for example for prioritizing testing sites and distributing protective gear.

Central to the infectious disease management, the test-trace-isolate-quarantine (TTIQ) strategy aims to rapidly identify and isolate infected individuals using methods such as contact tracing (Ashcroft, Lehtinen, and Bonhoeffer, 2022). For a well-rounded response to the epidemic, this strategy is usually complemented by other surveillance systems. These include genomic surveillance to understand the spread of the virus and identify circulating variants, syndromic surveillance for early detection of outbreaks, sentinel surveillance to track infection trends, serologic surveillance to assess population immunity, or environmental surveillance to monitor the presence of pathogens in common spaces such as water. Yet, within this broader surveillance framework, the importance of geographic surveillance is not fully recognized. This oversight may hinder its integration into national or regional surveillance programs.

In this chapter, we address this gap by exploring the synergy of geographic and genomic surveillance. Building on the results of the previous chapter, we conducted a genomic analysis of representative space-time scan clusters from the first epidemic wave. Our main objectives were: 1) to determine the overlap between geographic (or spatial) clusters and transmission (or epidemiologic) clusters, and 2) to propose a combined approach using both methods for enhanced infectious disease surveillance.

## Chapter 5. Combining spatio-temporal and genomic analyses to improve COVID-19 surveillance system

---

This chapter is a postprint version of the manuscript published in *Frontiers in Public Health*:

Choi, Y.<sup>1</sup>, Ladoy, A.<sup>1</sup>, De Ridder, D., Jacot, D., Vuilleumier, S., Bertelli, C., Guessous, I., Pillonel, T., Joost, S., & Greub, G. (2022). Detection of SARS-CoV-2 infection clusters: The useful combination of spatiotemporal clustering and genomic analyses. *Frontiers in Public Health*, 10. <https://www.frontiersin.org/articles/10.3389/fpubh.2022.1016169>

As co-first author of this publication, my contributions included study design, selection of spatial clusters for SARS-CoV-2 genome sequencing, coordination of statistical analyses, geographic visualization of results, and co-drafting of the manuscript.

### 5.1 Introduction

The extreme rapidity of the COVID-19 pandemic revealed the importance of developing, and strengthening, public health surveillance systems at both international, national, and regional levels (Krieger et al., 2020). Defined as “the ongoing, systematic collection, analysis, and interpretation of health data essential to the planning, implementation and evaluation of public health practice” (Thacker and Berkelman, 1988), an effective public health surveillance system must be able to monitor the spatial and temporal spread of a disease in a timely manner, to quickly detect emerging clusters of infection and cut chains of transmission (Budd et al., 2020).

In this context, spatiotemporal approaches that investigate disease clustering, such as prospective space-time scan statistics (Kulldorff, 2001), can constitute an integral part of such surveillance systems by systematically detecting emerging clusters of disease that require further investigations. Fundamentally, space-time scan statistics test whether the number of temporally close cases observed in a defined area exceeds the expected number according to the underlying at-risk population. In the context of the COVID-19 pandemic, several studies investigated how prospective space-time scan statistics could contribute to the ongoing surveillance of the pandemic at different spatial levels including a country-wide investigation using publicly available data across the United States of America (Desjardins, Hohl, and Delmelle, 2020), as well as investigations at higher spatio-temporal resolutions using laboratory test results to detect COVID-19 clusters in a Swiss state (Ladoy, Opota, et al., 2021) and in New York City (Greene et al., 2021). A limitation of using these approaches is that they rely on health data that are usually geocoded to a patient’s residential location, which constitutes only one part of virus transmission. Therefore, it may limit the ability of these scan statistics to depict epidemic trajectories and break the infection transmission chain. Some studies have investigated the interplay between geographical and transmission clusters in the context of sexually transmitted diseases (Skaathun et al., 2021; Lubelchek et al., 2015), but this research

---

<sup>1</sup>Yangji Choi and Anaïs Ladoy contributed equally to this work.



question has not been studied, to our knowledge, in the context of COVID-19.

At the same time, the role of genomics has become critical in the public health domain during the SARS-CoV-2 pandemic. The first SARS-CoV-2 genome sequences allowed the scientific community to characterize the virus and understand its zoonotic origin, infection and transmission mechanisms, as well as COVID-19 pathogenesis (Harrison, Lin, and P. Wang, 2020; Wu et al., 2020). Sequencing data also enabled biotechnology companies and pharmaceutical companies to quickly develop molecular diagnostic assays and vaccines. Virus genomes from infected individuals were constantly sequenced and submitted to public national (Attwood et al., 2020) and international (Maxmen, 2021) databases (e.g. GISAID database), forming hubs for SARS-CoV-2 genomic data sharing that assisted worldwide collaborations and standardized lineages definition (Lo and Jamroz, 2020). In parallel, many open-source bioinformatic tools were actively developed, to compare virus genomes, define and assign lineages, facilitating epidemiological investigations. Based on the plentiful open data and bioinformatic tools, numerous SARS-CoV-2 genome-based studies identified new variants of concern (Geoghegan et al., 2021; Di Giallonardo et al., 2020; Qutob et al., 2021) and tracked geographic transmission of the virus (Tegally et al., 2021; Zeller et al., 2021; Yi et al., 2021; Kraemer et al., 2021; Lai et al., 2022; Stange et al., 2021) in different countries. Although we found numerous studies tracing the origin and evolution dynamics of the COVID-19 pandemic, very few studies examined how genomic sequencing could be used for informed-decision making within an actionable time frame (Meredith et al., 2020; Lane et al., 2021).

In this context, our study aimed to investigate: (i) to what extent clusters identified by space-time scan analysis are confirmed as transmission clusters based on SARS-CoV-2 genome sequences, (ii) how genomic-based approaches can improve the epidemiological investigation associated with spatiotemporal clusters, and (iii) how can a combination of both complementary approaches be used in the context of infectious disease surveillance systems. To answer these questions, we sequenced the SARS-CoV-2 genomes of 172 cases contained in a set of spatiotemporal clusters identified in the Swiss state of Vaud during the first epidemic wave in Switzerland (Ladoy, Opota, et al., 2021). We then analyzed genetic similarity among cases within spatiotemporal clusters and spatiotemporal distribution across virus genotypes using different bioinformatic tools to better understand discrepancies and possible synergies between genomic-based and spatiotemporal clustering approaches.

## 5.2 Materials and methods

### 5.2.1 Study design

We previously described the spatiotemporal spread of COVID-19 during the first wave of the pandemic for the state of Vaud, Switzerland, using a prospective space-time scan analysis

## Chapter 5. Combining spatio-temporal and genomic analyses to improve COVID-19 surveillance system

---

(Ladoy, Opota, et al., 2021). Briefly, the analysis was performed on 3,317 individuals who were tested (RT-PCR) positive for SARS-CoV-2 between March 2 and June 30, 2020, geocoded to their residential address. The study was approved by the *Commission cantonale d'éthique de la recherche sur l'être humain (CER-VD)*, Switzerland (n°2020-01302). Spatiotemporal clusters were detected daily by comparing the number of observed cases to the expected number within and outside a circular window of varying sizes. Expected cases were estimated with a Poisson model adjusting for population size at the inhabited hectare level, and the analytical window was defined to contain a maximum of 0.5% of the population at-risk and last a maximum of 14 days.

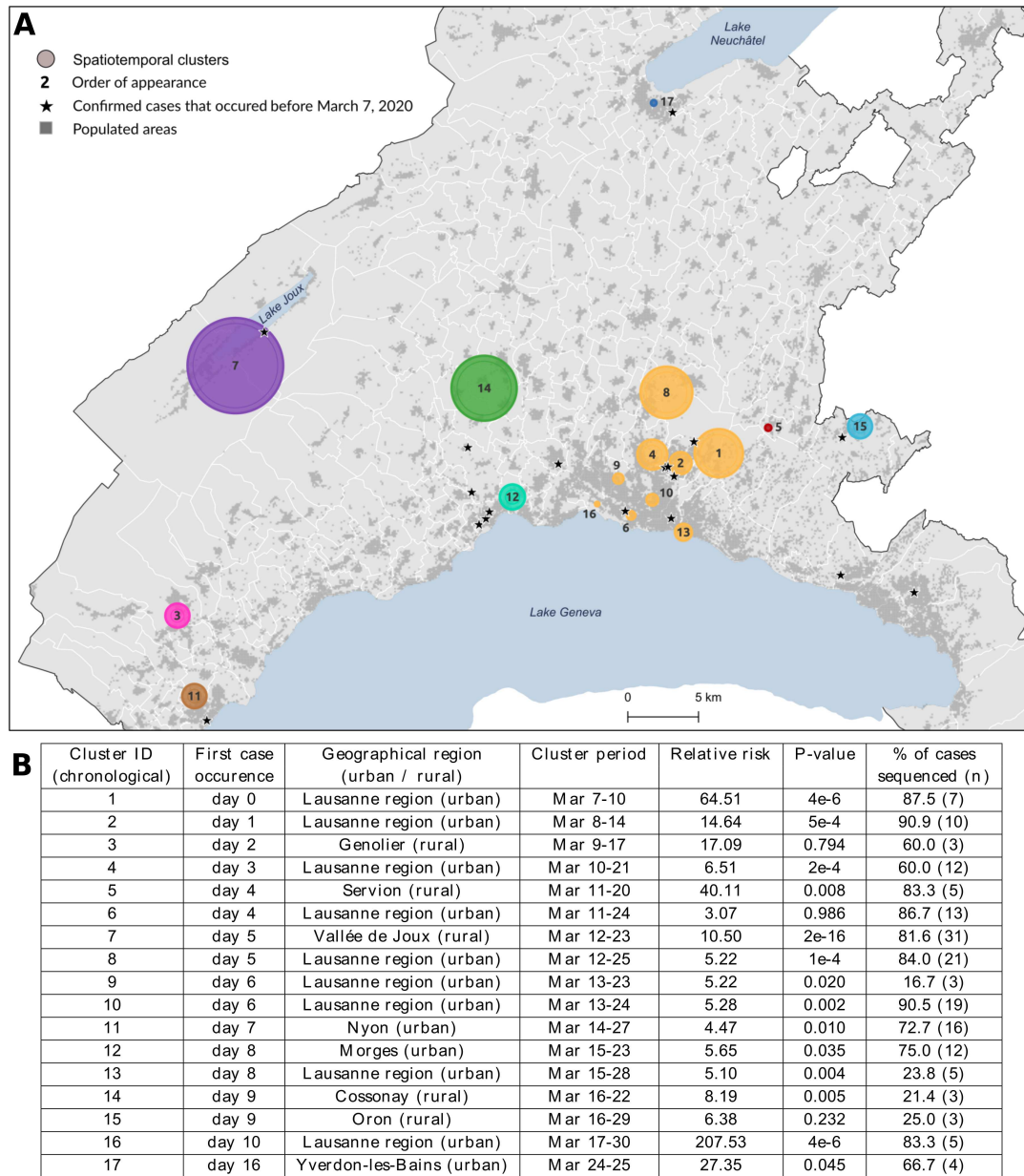
Of the 1,784 spatiotemporal clusters identified (454 with a p-value < 0.05), we selected 17 clusters for further investigation (Figure 5.1). This small number of clusters is partly explained by the many overlapping clusters due to analysis frequency. The selected clusters were chosen to be representatives of the spatial footprint and temporal variations obtained during the first wave of the pandemic, to allow for the comparison of different settings. We chose clusters from different geographical settings (urban vs. rural), of different sizes in terms of geographical coverage and number of cases, as well as some with unique particularities. Additionally, for clusters that were detected several days in a row (i.e., overlapping clusters), we selected the last appearance in order to increase the time span of analysis, even if the last occurrence was not necessarily significant (clusters #3, #6, #15 in Figure 5.1). The cluster selection process is depicted in Appendix, Figure A.5.1.

### 5.2.2 SARS-CoV-2 genome sequencing

We sequenced the SARS-CoV-2 genome of all cases presenting over 10,000 cp/ml from the 17 clusters to investigate the genetic similarity within spatiotemporal clusters. SARS-CoV-2 RNA was extracted from nasopharyngeal swabs (COPAN UTM medium, 3.5 ml) using the MagNA Pure 96 system (Roche, Basel, Switzerland). The viral genomes were amplified by the CleanPlex SARS-CoV-2 panel (Paragon Genomics, SKU 918011) following the manufacturer's instructions (Kubik et al., 2021). The quality of amplified products was assessed by Fragment Analyzer standard-sensitivity NGS (DNF-473; AATI) and quantified using Qubit standard-sensitivity double-stranded DNA (dsDNA) kit (Q32853; Invitrogen). The amplicons were sequenced by 150 bp paired-end reads on a MiSeq (Illumina, San Diego, CA). To evaluate sequencing quality, negative and positive internal controls were included in each run.

### 5.2.3 Reads processing and quality control

Reads were processed with GENCOV pipeline (<https://github.com/metagenlab/GENCOV>), modified from CoVpipe ([https://gitlab.com/RKIBioinformaticsPipelines/ncov\\_minipipe](https://gitlab.com/RKIBioinformaticsPipelines/ncov_minipipe)), in



**Figure 5.1 – Spatial distribution (A) and characteristics (B) of the 17 spatiotemporal clusters considered for genomic data analysis.** These clusters were identified using a space-time scan statistic run daily from March 2 to June 30 and implemented with SaTScan version 9.6.1 (Kulldorff, 2018). Characteristics include each cluster identifier with its corresponding geographical region, cluster period, relative risk of becoming infected to COVID-19 within the cluster compared to outside, significance evaluated with 999 Monte-Carlo permutations, and the proportion of sequenced cases within cluster. Clusters are colored according to the geographical region to which they belong.

## Chapter 5. Combining spatio-temporal and genomic analyses to improve COVID-19 surveillance system

---

order to perform sequence filtering with fastp (S. Chen et al., 2018), primer trimming with fgbio (Fennel, Homer, and Genomics, 2022), mapping to the reference genome NC\_045512.2 with bwa (H. Li, 2013), alignment evaluation with Qualimap (Okonechnikov, Conesa, and García-Alcalde, 2016), and variant calling with Freebayes (relative number of variant supporting reads = 0.1, minimal depth = 10, absolute number of variant supporting reads = 9) (Garrison and Marth, 2012). Variants were further filtered by bcftools (Westgard et al., 1981), determining the consensus based on the variants supported by more than 70% of mapped reads, whereas positions covered by fewer than 10 reads were masked with Ns. The consensus sequence was assigned to SARS-CoV-2 lineages using Pangolin (O'Toole et al., 2021). The quality of SARS-CoV-2 genome sequences was then manually evaluated according to quality criteria as described by Jacot, Pillonel, et al. (2021), including mutations supported by 10–70% of mapped reads termed “low-frequency variants”. Genome sequences that did not pass quality criteria were repeated.

### 5.2.4 Genomic analyses

Pairwise single nucleotide variant (SNV) distances were computed from quality-checked sequences using Nextstrain SARS-CoV-2 multiple sequence alignment (<https://github.com/nextstrain/ncov>) (Hadfield et al., 2018) and pairsnp (<https://github.com/gtonkinhill/pairsnp>). Based on the pairwise SNV matrix, we computed the Jaccard similarity index (Jaccard, 1912) to quantify genetic similarity within spatiotemporal clusters, by calculating the size of the intersection divided by the size of the union of SNVs. Jaccard similarity index was computed for each pair of genomes within the same cluster. Sets of samples with identical SARS-CoV-2 genome sequence (0 SNV distance) were defined as “genomic groups”.

### 5.2.5 Genomic and geographic visualization

Phylogenetic analysis and visualization were conducted with Augur and Auspice, respectively, which are parts of Nextstrain that allows for customization and interactive web visualization (Hadfield et al., 2018). The relationships among genomic groups and samples with unique genome sequences were visualized as minimum spanning trees (MST) on Cytoscape (Shannon et al., 2003), as demonstrated in Figure A.5.2. The network was computed with the optrees package in R (<https://github.com/cran/optrees>) adopting Prim's algorithm, which finds the shortest path by selecting a subset of the edge such that a spanning tree is formed with the minimal total weight of the edges (Chang et al., 2021). Each node represents either a genomic group or an individual sequence and the weight of the undirected edges reflects SNVs. The mapping of genomic groups within clusters was done using QGIS 3.22 (QGIS.org, 2022. QGIS Geographic Information System. QGIS Association. <http://www.qgis.org>).

## 5.3 Results

### 5.3.1 Description of selected spatiotemporal clusters

We investigated the genetic similarity within a set of 17 spatiotemporal clusters selected from a previous study (Ladoy, Opota, et al., 2021). Clusters were detected from March 7 to March 30, 2020, and lasted from 2 to 14 days, corresponding to the lower and upper bound values of the temporal window used in the analysis. The clusters' geographic location and characteristics are shown in Figure 5.1, where clusters are labeled according to their chronological occurrence. Forty percent of clusters ( $n = 7$ ) were in rural areas or intermediate-size cities, but the first cluster detected (#1) occurred in the Lausanne region, the capital of Vaud state. Cases of COVID-19 had already been declared in Vaud state a few days before the commencement of the study (the first case occurred on March 3), but this did not form any cluster. Their locations are starred in Figure 5.1 (A).

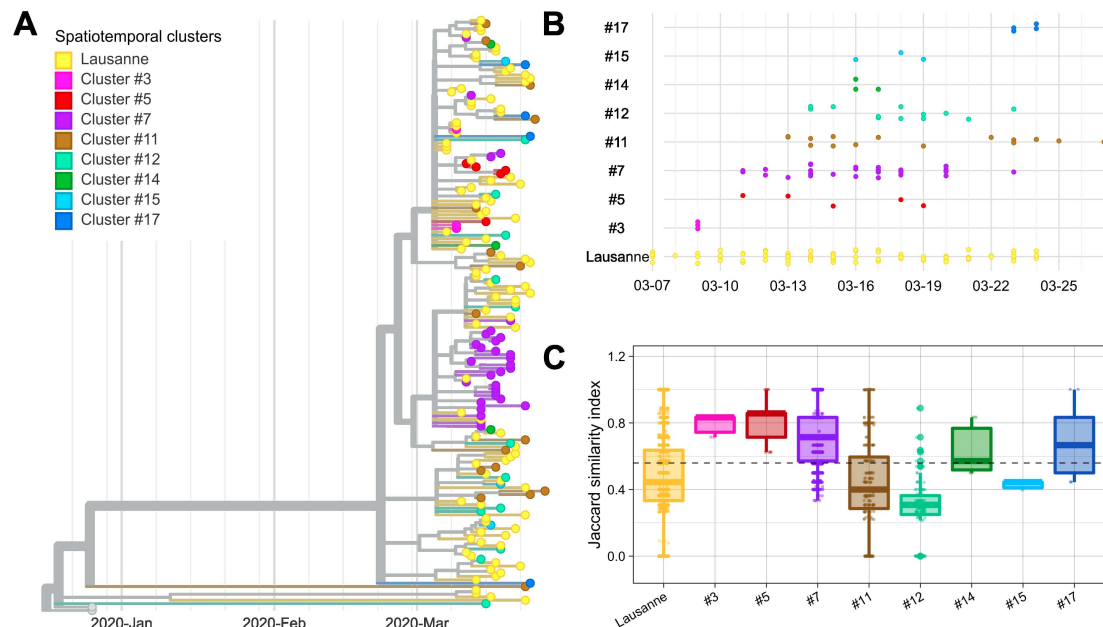
While the clusters included 264 lab-confirmed RT-PCR positive cases, only those with a viral load above 10,000 copies/ml ( $N = 172$ , 65.4%) could be sequenced (see the proportion by cluster in Figure 5.1 (B)), though this did not affect characterization of the affected populations. The number of cases within clusters varied from 3 to 38 (cluster #7), where individuals were 52.3% female, with a mean age of 57.2 years ( $\sigma = 20.2$ ). Detailed characteristics per cluster are provided in Table A.5.1. Infected individuals in rural areas tended to be older (median age 73 vs. 54 years,  $p$ -value  $< 0.001$ , Wilcoxon) with a lower mean viral load (230 vs. 590 million copies/ml,  $p$ -value = 0.04, Wilcoxon) when compared to individuals in urban areas.

The nine clusters within Lausanne metropolitan area (#1, #2, #4, #6, #8, #9, #10, #13, #16, a total of 94 cases) were labeled uniformly as the "Lausanne region" to reduce the complexity of representation. This choice was reinforced by the distinct patterns observed between these urban clusters and the rest of the state.

### 5.3.2 Genetic similarity within spatiotemporal clusters

In order to verify whether space-time clusters were transmission clusters based on SARS-CoV-2 genome sequences, we explored the genetic heterogeneity among 172 cases, within and between space-time clusters. The evolutionary relationships among SARS-CoV-2 genomes included in different spatiotemporal clusters were first examined using a phylogenetic tree (Figure 5.2 (A)). Overall, most spatiotemporal clusters did not appear as a monophyletic group on the phylogenetic tree. However, most cases in cluster #7 appeared on the same branch together, as did all cases in cluster #3 and cluster #5 that appeared at the very beginning of the outbreak, seven days or more before the peak of the epidemic curve (March 18) (Figure 5.2 (B)). Similarly, the sub-clusters within the Lausanne region did not show any clear clustering

on the phylogenetic tree, except for the last Lausanne cluster (cluster #16), which occurred after the lockdown (March 16).



**Figure 5.2 – Genetic distance between and within spatiotemporal clusters.** (A) Phylogenetic tree with 172 sequences and 2 Wuhan reference genomes. (B) Timeline of cases appearing in geographical regions. (C) Jaccard similarity within geographical regions. Jaccard distance was calculated between all pairs of samples in the same region. The overall median is indicated as a dotted line. The geographical regions were ordered based on the date of the first case in the region.

We compared the genetic homogeneity among spatiotemporal clusters, where the genetic similarity between pairs of samples was quantified with the Jaccard similarity index. In general, intra-cluster genetic similarity was higher in rural regions than in urban areas ( $p$ -value  $< 2.2 \times 10^{-16}$ , Wilcoxon). The genetic similarity was greater than the median in four clusters (clusters #3, #5, #7, #17) (Figure 5.2 (C)). Cluster #3, #5 and #7 are early-appearing clusters that aggregated in the phylogenetic tree and showed the highest Jaccard genetic similarities. They were followed by cluster #17, which occurred in the second largest city of Vaud at the end of the first epidemic wave, after the lockdown (March 16). Clusters #11, #12 and #15 with the lowest genetic similarity appeared after March 11 (close to the peak;  $< 7$  days).

The Lausanne region, the largest urban area of Vaud, showed low similarity among cases compared to the median (Figure 5.2 (C)). Although the genetic similarity of the nine clusters forming the Lausanne region remained relatively constant at low levels throughout the

timeline, the genetic similarity varied over time, showing a similar pattern as other clusters with a decrease in similarity toward the peak of contaminations, and an increase back the lockdown (Figure A.5.3). Interestingly, Lausanne cluster #8 exhibited a significantly lower Jaccard similarity compared to cluster #7, located in the mountainous areas in the north-west of the state, even though they appeared on the same day (p-value < 22e-16, Wilcoxon) (Figure 5.2 (C)).

### 5.3.3 Comparison of spatiotemporal clusters and genomic groups

We further investigated the genetic divergence of geographical clusters at single nucleotide variant (SNV) level (Figure 5.3 (A)). The distance in SNVs compared to the Wuhan reference genome varied between 2 to 13 mutations. The first cases in the Lausanne region (in cluster #1) harbored 5 SNVs, while some later cases showed fewer mutations (2 or 4 SNVs). Among the 172 SARS-CoV-2 genomes, we identified 20 sets of cases carrying identical genomes, hereafter referred to as “genomic groups” (Figure 5.3 (B)), in order to avoid confusion with geographical clusters. These 20 genomic groups include 101 of the 172 cases (group1: 37; group2: 12; group3: 6; group4 and group5: 5 each; group6 and group7: 4 each; group8 and group9: 3 each; group10-group20: 2 each). The other 71 genomes did not belong to any genomic group as they exhibited unique sequences (“singletons”). The genetic relationships among the 20 genomic groups and the 71 singletons were visualized on a minimum spanning tree network (Figure 5.4). This can be visualized with Figure 5.5, which shows the distribution of genomic groups within spatiotemporal clusters.

We identified 12 genomic groups of the 20 that were restricted to a single area (Lausanne region: 6, cluster #5: 2, cluster #7: 2, cluster #11: 2) and eight genomic groups that consisted of individuals living in two to seven different regions that always included at least one individual from the Lausanne region (Figure 5.4). For four spatiotemporal clusters, all cases were attributed to the same genomic group (clusters #1, #3, #15, #17) (Figure 5.5). We observed spatial heterogeneity within clusters, yet, unsurprisingly, cases that occurred in the same building usually shared the same genomic group (Figure A.5.4). The size of these multi-regional genomic groups varied between two to 37 cases. Group1 and group2 were the largest groups, with 37 and 12 cases, respectively. We investigated these groups further as there may have been superspreading events in each group. Group1 cases were split into seven geographical regions, connected to cases in the same cluster by one or two SNVs distance (Figure 5.4). Three of the Lausanne region cases in group1 and one case with 1 SNV distance from group1, which occurred on day 0, could represent the origin of the superspreading event that formed group1 cases. Group2 likely started with one case in the Lausanne region that was diagnosed on day 4, followed by 11 cases in spatiotemporal cluster #7 (in the mountainous north-west region). Although both group1 and group2 were identified as lineage B.1, sharing four nu-

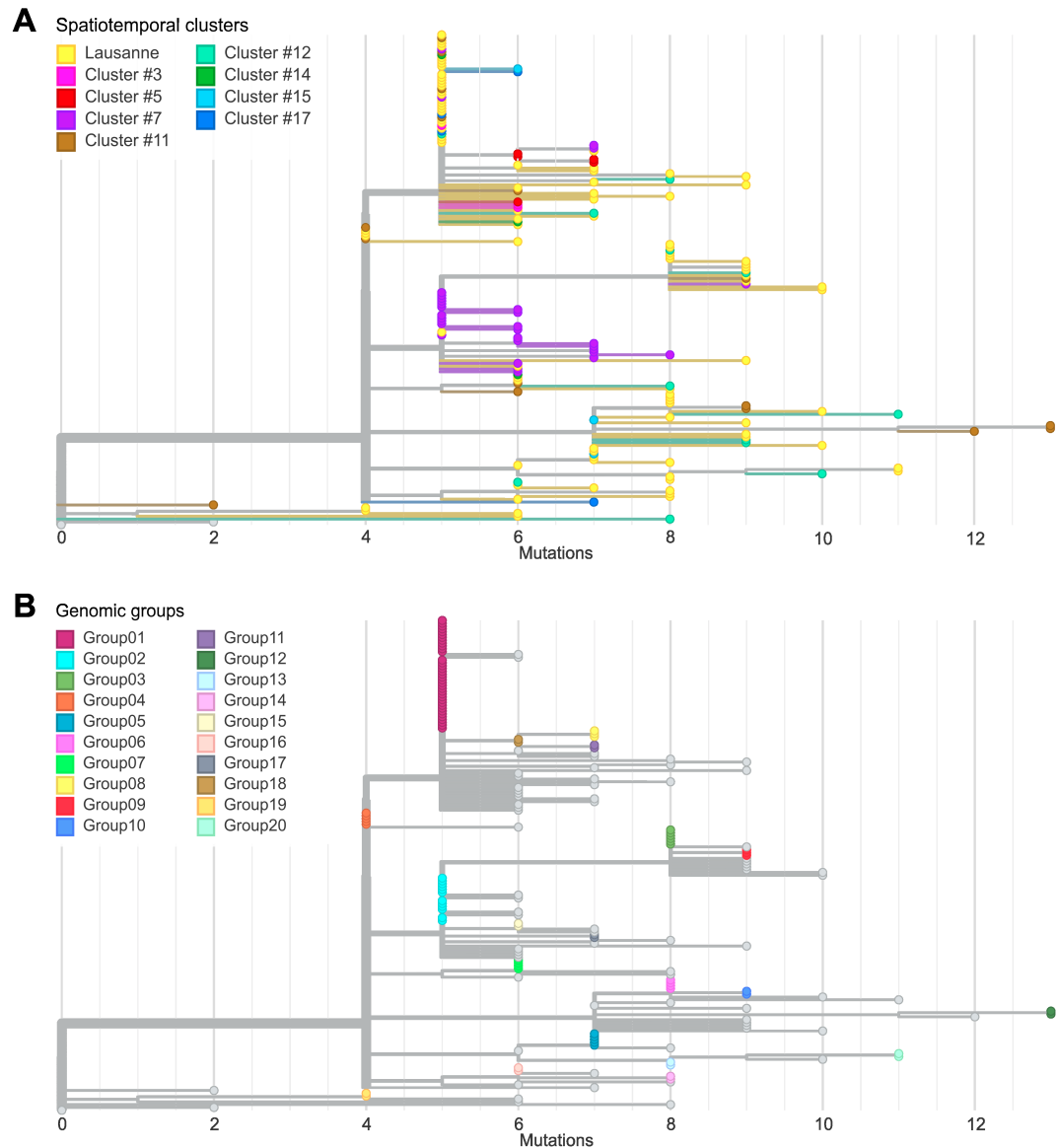


Figure 5.3 – **Comparison between spatiotemporal clusters and genomic groups in phylogenetic trees.** (A) Divergence of cases in different geographical regions at SNV level. (B) Definition of genomic groups. Cases with identical genome sequences were assigned to a genomic group. Overall 20 genomic groups were identified with varying numbers of cases within each group. The rest of the cases have unique sequences named singletons.



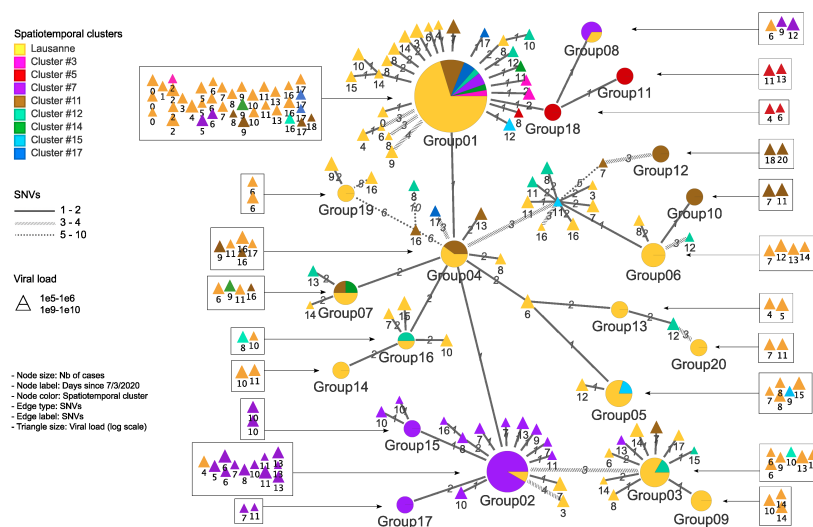


Figure 5.4 – **Genetic relationship among genomic groups and singletons in minimum spanning tree.** Nodes and edges indicate unique sequences and SNV distance. The number of cases in genomic groups is represented by the size of the pie. Genomic groups consist of cases in different geographical regions. The triangles are single cases with their size proportional to the log value of viral load detected by qPCR. They are represented in squares according to their occurrence in time.

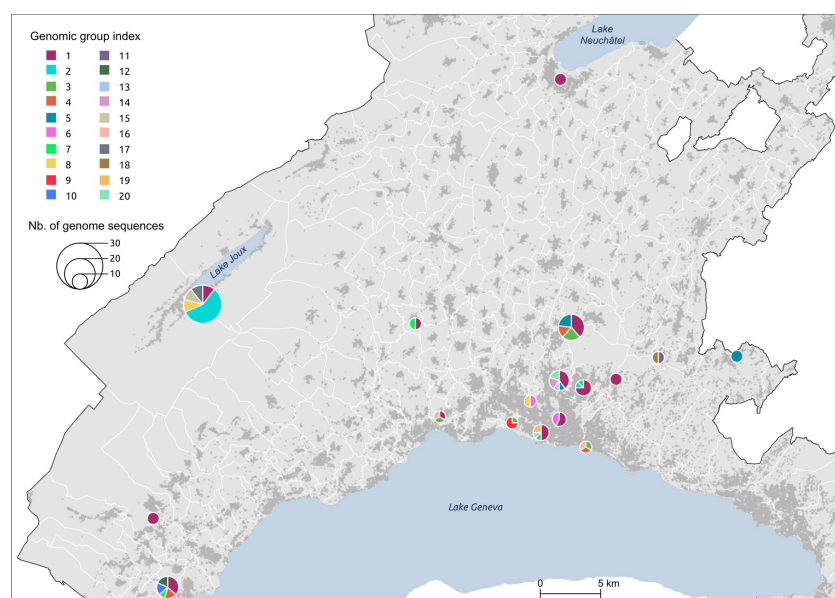


Figure 5.5 – **Distribution of genomic groups within clusters.** The size of the circle is proportional to the number of cases.

cleotide mutations (C241T, C3037T, C14408T, A23403G), each group was characterized by a specific mutation (Table A.5.2). The mutation C15324T characterized exclusively group1 and all group1-associated cases (or groups), except for group4. Likewise, the mutation A26530G featured only group2 and its neighbors (Figure 5.4).

## **5.4 Discussion**

### **5.4.1 Congruence between two approaches in different contexts**

Although the distinct use of spatiotemporal clustering and genomic-based approaches for COVID-19 management is recognized in the literature, we did not find any study investigating how the combined use of these two methods could compensate for their respective shortcomings in a surveillance context. By investigating the extent to which spatiotemporal clusters were confirmed as transmission clusters based on SARS-CoV-2 genome sequences, our results suggest that the consistency across the two methods might vary according to geographic characteristics of the area (rural/urban) and the epidemic context.

We often found less genetic similarity within clusters in urban areas compared to rural areas (p-value  $< 2.2e-16$ , Wilcoxon). This could be explained by differences in social activities and population mobility. In rural areas, we expect many close contacts to occur among a few people from the same village, where a single introduction event might spread quickly with fewer opportunities to acquire new variants. As infected individuals in rural clusters were significantly older (p-value  $< 0.001$ , Wilcoxon), the genetic similarity within spatiotemporal clusters could also possibly be associated with restricted mobility of elderly people. In contrast, urban areas have numerous factors that could multiply the risk of simultaneous circulations of multiple variants, such as more frequent use of public transportation and larger places of gathering (Chang et al., 2021). Within spatiotemporal clusters, cases located in the same building were generally epidemiologically linked, as they often stemmed from within-household transmission events. Transmission in densely inhabited structures, such as cluster #16 that occurred in a migrant center after lockdown, resulted in significantly higher genetic similarity than other clusters in the Lausanne region (Figure A.5.3).

Moreover, the congruence between spatiotemporal and transmission clusters appeared to vary along the epidemic curve. The genetic similarity was typically higher during the lockdown and at the very beginning of the pandemic, where only a few cases were detected, than during the epidemic peak. As no study to our knowledge has examined the congruence of space-time scan and genetic clustering for SARS-CoV-2, it is difficult to interpret our findings in light of other publications. However, several studies have investigated similar research questions in the context of sexually-transmitted diseases. For example, authors found that space-time scan clustering was less successful than genetic clustering in identifying HIV-transmission patterns

in small or urban HIV-endemic areas of Los Angeles County (Skaathun et al., 2021), while a study in the Netherlands observed a higher incidence of Hepatitis B associated with higher genetic clustering in rural areas (Soetens et al., 2015). However, even if similar patterns were observed in our study, the marked differences in disease characteristics do not permit a direct comparison.

In both genomic group1 and group2, the first cases from the Lausanne region seemed to spread in many neighboring areas, including a geographically isolated area (cluster #7), showing the significant impact of urban areas and superspreading events. Genomic group1 and group2, assigned to B.1 lineage, were differentially characterized by the mutations C15324T and A26530G, respectively. First, the mutation C15324T was suspected to originate from Mulhouse (France) according to Stange et al. (Stange et al., 2021), where the first case with an identified source of infection was from a religious gathering in Mulhouse. This mutation was the main feature of that local cluster ("Basel-city") in the early period of the first wave. Moreover, the mutation C15324T was found in other countries, mostly France and Luxembourg at considerable proportions (18.70% and 20.69% of population sequenced, respectively), but not in Italy (until 23rd March 2020). Second, the mutation A26530G was mentioned by Alteri et al. (2021) as a key feature of the early Lombardy (Italy) cluster, with >90% of intra-patient prevalence circulating mid-February. It was assumed to be the origin of the subsequent transmission chain in the Lombardy region based on its small number of foreign sequences at the bases of the transmission chain. Thus, we hypothesize that superspreading events in genomic group1 and group2 might stem from secondary cases of Mulhouse and Milan outbreaks, respectively.

The major strength of the present study lies in the fine-scale resolution of the analysis, and the high-quality dataset used to investigate the interplay between genomic and spatiotemporal clustering approaches. At the beginning of the pandemic, the Institute of Microbiology of Lausanne University Hospital received all samples from Vaud state ensuring a comprehensive coverage of all cases in the area within the time frame studied here. This was rarely achieved in most other regions that commonly had multiple testing and sequencing centers, which makes it difficult to obtain an in-depth overview of the local epidemiology. However, the sampling of individuals could be biased due to untested individuals, likely leading to underestimates of superspreading events. Indeed, at the beginning of the pandemic, only symptomatic individuals were tested, although asymptomatic but contagious individuals could have contributed to the spread of the virus. Furthermore, only a portion ( $n = 172$ ; 8%) of total positive cases were sequenced in the present study, which could affect the generalization of our results. In comparison, Brüningk et al. (2022) sequenced 40% ( $n = 247$ ) of the positive cases in the city of Basel, providing a much higher resolution but limited to a single town. As a tradeoff between the size of the study area and the sequencing density, our choice was partly dictated by the objective of comparing transmission within rural and urban settings, which is rarely done. In addition, the mobility restrictions (e.g., lockdown, homeworking, restaurants closure) and

the limited genomic distances observed during the early pandemic could inflate the genetic similarity observed within spatiotemporal clusters. Novel analyses using data from successive waves might refine our findings.

### **5.4.2 Combining genomic and spatiotemporal clustering approaches in infectious disease surveillance**

Timing is a crucial factor in any surveillance system. Space-time scan statistics can be run automatically as soon as new data arrive and in near real-time using the SaTScan software (Kulldorff, 2018) in batch mode. It constitutes, therefore, a powerful exploratory approach to detect high-incidence areas where authorities could prioritize cases for genome sequencing and contact tracing. The New York City Department of Health and Mental Hygiene already adopted this approach to prioritize interviews of patients and develop targeted actions for testing and prevention (Greene et al., 2021; Arnold, 2022). Our results suggest that one could restrict investigations to a smaller number of cases for clusters in rural areas or within the same building due to the high probability of epidemiological linkage, but also that during peak period, spatiotemporal clusters do not necessarily indicate transmission clusters. Because there are now multiple providers for COVID-19 testing, the space-time scan analysis should use newly reported infectious disease cases to regional authorities, a mandatory procedure in Switzerland. The input parameters should be fine-tuned following the recommendations from Greene et al. (2021), for example, by considering the number of tests rather than the total population as the underlying at-risk population to consider changes in testing rates.

An optimal framework for infectious disease surveillance may also be complemented by other approaches. Wastewater monitoring can give a reasonable estimate of infection level and circulating variants taking into account asymptomatic patients (Jahn et al., 2022), while epidemiological models can make projections about epidemic trajectories and healthcare capacity and estimate intervention scenarios (Lemaitre et al., 2021). Incorporating data from mobility patterns using, for example, aggregated mobile phone data (Kraemer et al., 2021), could also improve the spatiotemporal analysis of COVID-19 dynamics, allowing for the detection of infections outside the residential neighborhood, such as at work or activity sites. Even though our study was limited to SARS-CoV-2, we could imagine a similar framework for the Monkeypox virus surveillance, where space-time scan statistics (Mandja et al., 2019) and phylogeographic investigation (Nakazawa et al., 2015) were already used to disentangle disease dynamics.

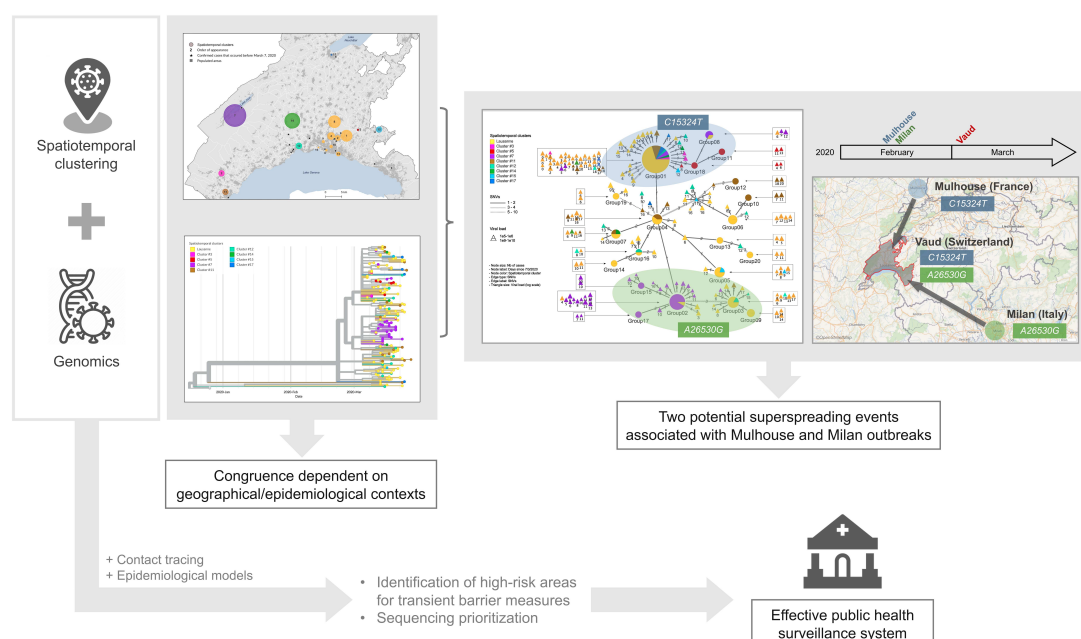


Figure 5.6 – Graphical representation of findings and conclusion.

## 5.5 Conclusion

Spatiotemporal clustering and genomic approaches have been extensively used during the COVID-19 pandemic. The former approach was mainly used to identify high-incidence areas to target immediate interventions and to draw hypotheses about vulnerable populations, while the latter allowed for tracking of the origin, transmission, and evolution of the SARS-CoV-2 virus globally, and to understand host susceptibility, response, disease severity, and outcomes. In addition to the silos existing between researchers mastering each approach, spatiotemporal methods are limited by the fact that they usually consider only one source of virus transmission (i.e., the residential setting), while genomic studies require significant resources and processing time, which could delay decision-making (Table A.5.3). Our genomic investigation of spatiotemporal clusters showed that the clusters identified by space-time scan statistics were more likely to be epidemiologically linked in rural areas and outside the epidemic peak. In addition, we identified two potential superspreading events, characterized by specific mutations indicating their respective origins from two major outbreaks in Europe at the beginning of the pandemic. These findings suggest that we could save considerable resources and improve the efficiency of the public health surveillance system by synergizing both approaches, and prioritizing genome sequencing and contact tracing in high-incidence areas detected using spatiotemporal clustering approaches (Figure 5.6).

## **Chapter 5. Combining spatio-temporal and genomic analyses to improve COVID-19 surveillance system**

---

Recently, SARS-CoV-2 genomic surveillance has gradually reduced (Nature, 2022). Without the ability to track the virus, and while much of the world remains unvaccinated, we are unlikely to make targeted public health decisions in the face of potentially threatening new variants. We must remember the lessons from the first wave of the pandemic, when lack of data and knowledge caused societal distress, and avoid returning to such a situation by maintaining genomic-based surveillance efforts, conjointly with spatiotemporal surveillance.

The previous two chapters have demonstrated the contribution of spatial epidemiology to COVID-19 monitoring, particularly in synergy with genomic surveillance. Spatial approaches should also be considered for their key role in vaccine deployment planning, particularly in guiding the location of vaccination centers to lean towards equitable access to vaccines for the whole population.

## 6 PRIORITIZATION OF POPULATIONS FOR MOBILE COVID-19 VACCINATION CAMPAIGNS

### 6.1 Introduction

The COVID-19 pandemic has exacerbated pre-existing social, racial, and economic health inequities (Bambra et al., 2020; Dorn, Cooney, and Sabin, 2020). In Switzerland, a comprehensive nationwide study investigated inequalities across the COVID-19 care continuum using surveillance data reported to the Swiss Federal Office of Public Health from March 2020 to April 2021 (Riou et al., 2021). The authors found that individuals living in socioeconomically disadvantaged neighborhoods were less likely to be tested for COVID-19, but more likely to test positive, be hospitalized, and succumb to the virus. These findings underscored the manifestation of the inverse law of care, which states that the populations most in need of health care are generally those who received the least (Hart, 1971).

The pathways leading to these inequities are multiple. Communities in socioeconomically disadvantaged neighborhoods, or socially marginalized groups, bear a disproportionately higher burden of the underlying clinical risk factors that exacerbate the severity and mortality of COVID-19 (Bambra et al., 2020). In addition, individuals living in deprived neighborhoods may face higher rates of infection due to overrepresentation in the essential workforce and more crowded living conditions. These challenges extend to impaired accessibility to testing due to limited availability of private transportation or inability to take time off work (Riou et al., 2021).

While the development of COVID-19 vaccines marked a critical milestone in controlling the course of the pandemic and protecting vulnerable populations (Watson et al., 2022), scientists quickly raised concerns about inequitable vaccination coverage both between and within countries (Bergen, Johns, et al., 2023). In particular, populations already burdened by the COVID-19 pandemic experienced low vaccination rates, potentially perpetuating health in-

## Chapter 6. Prioritization of populations for mobile COVID-19 vaccination campaigns

equities (Heiniger et al., 2022; Sacarny and Daw, 2021; J. Whitehead et al., 2022). Consequently, it became imperative for public health authorities to develop vaccination campaigns that comprehensively address barriers impeding vaccine access. These barriers are broadly categorized into two types: structural and attitudinal. Structural barriers encompass systemic obstacles that impede individuals' access to vaccination services, while attitudinal barriers, closely linked to vaccine hesitancy, include perceptions and beliefs that may influence patient willingness to pursue vaccination (Fisk, 2021; Kuehn et al., 2022). Table 6.1 enumerates examples of both structural and attitudinal barriers, specifically in the context of COVID-19 vaccination.

In the canton of Vaud, the vaccination campaign began on December 30, 2020, initially within nursing homes and then expanding to the broader resident population based on age thresholds (see Figure 6.1 for the campaign timeline). In Spring 2021, the authorities of the canton of Vaud pioneered an innovative vaccination strategy that involved deploying mobile vaccination centers throughout the canton. These centers were designed to be accessible to the entire population, with or without an appointment, and free of charge. The idea behind this campaign was to "go to meet eligible individuals rather than having them come to the 14 cantonal [stationary vaccination] centers, pharmacies, or doctors' offices" (Béatrice Métraux, State Councillor (Nicollier, 10.05.21)).

Barriers	References
STRUCTURAL BARRIERS	
Geographic and functional proximity to vaccines	(Kuehn et al., 2022; Bergen, Kirkby, et al., 2023)
Technology-related barriers (e.g., issues with scheduling appointments online)	(Kuehn et al., 2022; Bergen, Kirkby, et al., 2023)
Limited schedule flexibility for vaccination	(Kuehn et al., 2022; Bergen, Kirkby, et al., 2023)
Language barriers (e.g., lack of translated information)	(Kuehn et al., 2022; Crawshaw et al., 2022; Bergen, Kirkby, et al., 2023)
ATTITUDINAL BARRIERS	
Concerns about vaccine safety	(Heiniger et al., 2022; Crawshaw et al., 2022; Veys-Takeuchi et al., 2022)
Fear of side effects	(Heiniger et al., 2022; Crawshaw et al., 2022)
Belief in natural immunity superiority	(Heiniger et al., 2022; Veys-Takeuchi et al., 2022)
Mistrust in government and health authorities	(Heiniger et al., 2022; Crawshaw et al., 2022; Kadambari and Vanderslott, 2021; Veys-Takeuchi et al., 2022)

Table 6.1 – Potential barriers to COVID-19 vaccination

This campaign aimed to improve vaccine accessibility for hard-to-reach populations, reducing structural barriers listed in Table 6.1, but also addressing vaccine hesitancy. Identified by the World Health Organization (WHO) as one of the top ten threats to global health in 2019 (WHO, 2019), vaccine hesitancy was a significant obstacle in the fight against the COVID-19 pandemic (Veys-Takeuchi et al., 2022). Bringing vaccination services directly to the population was seen as a way to facilitate discussions with health practitioners and to disseminate scientifically accurate information to hesitant individuals.



## 6.2. Targeting vulnerable or deprived populations (phase 1)

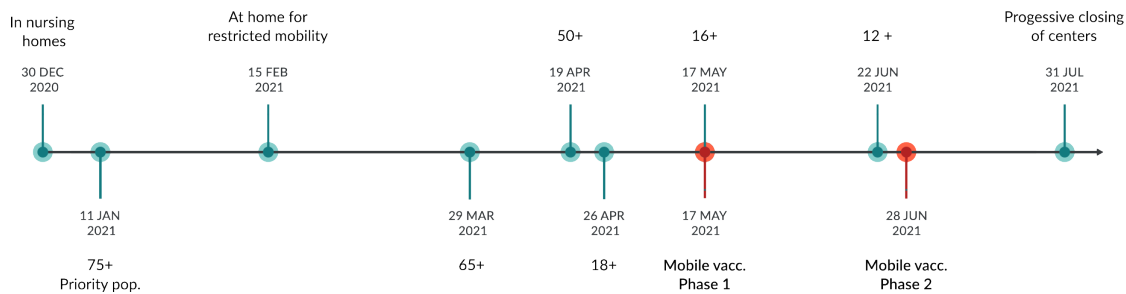


Figure 6.1 – Timeline of the vaccination campaign in the canton of Vaud.

The mobile vaccination campaign was divided into two phases, tailored to the evolving vaccination patterns of the population and overseen by the Vaud Civil Protection Office. The first phase, which began on May 17, involved the establishment of vaccination centers in community facilities with a capacity to administer 400 to 600 doses of vaccine per day. The next phase, which began on June 28, involved the use of a mobile vaccination center housed in a truck. The mobile facility was capable of administering 50 to 100 doses per day, and was stationed in parking lots of shopping centers and residential buildings. The specific vaccination settings for each phase is illustrated in Figure 6.2.

However, a critical challenge arose regarding the optimal placement of these mobile facilities. Financial constraints prevented comprehensive coverage of the canton. Therefore, we used spatial approaches to guide the strategic deployment of these mobile vaccination centers. The following sections describe the methodology behind each vaccination phase.

Before delving into the methodological details, it is important to highlight two key aspects of the study. First, the study was conducted in a relatively short timeframe of about 7 to 10 days for each phase of the campaign. Second, a group of experts was involved from the needs assessment to the selection of vaccination sites. This group was composed of two GIS specialists (the PhD student and the thesis director), two operational officers (commanders of the Vaud Civil Protection Office), and two public health administrators (a vaccination strategy expert and a communications manager). Recommendations formulated by the group of experts were subsequently relayed to the political authorities for final decision-making.

## 6.2 Targeting vulnerable or deprived populations (phase 1)

The process of spatial prioritization of vaccination sites was carried out between 17 March 2021 and 24 March 2021, two months before the start of the campaign. By March 17, approximately 8.88% of the population of Vaud had been vaccinated (compared with 8.60% at the national level), and this proportion increased to 28.03% by May 17 (28.79% at the national level).

## Chapter 6. Prioritization of populations for mobile COVID-19 vaccination campaigns



Figure 6.2 – **Vaccination settings employed during the Vaud mobile campaign.** Credits: Left - SSCM / Julie Masson, Right - SSCM / Valentin Dubach.

During an initial meeting, public health stakeholders identified several criteria to determine the populations to be targeted by the mobile vaccination centers. Priority populations were defined as those living far from vaccination facilities, the elderly population, especially those living alone, those with lower SES, and those living in regions with lower COVID-19 immunity. Therefore, the research question was to identify geographic areas within the canton that were most likely to contain the target populations. To prioritize areas, a composite indicator of area-level vulnerability was constructed. This composite indicator captured the multiple dimensions of the priority population in a single measure, facilitating the subsequent ranking of areas.

The first steps involved collecting relevant geodata and constructing nine indicators to serve as proxies for the target criteria. All indicators were constructed at the postcode level, with resident populations ranging from 20 to approximately 30,000.

As a proxy for geographic access to vaccines, information on existing and planned stationary vaccination centers was collected from the cantonal health department. The population had the option of being vaccinated at one of 16 stationary vaccination centers distributed throughout the canton, or at a limited number of doctors' offices (N=280) and pharmacies (N=150). Location data were available at the address level for vaccination centers and at the postcode level for pharmacies and doctors' offices. Therefore, we included a measure of the distance via road network to the nearest vaccination center and a measure of the density of all types of vaccination facilities (centers, pharmacies, and doctors' offices) per 1,000 inhabitants. In addition, the density of public transport in residential areas was included to reflect the accessibility for people without private transportation. Social and socioeconomic criteria were captured by including median household income per postcode, the percentage of the population in low socio-professional categories, the percentage of the population aged 65

## 6.2. Targeting vulnerable or deprived populations (phase 1)

and older, and the percentage of people aged 65 and older living alone. We also captured COVID-19 immunity at the area level by including the rate of COVID-19 cases and the rate of vaccinated population as of March 22, 2021.

Indicator	Data Source	Threshold	Weight
GEOGRAPHIC ACCESSIBILITY			
Distance to nearest vaccination center (by car)	DGS	Above Median (9.51 km)	2
Vaccination facilities per 1000 inhabitants	DGS	Above Zero	1
Public transport density in residential areas	OSM	Below Median (0.07)	1
SOCIOECONOMIC VULNERABILITY			
Median household income	MICROGIS	Below 60% Median (32.07 kCHF)	4
Population in low socio-professional category	MICROGIS	Above Median (16.7%)	3
Population aged 65 years and older	MICROGIS	Above Median (16.4%)	2
Population aged 65 years and older living alone	OFS	Above Median (4.1%)	3
COVID-19 IMMUNITY			
Confirmed COVID-19 cases per 1000 inhabitants	DGS	Below Median (84.59)	1
Administered COVID-19 vaccine doses per 1000 inhabitants	DGS	Below Median (86.49)	2

Table 6.2 – Indicators and associated parameters used to construct the composite indicator of area-level priority

The construction of the composite vulnerability indicator followed the methodology used by Nicodemo et al. (2020) to identify areas in England most vulnerable to the adverse effects of COVID-19. The process involved dichotomizing the nine indicators by comparing their values to established thresholds. For each indicator, a value of 1 (indicating increased vulnerability) was assigned to postcodes that: exceeded the cantonal median distance to vaccination centres; had no vaccination facilities of any type; had a lower than median density of public transport; had a median household income below 60% of the median; had higher proportions of low socio-professional workers and people aged 65 and over living alone; and had COVID-19 case or vaccination rates below the cantonal median. The threshold for median household income was set to 60% of the cantonal median to be in line with the European Commission's definition of relative poverty (European Union Statistical Office, 2019).

The resulting vulnerability indicator was obtained by summing these binary variables. However, in contrast to the equal weighting scheme used by Nicodemo et al. (2020), different weights (i.e., criteria weights) were assigned to the nine indicators based on their relative importance. Consequently, before summation, these weights were applied to the respective dummy variables. The specific indicators and associated parameters considered in the vulnerability index are detailed in Table 6.2. The final area-level indicator ranged from 0 to 16 (mean: 7.77), with higher values indicating areas of increased vulnerability based on the original target criteria.

## Chapter 6. Prioritization of populations for mobile COVID-19 vaccination campaigns

Results are presented in Figure 6.3, wherein each postcode is categorized by a priority level determined by the quantile classification of the composite vulnerability indicator. Notably, areas corresponding to the target population are predominantly situated at the cantonal boundaries.

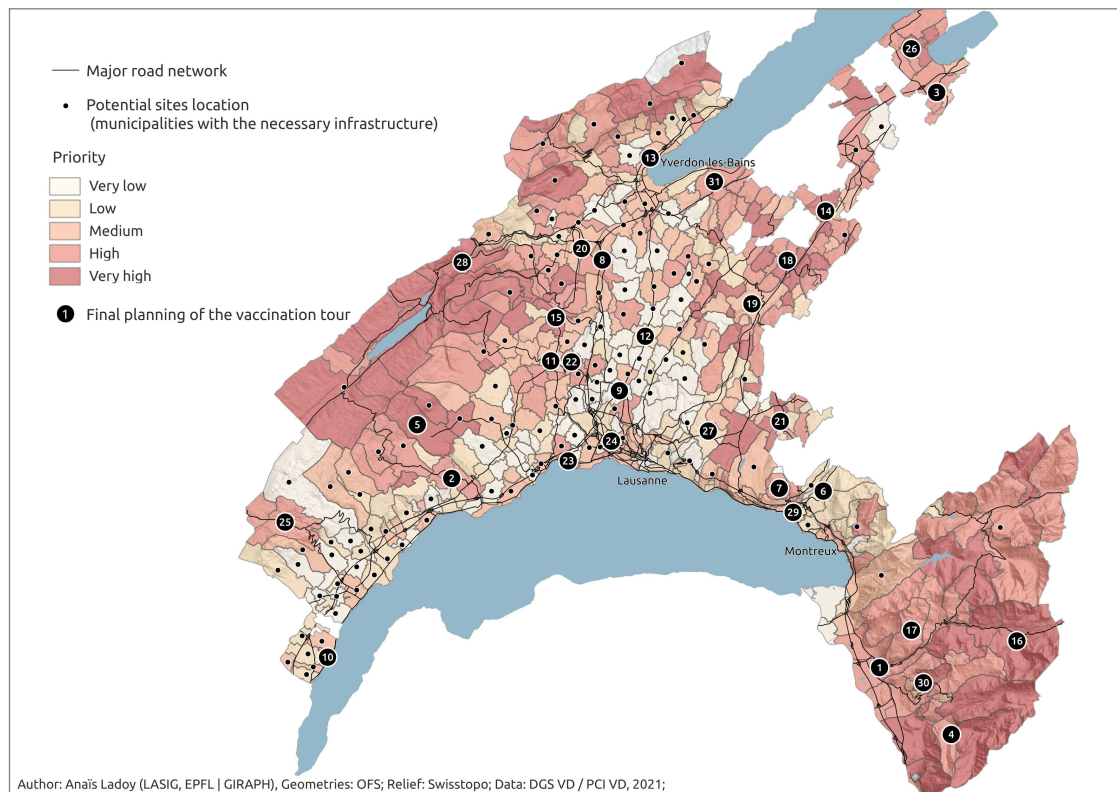


Figure 6.3 – **Optimal locations for the Phase 1 mobile vaccination campaign.** Postcodes are color-coded according to the priority level determined by the composite vulnerability indicator. Municipalities with the logistical capacity to host a vaccination center are indicated by dots, and the implemented itinerary is indicated by numerical sequences.

On March 24, a roundtable discussion was convened with the group of experts. The methodology was presented and discussions were held on refining the relative importance of the indicators through different weighting schemes. Both the geographic distributions of composite and individual indicators were examined. Although not included in the composite score, logistical constraints were incorporated by mapping the 161 municipalities identified by the Office of Civil Protection as ready to host temporary vaccination centers. These discussions resulted in a list of 30 municipalities that was proposed to the cantonal authorities and refined by political considerations. The final itinerary for the first phase of the mobile vaccination campaign was therefore a compromise between scientific, logistical and political considerations (see Figure 6.3).

## 6.3 Targeting areas with low vaccination rate (phase 2)

As of June 3, 2021, Switzerland had made significant progress in its vaccination efforts, with 37.92% of the population of Vaud having received their first dose of vaccine (compared with 37.71% at the national level).

At this stage, the eligible population seeking vaccination have had reasonable access to vaccination centers. As a result, the central objective of the second phase of the vaccination campaign had slightly shifted. It was less about filling gaps left by stationary vaccination sites, but rather about ensuring access to hard-to-reach populations and overcoming vaccine hesitancy. This meant setting up mobile centers with reduced capacity close to where people live, which required a finer spatial analysis. Importantly, it was unclear whether the remaining 60% of the Vaud population remained unvaccinated due to hesitancy, categorical refusal, or structural barriers. As a result, a different approach to vaccine site selection was needed, focusing on vaccination coverage analysis rather than targeting specific population characteristics.

Using data from the cantonal health department, we geocoded the home addresses of the ~350,000 individuals who had received at least one dose of vaccine by June 3<sup>1</sup>, and calculated the COVID-19 vaccination rates per inhabited hectare. The primary concern was to identify areas with low vaccination rates, indicative of communities that may have been underserved by the vaccination campaign.

A local spatial autocorrelation approach, the Getis-Ord  $G_i^*$  statistic (Getis and Ord, 1992), was employed for this purpose. This method identifies statistically significant clusters of high (hot spots) or low (cold spots) vaccination rates by comparing the weighted average of values around inhabited hectares to the overall sum of values across the entire study area. A spatial weights matrix, characterizing neighborhood structure, was defined using a radial buffer of 400 meters. The significance of Getis-Ord clusters was determined by 999 Monte Carlo permutations with a significance level set at 0.05.

Spatial clusters of both low and high vaccination rates were observed across the canton, as illustrated in Figure 6.4. Specifically, pockets of low vaccination rates were predominantly located in the Aigle region (southeast), the Jura region (north), the Broye region (northeast), and within urban agglomerations such as Lausanne and Yverdon-les-Bains. These cold spots could signify either vaccine-resistant communities or regions with barriers to vaccine accessibility.

Complementary maps were created to determine whether the populations in these areas had characteristics that might impede access to vaccines, based on context-specific hypotheses posed by public health experts and the scientific literature on barriers to vaccination (see Table 6.1). First, the spatial distribution of median income per inhabited hectare was mapped,

---

<sup>1</sup>Details of the geocoding algorithm used in this thesis are provided in Appendix Figure A.1.2.

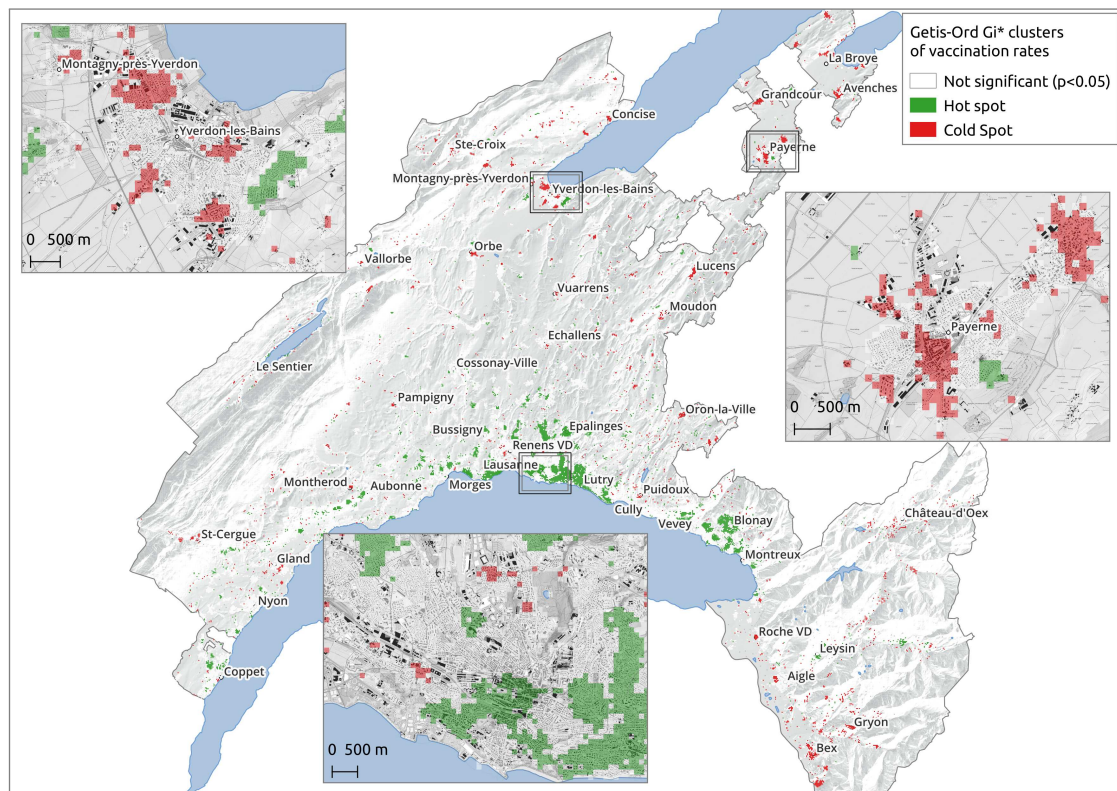


Figure 6.4 – **Spatial clusters of low (red) and high (green) vaccination rates identified using Getis-Ord Gi\* statistics.** Each square corresponds to a single inhabited hectare. Statistical significance was determined by 999 permutations at a significance level of 0.05 (Basemap layer: Digital Elevation Model, Géodonnées État de Vaud).

recognizing that low socioeconomic status can exacerbate structural barriers to vaccination, such as challenges in affording transportation or work-related scheduling conflicts that may prevent vaccination appointments. From a logistical perspective, calculating the number of people who would need to be vaccinated to reach the cantonal average was critical to justifying the establishment of vaccination centers in these areas. In light of studies from the US and UK showing lower vaccine uptake among ethnic minority populations (Kadambari and Vanderslott, 2021; Njoku, Joseph, and Felix, 2021), often linked to mistrust of government and health authorities due to historical discrimination, structural racism and unethical experimentation (e.g., the Tuskegee syphilis study in the US), mapping of sub-Saharan populations was undertaken to explore potential similar inequities in Switzerland. Language barriers were then explored by mapping the proportion of the Balkan-speaking population in the canton of Vaud, a significant demographic group that may face language challenges due to the limited availability of vaccination campaign communications in Balkan languages (e.g., Turkish, Albanian, Serbo-Croatian).



Collectively, these indicators were incorporated into GIS software and assessed by the group of experts to identify optimal locations for mobile vaccination facilities. Figure 6.5 outlines the campaign planning for the city of Yverdon-les-Bains (29,955 inhabitants as of November 29, 2021). The shaded regions signify areas with low vaccination rates, while the color-coded hectares represent different indicators: number of individuals to vaccinate (A), median annual income (B), Balkan-speaking population (C), and sub-Saharan origin population (D). Black symbols pinpoint locations where a mobile facility was deployed in July 2021 during the campaign's second phase. A notable cluster in the Southwest exemplifies potential structural barriers, characterized by low median income and high proportions of allophones and ethnic minorities. These complementary maps were also leveraged by communication teams to disseminate multilingual flyers in targeted neighborhoods, informing residents about the presence of the Civil Protection officers (Perroud, 19.08.21).

## 6.4 Discussion

The mobile vaccination campaign successfully vaccinated approximately 16,500 people.

During the first phase, an average of 468 people were vaccinated per day, for a total of 14,500 people in 31 municipalities visited. Considering the maximum daily capacity of 600 doses, these mobile units effectively served as a widely accepted alternative to stationary vaccination centers. In addition, the notable proportion of vaccinations provided without an appointment (36.9%) may indicate the participation of populations facing technology-related barriers. The need to establish walk-in vaccination opportunities to reduce disparities in vaccine access was highlighted by Michaels, Pirani, and Carrascal (2021), who found geographic correlations between limited internet access and low vaccination coverage in New York City.

In the second phase, an average of 56 people were vaccinated per day, for a total of 2,000 people in 36 visited sites. Despite a planned capacity of 100 doses per day for this phase, attendance was lower than expected but still significant at this advanced stage of the vaccination campaign (July-August 2021).

These statistics confirm the success of the mobile vaccination campaign and recognize that it met a need not necessarily covered by stationary vaccination centers. However, further evaluation of the approach is challenging because the addresses of the persons who came to be vaccinated at each mobile vaccination center could not be provided by the cantonal health department within the scope of this thesis. This limitation is regrettable, as these data would have been crucial to determine whether the vaccinated population resided in the selected locations, thus allowing validation and/or adaptation of spatial approaches for future vaccination campaigns.

## Chapter 6. Prioritization of populations for mobile COVID-19 vaccination campaigns

---

Nevertheless, several comments can be made about the methodology used.

First, although the primary objective of both phases was the same, i.e., to select sites for mobile vaccination centers that would ensure equitable access to COVID-19 vaccines, we used substantially different approaches. In the first phase, we used a "bottom-up" approach, starting with the characteristics of vulnerable populations (geographic distance, limited mobility, SES) to prioritize locations where these populations resided, creating a composite vulnerability indicator. In the second phase, we adopted a "top-bottom" approach by identifying spatial clusters of low vaccination rates using local clustering statistics. Characteristics of resident populations were considered afterward to prioritize areas.

To create the prioritization score, we used a relatively simple approach to aggregating indicators, which involves weighted summation of previously dichotomized indicators based on a threshold. For future applications, it may be possible to improve the score definition without adding complexity by following the approach used by Alemdar et al. (2021) to prioritize COVID-19 vaccination centers in Istanbul, Turkey. The approach is quite similar to the one used here, following the traditional procedure for creating composite indicators (i.e., defining criteria, creating indicators with GIS, normalization, defining criteria weights, aggregation), but with indicators normalized using min-max normalization and weights defined using Analytic Hierarchy Process (AHP) (Saaty, 1980). Compared to the approach used in our study, this would allow expert input to be collected directly in the definition of the criteria weights, not only in their subsequent adjustment.

By analyzing vaccination coverage at the hectare level, we identified pockets of low vaccination rates where allophone populations and/or low-income individuals resided in municipalities that were not prioritized in the first phase. This underscores the need to conduct hectare-level analyses when individual data are available, despite the time-consuming nature of the geocoding process.

The exigency surrounding the organization of the COVID-19 vaccination campaign provided a unique opportunity for interdisciplinary collaboration among academic researchers, field workers, and public health authorities. The involvement of various experts throughout the study proved crucial in developing spatial approaches that could (1) meet the needs identified by public health authorities, (2) be advocated to policy makers, and (3) be deployed in a very short timeframe. Using GIS software as a platform for roundtable discussions allowed for the integration of diverse expertise, resulting in a list of vaccination sites that was informed not only by scientific methods, but also by local territorial knowledge and logistical considerations. For example, this territorial insight enabled the identification of areas, such as those inhabited by evangelical Christian communities opposed to vaccination, where the deployment of a mobile facility would be unproductive.



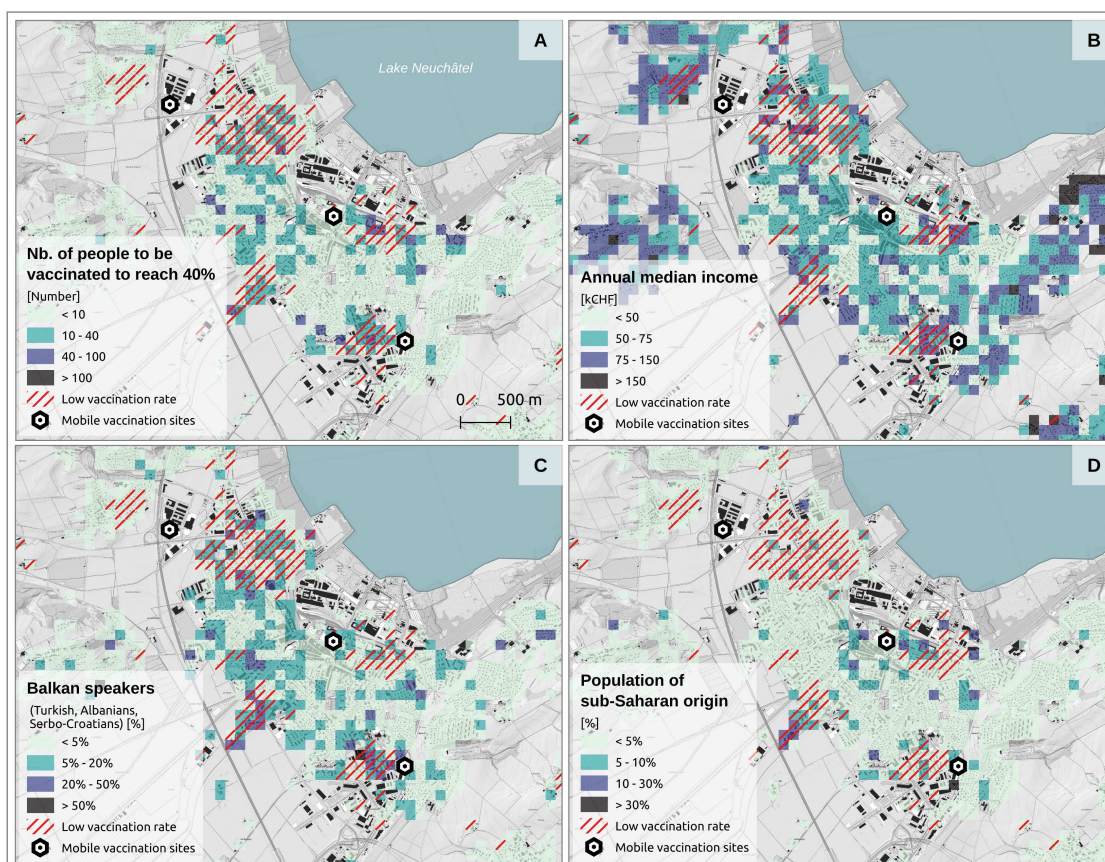


Figure 6.5 – **Optimal locations for the Phase 2 mobile vaccination campaign in Yverdon-les-Bains.** Color-coded cells represent specific hectare-level indicators: number of individuals to be vaccinated to reach the cantonal average (A), median annual income (B), prevalence of Balkan speakers (C), and sub-Saharan population (D). Shaded areas highlight regions with low vaccination rates. Black markers indicate locations where mobile facilities have been deployed.



*Mapping (...) do not permit us to see the world  
in an objective, value-free, wholly impartial fashion.  
But then, as science now insists, nothing really does.  
Instead, maps let us study interpretations of selected  
aspects of a spatially grounded, interrelated process.*

— Tom Koch

Cartographies of disease: maps, mapping, and medicine (2017)

In this thesis, spatial epidemiology methods were applied to fine-scale geodata to design public health and prevention initiatives. Although the research is grounded in the public health system of the canton of Vaud, Switzerland, the methodological framework is applicable to other geographical settings. The following sections summarize the main contributions of the thesis in relation to the original research questions. Recommendations are then articulated to facilitate the adoption of geographic lens *within* public health departments.

### **7.1 Summary of the contributions of spatial epidemiology to public health policy design**

In this section, the initial research questions serve as a guiding thread to explore the potential of fine-scale spatial epidemiology in the design of public health strategies. The primary

question concerns the ability of small-scale geodata to encapsulate the scope of SDOH.

*Can social determinants of health be accurately captured by small area-level indicators?*

In Chapter 2, publicly available spatial data were transformed using Geographic Information Systems (GIS) into meaningful proxies for contextual social determinants of health (SDOH) at the inhabited hectare level. These proxies included measures of environmental exposures (e.g., air pollution, noise), service accessibility (e.g., travel time to emergency facilities, pharmacies), housing conditions (e.g., household size, single-person housing), cultural diversity (e.g., foreign-born population), and socioeconomic status (e.g., median income, unemployment rate, education level). However, it is important to note that the same SDOH can have different impacts when measured at the individual level versus the neighborhood level. For example, individual-level income may influence the ability to afford a balanced diet, while neighborhood-level median income may influence social norms around smoking and diet. This distinction highlights the multifaceted applicability of the hectare-level indicators developed in Chapter 2, as they can serve both as aggregated summary representations of individual SDOH for identifying vulnerable populations (Chapters 2 and 6) and as unique attributes of the living environment for examining its impact on health (Chapter 3).

In studies focused on neighborhood effects, the combination of geocoded individual data and area-level indicators allows for a more comprehensive view of SDOH by incorporating both individual and contextual factors. This approach allows for control of individual-level confounders, including both non-modifiable (e.g., age, gender) and modifiable risk factors (e.g., education level, income, lifestyle behaviors). This approach mitigates a significant limitation of ecological studies, which is the inability to distinguish whether observed health disparities between areas are due to contextual factors or to the characteristics of individuals living in those areas (Diez Roux, 2001). This methodology was used in Chapter 3, where we found that neighborhood characteristics influenced the geographic distribution of hypertension, diabetes, and obesity among middle-aged and older adults in Lausanne, independent of individual-level factors.

However, the scope of GIS-derived SDOH indicators is inherently limited by the inability to integrate more qualitative dimensions, such as social cohesion, or the difficulty of constructing complex indicators at the spatial level, such as the availability of healthy food options. It is also difficult to incorporate macro-structural factors such as discrimination or economic restructuring, which may be upstream of health inequalities (see Figure 1.2) but are not captured by geodata alone (Diez Roux, 2001).

Area-level spatial indicators are highly effective in capturing a substantial range of SDOH. While they may not provide a complete picture of health determinants, their strengths lie in their ability to identify spatial trends and guide targeted interventions. Beyond the mapping of

## 7.1. Summary of the contributions of spatial epidemiology to public health policy design

health-related indicators, the next question is whether such fine-scale spatial analyses can be extended to more complex analytical tasks that inform public health and prevention strategies.

*Is it feasible to conduct fine-scale spatial epidemiology studies from an applied public health perspective?*

In this research, broad health issues such as cardiovascular disease (CVD) risk (Chapter 3), epidemic tracking (Chapters 4 and 5), and vaccination coverage (Chapter 6) have been examined using pseudonymized individual data geocoded to home addresses. On one hand, the availability of diverse health-related information at the individual level enables the application of fine-scale spatial epidemiology to inform public health strategies. On the other hand, the scarcity and the heterogeneous nature of these data across the canton, particularly with regard to the prevalence of chronic diseases, hinder the generalization of findings and the formulation of harmonized data-driven public health policies.

Turning to geographic contextual data, the standardization of socioeconomic, demographic, and environmental variables at the level of inhabited hectares has proven instrumental in mitigating the ecological biases often present in spatial epidemiology. Operating at a finer spatial scale than typical administrative boundaries is crucial, especially when spatial variation in the living environment (Chapter 2) and health outcomes (Chapters 3 and 6) manifest at an intra-municipal level. Thus, working with data aggregated at the postcode or municipality level masks significant health inequalities that exist across the territory. This was also observed outside our study area in an analysis of mortality data in the canton of Geneva (Ladoy, Vallarta-Robledo, et al., 2021). In this study, we analyzed a dataset of 22,751 individual mortality records spanning from 2009 to 2016, obtained from official death notices. Using Local Moran's I spatial autocorrelation statistics, our analysis revealed a marked spatial pattern in the Years of Potential Life Lost or Gained (YPLL<sub>G</sub>), i.e., the difference between an individual's age and their life expectancy at birth<sup>1</sup>.

From a methodological standpoint, the detection of COVID-19 clusters (Chapter 4) and the analysis of vaccination coverage (Chapter 6) at the cantonal level was possible by considering rates per inhabited hectare without encountering the small-number problem. For rarer diseases or limited sample sizes, point pattern analysis offers a viable approach to identifying spatial clusters, as demonstrated in our analysis of cardiometabolic risk factors (CMRFs) (Chapter 3). However, this scale of analysis has limitations when studying health-environment associations because of reduced statistical power and challenges in generalizing results. In such cases, an alternative strategy could consist of aggregating inhabited hectares into groups that guarantee a certain number of inhabitants (e.g., 50 inhabitants). This approach was

---

<sup>1</sup>The manuscript detailing this study of the spatial patterns of YPLL<sub>G</sub> within the canton of Geneva is available in Appendix A.6.

adopted in the "Projet Équité Vaud", a study commissioned by the Public Health Department of the canton of Vaud to examine the relationship between socioeconomic inequalities and potentially avoidable hospitalizations (Chevallereau et al., in prep.).

Having established the feasibility of conducting fine-scale spatial epidemiology to inform public health strategies, the next challenge is to translate these spatial insights into actionable interventions within the framework of noncommunicable disease (NCD) prevention programs.

*How can fine-scale spatial epidemiology approaches guide the design of NCD prevention programs?*

Through the application of intensity functions, spatial clusters of CMRFs - namely hypertension, obesity, diabetes, and dyslipidemia - were detected in Lausanne (Chapter 3). Using geographically weighted regressions, the research further identified characteristics related to the residential environment, such as poverty, unemployment and scarcity of green spaces, as likely influential factors in areas with elevated obesity risk. These insights challenge the effectiveness of universal prevention strategies by revealing spatially-structured health inequalities influenced by characteristics of the physical and social environments.

In line with the principle of proportionate universalism, these spatial insights are essential for tailoring population-level interventions. Specifically, the identified spatial clusters serve as focal points for the targeted allocation of resources to design contextually relevant prevention campaigns. However, while spatial regression methods are useful for hypothesis generation, they fall short in explaining the causal pathways through which social and physical environmental factors influence disease risk. This gap requires the incorporation of qualitative methods (e.g., surveys, focus groups) to achieve a comprehensive understanding of these mechanisms, a prerequisite for designing and implementing effective population health interventions.

Community health centers and neighborhood associations are strategic platforms for facilitating the implementation of prevention programs. Nurses, in particular, often serve as the first point of contact between the population and the health care system and can therefore gather experiential insights from communities, including hard-to-reach populations (Bischoff and Python, 2018). Consequently, nurses could play a critical role in the design, implementation, and refinement of tailored prevention initiatives and act as effective liaisons between communities, local organizations, and public health departments.

While the role of spatial epidemiology in the design of targeted interventions for NCDs has been outlined, we now focus on its applicability in infectious disease surveillance, an area of renewed importance underscored by the recent COVID-19 pandemic.

## 7.1. Summary of the contributions of spatial epidemiology to public health policy design

---

*How can fine-scale spatial epidemiology approaches enhance existing infectious disease surveillance systems?*

The role of spatial epidemiology in infectious disease surveillance was explored using the COVID-19 pandemic as a case study. The Swiss National COVID-19 Task Force, established in March 2020 and constituted of an interdisciplinary team of scientists, formulated strategies in reaction to escalating epidemics (Swiss National COVID-19 Science Task Force, 2020). These strategies transitioned from containment initiatives (such as test-trace-isolate-quarantine (TTIQ) and sentinel strategies) to regional mitigation measures (e.g., hygiene reinforcement, distance rules, selective mask wearing) to broader national mitigation measures (e.g., lockdowns, business closures, border control).

The results we obtained underscore the pivotal role of geographic surveillance during regional mitigation phases, especially when expansive outbreaks challenge the limits of contact tracing and hospital capacities. Although the Task Force guidelines did not explicitly mention geospatial approaches, they emphasized the importance of "detecting spatially confined increases in case numbers and transmission early, allowing response measures to be tailored to the specific outbreak situation" (Swiss National COVID-19 Science Task Force, 2020). Consequently, the prospective space-time scan statistics described in Chapter 4 are of paramount importance. This approach makes it possible to identify clusters of high COVID-19 incidence in a prospective setting, allowing prioritization of both regional mitigation measures and contact tracing.

There are two additional advantages to embedding geographic surveillance within a regional or national surveillance system: first, these analyses rely on mandatory reporting of negative and positive SARS-CoV-2 tests, which are already collected and centralized by the public health administration (Swiss Federal Office of Public Health, 2020). Thus, their implementation does not require additional financial resources from public authorities. In addition, prospective space-time scan clustering methods are highly adaptable. Algorithms can be modified to limit cluster sizes to match specific levels of intervention (e.g., neighborhoods, municipalities, etc.), to limit the number of clusters to a manageable number (e.g., limiting to the ten most significant daily clusters), or to adjust the minimum number of cases required to form a cluster (e.g., considering clusters with five or more cases to remove within-household transmission).

In addition, genetic analyses of space-time scan clusters (Chapter 5) suggest that fine-scale spatial epidemiology could improve the effectiveness of genomic surveillance. Given the resource constraints and extended turnaround times associated with genomic analyses, targeting sequencing efforts to clusters of increased incidence could optimize decision-making timelines, thereby accelerating the identification of variants of concern and superspreading events. However, to operationalize this approach, increased collaboration among key stakeholders is essential. These include the Office of the Cantonal Medical Officer within the

public health department, responsible for contact tracing and case reporting; the Institute of Microbiology at the CHUV, responsible for viral genome sequencing; and spatial analysts, responsible for geocoding cases and genomes and identifying high-risk areas. Importantly, the implementation of such a collaborative pipeline should be tested in anticipation of possible future infectious disease emergencies.

However, the effectiveness of these collaborative efforts in an urgent surveillance context depends on the direct applicability of spatial epidemiology findings to public health actions, an issue addressed in the following section.

*Can academic research findings be directly translated into public health actions?*

Our involvement as a GIS laboratory in planning the mobile vaccination campaign demonstrates that spatial epidemiology insights can be directly translated into public health interventions. This experience provided several lessons. The integration of a geographic lens would likely have been omitted in the urgent context of the vaccination campaign without having previously established close relationships with the public authorities. My presence within the Public Health Department strengthened the recognition and accessibility of spatial approaches. The successful outcome of the campaign was contingent upon collaboration among experts from operational, academic, and public health sectors. The public health experts articulated the needs, guiding the development of spatial strategies closely aligned with health priorities, while operational experts ensured that proposed solutions were feasible and grounded in local knowledge. It merits noting that the final list of vaccination sites did not entirely coincide with the expert-recommended list, influenced by political considerations that may not always align with health objectives (OFSP, 2019). This does not diminish the added value of this unique interdisciplinary collaboration but highlights the multifaceted nature of public health decision making.

### **7.2 Integration of geographic services within public health departments**

The potential of spatial approaches in public health decision-making have been substantiated in this thesis. The diverse case studies demonstrated the application of spatial epidemiology in different domains: the creation of health-promoting environments, cardiovascular disease prevention, COVID-19 epidemic surveillance, and the planning of a COVID-19 vaccination campaign.

However, to ensure the sustainable integration of the geographic lens into the public health decision-making framework, we stress the importance of creating a specialized geographic unit, focused on the use of geographic information *within* public health departments. Such



## 7.2. Integration of geographic services within public health departments

---

a unit would address common barriers to the integration of GIS into public health decision-making, such as the lack of technical skills, insufficient training and guidance, lack of data in appropriate formats, and lack of a clear GIS strategy (Boulos, 2004). Integrating this competence would facilitate the manipulation of geospatial data, the design of spatial analyses, and the interpretation of the results, thereby contributing to agile and effective decision making. This section outlines the roles and responsibilities that the geographic unit could shoulder.

### 7.2.1 Expanding fine-scale health data acquisition

To maximize the potential of spatial approaches in designing population health interventions, the acquisition and integration of fine-scale geographic data are essential.

The development of an extensive set of SDOH indicators for all inhabited hectares, as accomplished in this thesis, constitutes a foundational asset for future health-related applications. This spatial database will serve as a central resource for internal case studies, allowing for the dynamic selection, refinement, and expansion of indicators based on specific research objectives and available data. The long-term maintenance and enrichment of this database could be greatly enhanced by inter-departmental collaboration involving sectors such as environmental science, public health, transportation, cadastre, and social welfare.

Beyond the geographic contextual data, it is essential to supplement the existing health datasets with additional georeferenced health data. This can catalyze the formulation of tailored population health interventions across various health domains. The following data sources may provide a hierarchical understanding of health inequities tailored to different levels of interventions.

Mortality data from the Swiss National Cohort (Bopp et al., 2009) can help elucidate deeply rooted health inequities, as they encapsulate the ultimate health outcome influenced by multiple determinants. These data can also help to refine the composite indicators developed in Chapter 2, thereby ensuring their effectiveness in capturing health disparities (Allik et al., 2020; Kolak et al., 2020).

In addition, hospital discharge data from the Medical Statistics of Swiss Hospitals (Zellweger, Junker, and Bopp, 2019) can offer insights into the prevalence of chronic conditions requiring medical interventions. When these data are available at a finer spatial resolution than the Medstat regions, spatial analyses can reveal territorial inequities in healthcare access and inform tertiary prevention strategies, such as chronic disease management programs and support groups.

Lastly, survey-based sources like the Swiss Health Survey (*Étude suisse sur la santé* 2023) provide insights on disease prevalence and individual risk factors, acting as early warning

systems for identifying emerging inequalities. Specifically, the Swiss Health Survey (*Étude suisse sur la santé* 2023) is scheduled for nationwide implementation in 2024 and will assess the health status and behaviors of 100,000 Swiss residents through questionnaires and medical examinations. The data may exhibit low spatial density within the canton of Vaud; however, this limitation could be mitigated by supplementing the dataset with information from the participatory platform Métasanté (*Métasanté | Plateforme participative* 2023). This platform was developed to host population questionnaires on the impact of the environment on health, and participatory surveys have already been conducted in two municipalities in the canton of Vaud: Chavannes-près-Renens (240 participants) and Yverdon-les-Bains (135 participants). Extending Métasanté coverage to the whole territory could offer longitudinal health monitoring of the population, as it has recently been done with its Geneva counterpart, Specchio (*Specchio-Hub* 2023). These sources could be particularly valuable for guiding primary prevention strategies (e.g., health awareness, support groups) and health promotion measures.

The integration of these complementary data sources, alongside the health datasets used in this thesis (i.e., ColaUs-PsyColaUs study, molecular diagnostic tests, vaccination records) would enable the full leverage of spatial epidemiology to address diverse health concerns, geographic contexts, and populations. Such complementary data sources could also be used to extend spatial prediction of chronic disease risk to the entire territory, in line with the methods developed for the CMRFs in Lausanne.

### 7.2.2 Training and guidance for health practitioners

With the consolidation of a spatial database incorporating both contextual and health-related data, the new geographic unit could inform any project through a geographical lens, thereby serving as a decision-making support system.

Depending on the complexity and on the geographical relevance of health issues, the contribution of spatial epidemiology could range from simple visualizations of contextual elements to more complex analyses, such as those developed in this thesis, including clustering or regression approaches. As observed in the planning of the vaccination campaign, cartographic support significantly enhances interdisciplinary collaboration by synthesizing different types of information and enabling coordinated strategic decision-making (Endacott et al., 2009).

However, the cartographer's paradox suggests that in order to avoid hiding important information in a multitude of details, maps offer a selective and incomplete representation of reality (Monmonier, 2018). This limitation can lead public health practitioners to erroneous conclusions if they are not adequately guided through the correct interpretation and potential biases of spatial analyses. Therefore, the existing staff should be provided with GIS-related

training. Such training could extend beyond the practitioners of the public health department to include data managers and also nurses who, as previously discussed, will play a crucial role in the design, implementation, and evaluation of preventive actions in the field.

### 7.2.3 Operationalization and automation of spatial epidemiology pipelines

In this research, it was shown that spatial epidemiology is able to provide crucial insights in a public health crisis such as infectious disease outbreaks and vaccination campaigns. For practical implementation, the pipelines discussed herein, notably the geographic surveillance prototype introduced in Chapter 4, warrant operationalization and automation. However, the practical utility of these pipelines is contingent on real-time data access, which could be permitted through collaborations between health institutions, hospitals, and academic research centers.

Central to the operational pipeline is the geocoding of individual data, which must be balanced with stringent data confidentiality requirements. We introduced a home-made geocoding algorithm not relying on commercial geocoding services, thereby enhancing data privacy. Adherence to Swiss data protection legislation was maintained through the application of geomasking techniques such as data aggregation and randomized point displacements. For timely public health response in crisis situations, internal pipelines for automated geocoding are recommended. These pipelines could be modeled on the geocoding algorithm implemented in this research, and could incorporate sophisticated geomasking techniques to balance privacy concerns with analytical accuracy. These include data aggregation techniques, donut methods (Hampton et al., 2010), and network-based geomasking (Swanlund et al., 2020). In this context, active collaboration with IT departments could further optimize the pipeline for both analytical rapidity and ethical compliance.

The technological environment in which these pipelines operate has implications for data accessibility, cost management, and cross-sector collaboration. In this thesis, only free and open source software has been used, which is advisable for several reasons. FOSS not only ensures interoperability and cost-effectiveness, but also facilitates inter-institutional collaboration through universally accessible methods and codes.

## 7.3 Maintaining the link with academic research

While this work serves as a bridge between academic research and practical public health interventions, it will be important for public health departments to maintain collaborations with academia. This will ensure that operational frameworks remain aligned with major advances in spatial epidemiology. This section outlines two emerging fields that could enrich

the work presented here.

The first concerns advances in spatial data collection technologies. Rapid technological evolution in the form of wearable sensors, remote sensing, smartphone apps, and the Internet of Things has spawned the discipline of spatial lifecourse epidemiology (Jia, 2019). Using these advanced technologies to capture a wider range of spatial and temporal exposures associated with SDOH could ultimately make a significant contribution to reconstructing an individual's exposome, defined as all environmental exposures experienced from conception through life (M. C. Turner et al., 2017; Wild, 2005). While the complexities and individual-centric nature of these methods may pose challenges for direct translation into population-level interventions, they offer novel insights into health-environment associations that can be assimilated into existing spatial epidemiology frameworks.

The second area of focus concerns the intersection of COVID-19 and NCDs. The onset of the COVID-19 pandemic has led to a reorientation of public health priorities, which have largely focused on containing SARS-CoV-2 to the detriment of addressing the burden of NCDs. Nevertheless, the need for a syndemic approach to the study of infectious and noncommunicable diseases is compelling due to their interrelated nature (Horton, 2020; Mennis, Matthews, and Huston, 2022). Singer (1996) first introduced the theoretical concept of syndemics to describe the interaction among drug abuse, violence, and AIDS (SAVA). This concept can be characterized by "the presence of two or more disease states that adversely interact with each other, negatively affecting the mutual course of each disease trajectory, enhancing vulnerability, and which are made more deleterious by experienced inequities" (The Lancet, 2017). An empirical framework for investigating syndemics through spatial epidemiology has been proposed, involving methodologies such as local clustering statistics, bivariate mapping, and spatial regression (S. Shrestha et al., 2020). Future work should aimed at examining the syndemic relationship between COVID-19 and NCDs, which would offer invaluable insights into health inequities and the impact of pandemic-related policies on population health. Within the scope of this thesis, preliminary analyses were conducted to explore the syndemic nature of COVID-19 and CVD in Lausanne. However, these efforts were curtailed due to limitations in the representativeness of RT-PCR test data as a proxy for COVID-19 burden.

### 7.4 General conclusion

Leveraging fine-scale spatial data accessible to public authorities, this research demonstrates the potential of spatial epidemiology to guide population health strategies towards achieving proportionate universalism.

Addressing territorial health disparities requires the identification of locations where increased efforts should be directed. The fine-scale spatial approaches formulated in this thesis address

this need by geographically delineating areas of elevated incidence or acute vulnerability. When these approaches combine geographic contextual factors with specific health data, they enable the effective operationalization of proportionate universalism at the community level, by empowering public health practitioners to tailor preventive interventions to the unique characteristics of populations and their environments. Cartography is an invaluable tool in this endeavor. Maps synthesize complex and intricate information and act as a universal lexicon, effectively bridging disciplinary jargon and facilitating informed public health decision-making.

As we face the impending health threats of climate change (Costello et al., 2009), the development of population-level interventions focused on the living environment offers the opportunity to simultaneously improve individual and planetary health (Myers, 2017). From heightened exposure to disease vectors to increased mortality from heat waves, the effects of climate change are inextricably linked to geographic location, making it more imperative than ever to view health through a spatial lens.



## BIBLIOGRAPHY

- Abel, T., K. Hofmann, and D. Schori (2013). “Social and Regional Variations in Health Status and Health Behaviours among Swiss Young Adults”. In: *Swiss Med Wkly*. DOI: 10.4414/smw.2013.13901.
- Abolhassani, N. et al. (2017). “Determinants of Change in Polypharmacy Status in Switzerland: The Population-Based CoLaus Study”. In: *Eur J Clin Pharmacol* 73.9, pp. 1187–1194. DOI: 10.1007/s00228-017-2288-1.
- Aguayo, G. A. et al. (2020). “Identifying Hotspots of Cardiometabolic Outcomes Based on a Bayesian Approach: The Example of Chile”. In: *PLoS ONE* 15.6. Ed. by N. Ikeda, e0235009. DOI: 10.1371/journal.pone.0235009.
- Ahasan, R. and M. M. Hossain (2021). “Leveraging GIS and Spatial Analysis for Informed Decision-Making in COVID-19 Pandemic”. In: *Health Policy and Technology* 10.1, pp. 7–9. DOI: 10.1016/j.hlpt.2020.11.009.
- Alberti, K. G. M. M. and P. Z. Zimmet (1998). “Definition, Diagnosis and Classification of Diabetes Mellitus and Its Complications. Part 1: Diagnosis and Classification of Diabetes Mellitus. Provisional Report of a WHO Consultation”. In: *Diabetic medicine* 15.7, pp. 539–553.
- Alemdar, K. D. et al. (2021). “Accessibility of Vaccination Centers in COVID-19 Outbreak Control: A GIS-Based Multi-Criteria Decision Making Approach”. In: *IJGI* 10.10, p. 708. DOI: 10.3390/ijgi10100708.
- Allik, M. et al. (2020). “Creating Small-Area Deprivation Indices: A Guide for Stages and Options”. In: *J Epidemiol Community Health* 74.1, pp. 20–25. DOI: 10.1136/jech-2019-213255.
- Alteri, C. et al. (2021). “Genomic Epidemiology of SARS-CoV-2 Reveals Multiple Lineages and Early Spread of SARS-CoV-2 Infections in Lombardy, Italy”. In: *Nature communications* 12.1, p. 434.
- Amstutz, D., O. Villa, et al. (2020). “Health Promotion in Swiss Municipalities: Towards Proportionate Universalism”. In: *European Journal of Public Health* 30 (Supplement\_5), ckaa165.713. DOI: 10.1093/eurpub/ckaa165.713.

## Bibliography

---

- Amstutz, D. and O. Villa (2021). "Promouvoir la santé de la population pendant la crise de la COVID-19 : le rôle des communes sur le canton de Vaud, en Suisse". In: *Glob Health Promot* 28.1, pp. 103–107. DOI: 10.1177/1757975920977833.
- Andrew, M. K. and J. M. Keefe (2014). "Social Vulnerability from a Social Ecology Perspective: A Cohort Study of Older Adults from the National Population Health Survey of Canada". In: *BMC Geriatr* 14.1, p. 90. DOI: 10.1186/1471-2318-14-90.
- Anselin, L. (1995). "Local Indicators of Spatial Association-LISA". In: *Geographical Analysis* 27.2, pp. 93–115. DOI: 10.1111/j.1538-4632.1995.tb00338.x.
- Anselin, L. and S. J. Rey, eds. (2010). *Perspectives on Spatial Data Analysis*. Advances in Spatial Science. Heidelberg ; New York: Springer. 290 pp.
- ARE (2022). *Niveaux de Qualité de Desserte Par Les Transports Publics – Méthodologie de Calcul ARE*. Office fédéral du développement territorial ARE. URL: <https://www.are.admin.ch/are/fr/home/media-et-publications/publications/transports/ov-guteklassen-berechnungsmethodik-are.html>.
- Arnold, C. (2022). *SPURRED BY COVID, PUBLIC HEALTH GETS PRECISE*. NATURE PORTFOLIO HEIDELBERGER PLATZ 3, BERLIN, 14197, GERMANY.
- Ashcroft, P., S. Lehtinen, and S. Bonhoeffer (2022). "Test-Trace-Isolate-Quarantine (TTIQ) Intervention Strategies after Symptomatic COVID-19 Case Identification". In: *PLoS ONE* 17.2. Ed. by S. Blumberg, e0263597. DOI: 10.1371/journal.pone.0263597.
- Attwood, S. W. et al. (2020). "An Integrated National Scale SARS-CoV-2 Genomic Surveillance Network". In: *The Lancet Microbe* 1.3.
- Auchincloss, A. H. et al. (2012). "A Review of Spatial Methods in Epidemiology, 2000–2010". In: *Annual Review of Public Health* 33.1, pp. 107–122. DOI: 10.1146/annurev-publhealth-031811-124655.
- Babisch, W. et al. (1993). "Traffic Noise and Cardiovascular Risk: The Speedwell Study, First Phase. Outdoor Noise Levels and Risk Factors". In: *Archives of Environmental Health: An International Journal* 48.6, pp. 401–405.
- Baddeley, A., E. Rubak, and R. Turner (2016). "Spatial Point Patterns". In:
- Bakian, A. V. et al. (2015). "Spatial Relative Risk Patterns of Autism Spectrum Disorders in Utah". In: *J Autism Dev Disord* 45.4, pp. 988–1000. DOI: 10.1007/s10803-014-2253-0.
- Bambra, C. et al. (2020). "The COVID-19 Pandemic and Health Inequalities". In: *J Epidemiol Community Health*, jech-2020-214401. DOI: 10.1136/jech-2020-214401.
- Bejon, P. et al. (2010). "Stable and Unstable Malaria Hotspots in Longitudinal Cohort Studies in Kenya". In: *PLOS Medicine* 7.7, e1000304. DOI: 10.1371/journal.pmed.1000304.
- Beldomenico, P. M. (2020). "Do Superspreaders Generate New Superspreaders? A Hypothesis to Explain the Propagation Pattern of COVID-19". In: *International Journal of Infectious Diseases* 96, pp. 461–463. DOI: 10.1016/j.ijid.2020.05.025.



- Bentu -Mart nez, C., M. R. Mimbrero, and M. Z  niga-Ant n (2023). “Spatial Patterns in Sociodemographic Factors Explain to a Large Extent the Prevalence of Hypertension and Diabetes in Aragon (Spain)”. In: *Front. Med.* 10, p. 1016157. DOI: 10.3389/fmed.2023.1016157.
- Bergen, N., N. E. Johns, et al. (2023). “Within-Country Inequality in COVID-19 Vaccination Coverage: A Scoping Review of Academic Literature”. In: *Vaccines* 11.3, p. 517. DOI: 10.3390/vaccines11030517.
- Bergen, N., K. Kirkby, et al. (2023). “Global State of Education-Related Inequality in COVID-19 Vaccine Coverage, Structural Barriers, Vaccine Hesitancy, and Vaccine Refusal: Findings from the Global COVID-19 Trends and Impact Survey”. In: *The Lancet Global Health* 11.2, e207–e217. DOI: 10.1016/S2214-109X(22)00520-4.
- Besag, J. and P. J. Diggle (1977). “Simple Monte Carlo Tests for Spatial Pattern”. In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 26.3, pp. 327–333. DOI: 10.2307/2346974.
- Bischoff, A. and N. V. Python (2018). “M decins et soignants en sant  communautaire – o  en sommes-nous ?” In: *Revue M dicale Suisse* 14.619, pp. 1671–1673. DOI: 10.53738/REVMED.2018.14.619.1671.
- Blondel, V. D. et al. (2008). “Fast Unfolding of Communities in Large Networks”. In: *J. Stat. Mech.* 2008.10, P10008. DOI: 10.1088/1742-5468/2008/10/P10008.
- Bopp, M. et al. (2009). “Cohort Profile: The Swiss National Cohort–a Longitudinal Study of 6.8 Million People”. In: *International Journal of Epidemiology* 38.2, pp. 379–384. DOI: 10.1093/ije/dyn042.
- Bosch, M. et al. (2021). “A Spatially Explicit Approach to Simulate Urban Heat Mitigation with InVEST (v3.8.0)”. In: *Geosci. Model Dev.* 14.6, pp. 3521–3537. DOI: 10.5194/gmd-14-3521-2021.
- Boulos, M. N. K. (2004). “Towards Evidence-Based, GIS-driven National Spatial Health Information Infrastructure and Surveillance Services in the United Kingdom”. In: *International Journal of Health Geographics*, p. 50.
- Bousema, T. et al. (2012). “Hitting Hotspots: Spatial Targeting of Malaria for Control and Elimination”. In: *PLoS Med* 9.1, e1001165. DOI: 10.1371/journal.pmed.1001165.
- Braeye, T. et al. (2015). “A Large Community Outbreak of Gastroenteritis Associated with Consumption of Drinking Water Contaminated by River Water, Belgium, 2010”. In: *Epidemiol. Infect.* 143.4, pp. 711–719. DOI: 10.1017/S0950268814001629.
- Brody, H. et al. (2000). “Map-Making and Myth-Making in Broad Street: The London Cholera Epidemic, 1854”. In: *The Lancet* 356.9223, pp. 64–68. DOI: 10.1016/S0140-6736(00)02442-9.
- Brown, S. C. et al. (2016). “Neighborhood Greenness and Chronic Health Conditions in Medicare Beneficiaries”. In: *American Journal of Preventive Medicine* 51.1, pp. 78–89. DOI: 10.1016/j.amepre.2016.02.008.

## Bibliography

---

- Browning, M. and K. Lee (2017). “Within What Distance Does “Greenness” Best Predict Physical Health? A Systematic Review of Articles with GIS Buffer Analyses across the Lifespan”. In: *IJERPH* 14.7, p. 675. DOI: 10.3390/ijerph14070675.
- Brüningk, S. C. et al. (2022). “Determinants of SARS-CoV-2 Transmission to Guide Vaccination Strategy in an Urban Area”. In: *Virus Evolution* 8.1, veac002. DOI: 10.1093/ve/veac002.
- Budd, J. et al. (2020). “Digital Technologies in the Public-Health Response to COVID-19”. In: *Nat Med* 26.8, pp. 1183–1192. DOI: 10.1038/s41591-020-1011-4.
- Canton de Vaud (2023a). *Institutions médicales et sociales*. Viageo. URL: <https://viageo.ch/md/cf721fc7-23e5-ad34-b5ad-58b326d693dd> (visited on 10/08/2023).
- (2023b). *Plans d'affectation - Zones d'affectation et contenus superposés*. Viageo. URL: <https://viageo.ch/md/3b83b35f-6f66-4e5c-88ec-c1bbb8e7efa4> (visited on 10/08/2023).
- Caruana, G. et al. (2021). “Implementing SARS-CoV-2 Rapid Antigen Testing in the Emergency Ward of a Swiss University Hospital: The INCREASE Study”. In: *Microorganisms* 9.4 (4), p. 798. DOI: 10.3390/microorganisms9040798.
- Castro, A., N. Künzli, and T. Götschi (2017). “Health Benefits of a Reduction of PM10 and NO2 Exposure after Implementing a Clean Air Plan in the Agglomeration Lausanne-Morges”. In: *International Journal of Hygiene and Environmental Health* 220.5, pp. 829–839. DOI: 10.1016/j.ijheh.2017.03.012.
- Chaix, B. (2005). “Comparison of a Spatial Approach with the Multilevel Approach for Investigating Place Effects on Health: The Example of Healthcare Utilisation in France”. In: *Journal of Epidemiology & Community Health* 59.6, pp. 517–526. DOI: 10.1136/jech.2004.025478.
- Chaix, B. et al. (2010). “Individual/Neighborhood Social Factors and Blood Pressure in the RECORD Cohort Study: Which Risk Factors Explain the Associations?” In: *Hypertension* 55.3, pp. 769–775. DOI: 10.1161/HYPERTENSIONAHA.109.143206.
- Chakraborty, J. (2011). “Revisiting Tobler’s First Law of Geography: Spatial Regression Models for Assessing Environmental Justice and Health Risk Disparities”. In: *Geospatial analysis of environmental health*, pp. 337–356.
- Chang, S. et al. (2021). “Mobility Network Models of COVID-19 Explain Inequities and Inform Reopening”. In: *Nature* 589.7840, pp. 82–87. DOI: 10.1038/s41586-020-2923-3.
- Chen, J. et al. (2008). “Geovisual Analytics to Enhance Spatial Scan Statistic Interpretation: An Analysis of U.S. Cervical Cancer Mortality”. In: *Int J Health Geogr* 7.1, p. 57. DOI: 10.1186/1476-072X-7-57.
- Chen, S. et al. (2018). “Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor”. In: *Bioinformatics* 34.17, pp. i884–i890.
- Christian, W. J. et al. (2011). “Exploring Geographic Variation in Lung Cancer Incidence in Kentucky Using a Spatial Scan Statistic: Elevated Risk in the Appalachian Coal-Mining Region”. In: *Public Health Rep* 126.6, pp. 789–796. DOI: 10.1177/003335491112600604.

- Chu, I. Y.-H. et al. (2020). “Social Consequences of Mass Quarantine during Epidemics: A Systematic Review with Implications for the COVID-19 Response”. In: *J Travel Med* 27.7. DOI: 10.1093/jtm/taaa192.
- Clark, A. et al. (2020). “Global, Regional, and National Estimates of the Population at Increased Risk of Severe COVID-19 Due to Underlying Health Conditions in 2020: A Modelling Study”. In: *The Lancet Global Health* 8.8, e1003–e1017. DOI: 10.1016/S2214-109X(20)30264-3.
- Corman, V. M. et al. (2020). “Detection of 2019 Novel Coronavirus (2019-nCoV) by Real-Time RT-PCR”. In: *Eurosurveillance* 25.3, p. 2000045. DOI: 10.2807/1560-7917.ES.2020.25.3.2000045.
- Costello, A. et al. (2009). “Managing the Health Effects of Climate Change”. In: *The Lancet* 373.9676, pp. 1693–1733. DOI: 10.1016/S0140-6736(09)60935-1.
- Crawshaw, A. F. et al. (2022). “Defining the Determinants of Vaccine Uptake and Undervaccination in Migrant Populations in Europe to Improve Routine and COVID-19 Vaccine Uptake: A Systematic Review”. In: *The Lancet Infectious Diseases* 22.9, e254–e266. DOI: 10.1016/S1473-3099(22)00066-4.
- Cromley, E. K. and S. L. McLafferty (2011). *GIS and Public Health*. Guilford Press.
- Cromley, E. K. (2019). “Using GIS to Address Epidemiologic Research Questions”. In: *Curr Epidemiol Rep* 6.2, pp. 162–173. DOI: 10.1007/s40471-019-00193-6.
- Curtis, S. (2004). *Health and Inequality: Geographical Perspectives*. Sage.
- Dahlgren, G. and M. Whitehead (2021). “The Dahlgren-Whitehead Model of Health Determinants: 30 Years on and Still Chasing Rainbows”. In: *Public Health* 199, pp. 20–24. DOI: 10.1016/j.puhe.2021.08.009.
- Danis, K. et al. (2020). “Cluster of Coronavirus Disease 2019 (COVID-19) in the French Alps, February 2020”. In: *Clinical Infectious Diseases* 71.15, pp. 825–832. DOI: 10.1093/cid/ciaa424.
- Davies, T. M. (2013). “Jointly Optimal Bandwidth Selection for the Planar Kernel-Smoothed Density-Ratio”. In: *Spatial and Spatio-Temporal Epidemiology* 5, pp. 51–65.
- Davies, T. M., M. L. Hazelton, and J. C. Marshall (2011). “**Sparr** : Analyzing Spatial Relative Risk Using Fixed and Adaptive Kernel Density Estimation in R”. In: *J. Stat. Soft.* 39.1. DOI: 10.18637/jss.v039.i01.
- Davies, T. M., J. C. Marshall, and M. L. Hazelton (2018). “Tutorial on Kernel Estimation of Continuous Spatial and Spatiotemporal Relative Risk: Spatial and Spatiotemporal Relative Risk”. In: *Statistics in Medicine* 37.7, pp. 1191–1221. DOI: 10.1002/sim.7577.
- De Courrèges, A. et al. (2021). “The Relationship between Neighbourhood Walkability and Cardiovascular Risk Factors in Northern France”. In: *Science of The Total Environment* 772, p. 144877. DOI: 10.1016/j.scitotenv.2020.144877.
- De Groot, R. et al. (2019). “Urban-Rural Differences in the Association between Blood Lipids and Characteristics of the Built Environment: A Systematic Review and Meta-Analysis”. In: *BMJ Glob Health* 4.1, e001017. DOI: 10.1136/bmjgh-2018-001017.
- De Pietro, C. et al. (2015). “Switzerland: Health System Review”. In:

## Bibliography

---

- De Ridder, D., J. Sandoval, N. Vuilleumier, A. S. Azman, et al. (2021). “Socioeconomically Disadvantaged Neighborhoods Face Increased Persistence of SARS-CoV-2 Clusters”. In: *Front. Public Health* 8. DOI: 10.3389/fpubh.2020.626090.
- De Ridder, D., J. Sandoval, N. Vuilleumier, S. Stringhini, et al. (2020). “Geospatial Digital Monitoring of COVID-19 Cases at High Spatiotemporal Resolution”. In: *The Lancet Digital Health* 2.8, e393–e394. DOI: 10.1016/S2589-7500(20)30139-4.
- Den Braver, N. R. et al. (2018). “Built Environmental Characteristics and Diabetes: A Systematic Review and Meta-Analysis”. In: *BMC Med* 16.1, p. 12. DOI: 10.1186/s12916-017-0997-z.
- Desjardins, M., A. Hohl, and E. Delmelle (2020). “Rapid Surveillance of COVID-19 in the United States Using a Prospective Space-Time Scan Statistic: Detecting and Evaluating Emerging Clusters”. In: *Applied Geography* 118, p. 102202. DOI: 10.1016/j.apgeog.2020.102202.
- DGS Vaud (2018). *Rapport Sur La Politique de Santé Publique Du Canton de Vaud 2018-2022*. URL: [https://www.vd.ch/fileadmin/user\\_upload/accueil/Communique\\_presse/documents/Rapport\\_sur\\_la\\_politique\\_de\\_sant%C3%A9\\_publique\\_VD\\_2018-2022.pdf](https://www.vd.ch/fileadmin/user_upload/accueil/Communique_presse/documents/Rapport_sur_la_politique_de_sant%C3%A9_publique_VD_2018-2022.pdf) (visited on 01/03/2020).
- Di Giallonardo, F. et al. (2020). “Genomic Epidemiology of the First Wave of SARS-CoV-2 in Italy”. In: *Viruses* 12.12, p. 1438.
- Diez Roux, A. V. (2003). “Residential Environments and Cardiovascular Risk”. In: *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 80.4, pp. 569–589. DOI: 10.1093/jurban/jtg065.
- Diez Roux, A. V. (2001). “Investigating Neighborhood and Area Effects on Health”. In: *American Journal of Public Health* 91.11, pp. 1783–1789. DOI: 10.2105/AJPH.91.11.1783.
- Diez Roux, A. V. and C. Mair (2010). “Neighborhoods and Health: Neighborhoods and Health”. In: *Annals of the New York Academy of Sciences* 1186.1, pp. 125–145. DOI: 10.1111/j.1749-6632.2009.05333.x.
- Diez-Roux, A. V. (2000). “Multilevel Analysis in Public Health Research”. In: *Annual Review of Public Health* 21.1, pp. 171–192. DOI: 10.1146/annurev.publhealth.21.1.171.
- Diez-Roux, A. V., B. G. Link, and M. E. Northridge (2000). “A Multilevel Analysis of Income Inequality and Cardiovascular Disease Risk Factors”. In: *Social Science & Medicine* 50.5, pp. 673–687. DOI: 10.1016/S0277-9536(99)00320-2.
- Django Software Foundation (2019). *Django*. Version 2.2. URL: <https://djangoproject.com>.
- Dong, E., H. Du, and L. Gardner (2020). “An Interactive Web-Based Dashboard to Track COVID-19 in Real Time”. In: *The Lancet Infectious Diseases*, S1473309920301201. DOI: 10.1016/S1473-3099(20)30120-1.
- Dorn, A. V., R. E. Cooney, and M. L. Sabin (2020). “COVID-19 Exacerbating Inequalities in the US”. In: *The Lancet* 395.10232, pp. 1243–1244. DOI: 10.1016/S0140-6736(20)30893-X.
- Du, M., X. Zhang, and L. Mora (2021). “Strategic Planning for Smart City Development: Assessing Spatial Inequalities in the Basic Service Provision of Metropolitan Cities”. In: *Journal of Urban Technology* 28.1-2, pp. 115–134. DOI: 10.1080/10630732.2020.1803715.

- Dumondel, M. (1985). “Typologies Des Communes Suisses 1970-1980: Étude Dynamique Des Disparités Socio-Économiques Au Niveau Local”. PhD thesis. ETH Zurich.
- Durfey, S. N. M. et al. (2019). “Neighborhood Disadvantage and Chronic Disease Management”. In: *Health Serv Res* 54, pp. 206–216. DOI: 10.1111/1475-6773.13092.
- Edwards, K. L. et al. (2010). “The Neighbourhood Matters: Studying Exposures Relevant to Childhood Obesity and the Policy Implications in Leeds, UK”. In: *Journal of Epidemiology & Community Health* 64.3, pp. 194–201. DOI: 10.1136/jech.2009.088906.
- Ellaway, A., S. Macintyre, and X. Bonnefoy (2005). “Graffiti, Greenery, and Obesity in Adults: Secondary Analysis of European Cross Sectional Survey”. In: *BMJ* 331.7517, pp. 611–612. DOI: 10.1136/bmj.38575.664549.F7.
- Endacott, R. et al. (2009). “Geographic Information Systems for Healthcare Organizations: A Primer for Nursing Professions”. In: *CIN: Computers, Informatics, Nursing* 27.1, pp. 50–56. DOI: 10.1097/NCN.0b013e31818e4660.
- England, P. H. (2020). *COVID-19: Epidemiological Definitions of Outbreaks and Clusters in Particular Settings*. GOV.UK. URL: <https://www.gov.uk/government/publications/covid-19-epidemiological-definitions-of-outbreaks-and-clusters/covid-19-epidemiological-definitions-of-outbreaks-and-clusters-in-particular-settings> (visited on 12/01/2020).
- Ester, M., H.-P. Kriegel, and X. Xu (1996). “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. Evangelos Simoudis, Jiawei Han, and Usama Fayyad. Menlo Park, California: AAAI Press, p. 6. URL: <https://aaai.org/Press/Proceedings/kdd96.php>.
- Étude suisse sur la santé (2023). Étude suisse sur la santé. URL: <https://www.etude-sur-la-sante.ch/> (visited on 09/08/2023).
- European Union Statistical Office (2019). *Archive:Income Poverty Statistics*. Eurostat. URL: [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Income\\_poverty\\_statistics](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Income_poverty_statistics) (visited on 09/23/2023).
- Evrard, A.-S. et al. (2017). “Does Aircraft Noise Exposure Increase the Risk of Hypertension in the Population Living near Airports in France?” In: *Occupational and environmental medicine* 74.2, pp. 123–129.
- Faber, M., A. Ghisletta, and K. Schmidheiny (2020). “A Lockdown Index to Assess the Economic Impact of the Coronavirus”. In: *Swiss J Econ Stat* 156.1, p. 11. DOI: 10.1186/s41937-020-00056-8.
- Federal Council (2020). *Coronavirus: Federal Council Declares ‘Extraordinary Situation’ and Introduces More Stringent Measures - 16 March 2020*. URL: <https://www.admin.ch/gov/en/start/documentation/media-releases.msg-id-78454.html> (visited on 06/29/2023).
- Fennel, T., N. Homer, and F. Genomics (2022). “Fgbio: Tools for Working with Genomic and High Throughput Sequencing Data”. In: URL <https://github.com/fulcrumgenomics/fgbio>. [15, 40, 81].

## Bibliography

---

- Feuillet, T. et al. (2015). "Spatial Heterogeneity of the Relationships between Environmental Characteristics and Active Commuting: Towards a Locally Varying Social Ecological Model". In: *Int J Health Geogr* 14.1, p. 12. DOI: 10.1186/s12942-015-0002-z.
- Firmann, M. et al. (2008). "The CoLaus Study: A Population-Based Study to Investigate the Epidemiology and Genetic Determinants of Cardiovascular Risk Factors and Metabolic Syndrome". In: *BMC Cardiovascular Disorders* 8.1. DOI: 10.1186/1471-2261-8-6.
- Fischer, M. M. and A. Getis, eds. (2010). *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Berlin ; New York: Springer. 811 pp.
- Fisk, R. J. (2021). "Barriers to Vaccination for Coronavirus Disease 2019 (COVID-19) Control: Experience from the United States". In: *Global Health Journal* 5.1, pp. 51–55. DOI: 10.1016/j.glohj.2021.02.005.
- Fleury, V. et al. (2021). "Geospatial Analysis of Individual-Based Parkinson's Disease Data Supports a Link with Air Pollution: A Case-Control Study". In: *Parkinsonism & Related Disorders* 83, pp. 41–48. DOI: 10.1016/j.parkreldis.2020.12.013.
- FOEN (2009). *SonBase – The GIS Noise Database of Switzerland. Technical Bases*. Environmental studies no. 0908. Federal Office for the Environment, Bern: 61 pp.
- Fotheringham, A. S. and C. Brunsdon (1999). "Local Forms of Spatial Analysis". In: *Geographical analysis* 31.4, pp. 340–358.
- Fotheringham, A. S., C. Brunsdon, and M. Charlton (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.
- Franch-Pardo, I. et al. (2020). "Spatial Analysis and GIS in the Study of COVID-19. A Review". In: *Science of The Total Environment* 739, p. 140033. DOI: 10.1016/j.scitotenv.2020.140033.
- Franch-Pardo, I. et al. (2021). "A Review of GIS Methodologies to Analyze the Dynamics of COVID-19 in the Second Half of 2020". In: *Transactions in GIS* 25.5, pp. 2191–2239. DOI: 10.1111/tgis.12792.
- Frank, L. D. et al. (2022). "Chronic Disease and Where You Live: Built and Natural Environment Relationships with Physical Activity, Obesity, and Diabetes". In: *Environment International* 158, p. 106959. DOI: 10.1016/j.envint.2021.106959.
- French, J. (2022). *Smapod: Statistical Methods for the Analysis of Case-Control Point Data*. manual. URL: <https://CRAN.R-project.org/package=smapod>.
- Frieden, T. R. and C. T. Lee (2020). "Identifying and Interrupting Superspreading Events-Implications for Control of Severe Acute Respiratory Syndrome Coronavirus 2". In: *Emerg Infect Dis* 26.6, pp. 1059–1066. DOI: 10.3201/eid2606.200495.
- Garrison, E. and G. Marth (2012). "Haplotype-Based Variant Detection from Short-Read Sequencing".
- Gatrell, A. C. et al. (1996). "Spatial Point Pattern Analysis and Its Application in Geographical Epidemiology". In: *Transactions of the Institute of British Geographers* 21.1, pp. 256–274. DOI: 10.2307/622936.

- Geoghegan, J. L. et al. (2021). "Use of Genomics to Track Coronavirus Disease Outbreaks, New Zealand". In: *Emerging Infectious Diseases* 27.5, p. 1317.
- Getis, A. (2008). "A History of the Concept of Spatial Autocorrelation: A Geographer's Perspective". In: *Geographical analysis* 40.3, pp. 297–309.
- Getis, A. and J. K. Ord (1992). "The Analysis of Spatial Association by Use of Distance Statistics". In: *Geographical analysis* 24.3, pp. 189–206.
- Greene, S. K. et al. (2021). "Detecting COVID-19 Clusters at High Spatiotemporal Resolution, New York City, New York, USA, June–July 2020". In: *Emerg. Infect. Dis.* 27.5. DOI: 10.3201/eid2705.203583.
- Greub, G. et al. (2016). "Ten Years of R&D and Full Automation in Molecular Diagnosis". In: *Future Microbiology* 11.3, pp. 403–425. DOI: 10.2217/fmb.15.152.
- Hadfield, J. et al. (2018). "Nextstrain: Real-Time Tracking of Pathogen Evolution". In: *Bioinformatics* 34.23, pp. 4121–4123.
- Hampton, K. H. et al. (2010). "Mapping Health Data: Improved Privacy Protection With Donut Method Geomasking". In: *American Journal of Epidemiology* 172.9, pp. 1062–1069. DOI: 10.1093/aje/kwq248.
- Harrison, A. G., T. Lin, and P. Wang (2020). "Mechanisms of SARS-CoV-2 Transmission and Pathogenesis". In: *Trends in immunology* 41.12, pp. 1100–1115.
- Hart, J. T. (1971). "The Inverse Care Law". In: *The Lancet* 297.7696, pp. 405–412.
- Hawe, P. and L. Potvin (2009). "What Is Population Health Intervention Research?" In: *Can J Public Health* 100.1, pp. I8–I14. DOI: 10.1007/BF03405503.
- Heiniger, S. et al. (2022). "Differences in COVID-19 Vaccination Uptake in the First 12 Months of Vaccine Availability in Switzerland – a Prospective Cohort Study". In: *Swiss Med Wkly* 152.1314, w30162. DOI: 10.4414/SMW.2022.w30162.
- Heldstab, J., B. Schächli, and T. Künzle (2020). "Aktualisierung des PolluMap- Modells für 2015, 2020 und 2030". In: *INFRAS, Meteotest*.
- Héritier, H. et al. (2017). "Transportation Noise Exposure and Cardiovascular Mortality: A Nationwide Cohort Study from Switzerland". In: *European journal of epidemiology* 32, pp. 307–315.
- Hohl, A. et al. (2020). "Daily Surveillance of COVID-19 Using the Prospective Space-Time Scan Statistic in the United States". In: *Spatial and Spatio-temporal Epidemiology* 34, p. 100354. DOI: 10.1016/j.sste.2020.100354.
- Hollands, S. et al. (2013). "A Spatial Analysis of the Association between Restaurant Density and Body Mass Index in Canadian Adults". In: *Preventive Medicine* 57.4, pp. 258–264. DOI: 10.1016/j.ypmed.2013.07.002.
- Horton, R. (2020). "Offline: COVID-19 Is Not a Pandemic". In: *The Lancet* 396.10255, p. 874. DOI: 10.1016/S0140-6736(20)32000-6.
- Huang, G., W. Zhou, and M. Cadenasso (2011). "Is Everyone Hot in the City? Spatial Pattern of Land Surface Temperatures, Land Cover and Neighborhood Socioeconomic Characteristics

## Bibliography

---

- in Baltimore, MD". In: *Journal of Environmental Management* 92.7, pp. 1753–1759. DOI: 10.1016/j.jenvman.2011.02.006.
- Huang, R. et al. (2015). "The Spatial Clustering of Obesity: Does the Built Environment Matter?" In: *J Hum Nutr Diet* 28.6, pp. 604–612. DOI: 10.1111/jhn.12279.
- Hunter, D. J. et al. (2009). "Learning Lessons from the Past: Shaping a Different Future". In: *Marmot Review Working Committee* 3, pp. 1–11.
- Jaafar, R. et al. (2020). "Correlation between 3790 qPCR Positives Samples and Positive Cell Cultures Including 1941 SARS-CoV-2 Isolates". In: *Clin Infect Dis*. DOI: 10.1093/cid/ciaa1491.
- Jaccard, P. (1912). "The Distribution of the Flora in the Alpine Zone. 1". In: *New phytologist* 11.2, pp. 37–50.
- Jacot, D., G. Greub, et al. (2020). "Viral Load of SARS-CoV-2 across Patients and Compared to Other Respiratory Viruses". In: *Microbes and Infection* 22.10, pp. 617–621. DOI: 10.1016/j.micinf.2020.08.004.
- Jacot, D., T. Pillonel, et al. (2021). "Assessment of SARS-CoV-2 Genome Sequencing: Quality Criteria and Low-Frequency Variants". In: *Journal of Clinical Microbiology* 59.10, pp. 10–1128.
- Jacquez, G. M. and D. A. Greiling (2003). "Local Clustering in Breast, Lung and Colorectal Cancer in Long Island, New York". In: *Int J Health Geogr* 2.1, p. 3. DOI: 10.1186/1476-072X-2-3.
- Jahn, K. et al. (2022). "Early Detection and Surveillance of SARS-CoV-2 Genomic Variants in Wastewater Using COJAC". In: *Nat Microbiol*. DOI: 10.1038/s41564-022-01185-x.
- Jeong, H. W. et al. (2020). "Viable SARS-CoV-2 in Various Specimens from COVID-19 Patients". In: *Clin Microbiol Infect* 26.11, pp. 1520–1524. DOI: 10.1016/j.cmi.2020.07.020.
- Jia, P. (2019). "Spatial Lifecourse Epidemiology". In: *The Lancet Planetary Health* 3.2, e57–e59. DOI: 10.1016/S2542-5196(18)30245-6.
- Jones, R. C. et al. (2006). "Use of a Prospective Space-Time Scan Statistic to Prioritize Shigellosis Case Investigations in an Urban Jurisdiction". In: *Public Health Rep* 121.2, pp. 133–139. DOI: 10.1177/003335490612100206.
- Joost, S., S. Duruz, et al. (2016). "Persistent Spatial Clusters of High Body Mass Index in a Swiss Urban Population as Revealed by the 5-Year GeoCoLaus Longitudinal Study". In: *BMJ Open* 6.1, e010145. DOI: 10.1136/bmjopen-2015-010145.
- Joost, S., J. Haba-Rubio, et al. (2018). "Spatial Clusters of Daytime Sleepiness and Association with Nighttime Noise Levels in a Swiss General Population (GeoHypnoLaus)". In: *International Journal of Hygiene and Environmental Health* 221.6, pp. 951–957. DOI: 10.1016/j.ijheh.2018.05.004.
- Kadambari, S. and S. Vanderslott (2021). "Lessons about COVID-19 Vaccine Hesitancy among Minority Ethnic People in the UK". In: *The Lancet Infectious Diseases* 21.9, pp. 1204–1206. DOI: 10.1016/S1473-3099(21)00404-7.



- Kamel Boulos, M. N. and E. M. Geraghty (2020). "Geographical Tracking and Mapping of Coronavirus Disease COVID-19/Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Epidemic and Associated Events around the World: How 21st Century GIS Technologies Are Supporting the Global Fight against Outbreaks and Epidemics". In: *Int J Health Geogr* 19.1, 8, s12942-020-00202-8. DOI: 10.1186/s12942-020-00202-8.
- Kaufman, L. and P. J. Rousseeuw (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.
- Kauhl, B. et al. (2016). "Do the Risk Factors for Type 2 Diabetes Mellitus Vary by Location? A Spatial Analysis of Health Insurance Claims in Northeastern Germany Using Kernel Density Estimation and Geographically Weighted Regression". In: *International Journal of Health Geographics* 15.1. DOI: 10.1186/s12942-016-0068-2.
- Keesara, S., A. Jonas, and K. Schulman (2020). "Covid-19 and Health Care's Digital Revolution". In: *New England Journal of Medicine* 382.23, e82. DOI: 10.1056/NEJMp2005835.
- Kelsall, J. E. and P. J. Diggle (1995a). "Kernel Estimation of Relative Risk". In: *Bernoulli* 1.1/2, p. 3. DOI: 10.2307/3318678.
- (1995b). "Non-Parametric Estimation of Spatial Variation in Relative Risk". In: *Statistics in medicine* 14.21-22, pp. 2335-2342.
- Kirby, R. S., E. Delmelle, and J. M. Eberth (2017). "Advances in Spatial Epidemiology and Geographic Information Systems". In: *Annals of Epidemiology* 27.1, pp. 1-9. DOI: 10.1016/j.annepidem.2016.12.001.
- Klauss, G. et al. (2005). "Hospital Service Areas – a New Tool for Health Care Planning in Switzerland". In: *BMC Health Serv Res* 5.1, p. 33. DOI: 10.1186/1472-6963-5-33.
- Klompmaier, J. O. et al. (2019). "Associations of Combined Exposures to Surrounding Green, Air Pollution, and Road Traffic Noise with Cardiometabolic Diseases". In: *Environ Health Perspect* 127.8, p. 087003. DOI: 10.1289/EHP3857.
- Kluge, H. H. P. et al. (2020). "Prevention and Control of Non-Communicable Diseases in the COVID-19 Response". In: *The Lancet* 395.10238, pp. 1678-1680. DOI: 10.1016/S0140-6736(20)31067-9.
- Knox, E. G. (1989). "Detection of Clusters". In: *Methodology of enquiries into disease clustering. London: Small Area Health Statistics Unit* 17, p. 20.
- Kolak, M. et al. (2020). "Quantification of Neighborhood-Level Social Determinants of Health in the Continental United States". In: *JAMA Netw Open* 3.1, e1919928. DOI: 10.1001/jamanetworkopen.2019.19928.
- Konstantinoudis, G. et al. (2017). "Spatial Clustering of Childhood Leukaemia in Switzerland: A Nationwide Study". In: *International journal of cancer* 141.7, pp. 1324-1332.
- Koschinsky, J. (2013). "The Case for Spatial Analysis in Evaluation to Reduce Health Inequities". In: *Evaluation and Program Planning* 36.1, pp. 172-176. DOI: 10.1016/j.evalprogplan.2012.03.004.

## Bibliography

---

- Kraemer, M. U. G. et al. (2021). "Spatiotemporal Invasion Dynamics of SARS-CoV-2 Lineage B.1.1.7 Emergence". In: *Science* 373.6557, pp. 889–895. DOI: 10.1126/science.abj0113.
- Krieger, N. et al. (2020). "The Fierce Urgency of Now: Closing Glaring Gaps in US Surveillance Data on COVID-19". In: *Health Affairs Blog*, p. 6.
- Kubik, S. et al. (2021). "Recommendations for Accurate Genotyping of SARS-CoV-2 Using Amplicon-Based Sequencing of Clinical Samples". In: *Clinical Microbiology and Infection* 27.7, 1036–e1.
- Kuehn, M. et al. (2022). "Assessing Barriers to Access and Equity for COVID-19 Vaccination in the US". In: *BMC Public Health* 22.1, p. 2263. DOI: 10.1186/s12889-022-14636-1.
- Kulldorff, M. (2018). "SaTScan™ User Guide for Version 9.6. 2018". In:
- Kulldorff, M. et al. (1998). "Evaluating Cluster Alarms: A Space-Time Scan Statistic and Brain Cancer in Los Alamos, New Mexico." In: *Am J Public Health* 88.9, pp. 1377–1380. DOI: 10.2105/AJPH.88.9.1377.
- Kulldorff, M. (1997). "A Spatial Scan Statistic". In: *Communications in Statistics - Theory and Methods* 26.6, pp. 1481–1496. DOI: 10.1080/03610929708831995.
- (2001). "Prospective Time Periodic Geographical Disease Surveillance Using a Scan Statistic". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 164.1, pp. 61–72. DOI: 10.1111/1467-985X.00186.
- Kuo, F.-Y., T.-H. Wen, and C. E. Sabel (2018). "Characterizing Diffusion Dynamics of Disease Clustering: A Modified Space-Time DBSCAN (MST-DBSCAN) Algorithm". In: *Annals of the American Association of Geographers* 108.4, pp. 1168–1186. DOI: 10.1080/24694452.2017.1407630.
- Ladoy, A., O. Opota, et al. (2021). "Size and Duration of COVID-19 Clusters Go along with a High SARS-CoV-2 Viral Load: A Spatio-Temporal Investigation in Vaud State, Switzerland". In: *Science of The Total Environment* 787. DOI: 10.1016/j.scitotenv.2021.147483.
- Ladoy, A., J. R. Vallarta-Robledo, et al. (2021). "Geographic Footprints of Life Expectancy Inequalities in the State of Geneva, Switzerland". In: *Sci Rep* 11.1, p. 23326. DOI: 10.1038/s41598-021-02733-x.
- Lai, A. et al. (2022). "Phylogeography and Genomic Epidemiology of SARS-CoV-2 in Italy and Europe with Newly Characterized Italian Genomes between February-June 2020". In: *Sci Rep* 12.1, p. 5736. DOI: 10.1038/s41598-022-09738-0.
- Lalloué, B. et al. (2013). "A Statistical Procedure to Create a Neighborhood Socioeconomic Index for Health Inequalities Analysis". In: *Int J Equity Health* 12.1, p. 21. DOI: 10.1186/1475-9276-12-21.
- Lambio, C. et al. (2023). "Exploring the Spatial Relative Risk of COVID-19 in Berlin-Neukölln". In: *IJERPH* 20.10, p. 5830. DOI: 10.3390/ijerph20105830.
- Lane, C. R. et al. (2021). "Genomics-Informed Responses in the Elimination of COVID-19 in Victoria, Australia: An Observational, Genomic Epidemiological Study". In: *The Lancet Public Health* 6.8, e547–e556. DOI: 10.1016/S2468-2667(21)00133-X.

- Lau, M. S. Y. et al. (2020). "Characterizing Superspreading Events and Age-Specific Infectiousness of SARS-CoV-2 Transmission in Georgia, USA". In: *PNAS* 117.36, pp. 22430–22435. DOI: 10.1073/pnas.2011802117.
- Lawson, A. (2006). *Statistical Methods in Spatial Epidemiology*. 2nd ed. Wiley Series in Probability and Statistics. Chichester, England ; Hoboken, NJ: Wiley. 398 pp.
- Leal, C. and B. Chaix (2011). "The Influence of Geographic Life Environments on Cardiometabolic Risk Factors: A Systematic Review, a Methodological Assessment and a Research Agenda: Geographic Life Environments and Cardiometabolic Risk Factors". In: *Obesity Reviews* 12.3, pp. 217–230. DOI: 10.1111/j.1467-789X.2010.00726.x.
- Lebel, G. et al. (2021). "Detection of COVID-19 Case Clusters in Québec, May–October 2020". In: *Can J Public Health* 112.5, pp. 807–817. DOI: 10.17269/s41997-021-00560-1.
- Lee, A., H. Jordan, and J. Horsley (2015). "Value of Urban Green Spaces in Promoting Healthy Living and Wellbeing: Prospects for Planning". In: *RMHP*, p. 131. DOI: 10.2147/RMHP.S61654.
- Lemaitre, J. C. et al. (2021). "A Scenario Modeling Pipeline for COVID-19 Emergency Planning". In: *Sci Rep* 11.1, p. 7534. DOI: 10.1038/s41598-021-86811-0.
- Lessler, J. et al. (2017). "What Is a Hotspot Anyway?" In: *The American Journal of Tropical Medicine and Hygiene* 96.6, pp. 1270–1273. DOI: 10.4269/ajtmh.16-0427.
- Li, H. (2013). "Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM".
- Li, Y.-S. and Y.-C. Chuang (2009). "Neighborhood Effects on an Individual's Health Using Neighborhood Measurements Developed by Factor Analysis and Cluster Analysis". In: *J Urban Health* 86.1, pp. 5–18. DOI: 10.1007/s11524-008-9306-7.
- Liu, Y. et al. (2020). "The Reproductive Number of COVID-19 Is Higher Compared to SARS Coronavirus". In: *Journal of Travel Medicine* 27 (taaa021). DOI: 10.1093/jtm/taaa021.
- Lloyd-Smith, J. O. et al. (2005). "Superspreading and the Effect of Individual Variation on Disease Emergence". In: *Nature* 438.7066 (7066), pp. 355–359. DOI: 10.1038/nature04153.
- Lo, S. W. and D. Jamroz (2020). "Author Correction: Genomics and Epidemiological Surveillance". In: *Nature reviews. Microbiology* 18.9, p. 539.
- Lowe, M., C. Whitzman, and B. Giles-Corti (2018). "Health-Promoting Spatial Planning: Approaches for Strengthening Urban Policy Integration". In: *Planning theory & practice* 19.2, pp. 180–197.
- Lubelchek, R. J. et al. (2015). "Transmission Clustering Among Newly Diagnosed HIV Patients in Chicago, 2008 to 2011: Using Phylogenetics to Expand Knowledge of Regional HIV Transmission Patterns". In: *JAIDS Journal of Acquired Immune Deficiency Syndromes* 68.1, pp. 46–54. DOI: 10.1097/QAI.0000000000000404.
- Lutz, A. (2019). "Revue Médicale Suisse : Vers un universalisme proportionné en promotion de la santé et prévention : réflexions et pistes d'action". Trans. by M. Pasche. In: *Revue*

## Bibliography

---

- Médicale Suisse* 15.669. Ed. by K. Zuercher, pp. 1987–1990. DOI: 10.53738/REVMED.2019.15.669.1987.
- Macintyre, S., A. Ellaway, and S. Cummins (2002). “Place Effects on Health: How Can We Conceptualise, Operationalise and Measure Them?” In: *Social Science & Medicine* 55.1, pp. 125–139. DOI: 10.1016/S0277-9536(01)00214-3.
- MacMahon, B., T. F. Pugh, et al. (1970). “Epidemiology: Principles and Methods.” In: *Epidemiology: principles and methods*.
- Mandja, B.-A. M. et al. (2019). “Temporal and Spatial Dynamics of Monkeypox in Democratic Republic of Congo, 2000–2015”. In: *EcoHealth* 16.3, pp. 476–487. DOI: 10.1007/s10393-019-01435-1.
- Manisalidis, I. et al. (2020). “Environmental and Health Impacts of Air Pollution: A Review”. In: *Front. Public Health* 8, p. 14. DOI: 10.3389/fpubh.2020.00014.
- Marmot, M., S. Stansfeld, et al. (1991). “Health Inequalities among British Civil Servants: The Whitehall II Study”. In: *The Lancet* 337.8754, pp. 1387–1393. DOI: 10.1016/0140-6736(91)93068-K.
- Marmot, M., J. Allen, et al. (2012). “WHO European Review of Social Determinants of Health and the Health Divide”. In: *The Lancet* 380.9846, pp. 1011–1029. DOI: 10.1016/S0140-6736(12)61228-8.
- Marmot, M. and J. J. Allen (2014). “Social Determinants of Health Equity”. In: *Am J Public Health* 104.S4, S517–S519. DOI: 10.2105/AJPH.2014.302200.
- Marmot Review et al. (2010). “Fair Society, Healthy Lives: Strategic Review of Health Inequalities in England Post 2010”. In: *London: University College London*.
- Maroko, A. R., D. Nash, and B. T. Pavidonis (2020). “COVID-19 and Inequity: A Comparative Spatial Analysis of New York City and Chicago Hot Spots”. In: *J Urban Health* 97.4, pp. 461–470. DOI: 10.1007/s11524-020-00468-0.
- Matter-Walstra, K., M. Widmer, and A. Busato (2006). “Analysis of Patient Flows for Orthopedic Procedures Using Small Area Analysis in Switzerland”. In: *BMC Health Serv Res* 6.1, p. 119. DOI: 10.1186/1472-6963-6-119.
- Mattig, T. (2013). *L'échec de La Loi Sur La Prévention: Un Enseignement. Document de Travail* 9. URL: <https://promozionesalute.ch> (visited on 06/09/2023).
- Maung, K. K. and P. Marques-Vidal (2023). “Impact of COVID-19 Pandemic on Cardiovascular Diseases Hospitalisation, Management and Mortality in Switzerland”. In: *Open Heart* 10.1, e002259. DOI: 10.1136/openhrt-2023-002259.
- Maxmen, A. (2021). “One Million Coronavirus Sequences: Popular Genome Site Hits Mega Milestone”. In: *Nature* 593.7857, pp. 21–21.
- Mayer, J. D. (1983). “The Role of Spatial Analysis and Geographic Data in the Detection of Disease Causation”. In: *Social Science & Medicine* 17.16, pp. 1213–1221. DOI: 10.1016/0277-9536(83)90014-X.

- Mennis, J., K. A. Matthews, and S. L. Huston (2022). “Geospatial Perspectives on the Intersection of Chronic Disease and COVID-19”. In: *Prev. Chronic Dis.* 19, p. 220145. DOI: 10.5888/pcd19.220145.
- Meredith, L. W. et al. (2020). “Rapid Implementation of SARS-CoV-2 Sequencing to Investigate Cases of Health-Care Associated COVID-19: A Prospective Genomic Surveillance Study”. In: *The Lancet Infectious Diseases* 20.11, pp. 1263–1271. DOI: 10.1016/S1473-3099(20)30562-4.
- Métasanté | Plateforme participative (2023). Métasanté. URL: <https://metasante.ch/> (visited on 09/08/2023).
- Michaels, I. H., S. J. Pirani, and A. Carrascal (2021). “Peer Reviewed: Disparities in Internet Access and COVID-19 Vaccination in New York City”. In: *Preventing Chronic Disease* 18.
- MICROGIS (2019). *LC Swiss Localities - Swiss Postcode Database*.
- (2023). *Données Géographiques - Swiss Areas*. URL: <https://microgis.ch/prestations/donnees-geographiques-statistiques/> (visited on 08/24/2023).
- Monin, M. et al. (2020). “Évaluation du label «Commune en santé»”. Promotion Santé Suisse”. In: *EESP*.
- Monmonier, M. (2018). *How to Lie with Maps*. University of Chicago Press.
- Monod, S. et al. (2023). “Système de santé suisse : y a-t-il un pilote dans la machine ?” In: *Rev Med Suisse* 819, pp. 583–588. URL: <https://www.revmed.ch/revue-medicale-suisse/2023/revue-medicale-suisse-819/systeme-de-sante-suisse-y-a-t-il-un-pilote-dans-la-machine>.
- Monteverde, S. and M. Eicher (2023). “Fighting the Virus Is Not Enough—Pandemics, Social Justice, and the Role of Nurses in Switzerland”. In: *J of Nursing Scholarship* 55.1, pp. 8–10. DOI: 10.1111/jnu.12801.
- Morabia, A. et al. (1997). “Community-Based Surveillance of Cardiovascular Risk Factors in Geneva: Methods, Resulting Distributions, and Comparisons with Other Populations”. In: *Preventive Medicine* 26.3, pp. 311–319. DOI: 10.1006/pmed.1997.0146.
- Moraga, P. and F. Montes (2011). “Detection of Spatial Disease Clusters with LISA Functions”. In: *Stat Med* 30.10, pp. 1057–1071. DOI: 10.1002/sim.4160.
- Moran, P. A. (1950). “Notes on Continuous Stochastic Phenomena”. In: *Biometrika* 37.1/2, pp. 17–23.
- Moraz, M. et al. (2020). “Universal Admission Screening Strategy for COVID-19 Highlighted the Clinical Importance of Reporting SARS-CoV-2 Viral Loads”. In: *New Microbes and New Infections* 38, p. 100820. DOI: 10.1016/j.nmni.2020.100820.
- Mueller, L. et al. (2020). “Computer-Aided Medical Microbiology Monitoring Tool: A Strategy to Adapt to the SARS-CoV-2 Epidemic and That Highlights RT-PCR Consistency”. In: *medRxiv*, p. 2020.07.27.20162123. DOI: 10.1101/2020.07.27.20162123.
- Myers, S. S. (2017). “Planetary Health: Protecting Human Health on a Rapidly Changing Planet”. In: *The Lancet* 390.10114, pp. 2860–2868. DOI: 10.1016/S0140-6736(17)32846-5.

## Bibliography

---

- Nakazawa, Y. et al. (2015). "A Phylogeographic Investigation of African Monkeypox". In: *Viruses* 7.4, pp. 2168–2184.
- Nature (2022). "This Is No Time to Stop Tracking COVID-19". In: *Nature* 603, p. 550.
- Nicodemo, C. et al. (2020). "Measuring Geographical Disparities in England at the Time of COVID-19: Results Using a Composite Indicator of Population Vulnerability". In: *BMJ Open* 10.9, e039749. DOI: 10.1136/bmjopen-2020-039749.
- Nicola, M. et al. (2020). "Health Policy and Leadership Models during the COVID-19 Pandemic: A Review". In: *International Journal of Surgery* 81, pp. 122–129. DOI: 10.1016/j.ijssu.2020.07.026.
- Nicollier, M. (10.05.21). *L'injection près de chez soi – Vaud innove en lançant la vaccination itinérante*. 24 heures. URL: <https://www.24heures.ch/naud-innove-en-lancant-la-vaccination-itinerante-524147763858> (visited on 09/23/2023).
- Njoku, A., M. Joseph, and R. Felix (2021). "Changing the Narrative: Structural Barriers and Racial and Ethnic Inequities in COVID-19 Vaccination". In: *IJERPH* 18.18, p. 9904. DOI: 10.3390/ijerph18189904.
- Nykiforuk, C. I. J. and L. M. Flaman (2011). "Geographic Information Systems (GIS) for Health Promotion and Public Health: A Review". In: *Health Promotion Practice* 12.1, pp. 63–73. DOI: 10.1177/1524839909334624.
- O'Brien, R. M. (2007). "A Caution Regarding Rules of Thumb for Variance Inflation Factors". In: *Qual Quant* 41.5, pp. 673–690. DOI: 10.1007/s11135-006-9018-6.
- O'Toole, Á. et al. (2021). "Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pangolin Tool". In: *Virus evolution* 7.2, veab064.
- OBSAN (2015). *La Santé En Suisse – Le Point Sur Les Maladies Chroniques. Rapport National Sur La Santé 2015*.
- OCDE (2017). *Panorama de la santé 2017: Les indicateurs de l'OCDE*. Panorama de la santé. OECD. DOI: 10.1787/health\_glance-2017-fr.
- OECD (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD. DOI: 10.1787/9789264043466-en.
- OFEV (2018). *Pollution sonore en Suisse*. Office Fédéral de l'Environnement OFEV.
- (2023). *Valeurs limites d'exposition au bruit*. Office fédéral de l'environnement OFEV. URL: <https://www.bafu.admin.ch/bafu/fr/home/themen/thema-laerm/laerm-fachinformationen/laermbelastung/grenzwerte-fuer-laerm/belastungsgrenzwerte-fuer-laerm.html> (visited on 09/24/2023).
- OFS (2018). *Services à La Population et Disparités Régionales: Distances d'accès Aux Services de La Vie Courante, En 2015*. Office fédéral de la statistique (OFS).
- (2022a). *Accidents de la route*. URL: <https://www.bfs.admin.ch/bfs/fr/home/statistiken/mobilitaet-verkehr/unfaelle-umweltauswirkungen/verkehrsunfaelle/strassenverkehr.html> (visited on 10/08/2023).

- (2022b). *CIM-10-GM 2022. Index Systématique – Version Française*. Office fédéral de la statistique OFS.
- (2023a). *Maladies cardiovasculaires*. Office fédéral de la statistique. URL: <https://www.bfs.admin.ch/bfs/fr/home/statistiken/gesundheit/gesundheitszustand/krankheiten/herz-kreislauf-erkrankungen.html> (visited on 06/23/2023).
- (2023b). *Portraits City Statistics 2022: villes*. Office fédéral de la statistique (OFS).
- (2023c). *Registre Fédéral Des Bâtiments et Des Logements (RegBL)*. Office fédéral de la statistique (OFS). URL: <https://www.housing-stat.ch> (visited on 08/24/2023).
- OFSP (2019). *Politique de la santé; stratégie du Conseil fédéral 2020-2030*. Office fédéral de la santé publique.
- Okonechnikov, K., A. Conesa, and F. García-Alcalde (2016). “Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data”. In: *Bioinformatics* 32.2, pp. 292–294.
- Openshaw, S. (1984). “The Modifiable Areal Unit Problem”. In: *Concepts and techniques in modern geography*.
- Opota, O. et al. (2020). “Comparison of SARS-CoV-2 RT-PCR on a High-Throughput Molecular Diagnostic Platform and the Cobas SARS-CoV-2 Test for the Diagnostic of COVID-19 on Various Clinical Samples”. In: *Pathog Dis* 78.8. DOI: 10.1093/femspd/ftaa061.
- Oshan, T. et al. (2019). “Mgwr: A Python Implementation of Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale”. In: *IJGI* 8.6, p. 269. DOI: 10.3390/ijgi8060269.
- Oshan, T. M., J. P. Smith, and A. S. Fotheringham (2020). “Targeting the Spatial Context of Obesity Determinants via Multiscale Geographically Weighted Regression”. In: *Int J Health Geogr* 19.1, p. 11. DOI: 10.1186/s12942-020-00204-6.
- OSM contributors (2023). *Playgrounds Retrieved from https://overpass-turbo.eu/*. OpenStreetMap. URL: <https://www.openstreetmap.org>.
- Owen, G., R. Harris, and K. Jones (2016). “Under Examination: Multilevel Models, Geography and Health Research”. In: *Progress in Human Geography* 40.3, pp. 394–412. DOI: 10.1177/0309132515580814.
- Pahud, O. and J. Zufferey (2019). *Rapport de Base Sur La Santé Pour Le Canton de Vaud. Exploitations Standardisées Des Données de l'Enquête Suisse Sur La Santé 2017 et d'autres Bases de Données*. Observatoire suisse de la santé (Obsan). URL: [https://www.obsan.admin.ch/sites/default/files/2021-08/obsan\\_rapport\\_12-2019-vaud.pdf](https://www.obsan.admin.ch/sites/default/files/2021-08/obsan_rapport_12-2019-vaud.pdf).
- Pampalon, R. et al. (2009). “A Deprivation Index for Health Planning in Canada”. In: *Chronic Dis Can* 29.4, pp. 178–191. DOI: 10.24095/hpcdp.29.4.05.
- Panczak, R. et al. (2012). “A Swiss Neighbourhood Index of Socioeconomic Position: Development and Association with Mortality”. In: *J Epidemiol Community Health* 66.12, pp. 1129–1136. DOI: 10.1136/jech-2011-200699.

## Bibliography

---

- Paquet, C. et al. (2016). “Geographic Clustering of Cardiometabolic Risk Factors in Metropolitan Centres in France and Australia”. In: *IJERPH* 13.5, p. 519. DOI: 10.3390/ijerph13050519.
- Paskett, E. et al. (2016). “Multilevel Interventions to Address Health Disparities Show Promise in Improving Population Health”. In: *Health Affairs* 35.8, pp. 1429–1434.
- Pearce, J. R. et al. (2010). “Environmental Justice and Health: The Implications of the Socio-Spatial Distribution of Multiple Environmental Deprivation for Health Inequalities in the United Kingdom: Environmental Justice and Health”. In: *Transactions of the Institute of British Geographers* 35.4, pp. 522–539. DOI: 10.1111/j.1475-5661.2010.00399.x.
- Perroud, S. (19.08.21). *Improving Mobile COVID Vaccination Efforts through GIS Mapping*. EPFL. URL: <https://actu.epfl.ch/news/improving-mobile-covid-vaccination-efforts-throu-3/> (visited on 09/01/2023).
- Petri, D. et al. (2021). “Effects of Exposure to Road, Railway, Airport and Recreational Noise on Blood Pressure and Hypertension”. In: *IJERPH* 18.17, p. 9145. DOI: 10.3390/ijerph18179145.
- Picheral, H. (1982). “Géographie médicale, géographie des maladies, géographie de la santé”. In: *Espace géographique* 11.3, pp. 161–175. DOI: 10.3406/spgeo.1982.3751.
- Pillonel, T. et al. (2020). “Letter to the Editor: SARS-CoV-2 Detection by Real-Time RT-PCR”. In: *Euro Surveill* 25.21. DOI: 10.2807/1560-7917.ES.2020.25.21.2000880.
- Qutob, N. et al. (2021). “Genomic Epidemiology of the First Epidemic Wave of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in Palestine”. In: *Microbial Genomics* 7.6.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. manual. Vienna, Austria: R Foundation for Statistical Computing. URL: <https://www.R-project.org/>.
- Rey, G. et al. (2009). “Ecological Association between a Deprivation Index and Mortality in France over the Period 1997 – 2001: Variations with Spatial Scale, Degree of Urbanicity, Age, Gender and Cause of Death”. In: *BMC Public Health* 9.1, p. 33. DOI: 10.1186/1471-2458-9-33.
- Rey, S. J. and L. Anselin (2007). “PySAL: A Python Library of Spatial Analytical Methods”. In: *The Review of Regional Studies* 37.1, pp. 5–27.
- Rezaeian, M. et al. (2007). “Geographical Epidemiology, Spatial Analysis and Geographical Information Systems: A Multidisciplinary Glossary”. In: *Journal of Epidemiology & Community Health* 61.2, pp. 98–102. DOI: 10.1136/jech.2005.043117.
- Ribeiro, A. I., M. d. F. de Pina, and R. Mitchell (2015). “Development of a Measure of Multiple Physical Environmental Deprivation. After United Kingdom and New Zealand, Portugal”. In: *Eur J Public Health* 25.4, pp. 610–617. DOI: 10.1093/eurpub/cku242.
- Richardson, E. A. et al. (2010). “Developing Summary Measures of Health-Related Multiple Physical Environmental Deprivation for Epidemiological Research”. In: *Environ Plan A* 42.7, pp. 1650–1668. DOI: 10.1068/a42459.



- Riou, J. et al. (2021). "Socioeconomic Position and the COVID-19 Care Cascade from Testing to Mortality in Switzerland: A Population-Based Analysis". In: *The Lancet Public Health* 6.9, e683–e691. DOI: 10.1016/S2468-2667(21)00160-2.
- Rose, G. A., K.-T. Khaw, and M. G. Marmot (2008). *Rose's Strategy of Preventive Medicine: The Complete Original Text*. New ed. Oxford ; New York: Oxford University Press. 171 pp.
- Rose, G. (1992). "The Population Strategy of Prevention". In: *The strategy of preventive medicine*, pp. 95–106.
- Ruktanonchai, N. W. et al. (2020). "Assessing the Impact of Coordinated COVID-19 Exit Strategies across Europe". In: *Science* 369.6510, pp. 1465–1470. DOI: 10.1126/science.abc5096.
- Rushton, G. (2003). "Public Health, GIS, and Spatial Analytic Tools". In: *Annu. Rev. Public Health* 24.1, pp. 43–56. DOI: 10.1146/annurev.publhealth.24.012902.140843.
- Saaty, T. L. (1980). "The Analytic Hierarchy Process, New York: Mcgrew Hill". In: *International, Translated to Russian, Portugueses and Chinese, Revised edition, Paperback (1996, 2000), Pittsburgh: RWS Publications* 9, pp. 19–22.
- Sacarny, A. and J. R. Daw (2021). "Inequities in COVID-19 Vaccination Rates in the 9 Largest US Cities". In: *JAMA Health Forum* 2.9, e212415. DOI: 10.1001/jamahealthforum.2021.2415.
- Schüle, S. A. et al. (2019). "Social Inequalities in Environmental Resources of Green and Blue Spaces: A Review of Evidence in the WHO European Region". In: *IJERPH* 16.7, p. 1216. DOI: 10.3390/ijerph16071216.
- Schwob, J. M. et al. (2020). "Antigen Rapid Tests, Nasopharyngeal PCR and Saliva PCR to Detect SARS-CoV-2: A Prospective Comparative Clinical Trial". In: *medRxiv*, p. 2020.11.23.20237057. DOI: 10.1101/2020.11.23.20237057.
- SFSO (2017). "Population and Household Statistics (STATPOP)". In: *Spatial Data*.
- (2019). *STATPOP - Population and Households Statistics*. Federal Statistical Office. URL: /content/bfs/fr/home/statistiken/bevoelkerung/erhebungen/statpop.assetdetail.7766422.html (visited on 01/18/2021).
- (2020). *Population and Household Statistics (STATPOP), Spatial Data 2019*. Federal Statistical Office. URL: /content/bfs/de/home/dienstleistungen/geostat/geodaten-bundesstatistik/gebaeude-wohnungen-haushalte-personen/bevoelkerung-haushalte-ab-2010.assetdetail.14716365.html (visited on 01/18/2021).
- Shannon, P. et al. (2003). "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks". In: *Genome research* 13.11, pp. 2498–2504.
- Shrestha, R. et al. (2016). "Environmental Health Related Socio-Spatial Inequalities: Identifying "Hotspots" of Environmental Burdens and Social Vulnerability". In: *IJERPH* 13.7, p. 691. DOI: 10.3390/ijerph13070691.
- Shrestha, S. et al. (2020). "Spatial Epidemiology: An Empirical Framework for Syndemics Research". In: *Social Science & Medicine*, p. 113352. DOI: 10.1016/j.socscimed.2020.113352.
- Singer, M. (1996). "A Dose of Drugs, a Touch of Violence, a Case of AIDS: Conceptualizing the SAVA Syndemic". In: p. 12.

## Bibliography

---

- Skaathun, B. et al. (2021). "Interplay Between Geography and HIV Transmission Clusters in Los Angeles County". In: *Open Forum Infectious Diseases* 8.6, ofab211. DOI: 10.1093/ofid/ofab211.
- Snow, J. (1855). *On the Mode of Communication of Cholera*. John Churchill.
- Soetens, L. C. et al. (2015). "Ongoing Transmission of Hepatitis B Virus in Rural Parts of the Netherlands, 2009–2013". In: *PLoS ONE* 10.2. Ed. by O. Schildgen, e0117703. DOI: 10.1371/journal.pone.0117703.
- Specchio-Hub* (2023). Specchio-Hub. URL: <https://www.specchio-hub.ch/> (visited on 10/03/2023).
- Spiess, M. and K. Schnyder-Walser (2018). "Égalité Des Chances et Santé—Chiffres et Données Pour La Suisse". In: *Document de base*. Berne: Socialdesign SA, sur mandat de l'Office fédéral de la santé publique (OFSP).
- Spoerri, A. et al. (2010). "The Swiss National Cohort: A Unique Database for National and International Researchers". In: *Int J Public Health* 55.4, pp. 239–242. DOI: 10.1007/s00038-010-0160-5.
- Stange, M. et al. (2021). "SARS-CoV-2 Outbreak in a Tri-National Urban Area Is Dominated by a B.1 Lineage Variant Linked to a Mass Gathering Event". In: *PLoS Pathog* 17.3. Ed. by K. Subbarao, e1009374. DOI: 10.1371/journal.ppat.1009374.
- Stein, R. A. (2011). "Super-Spreaders in Infectious Diseases". In: *International Journal of Infectious Diseases* 15.8, e510–e513. DOI: 10.1016/j.ijid.2010.06.020.
- Swanlund, D. et al. (2020). "Street Masking: A Network-Based Geographic Mask for Easily Protecting Geoprivacy". In: *Int J Health Geogr* 19.1, p. 26. DOI: 10.1186/s12942-020-00219-z.
- Swiss Federal Office of Public Health (2020). *Infectious Diseases Requiring Notification—Key Facts*. URL: <https://www.bag.admin.ch/bag/en/home/krankheiten/infektionskrankheiten-bekaempfen/meldesysteme-infektionskrankheiten/meldepflichtige-ik.html>.
- Swiss National COVID-19 Science Task Force (2020). *Strategy to React to Substantial Increases in the Numbers of SARS-CoV-2 Infections in Switzerland*. URL: [https://scienctaskforce.ch/wp-content/uploads/2020/10/Strategy\\_to\\_react\\_to\\_substantial\\_increases\\_in\\_the\\_numbers\\_of\\_SARS-CoV-2\\_26\\_June\\_-\\_EN\\_old.pdf](https://scienctaskforce.ch/wp-content/uploads/2020/10/Strategy_to_react_to_substantial_increases_in_the_numbers_of_SARS-CoV-2_26_June_-_EN_old.pdf).
- swisstopo (2023). *swissTLMRegio*. URL: <https://www.swisstopo.admin.ch/en/geodata/landscape/tlmregio.html> (visited on 10/08/2023).
- Syme, S. L. (1986). "Strategies for Health Promotion". In: *Preventive Medicine* 15.5, pp. 492–507.
- Tegally, H. et al. (2021). "Detection of a SARS-CoV-2 Variant of Concern in South Africa". In: *Nature* 592.7854, pp. 438–443.
- Thacker, S. B. and R. L. Berkelman (1988). "Public Health Surveillance in the United States". In: *Epidemiologic reviews* 10.1, pp. 164–190.
- The Lancet (2017). "Syndemics: Health in Context". In: *The Lancet* 389.10072, p. 881. DOI: 10.1016/S0140-6736(17)30640-2.

- Thommen, O. (2005). "Heat Wave 2003 and Mortality in Switzerland". In: *Swiss medical weekly* 135.1314, pp. 200–205.
- To, K. K.-W. et al. (2020). "Temporal Profiles of Viral Load in Posterior Oropharyngeal Saliva Samples and Serum Antibody Responses during Infection by SARS-CoV-2: An Observational Cohort Study". In: *Lancet Infect Dis* 20.5, pp. 565–574. DOI: 10.1016/S1473-3099(20)30196-1.
- Tobler, W. R. (1970). "A Computer Movie Simulating Urban Growth in the Detroit Region". In: *Economic geography* 46 (sup1), pp. 234–240.
- Toms, R. et al. (2019). "Geographic and Area-Level Socioeconomic Variation in Cardiometabolic Risk Factor Distribution: A Systematic Review of the Literature". In: *Int J Health Geogr* 18.1, p. 1. DOI: 10.1186/s12942-018-0165-5.
- Townsend, P. (1987). "Deprivation". In: *Journal of social policy* 16.2, pp. 125–146.
- Turner, M. C. et al. (2017). "Assessing the Exposome with External Measures: Commentary on the State of the Science and Research Recommendations". In: *Annu. Rev. Public Health* 38.1, pp. 215–239. DOI: 10.1146/annurev-publhealth-082516-012802.
- USGS (2022). *Landsat 8-9 Collection 2 (C2) Level 2 Science Product (L2SP) Guide (Version 4.0)*. United States Geological Survey. URL: [https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/media/files/LSDS-1619\\_Landsat-8-9-C2-L2-ScienceProductGuide-v4.pdf](https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/media/files/LSDS-1619_Landsat-8-9-C2-L2-ScienceProductGuide-v4.pdf).
- Vallarta-Robledo, J. R. et al. (2022). "The Neighborhood Environment and Its Association with the Spatio-Temporal Footprint of Tobacco Consumption and Changes in Smoking-Related Behaviors in a Swiss Urban Area". In: *Health & Place* 76, p. 102845. DOI: 10.1016/j.healthplace.2022.102845.
- Verhoog, S. et al. (2019). "Comparison of the Physical Activity Frequency Questionnaire (PAFQ) with Accelerometry in a Middle-Aged and Elderly Population: The CoLaus Study". In: *Maturitas* 129, pp. 68–75. DOI: 10.1016/j.maturitas.2019.08.004.
- Veys-Takeuchi, C. et al. (2022). "Determinants of COVID-19 Vaccine Hesitancy During the Pandemic: A Cross-Sectional Survey in the Canton of Vaud, Switzerland". In: *Int J Public Health* 67, p. 1604987. DOI: 10.3389/ijph.2022.1604987.
- Vienneau, D., I. C. Eze, et al. (2019). "Association between Transportation Noise and Cardio-Metabolic Diseases: An Update of the WHO Meta-Analysis". In:
- Vienneau, D., L. Perez, et al. (2015). "Years of Life Lost and Morbidity Cases Attributable to Transportation Noise and Air Pollution: A Comparative Health Risk Assessment for Switzerland in 2010". In: *International Journal of Hygiene and Environmental Health* 218.6, pp. 514–521. DOI: 10.1016/j.ijheh.2015.05.003.
- Visseren, F. L. J. et al. (2021). "2021 ESC Guidelines on Cardiovascular Disease Prevention in Clinical Practice". In: *European Heart Journal* 42.34, pp. 3227–3337. DOI: 10.1093/eurheartj/ehab484.

## Bibliography

---

- Wagner, K. J. P. et al. (2016). “Effects of Neighborhood Socioeconomic Status on Blood Pressure in Older Adults”. In: *Rev. Saúde Pública* 50.0. DOI: 10.1590/s1518-8787.2016050006595.
- Waller, L. A. (2021). “Spatial Clustering and Autocorrelation of Health Events”. In: *Handbook of Regional Science*. Ed. by M. M. Fischer and P. Nijkamp. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 2035–2051. DOI: 10.1007/978-3-662-60723-7\_80.
- Waller, L. A. and C. A. Gotway (2004). *Applied Spatial Statistics for Public Health Data*. Wiley Series in Probability and Statistics. Hoboken, N.J: John Wiley & Sons. 494 pp.
- Wang, F., M. Wen, and Y. Xu (2013). “Population-Adjusted Street Connectivity, Urbanicity and Risk of Obesity in the U.S.” In: *Applied Geography* 41, pp. 1–14. DOI: 10.1016/j.apgeog.2013.03.006.
- Watson, O. J. et al. (2022). “Global Impact of the First Year of COVID-19 Vaccination: A Mathematical Modelling Study”. In: *The Lancet Infectious Diseases* 22.9, pp. 1293–1302. DOI: 10.1016/S1473-3099(22)00320-6.
- Weber, D. (2020). *Égalité Des Chances Dans La Promotion de La Santé et La Prévention En Suisse: Définitions, Introduction Théorique, Recommandations Pratiques*. Récupéré sur Promotion Santé Suisse: <https://promotionsante.ch/recherche.html>.
- Westgard, J. O. et al. (1981). “A Multi-Rule Shewhart Chart for Quality Control in Clinical Chemistry”. In: *Clin Chem* 27.3, pp. 493–501.
- Whitehead, J. et al. (2022). “Structural Disadvantage for Priority Populations: The Spatial Inequity of COVID-19 Vaccination Services in Aotearoa”. In: *New Zealand Medical Journal*.
- Whitehead, M. (1992). “The Concepts and Principles of Equity and Health”. In: *International journal of health services* 22.3, pp. 429–445.
- WHO (2000). *Obesity: Preventing and Managing the Global Epidemic: Report of a WHO Consultation*. World Health Organization.
- (2011). *Burden of Disease from Environmental Noise: Quantification of Healthy Life Years Lost in Europe*. World Health Organization, Regional Office for Europe.
- (2019). *Ten Health Issues WHO Will Tackle This Year*. World Health Organization. URL: <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019> (visited on 08/28/2023).
- WHO Regional Office for Europe (2009). *Night Noise Guidelines for Europe*. World Health Organization.
- Wild, C. P. (2005). “Complementing the Genome with an “Exposome”: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology”. In: *Cancer Epidemiology, Biomarkers & Prevention* 14.8, pp. 1847–1850. DOI: 10.1158/1055-9965.EPI-05-0456.
- World Health Organization (2013). “Guidelines for Second Generation HIV Surveillance: An Update: Know Your Epidemic”. In:
- (2020). *WHO Director-General’s Opening Remarks at the Media Briefing on COVID-19 - 11 March 2020*. URL: <https://www.who.int/director-general/speeches/detail/who-director->

- general - s - opening - remarks - at - the - media - briefing - on - covid - 19 --- 11 - march - 2020 (visited on 06/29/2023).
- (2021a). *Weekly Epidemiological Update - 12 January 2021*. Weekly epidemiological update - 12 January 2021. URL: <https://www.who.int/publications/m/item/weekly-epidemiological-update--12-january-2021> (visited on 01/18/2021).
- (2021b). *WHO Coronavirus Disease (COVID-19) Dashboard*. URL: <https://covid19.who.int/>.
- Wu, F. et al. (2020). “A New Coronavirus Associated with Human Respiratory Disease in China”. In: *Nature* 579.7798, pp. 265–269.
- Yang, T.-C., C. Shoff, and A. J. Noah (2014). “Spatializing Health Research: What We Know and Where We Are Heading”. In: p. 12.
- Yi, B. et al. (2021). “Phylogenetic Analysis of SARS-CoV-2 Lineage Development across the First and Second Waves in Eastern Germany in 2020: Insights into the Cause of the Second Wave”. In: *Epidemiology & Infection* 149, e177.
- Zecha, L., F. Kohler, and V. Goebel (2017). *Niveaux Géographiques de La Suisse. Typologie Des Communes et Typologie Urbain-Rural 2012*. Technical report, Office fédéral de la statistique (OFS).
- Zeller, M. et al. (2021). “Emergence of an Early SARS-CoV-2 Epidemic in the United States”. In: *Cell* 184.19, pp. 4939–4952.
- Zellweger, U., M. Bopp, et al. (2014). “Prevalence of Chronic Medical Conditions in Switzerland: Exploring Estimates Validity by Comparing Complementary Data Sources”. In: *BMC Public Health* 14.1, p. 1157. DOI: 10.1186/1471-2458-14-1157.
- Zellweger, U., C. Junker, and M. Bopp (2019). “Cause of Death Coding in Switzerland: Evaluation Based on a Nationwide Individual Linkage of Mortality and Hospital in-Patient Records”. In: *Popul Health Metrics* 17.1, p. 2. DOI: 10.1186/s12963-019-0182-z.
- Zhang, X. et al. (2020). “Comparison of Spatiotemporal Characteristics of the COVID-19 and SARS Outbreaks in Mainland China”. In: *BMC Infectious Diseases* 20.1, p. 805. DOI: 10.1186/s12879-020-05537-y.
- Zhang, Y. et al. (2023). “Neighborhood Infrastructure-Related Risk Factors and Non-Communicable Diseases: A Systematic Meta-Review”. In: *Environ Health* 22.1, p. 2. DOI: 10.1186/s12940-022-00955-8.

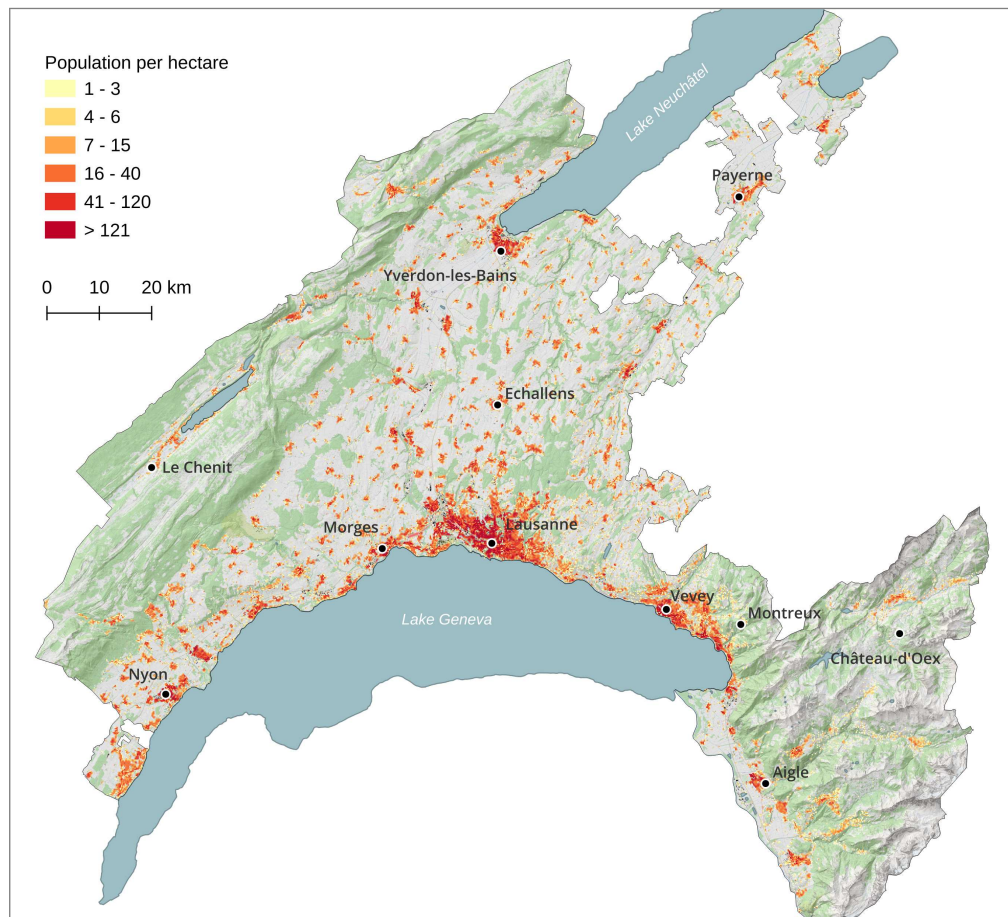




## APPENDIX

## A.1 Supplementary materials for Chapter 1

Figure A.1.1 – **Population distribution in the canton of Vaud, Switzerland** (Basemap layer: swissTLM-Map, swisstopo).



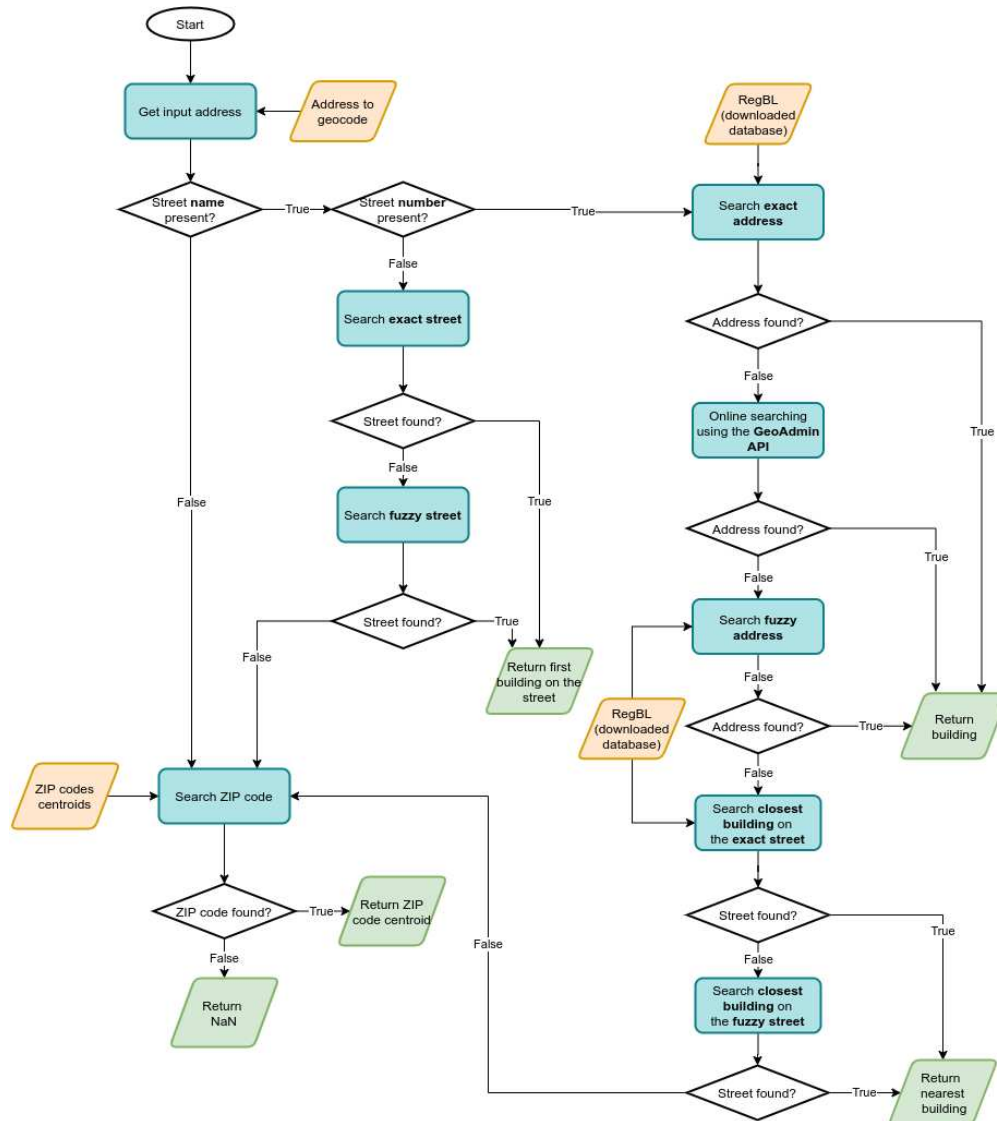


## A.1. Supplementary materials for Chapter 1

Table A.1.1 – **List of public health practitioners interviewed with their respective fields of expertise and institutional affiliations.** Interviews were conducted between August and September 2019 with professionals from the Département de la santé et de l'action sociale (DSAS) — the cantonal health authority — and Unisanté, a center responsible for regional coordination and the execution of health promotion and prevention strategies as delegated by public authorities.

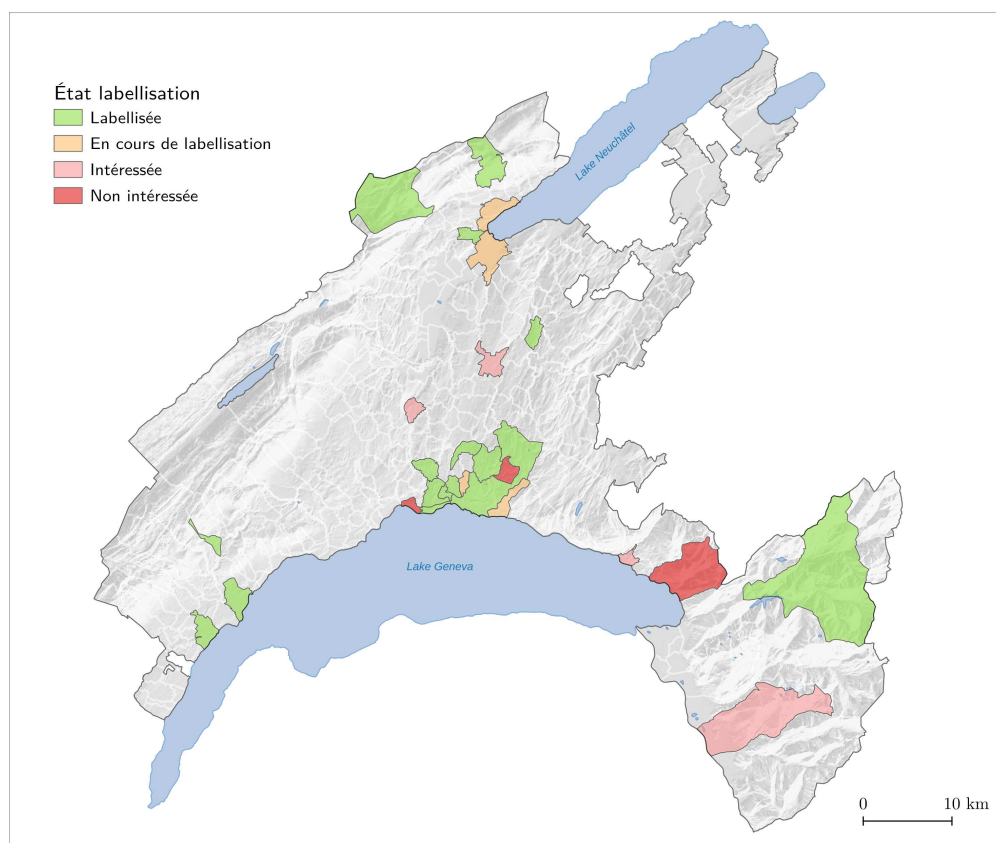
NAMES	FIELD OF EXPERTISE	AFFILIATION
Tania Larequi	Health promotion and prevention, cancers, physical activity and nutrition, public spaces	Direction de la santé communautaire DSC, Direction générale de la santé DGS, DSAS
Marta Pinto	Social determinants of health, social vulnerability	Secrétariat général, DSAS
Ahmed Berzig & Hugues Balthasar	Addictions, migrant populations, vulnerable population, sexually transmitted diseases	Office du médecin cantonal OMC, Direction générale de la santé DGS, DSAS
Marjorie Audard	Hospital planning, quality monitoring, specialised practice nurses	Centre qualité et systèmes CQS, Direction générale de la santé DGS, DSAS
Isabelle Rossi	Environmental health	Office du médecin cantonal OMC, Direction générale de la santé DGS, DSAS
Antonello Spagnolo	Social integration, social vulnerability	Direction de l'insertion et des solidarités DIRIS, DSAS
Thierry Blanc, Carina Hum, Fabio Peduzzi	Community health, emergency response, hospitals frequent users	Direction de la santé communautaire DSC, Direction générale de la santé DGS, DSAS
Déboray Ogay	Early childhood, children and youth, sexual health	Direction de la santé communautaire DSC, Direction générale de la santé DGS, DSAS
Olivier Linder	Hospital planning, health poles	Direction hôpitaux et préhospitalier DHPH, Direction générale de la santé DGS, DSAS
Myriam Pasche & Karine Zurcher	Health prevention and promotion	Département promotion de la santé et préventions DPSP, Unisanté

Figure A.1.2 – **Flowchart of the geocoding algorithm developed to retrieve coordinates of individuals' home addresses.** The algorithm is represented through a series of procedural steps (rectangles) and decision nodes (diamonds), employing input data (orange parallelograms) to achieve the desired outputs (green parallelograms). Initiated with a given address, the algorithm seeks the corresponding coordinates using the official Swiss address directory (RegBL). Address matching is carried out either through the pre-downloaded database or via requests to the GeoAdmin API. A fuzzy-matching procedure is employed at several steps to accommodate potential discrepancies or slight errors in address spellings. This method identifies strings of close similarity based on a set tolerance level, typically 0.8, ensuring matches below this threshold are discarded to prevent inaccuracies. The final coordinates' precision spans from the exact building location to the nearest building, down to street coordinates or the ZIP code centroid. If the location remains unidentified, NaN is returned. These outputs (shown in green) mark the algorithm's end points. The code underlying this algorithm is accessible in the `geocoding_utils.py` file within the GitHub repository: <https://github.com/aladoy/GIRAPH-functions>.



## A.2 Supplementary materials for Chapter 2

Figure A.2.1 – Status of the "Commune en santé" label in the Canton of Vaud as of October 5, 2021. Municipalities already certified are depicted in green, those currently undergoing the process are shown in yellow. Municipalities that have expressed interest are represented in pink, while those that have expressed no interest are shown in red.



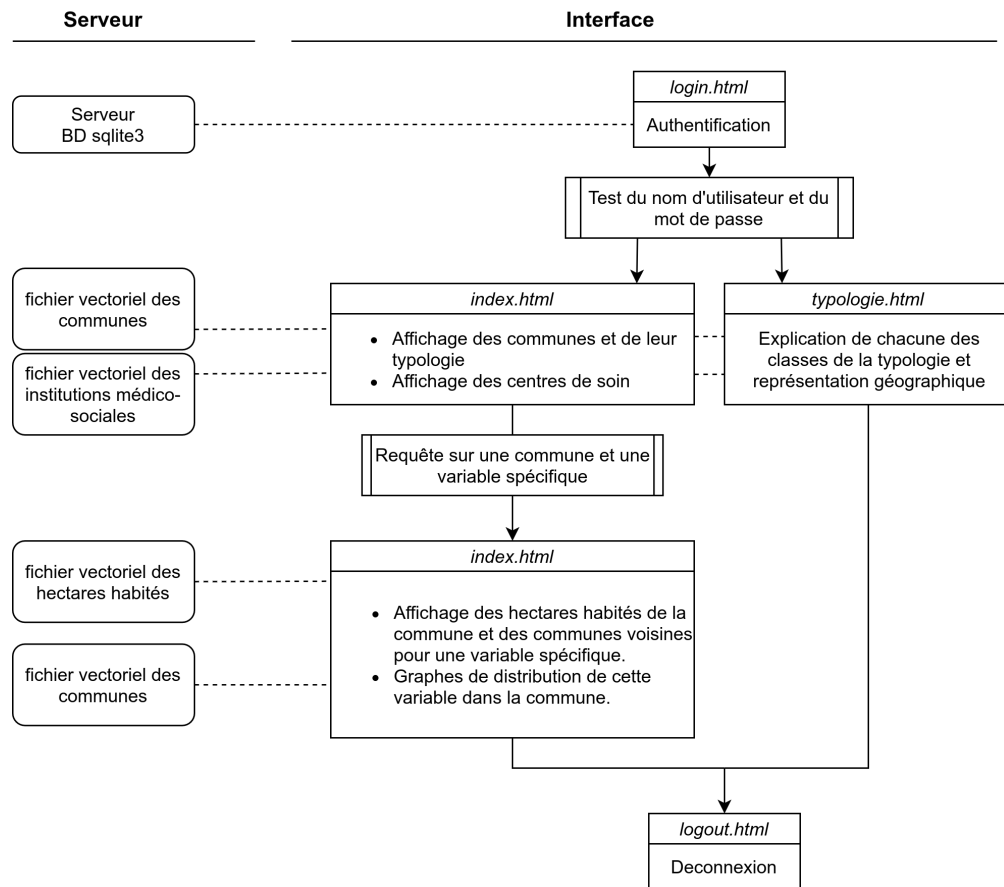


Figure A.2.2 – Architecture of the GEOSAN web application.

## A.2. Supplementary materials for Chapter 2

Table A.2.1 – Description of the 39 variables used in the typology of municipalities.

Domain	Variable	Abbrev.	Time period	Source
DEMOGRAPHY				
Age structure	Total population	PTOT	2021	Microgis
	0 – 15 years	RP0014	2021	Microgis
	15 – 24 years	RP1524	2021	Microgis
	25 – 64 years	RP2564	2021	Microgis
	65 – 79 years	RP6579	2021	Microgis
Migration	80 years and over	RP80M	2021	Microgis
	Swiss born	RPBCH	2019	OFS
	Foreigners from EU / EFTA	RPBEUR	2019	OFS
	Foreigners from outside EU / EFTA	RPBOUT	2019	OFS
	Arrivals (excluding births)	MEDNI	2015-2019	OFS
	Allophones	RPA	2021	Microgis
SOCIO-ECONOMY				
Income	Annual household median income	MEDINCM	2021	Microgis
	Income inequality (interquartile range)	IQRINCM	2021	Microgis
	Municipal finances – accounts	MEDACC	2012-2014	Statistique Vaud
	Municipal finances – taxes	MEDTPVI	2015-2019	Statistique Vaud
Education level	Non-formal education	RPF0	2021	Microgis
	Primary or lower secondary school	RPF1	2021	Microgis
	Higher secondary school	RPF2	2021	Microgis
	Tertiary education	RPF3	2021	Microgis
Employment	Actives in the primary sector	RPAPRIM	2021	Microgis
	Actives in the secondary sector	RPASEC	2021	Microgis
	Actives in the tertiary sector	RPATER	2021	Microgis
	Commuters	RAPC	2018	Mobility Atlas, Vaud
Political environment	Left-wing (Social Democratic Party, Green Party) voters	PS_VERTS	2019	OFS
	Centrist (Christian Democratic Party) voters	PDC	2019	OFS
	Liberal (Radical-Liberal Party, Liberal Green Party) voters	PLR_PVL	2019	OFS
	Right-wing (Swiss People's Party) voters	UDC	2019	OFS
TERRITORY				
Housing	Owner-occupancy	RDOO	2017-2019	OFS
	Population density in residential areas	DENS	2021	DGTL, Microgis
Land-use	Agriculture	RSA	2021	DGTL
	Industrial and commercial	RSI	2021	DGTL
	Institutional	RSDP	2021	DGTL
	Leisure	RSH	2021	DGTL
	Green space	RSN	2021	DGTL
	Residential	RSR	2021	DGTL
	Parks	RSP	2021	DGTL
Geography	Surface area	SURF	2021	swisstopo
	Altitude	MEDALT	2021	OFS
	Travel time by public transport to the nearest urban area	PTTAGGL	2017	ARE

## Appendix . Appendix

Table A.2.2 – Description of the Social Determinants of Health (SDOH) indicators used to characterize inhabited hectares. \*: indicators not available in the web application, but available in the QGIS project.

Domain	Variable	Abbrev.	Time period	Source
ENV. EXPOSURES				
Air	Exposure to nitrogen dioxide (NO2) [ug/m3]	NO2	2020	OFEV
	Exposure to fine particulate matter PM10 [ug/m3]	PM10	2020	OFEV
	Exposure to fine particulate matter PM2.5 [ug/m3]	PM25	2020	OFEV
Noise	Nighttime road and railway noise [dB]	NOISE	2015	OFEV
Heat	Land Surface temperature - average of three summer days [°C]	LST	2019-2021	USGS
ACCESSIBILITY				
Healthcare	Emergency facilities - access time [mn]	ER_TIME	2023	DGS
	Dental clinics - access time [mn]	DENTAL_TIME	2023	DGS
	Pharmacies - access time [mn]	PHARMA_TIME	2023	DGS
	Ambulances - access time [mn]	AMBU_TIME	2023	DGS
	Medical offices and outpatient centers - distance via road network [km]	D_MEDIC	2018	OFS
Education	Compulsory education - distance via road network [km]	D_SCHOOL_O	2018	OFS
	Secondary level II - distance via road network [km]	D_SCHOOL_S	2018	OFS
	Tertiary level - distance via road network [km]	D_EDUC	2018	OFS
Safety	Police services - distance via road network [km]	D_SECURITY	2018	OFS
	Road accidents involving pedestrians or cyclists within an 800m radius [nb btw. 2011-2021]	NB_ACDNT	2011-2021	OFROU
Food	Grocery stores - distance via road network [km]	D_GROCERY	2018	OFS
Leisure	Playgrounds - access time [mn]	PLAYGRD_TIME	2022	OSM
	Forests - distance via road network [km]	D_FOREST	2018	OFS
	Lakes - distance via road network [km]	D_LAKE	2018	OFS
	Sports and fitness facilities - distance via road network [km]	D_SPORT	2018	OFS
	Swimming pools and beaches - distance via road network [km]	D_SWIM	2018	OFS
Transportation	Public transport stop - distance via road network [km]	D_SPOT_TOT	2018	OFS
	Quality of public transport service - A: very good to E: very poor/non-existent [qualitative]*	QPTS	2022	ARE
Nature	Green spaces within an 800m radius [%]	GREEN_SP	2022	DGTL
	Blue spaces within an 800m radius [%]	BLUE_SP	2022	DGTL
DEMOGRAPHY				
Total Population	Total resident population [#]*	PTOT	2019	OFS
	Population aged 0-14 [#]*	P0014	2019	OFS
	Population aged 15-24 [#]*	P1524	2019	OFS
	Population aged 25-64 [#]*	P2564	2019	OFS
	Population aged 65 and older [#]*	P65M	2019	OFS
	Population aged 80 and older [#]*	P80M	2019	OFS
	Percentage of population aged 0-14 [%]	RP0014	2019	OFS
	Percentage of population aged 15-24 [%]	RP1524	2019	OFS
	Percentage of population aged 25-64 [%]	RP2564	2019	OFS
	Percentage of population aged 65 and older [%]	RP65M	2019	OFS
	Percentage of population aged 80 and older [%]	RP80M	2019	OFS
Women	Females aged 0-14 [#]*	P0014_F	2021	OFS
	Females aged 15-24 [#]*	P1524_F	2021	OFS
	Females aged 25-64 [#]*	P2564_F	2021	OFS
	Females aged 65 and older [#]*	P65M_F	2021	OFS
	Females aged 80 and older [#]*	P80M_F	2021	OFS

## A.2. Supplementary materials for Chapter 2

Table A.2.3 – Description of SDOH indicators used to characterize inhabited hectares (continued)

Domain	Variable	Abbrev.	Time period	Source
Men	Percentage of females aged 0-14 [%]	RP0014_F	2021	OFS
	Percentage of females aged 15-24 [%]	RP1524_F	2021	OFS
	Percentage of females aged 25-64 [%]	RP2564_F	2021	OFS
	Percentage of females aged 65 and older [%]	RP65M_F	2021	OFS
	Percentage of females aged 80 and older [%]	RP80M_F	2021	OFS
	Males aged 0-14 [#]*	P0014_M	2021	OFS
	Males aged 15-24 [#]*	P1524_M	2021	OFS
	Males aged 25-64 [#]*	P2564_M	2021	OFS
	Males aged 65 and older [#]*	P65M_M	2021	OFS
	Males aged 80 and older [#]*	P80M_M	2021	OFS
	Percentage of males aged 0-14 [%]	RP0014_M	2021	OFS
	Percentage of males aged 15-24 [%]	RP1524_M	2021	OFS
	Percentage of males aged 25-64 [%]	RP2564_M	2021	OFS
	Percentage of males aged 65 and older [%]	RP65M_M	2021	OFS
	Percentage of males aged 80 and older [%]	RP80M_M	2021	OFS
SOCIO-ECO				
	Percentage of population aged 15 and older living alone [%]	R_PLA	2019	microgis
	Average number of persons per household [#]	AVG_PPH	2019	microgis
	Median annual income per household [kCHF]	MEDREV	2019	microgis
	Percentage of population aged 15 and older divorced or widowed [%]	R_DIV_WID	2019	microgis
	Percentage of foreign population (foreign-born) [%]	R_FFB	2019	microgis
	Percentage of non-French speaking population [%]	R_NN_FRA	2019	microgis
	Percentage of population aged 15 and older with compulsory education [%]	R_NN_POBL	2021	microgis
	Percentage of population with disabilities [%]	R_DIS	2019	microgis
	Unemployment rate for individuals aged 15 and older [%]	R_UNEMP	2019	microgis
INDICES				
	Index of cumulative environmental burdens	ENV_INDEX		
	Index of spatial accessibility to primary care services	HEALTHCARE_INDEX		
	Index of material deprivation	SOC_ECO_INDEX		

## Appendix . Appendix

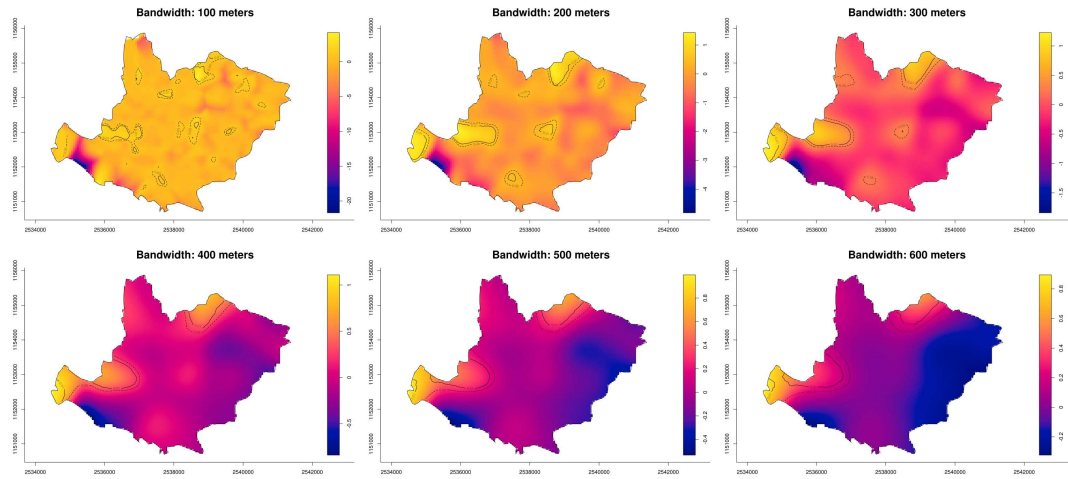
Figure A.2.3 – Program of the workshop with public health practioners held on January 16, 2023 (in French).

9h30	Mots de bienvenue (Myriam Pasche, Oriana Villa)
9h45	Présentation du projet et du Système d'Information Géographique GEOSAN (Anaïs Ladoy, Stéphane Joost)  <ul style="list-style-type: none"> <li>- Vers une santé publique de précision</li> <li>- Présentation et description de la base de données GEOSAN</li> <li>- Pour une meilleure compréhension du territoire et pour favoriser l'adoption du label: une typologie des communes vaudoises</li> <li>- Comment accompagner le développement de nouvelles actions de promotion de la santé : cartographie d'une sélection de déterminants de la santé</li> </ul>
10h30	Pause café
11h00	Présentation des données et des indicateurs à travers l'utilisation de deux outils dans le cadre de scénarios prédéfinis: <ul style="list-style-type: none"> <li>- l'application de cartographie sur Internet (WebGIS)</li> <li>- le Système d'Information Géographique GEOSAN (logiciel QGIS)</li> </ul>
11h45	Questions et réponses - discussion
12h15	Pause - Repas
13h30	Formation aux Systèmes d'Information Géographique (SIG). (Anaïs Ladoy, Stéphane Joost)  <p>La formation se déroulera sur le logiciel libre QGIS (version 3.22 LTR), qui peut être installé gratuitement sur tous les systèmes d'exploitation. Le cours sera centré sur la manipulation du projet SIG présenté dans la matinée et inclura les aspects suivants :</p> <ul style="list-style-type: none"> <li>- Accès et manipulation des différents indicateurs</li> <li>- Sélections et requêtes</li> <li>- Importation de vos propres données, exportation de tables de données</li> <li>- Création d'une carte</li> </ul>
16h30	Fin de la journée



### A.3 Supplementary materials for Chapter 3

Figure A.3.1 – **Spatial variation of obesity risk using a Gaussian Kernel with incremental bandwidths.** Tolerance contours indicate high-risk areas. The disease exhibited a global clustering for all bandwidths values (p-values are 100m:0.042, 200m:0.030, 300m:0.022, 400m:0.025, 500m:0.02, 600m:0.012). Note that the legend scale for the log relative risk of obesity is different for each bandwidth, which contributes to a shift in the color representation.



## Appendix . Appendix

Table A.3.1 – Comparative statistics of baseline, study sample (follow-up 2, Lausanne urban districts), and excluded population (follow-up 2, outside Lausanne urban districts).

	Baseline population	Study sample	Excluded population
<b>N</b>	6733	3695	1186
<b>Age, years</b>	52.6 (10.7)	64.1 (10.5)	59.2 (9.4)
<b>Sex</b>			
Female	3544 (52.6%)	2097 (56.8%)	592 (49.9%)
Male	3189 (47.4%)	1598 (43.2%)	594 (50.1%)
<b>Born in Switzerland</b>	4031 (59.9%)	2296 (62.1%)	766 (64.6%)
Missing	0	0	1 (0.1%)
<b>Living in couple / cohabiting</b>	4512 (67.0%)	2129 (57.6%)	623 (52.5%)
Missing	0	0	215 (18.1%)
<b>Education</b>			
Low	3774 (56.0%)	2011 (54.4%)	577 (48.6%)
Medium	1625 (24.1%)	919 (24.9%)	339 (28.6%)
High	1320 (19.6%)	765 (20.7%)	266 (22.4%)
Missing	0	0	4 (0.34%)
<b>Professional activity</b>	4771 (70.9%)	1849 (50.0%)	728 (61.4%)
Missing	0	0	77 (6.5%)
<b>Financial difficulties</b>		990 (26.8%)	324 (27.3%)
<b>Current smoking</b>	1818 (27.0%)	913 (24.7%)	328 (27.7%)
<b>Hazardous drinking</b>	1165 (17.3%)	931 (25.2%)	287 (24.2%)
<b>Sedentary lifestyle</b>		1258 (34.0%)	432 (36.4%)
Missing		1353 (36.6%)	371 (31.3%)

### A.3. Supplementary materials for Chapter 3

Figure A.3.2 – Flow charts for the sample selection of the baseline (2003-2006), follow-up 1 (2009-2012), and follow-up 2 (2014-2017) participants.

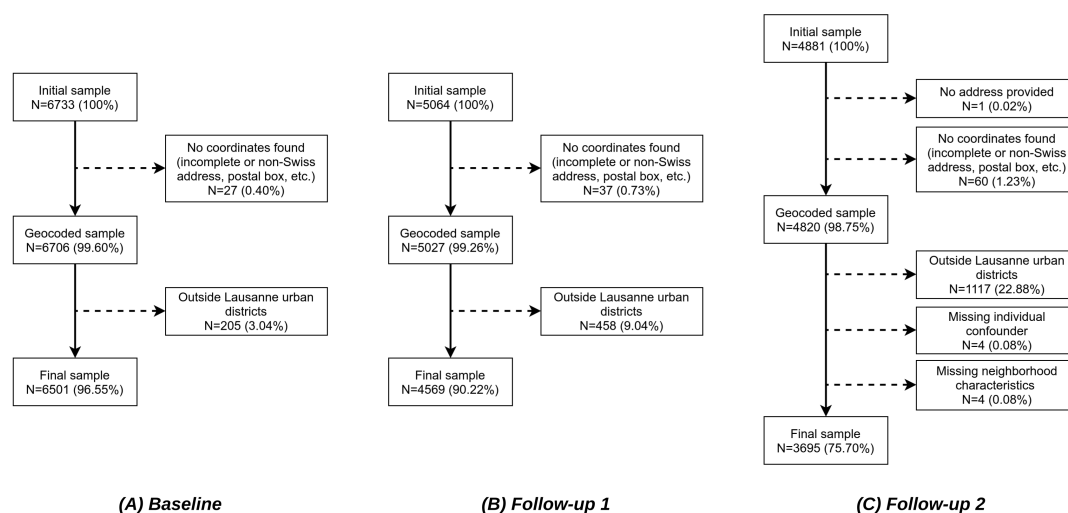
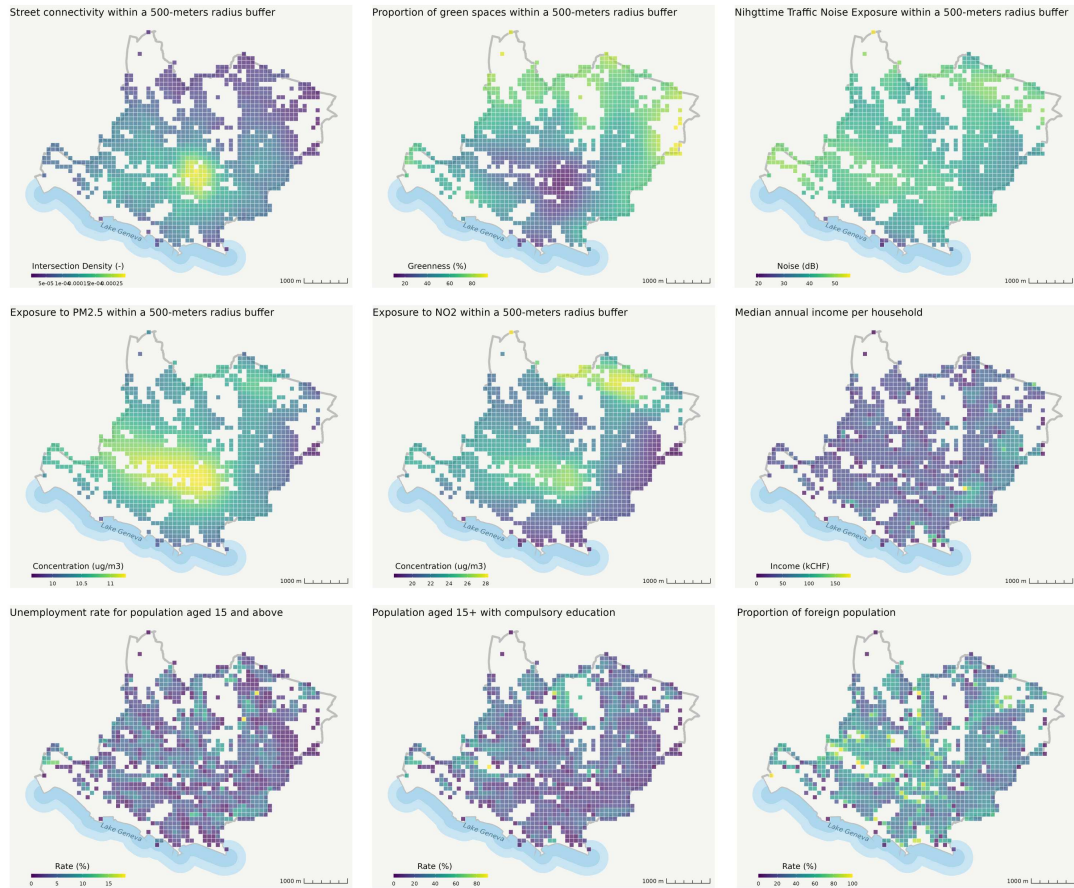


Table A.3.2 – Frequency distribution of outcome variables for the baseline, follow-up, and follow-up 2 samples.

	Baseline sample (N=6501)	Follow-up 1 sample (N=4569)	Follow-up 2 sample (N=3695)
<b>Hypertension</b>			
Yes	2427 (37.33%)	1934 (42.33%)	1780 (48.17%)
No	4067 (62.56%)	2627 (57.5%)	1789 (48.42%)
Missing	7 (0.11%)	8 (0.18%)	126 (3.41%)
<b>Obesity</b>			
Yes	1003 (15.43%)	787 (17.22%)	654 (17.70%)
No	5495 (84.53%)	3724 (81.51%)	2757 (74.61%)
Missing	3 (0.05%)	58 (1.27%)	284 (7.69%)
<b>Diabetes</b>			
Yes	423 (6.51%)	505 (11.05%)	396 (10.72%)
No	6062 (93.25%)	4046 (88.55%)	3064 (82.92%)
Missing	16 (0.25%)	18 (0.39%)	235 (6.36%)
<b>Dyslipidemia</b>			
Yes	2058 (31.66%)	1946 (42.59%)	1225 (33.15%)
No	4394 (67.59%)	2596 (56.82%)	2180 (59.00%)
Missing	49 (0.75%)	27 (0.59%)	290 (7.85%)

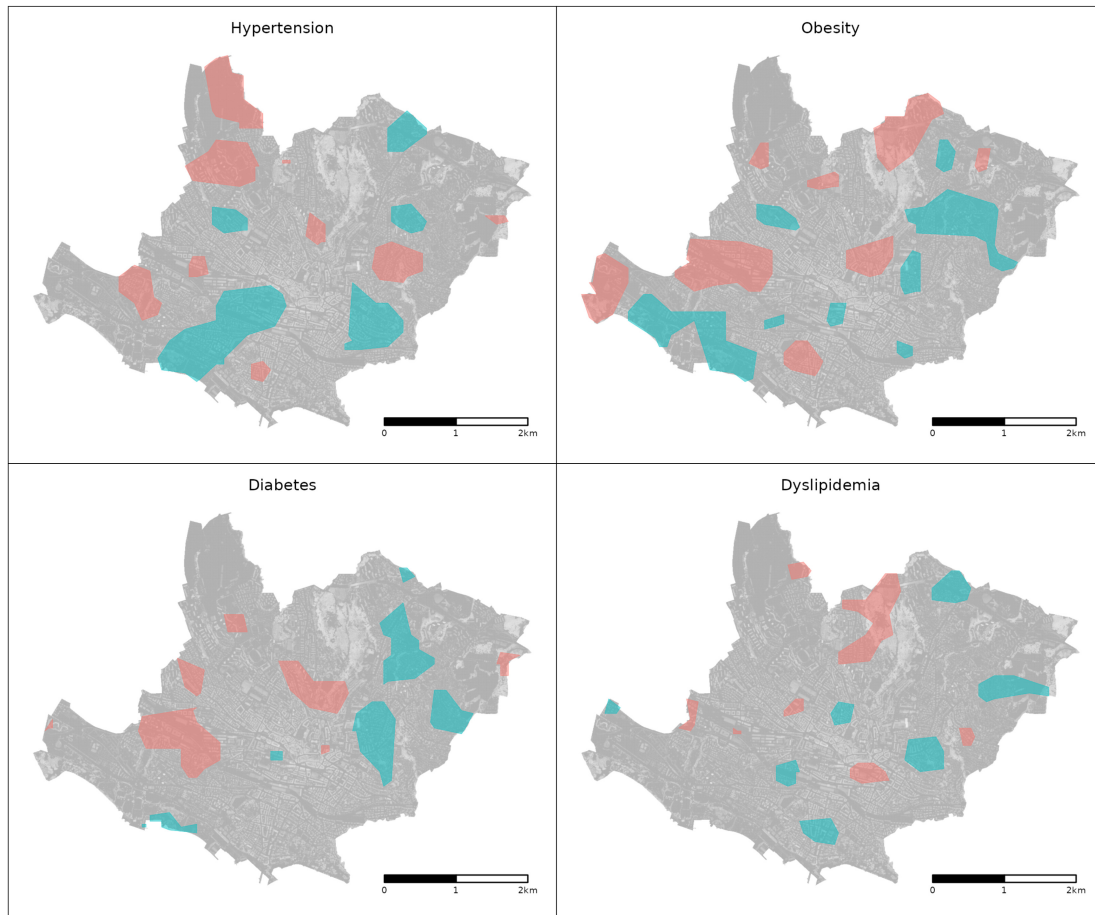
## Appendix . Appendix

Figure A.3.3 – Spatial distribution of contextual factors across the study area (Lausanne, Switzerland).



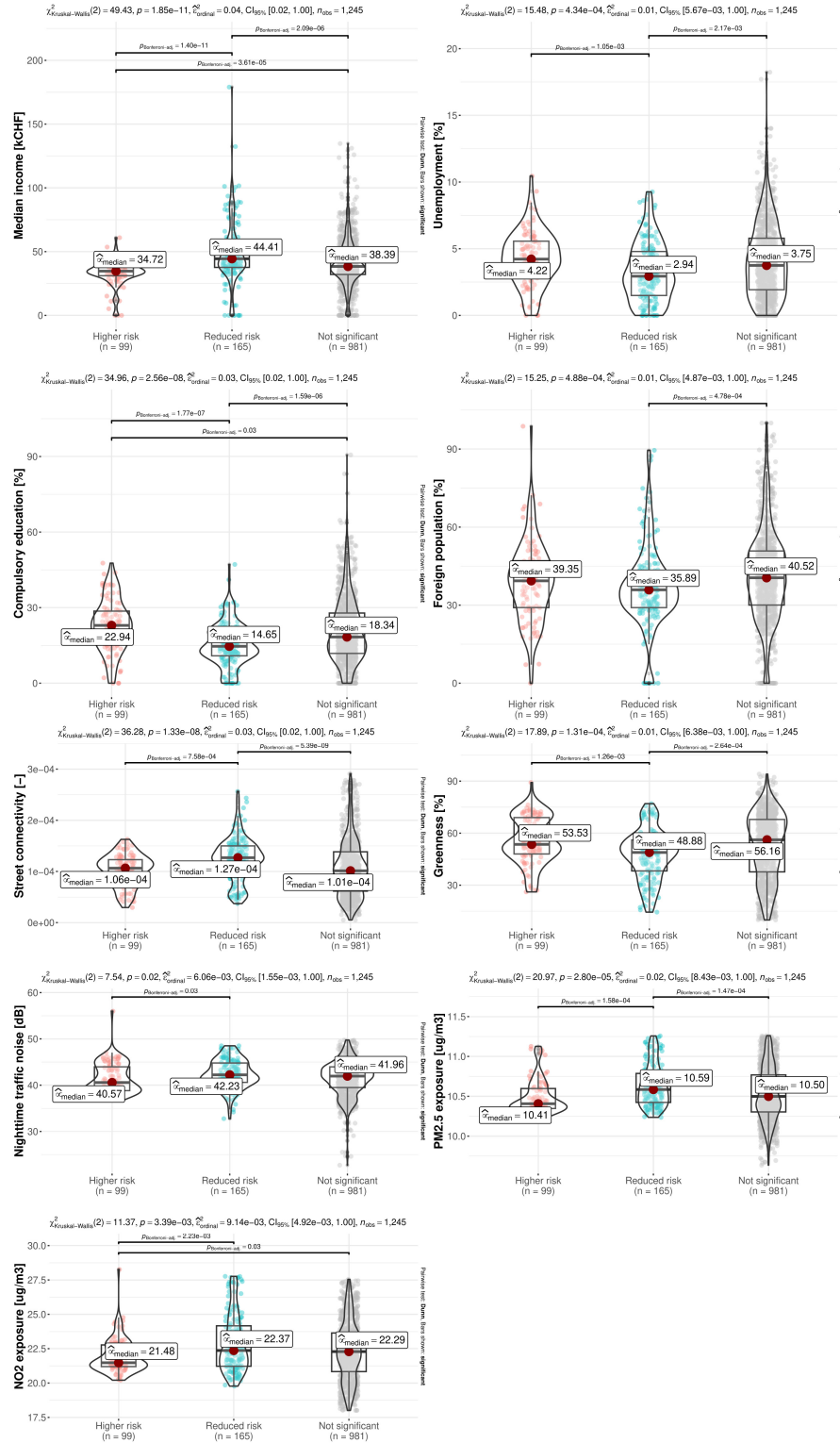
### A.3. Supplementary materials for Chapter 3

Figure A.3.4 – **Delimitation of areas presenting a significant high-risk (in red) and low-risk (in blue) for hypertension, obesity, diabetes, and dyslipidemia.** The remaining study area is represented in grey, indicating areas with a non-significant risk. The boundaries of these areas were determined through 999 Monte-Carlo simulations on the log relative risk surface for each disease, utilizing a Gaussian kernel with a bandwidth of 200 meters. The identified areas were subsequently used to generate the boxplots presented in Figures S5 to S7.



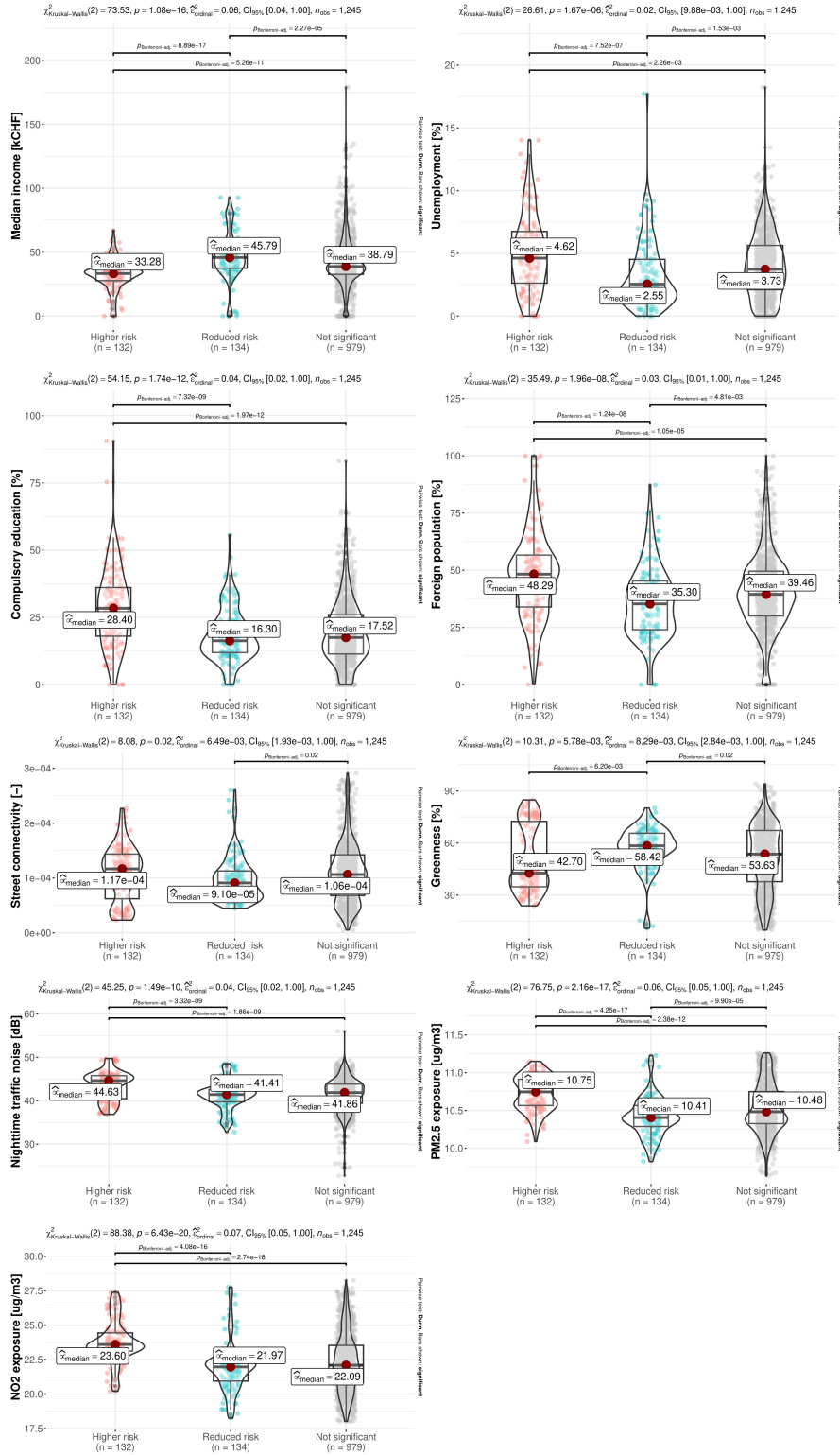
## Appendix . Appendix

Figure A.3.5 – Comparisons of contextual factors among areas categorized as high-risk, low-risk and neutral risk (i.e., not significant) for hypertension. Plots were generated using the ggstatsplot package.



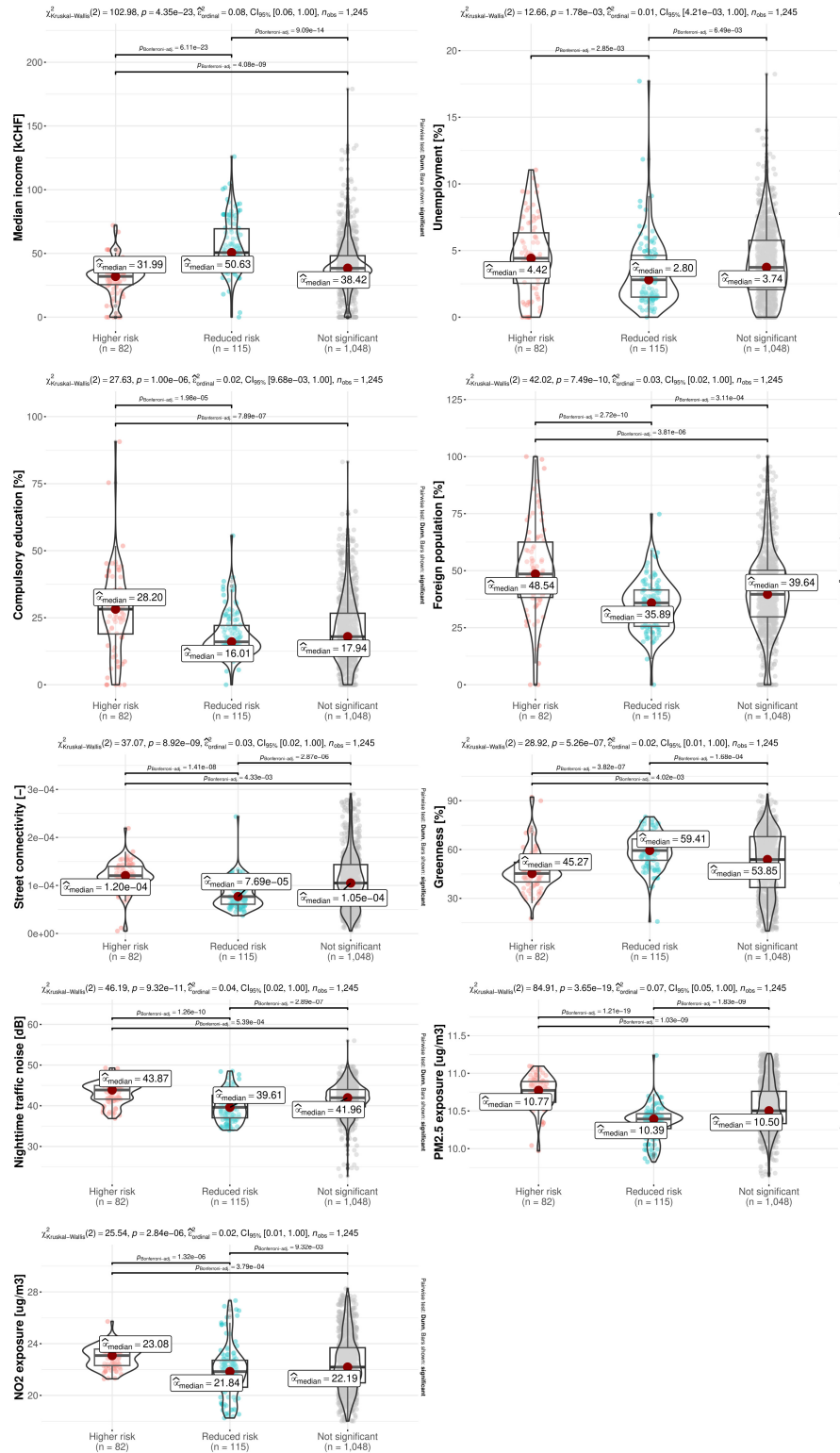
### A.3. Supplementary materials for Chapter 3

Figure A.3.6 – Comparisons of contextual factors among areas categorized as high-risk, low-risk and neutral risk (i.e., not significant) for obesity. Plots were generated using the ggstatplot package.



## Appendix . Appendix

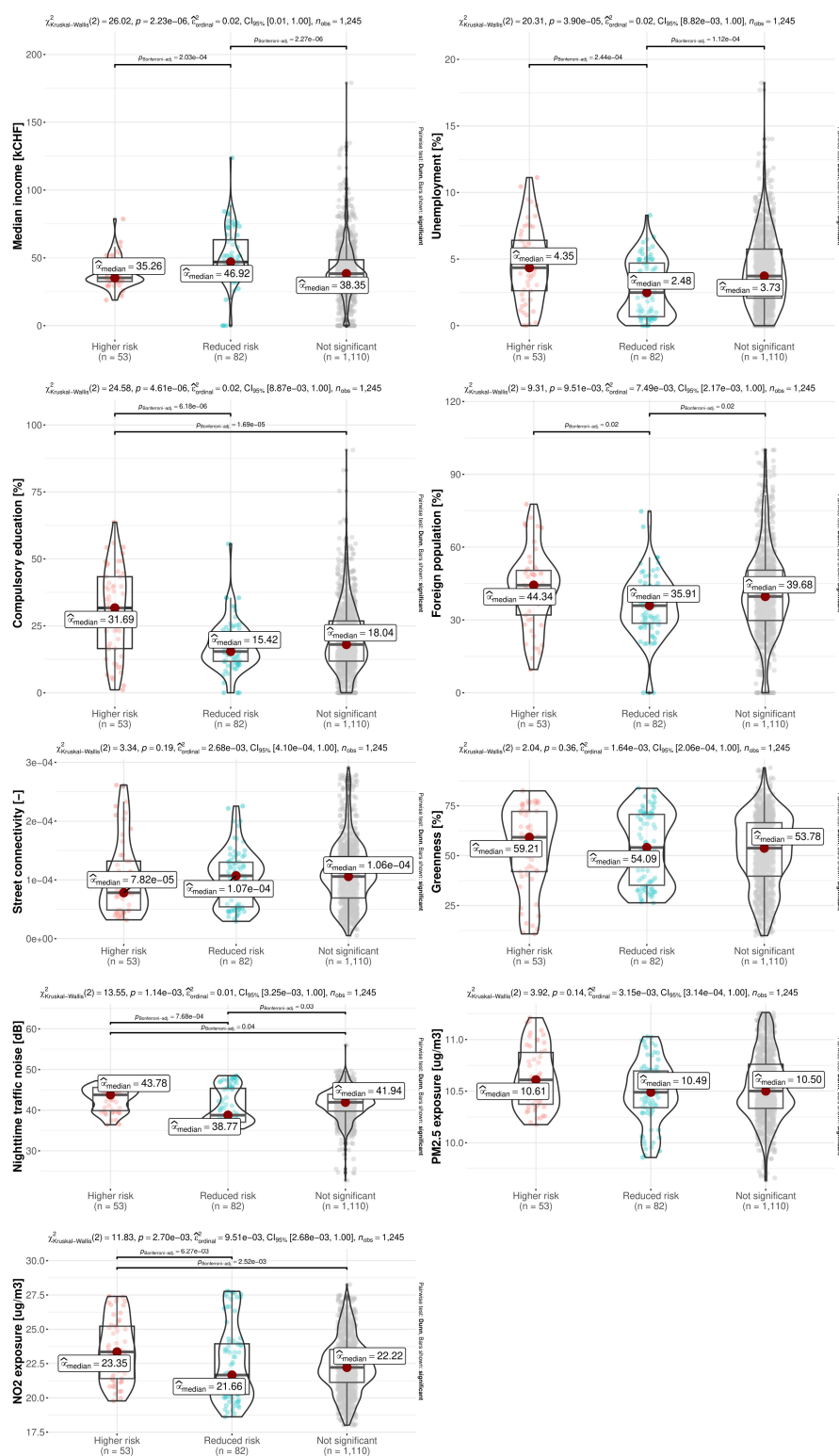
Figure A.3.7 – Comparisons of contextual factors among areas categorized as high-risk, low-risk and neutral risk (i.e., not significant) for diabetes. Plots were generated using the ggstatplot package.





## A.3. Supplementary materials for Chapter 3

Figure A.3.8 – Comparisons of contextual factors among areas categorized as high-risk, low-risk and neutral risk (i.e., not significant) for dyslipidemia. Plots were generated using the ggstatsplot package.



## Appendix . Appendix

Figure A.3.9 – **Spatial variation of hypertension risk during (A) baseline (2003-2006), (B) follow-up 1 (2009-2012), and (C) follow-up 2 (2014-2017) periods.** Prevalence of hypertension was 37.37%, 42.40%, and 49.87%, respectively. High-risk areas are delineated at the 5% and 1% significance levels. Basemap layer: DEM, Géodonnées État de Vaud (2012).

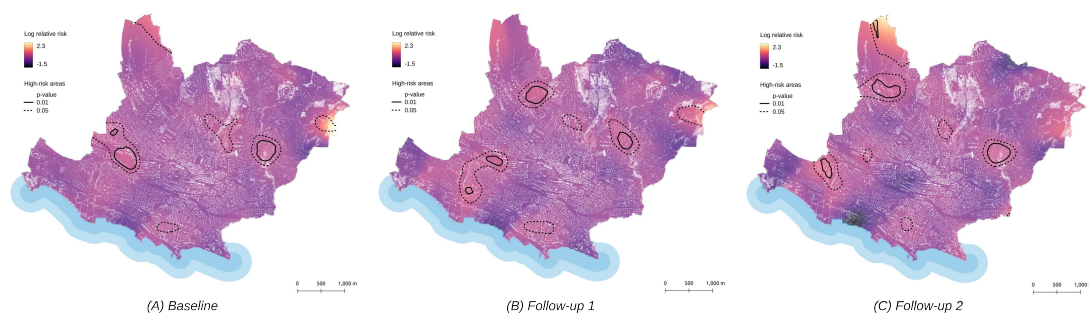
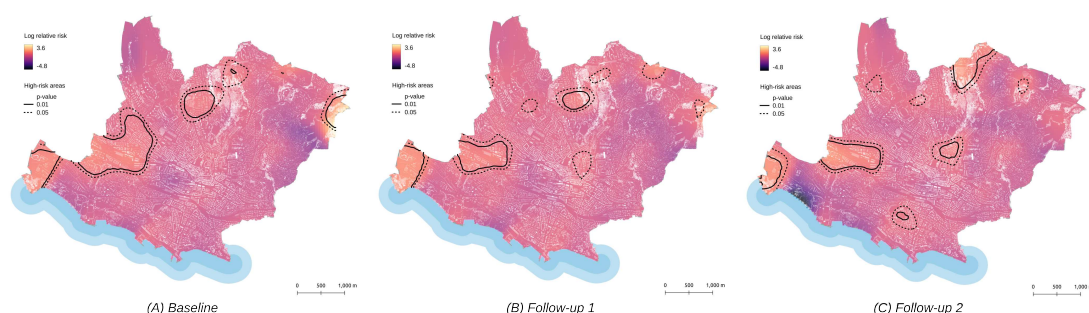


Figure A.3.10 – **Spatial variation of obesity risk during (A) baseline (2003-2006), (B) follow-up 1 (2009-2012), and (C) follow-up 2 (2014-2017) periods.** Prevalence of obesity was 15.44%, 17.45%, and 19.17%, respectively. High-risk areas are delineated at the 5% and 1% significance levels. Basemap layer: DEM, Géodonnées État de Vaud (2012).



### A.3. Supplementary materials for Chapter 3

Figure A.3.11 – **Spatial variation of diabetes risk during (A) baseline (2003-2006), (B) follow-up 1 (2009-2012), and (C) follow-up 2 (2014-2017) periods.** Prevalence of dyslipidemia was 31.9%, 42.84%, and 35.98%, respectively. High-risk areas are delineated at the 5% and 1% significance levels. Basemap layer: DEM, Géodonnées État de Vaud (2012).

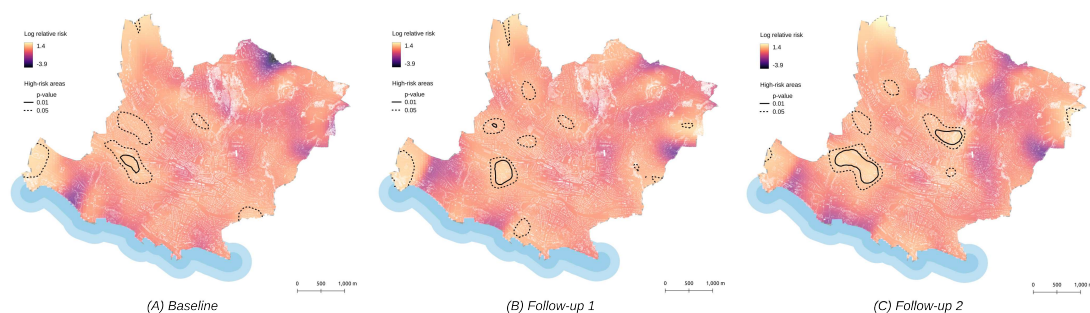
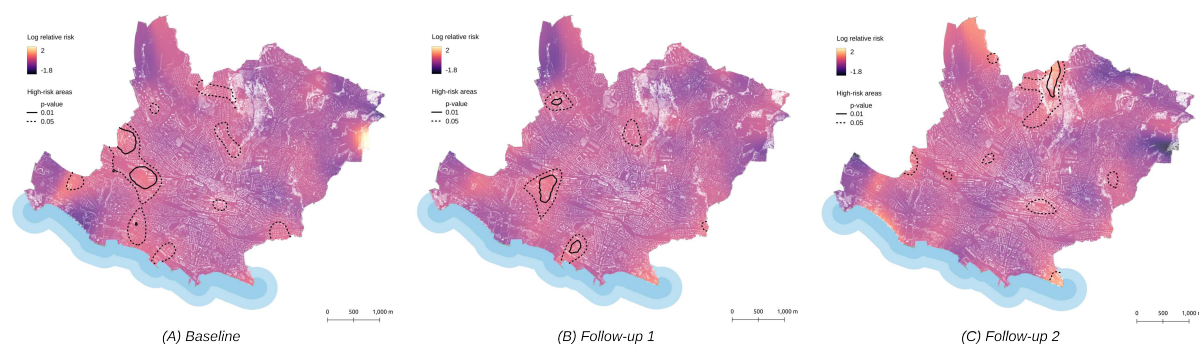


Figure A.3.12 – **Spatial variation of dyslipidemia risk during (A) baseline (2003-2006), (B) follow-up 1 (2009-2012), and (C) follow-up 2 (2014-2017) periods.** Prevalence of obesity was 6.52%, 11.10%, and 11.4%, respectively. High-risk areas are delineated at the 5% and 1% significance levels. Basemap layer: DEM, Géodonnées État de Vaud (2012).



## Appendix . Appendix

**Table A.3.3 – Results of the ordinary least squares (OLS) regression model.** Health outcomes were previously adjusted for individual confounders, including age, sex, education, income, and country of birth, using a logistic regression model (see Table A.3.3 for details). n/a: not available, I: Global Moran's I of residuals, p: p-value, Coeff.: Coefficient estimates, VIF: Variance Inflation Factor.

	Hypertension			Obesity			Diabetes			Dyslipidemia		
	N=3569			N=3411			N=3460			N=3405		
	I=0.033 (p=0)			I=0.041 (p=0)			I=0.031 (p=0)			I=0.026 (p=0)		
Variable	Coeff.	p	VIF	Coeff.	p	VIF	Coeff.	p	VIF	Coeff.	p	VIF
Intercept	0.000	0.984	n/a	0.000	1.000	n/a	-0.002	0.926	n/a	-0.001	0.945	n/a
Street connectivity	0.042	0.276	5.34	-0.023	0.567	5.29	-0.031	0.426	5.310	-0.011	0.779	5.26
Greenness	0.001	0.984	4.52	-0.018	0.612	4.48	-0.042	0.239	4.490	-0.023	0.527	4.44
Nighttime noise	0.005	0.848	2.72	-0.013	0.653	2.73	0.024	0.390	2.720	0.016	0.569	2.72
PM2.5	-0.076	0.148	9.71	0.012	0.820	9.76	0.009	0.868	9.710	0.008	0.880	9.60
NO2	-0.018	0.622	4.86	0.000	0.995	4.89	-0.053	0.149	4.840	-0.027	0.475	4.82
Median income	-0.061	0.002	1.35	-0.046	0.019	1.35	-0.044	0.023	1.340	-0.008	0.694	1.34
Unemployment rate	0.025	0.208	1.41	0.031	0.120	1.40	0.018	0.375	1.400	-0.003	0.899	1.40
Compulsory education	0.010	0.648	1.61	0.011	0.616	1.60	0.014	0.517	1.610	-0.003	-0.142	1.61
Foreign population	0.004	0.846	1.41	0.006	0.763	1.41	0.037	0.062	1.410	0.012	0.578	1.40

**Table A.3.4 – Model fit metrics for ordinary least squares (OLS), geographically weighted regression (GWR), and multiscale geographically weighted regression (MGWR) models.** Health outcomes were adjusted beforehand for individual confounders, including age, sex, education, income, and nationality, using a logistic regression model (see Table 3.1).

	Hypertension			Obesity			Diabetes			Dyslipidemia		
	OLS	GWR	MGWR	OLS	GWR	MGWR	OLS	GWR	MGWR	OLS	GWR	MGWR
R <sup>2</sup>	0.006	0.010	0.013	0.005	0.025	0.025	0.007	0.010	0.017	0.001	0.004	0.003
Adjusted R <sup>2</sup>	0.004	0.005	0.008	0.002	0.016	0.017	0.005	0.006	0.009	-0.002	-0.001	-0.001
AIC	10139	10141	10135	9676	9654	9648	9742	9745	9742	9665	9670	9671
AICc		10141	10135		9654	9649		9745	9743		9671	9671

### A.3. Supplementary materials for Chapter 3

Figure A.3.13 – **GWR results for adjusted hypertension.** The figure shows coefficient estimates (left) and spatial associations (right) for physical and social environment characteristics associated with risk of hypertension, adjusted for individual confounders. Indicative landmarks are shown on the maps to help interpret the results (#1- #8). Only variables that showed statistically significant variation (based on corrected critical t-values) are shown.

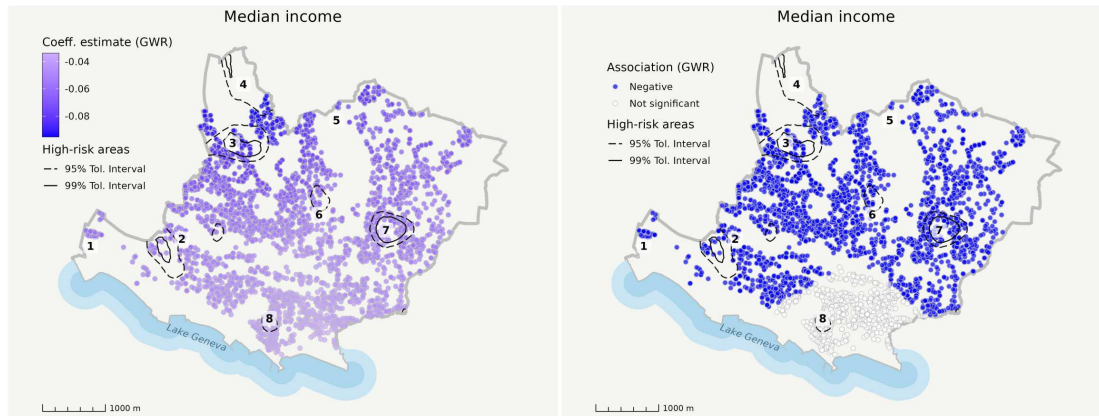
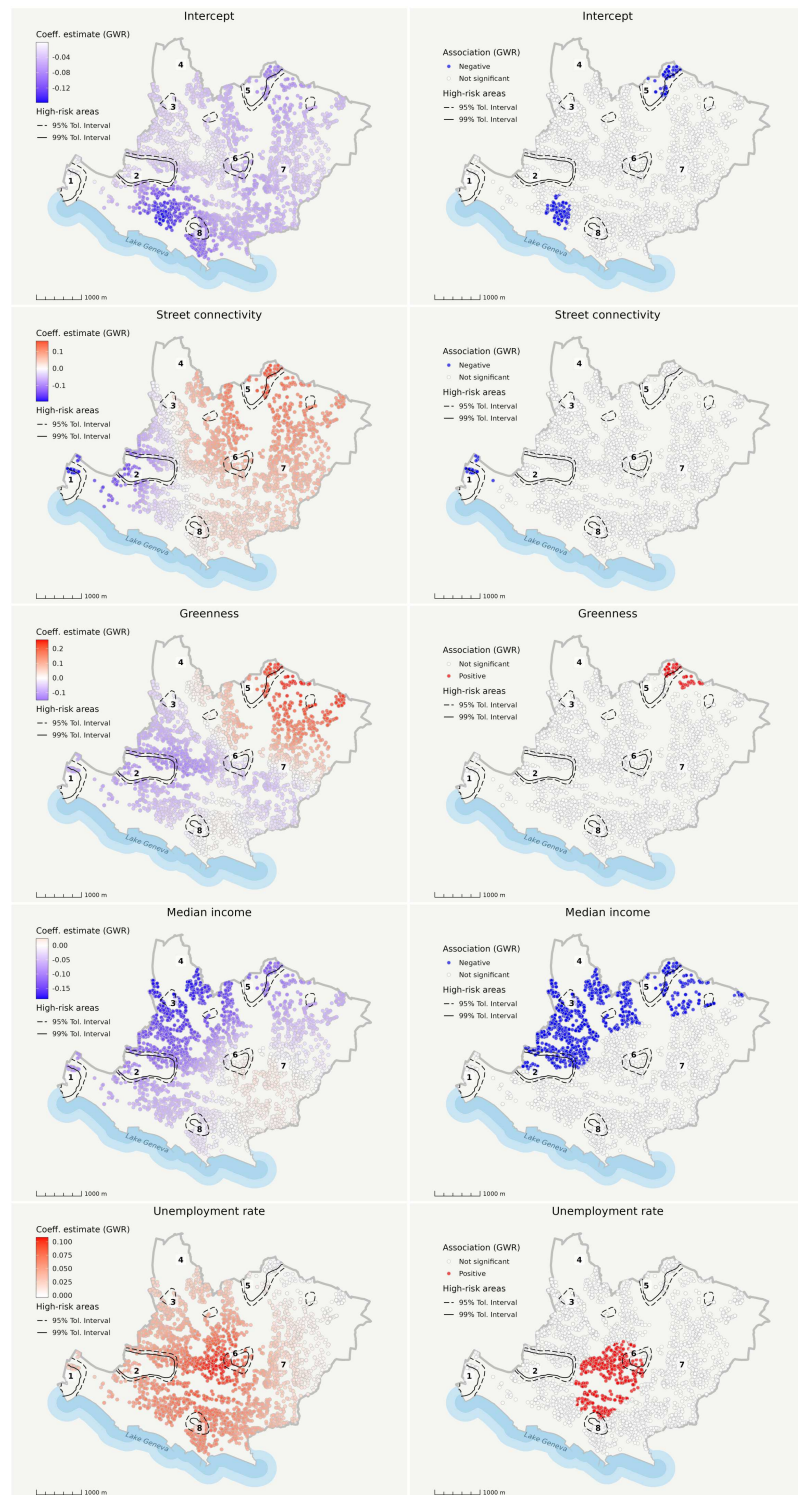


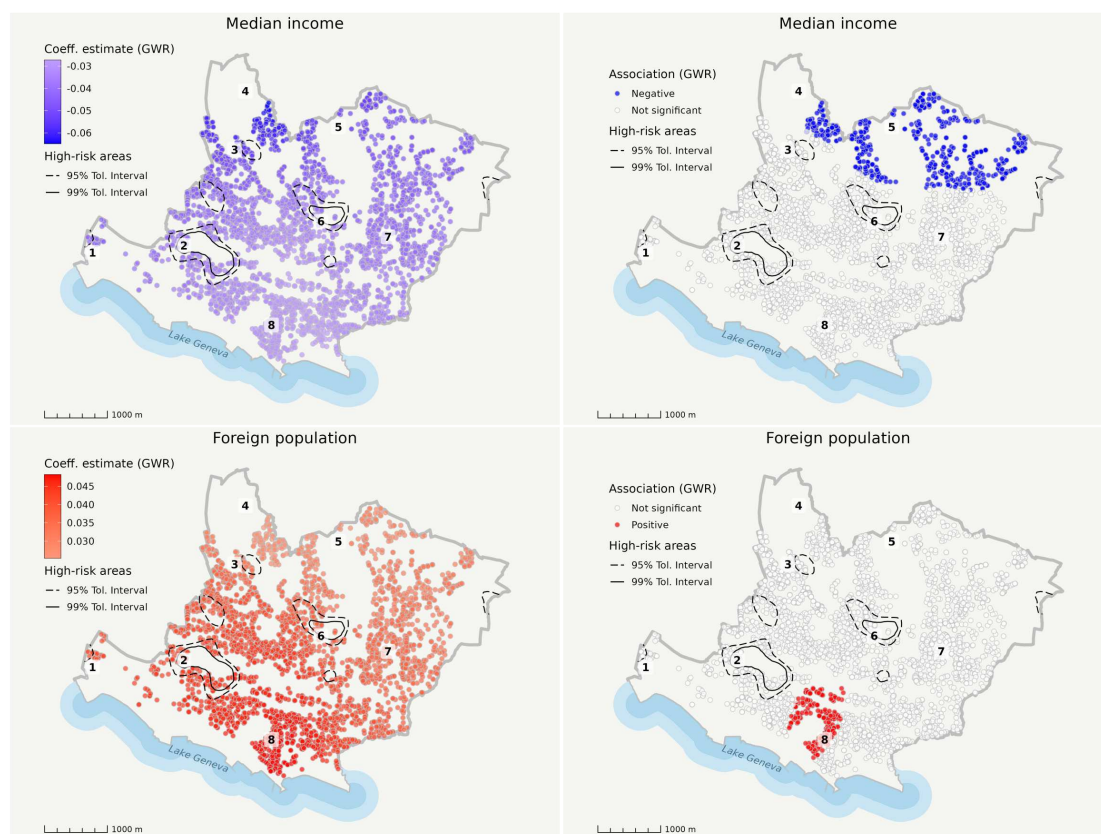
Figure A.3.14 – **GWR results for adjusted obesity.** The figure shows coefficient estimates (left) and spatial associations (right) for physical and social environment characteristics associated with risk of obesity, adjusted for individual confounders. Indicative landmarks are shown on the maps to help interpret the results (#1-#8). Only variables that showed statistically significant variation (based on corrected critical t-values) are shown.





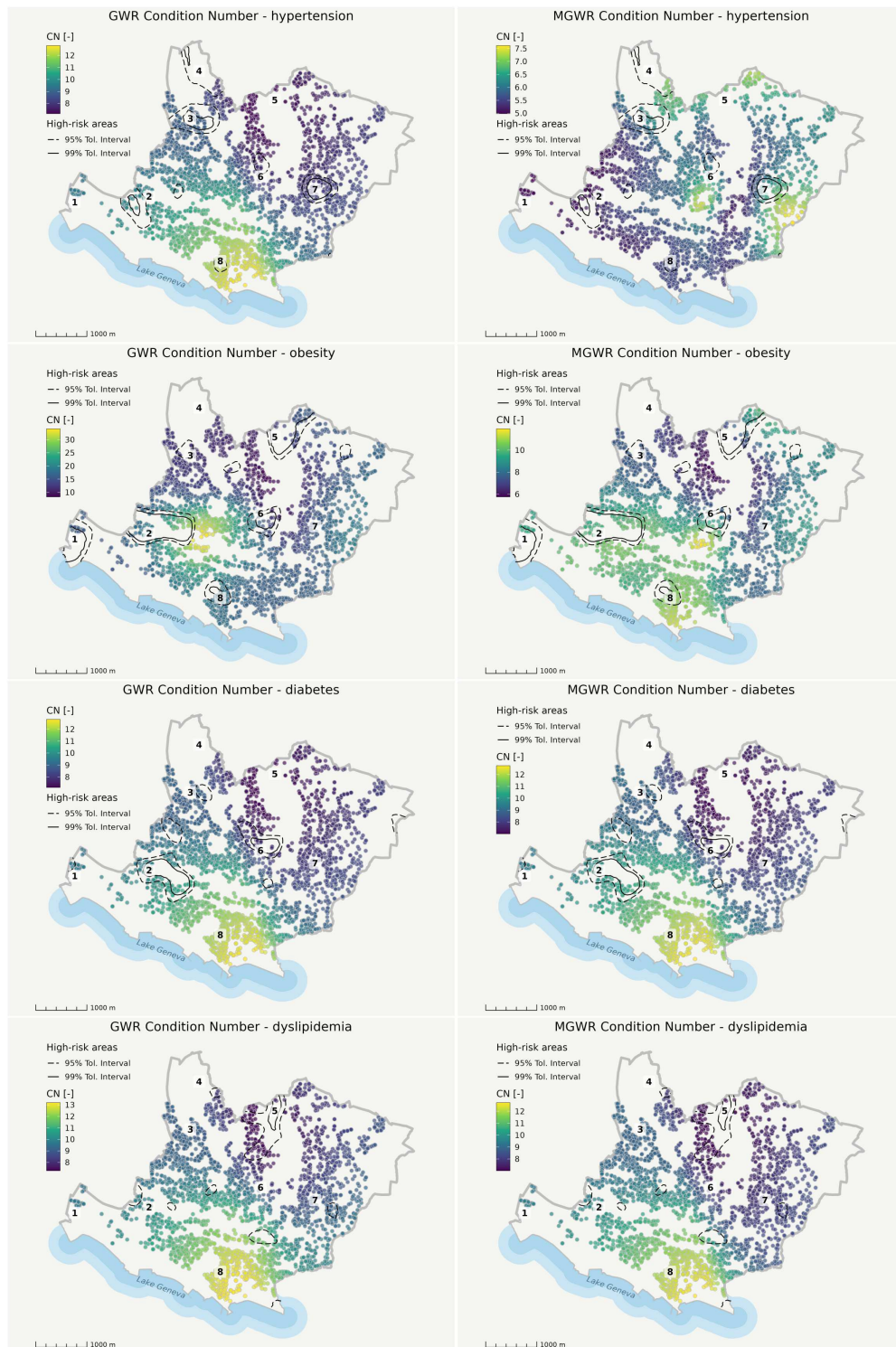
### A.3. Supplementary materials for Chapter 3

Figure A.3.15 – **GWR results for adjusted diabetes.** The figure shows coefficient estimates (left) and spatial associations (right) for physical and social environment characteristics associated with risk of diabetes, adjusted for individual confounders. Indicative landmarks are shown on the maps to help interpret the results (#1-#8). Only variables that showed statistically significant variation (based on corrected critical t-values) are shown.



## Appendix . Appendix

Figure A.3.16 – Maps of local condition number for GWR (left) and MGWR (right) models. Condition number values > 30 suggest local multicollinearity issues. Indicative landmarks are shown on the maps to help interpret the results (#1-#8).



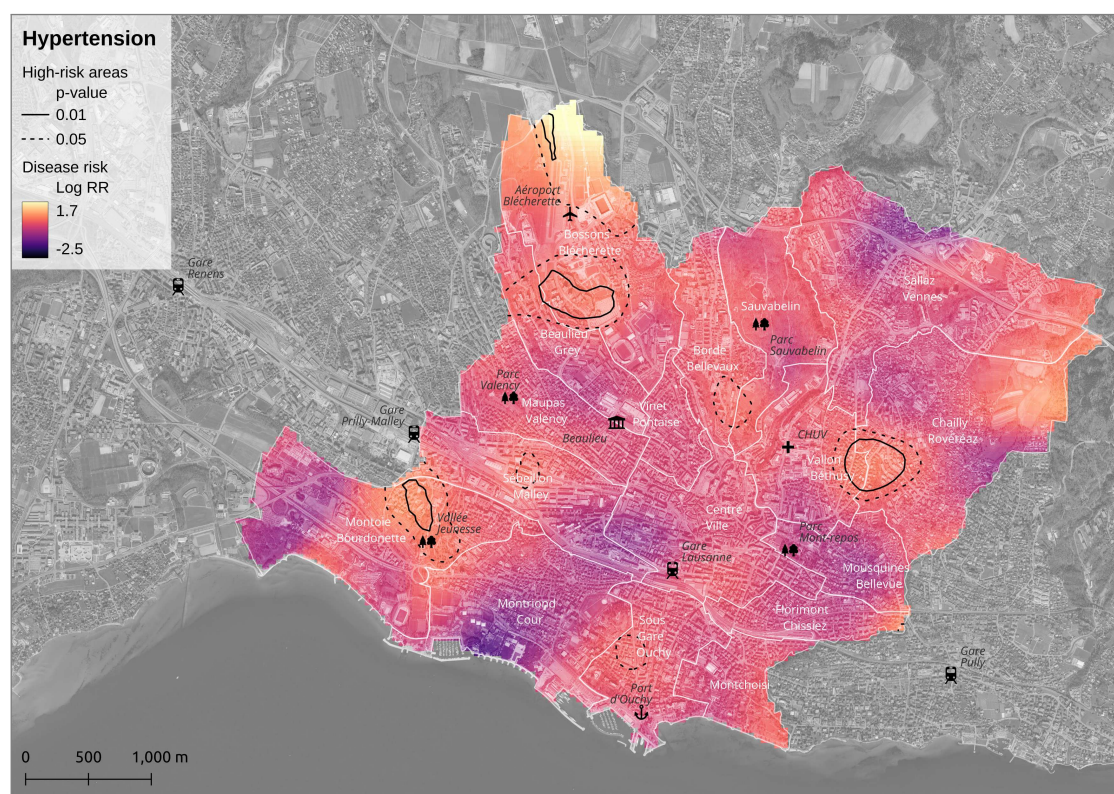


### A.3. Supplementary materials for Chapter 3

Table A.3.5 – Comparisons of the bandwidths used in the GWR and MGWR models.

	Hypertension		Obesity		Diabetes		Dyslipidemia	
	GWR	MGWR	GWR	MGWR	GWR	MGWR	GWR	MGWR
Intercept	3568	3568	2173	3409	3458	3458	3378	3403
Street connectivity	3568	3568	2173	3409	3458	3458	3378	3363
Greenness	3568	3568	2173	1010	3458	3458	3378	3403
Nighttime noise	3568	3568	2173	3409	3458	3458	3378	3403
PM2.5	3568	1137	2173	3409	3458	3458	3378	3403
NO2	3568	3568	2173	3409	3458	3458	3378	3403
Median income	3568	3568	2173	2792	3458	659	3378	3339
Unemployment rate	3568	3567	2173	3350	3458	3458	3378	2958
% Low education	3568	3568	2173	1077	3458	3458	3378	3373
% Foreign population	3568	3566	2173	1579	3458	3458	3378	3403

Figure A.3.17 – **Spatial variation of hypertension risk among middle-aged and older adults in Lausanne.** The disease relative risk was estimated via intensity functions with a 200m Gaussian kernel bandwidth and high-risk areas were delineated using tolerance contours from 999 Monte Carlo permutations (Basemap layer: swissimage, swisstopo).

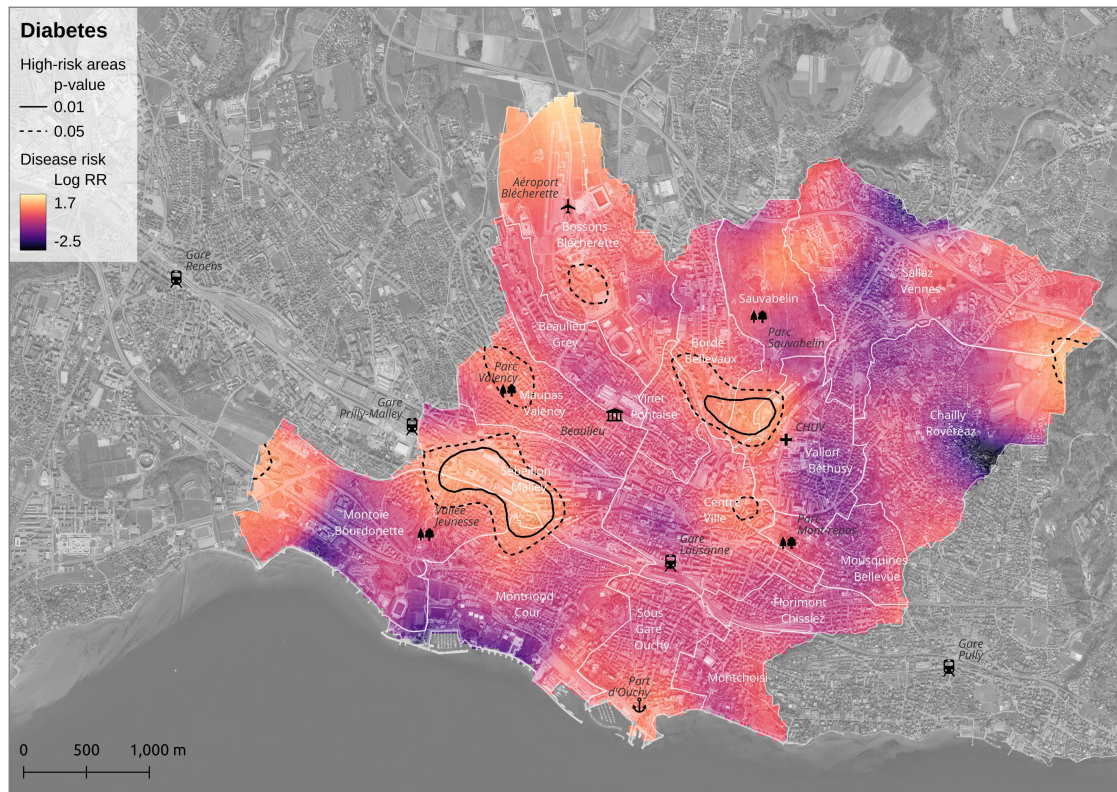






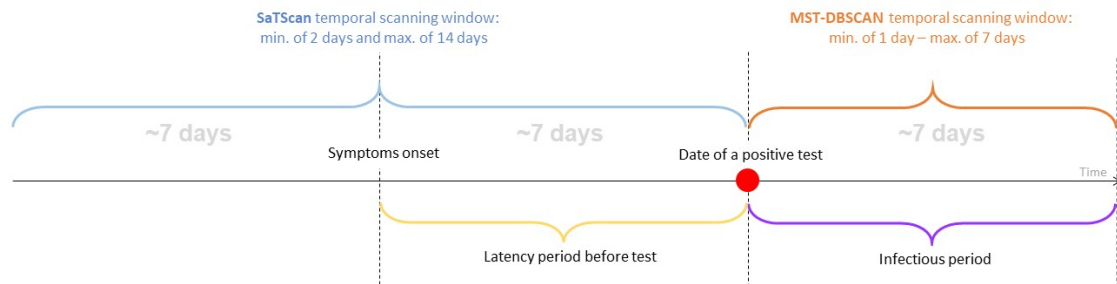
### A.3. Supplementary materials for Chapter 3

Figure A.3.19 – **Spatial variation of diabetes risk among middle-aged and older adults in Lausanne.** The disease relative risk was estimated via intensity functions with a 200m Gaussian kernel bandwidth and high-risk areas were delineated using tolerance contours from 999 Monte Carlo permutations (Basemap layer: swissimage, swisstopo).



## A.4 Supplementary materials for Chapter 4

Figure A.4.1 – **Definition of time windows with SaTScan and MST-DBSCAN.** In this study we have adopted a posture that allows us to early detect alive clusters only. The detection of clusters that emerge all the way to current time is carried out by SaTScan with parameters set to process a prospective Poisson space-time scan analysis considering a window of 2-14 days in the past. The analysis is repeated each day in order to gradually monitor the progression of the epidemic. To complete the results obtained, we used MST-DBSCAN in order to simultaneously assess the transmission behavior of clusters and understand their evolution in time. SaTScan and MST-DBSCAN work differently. While SaTScan looks backwards in time, i.e. from the date of analysis and considering the 14 previous days, MST-DBSCAN looks for positive cases in the  $N$  following days. Indeed, in this case a cluster is defined by a core point having a minimum number of 3 neighbors, and then a positive case must appear close to it ( $<1000\text{m}$ ) and later (1-7 days) than the core point to be considered a spatio-temporal neighbor. This difference (7 days temporal window for MST-DBSCAN compared to 14 days temporal window for SaTScan) is due to the fact that we take into account a latency period of ~7 days since the gradual onset of the first symptoms for a person to decide to take the steps to get tested for COVID-19. Thus the date of the test we have in the database is approximately in the middle of a period of 14 days (like SaTScan), meaning that we need to consider the 7 next days as input parameter for MST-DBSCAN.



### Note A.4: Main stages of the MST-DBSCAN algorithm

#### Stage 1: Identifying transmission clusters

An MST-DBSCAN transmission cluster (Kuo et al., 2018) is formed by a set of case points, including a core point and a minimum number of spatio-temporal neighbors (at least 3 in our study).

The definition of the spatio-temporal neighbors, given below, models the transmission relationship between cases by ensuring that the neighbors  $j$  of a case  $i$  appear spatially near but

later so that  $j$  could be infected by  $i$ .

$$STNB_{i,t} = \{j \in D_t \mid dist(i, j) \leq EpsS \cap EpsT2 \leq \Delta T(i, j) \leq EpsT1 \cap i \neq j\} \quad (1)$$

where

$D_t$  denotes the set of case points included at time  $t$ ,

$dist(i, j)$  denotes the spatial distance from  $i$  to  $j$ ,

$EpsS$  is the threshold value of the spatial adjacency (1000m in our study),

$EpsT1$  and  $EpsT2$  are the threshold values of the possible shortest and longest transmission (respectively 1 and 7 days in our study)

and  $\Delta T(i, j)$  denotes the difference between the appearance time of  $i$  and  $j$

### Stage 2: Classifying the diffusion type of transmission clusters at each epoch

The algorithm identifies the diffusion type (e.g., emerge, grow, merge, steady, move, split, reduce) of each transmission cluster by calculating the movement distance of the mean center and the proportion of the area change between two epochs, according to the formulas below:

$$\Delta MC_{i,t} = dist(MC_{i,t}, MC_{i,t-1}) \quad (2)$$

$$\Delta A_{i,t} = \frac{A_{i,t} - A_{i,t-1}}{A_{i,t-1}} \quad (3)$$

where  $MC_{i,t}$  and  $A_{i,t}$  are, respectively, the spatial mean center and the area of the transmission cluster  $i$  at time  $t$ .

For example, the diffusion type "Move" corresponds to a transmission cluster for which the area remains the same, but the center moves.

### Stage 3: Delineating diffusion zones

To compare the diffusion dynamics across our study area, we first project the evolution patterns of transmission clusters into administrative units (i.e., postal codes in our study). Then, a

diffusion similarity coefficient  $S$  is computed for any two areas according to the formula below:

$$S_{i,j} = \frac{E_{i,j}}{E_a} \quad (4)$$

where  $E_a$  denotes the total number of epochs of the epidemic and  $E_{i,j}$  denotes the number of epochs in which units  $i$  and  $j$  have the same evolution pattern.

Finally, the Louvain method (Blondel et al., 2008) groups administrative units by finding a set that has high similarities within each group and low similarities between groups. For more details, please refer to the publication of Kuo et al. (2018)

## References

- Kuo F-Y, Wen T-H, Sabel CE. 2018. Characterizing Diffusion Dynamics of Disease Clustering: A Modified Space–Time DBSCAN (MST-DBSCAN) Algorithm. *Ann Am Assoc Geogr* 108:1168–1186.
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. 2008. Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008:P10008.

Figure A.4.2 – Distribution of viral loads for individuals within significant clusters

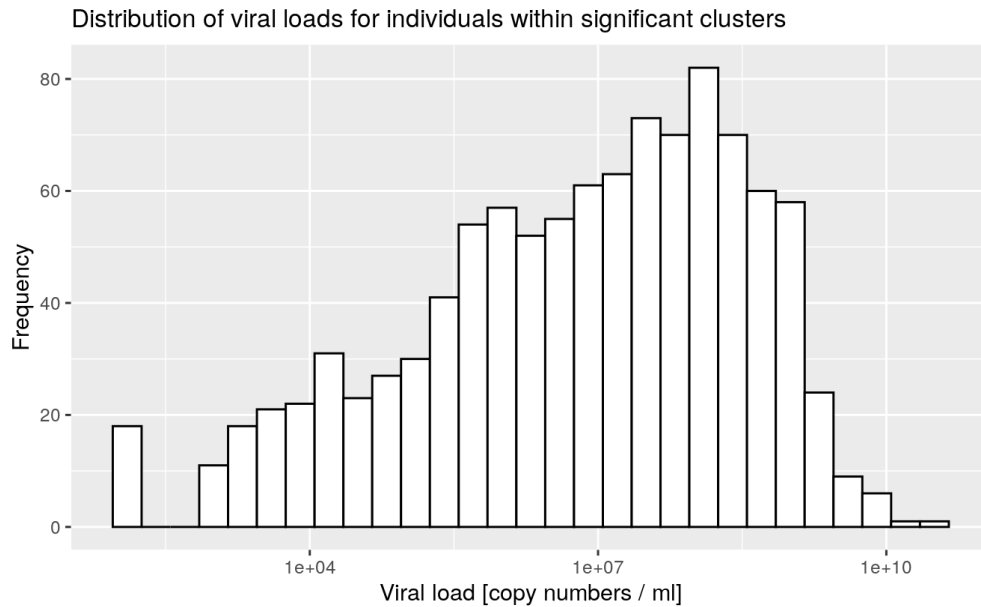


Figure A.4.3 – Distribution of viral loads for individuals outside significant clusters

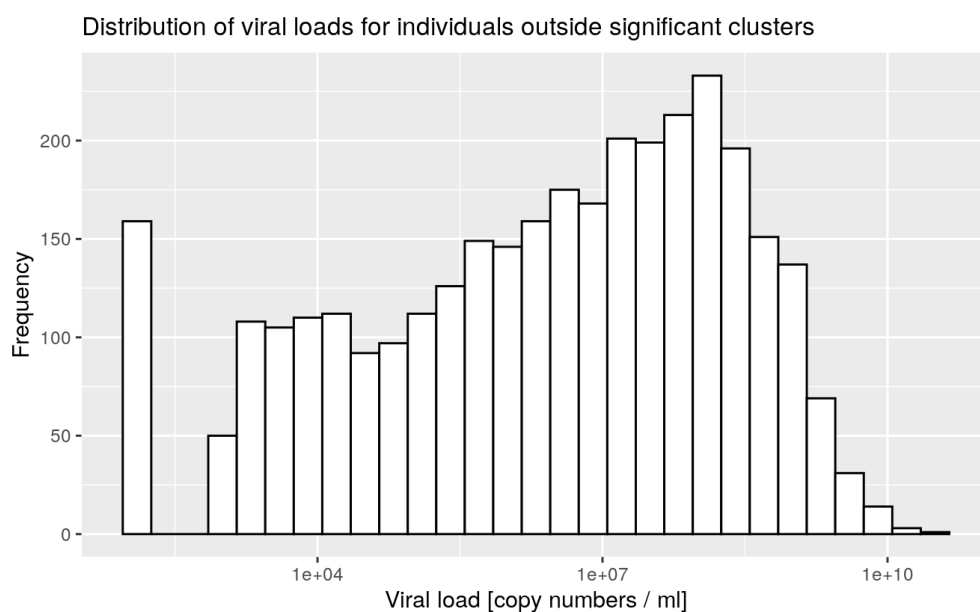


Table A.4.1 – **Mean viral load of the first three cases within clusters.** The numbers in brackets correspond to the number of clusters per category if we consider all individuals and not just the first three cases.

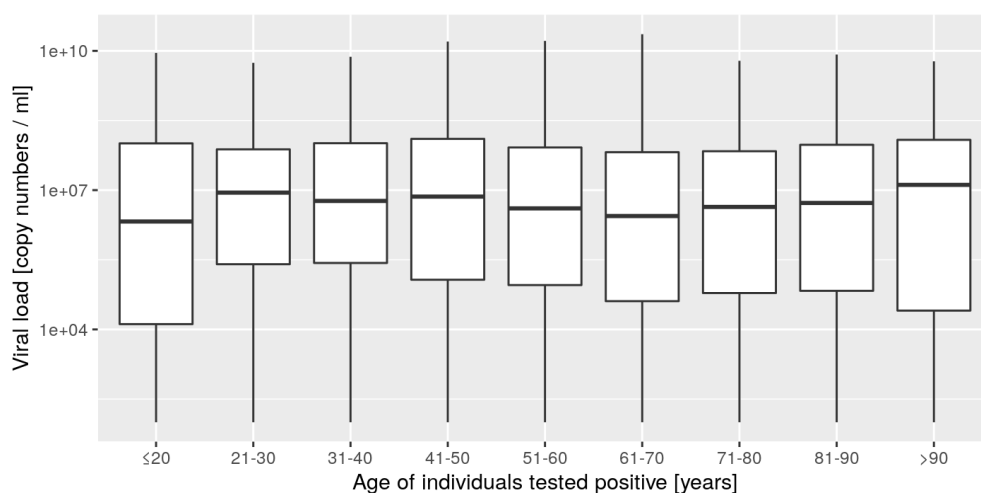
Mean viral load of the first three cases [copy numbers / ml]	Within-clusters cases						
	<5	5-9	10-14	15-19	20-24	25-29	≥30
<1e4	15 (15)	-	-	-	-	-	-
1e4-1e5	17 (17)	4 (-)	-	-	-	-	-
1e5-1e6	1 (1)	1 (1)	6 (-)	3 (-)	5 (-)	2 (-)	-
1e6-1e7	14 (13)	28 (15)	6 (-)	8 (-)	10 (-)	4 (-)	-
1e7-1e8	8 (10)	22 (33)	25 (10)	31 (16)	34 (13)	5 (-)	2 (-)
1e8-1e9	30 (30)	50 (53)	26 (50)	33 (61)	38 (76)	3 (15)	10 (12)
≥ 1e9	1 (0)	3 (6)	4 (7)	5 (3)	2 (-)	1 (-)	-

## Appendix . Appendix

Table A.4.2 – **Maximal viral load of the first three cases within clusters.** The numbers in brackets correspond to the number of clusters per category if we consider all individuals and not just the first three cases.

Maximal viral load of the first three cases [copy numbers /ml]	Within-clusters cases						
	<5	5-9	10-14	15-19	20-24	25-29	≥30
<1e4	5 (5)	-	-	-	-	-	-
1e4-1e5	13 (13)	2 (-)	-	-	-	-	-
1e5-1e6	15 (15)	2 (-)	2 (-)	-	1 (-)	-	-
1e6-1e7	14 (13)	17 (10)	5 (-)	9 (-)	11 (-)	6 (-)	-
1e7-1e8	1 (2)	19 (10)	17 (1)	14 (-)	18 (-)	2 (-)	1 (-)
1e8-1e9	29 (27)	42 (42)	34 (22)	39 (20)	40 (17)	5 (-)	9 (-)
≥ 1e9	9 (11)	26 (46)	9 (44)	18 (60)	19 (72)	2 (15)	2 (-)

Figure A.4.4 – Boxplots of viral load per age class of individuals tested positive



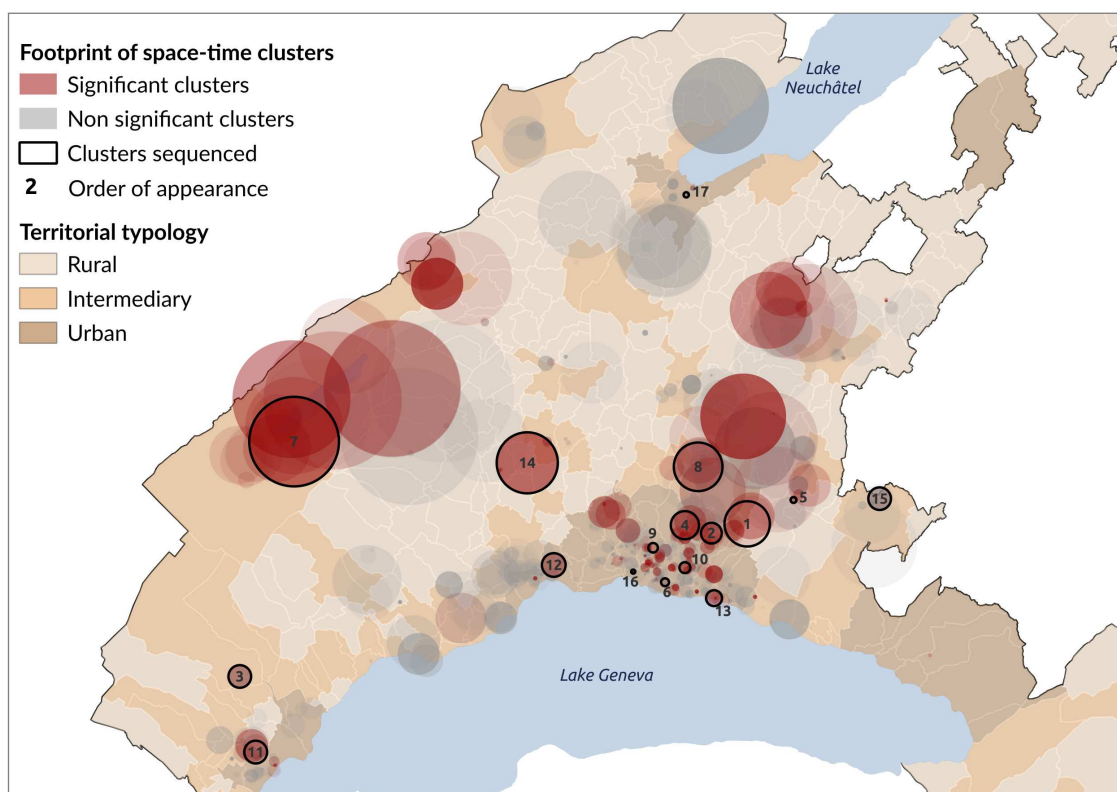


## A.5 Supplementary materials for Chapter 5

Table A.5.1 – **Characteristics of the individuals for whom we analyzed the SARS-CoV-2 genomic sequence (N=172).** Viral load is expressed as number of copies/ml.

Cluster	N	Age (SD)	Females (%)	Viral load (SD)
Lausanne region	95	53.1 (18.8)	47 (49.5)	5.6e+08 (1.1e+09)
#1	7	58.6 (14.8)	2 (28.6)	2.7e+08 (4.2e+08)
#2	10	47.7 (17.1)	7 (70)	1.2e+09 (2e+09)
#4	12	66.7 (17.1)	6 (50)	4e+08 (6.6e+08)
#6	13	52.9 (20.5)	7 (53.8)	2.5e+08 (3.8e+08)
#8	21	57.6 (18.3)	13 (61.9)	7.4e+08 (1.3e+09)
#9	3	55.6 (22.4)	2 (66.7)	1.6e+08 (1.6e+08)
#10	19	46.4 (16.3)	6 (31.6)	6.4e+08 (1.3e+09)
#13	5	44 (18.6)	1 (20)	2.7e+07 (2.4e+07)
#16	5	50.3 (6.83)	3 (60)	4.4e+08 (4.4e+08)
#3	3	62.1 (7.12)	1 (33.3)	1e+08 (1.6e+08)
#5	5	50.9 (26.3)	2 (40)	2.1e+08 (4e+08)
#7	31	71.6 (19.3)	20 (64.5)	2.7e+08 (5.9e+08)
#11	16	54.8 (17.3)	8 (50)	1.1e+09 (2.4e+09)
#12	12	50.7 (20.9)	5 (41.7)	1.1e+08 (2e+08)
#14	3	65.3 (9.93)	1 (33.3)	3.7e+08 (5.4e+08)
#15	3	45.8 (3.05)	2 (66.7)	9.7e+07 (1.2e+08)
#17	4	78.0 (27.1)	4 (100)	3.5e+07 (4.5e+07)

Figure A.5.1 – **Selection of the 17 spatiotemporal clusters considered for genomic data analysis.** Overlaid on this map are all clusters of SARS-CoV-2 cases detected daily from March 2 to June 30, 2020 period using space-time scan statistics [1]. The 17 geographical clusters selected for genome sequencing are labeled according to their order of apparition. The clusters were selected from different urban-rural contexts that are represented here using the territorial typology developed by the Federal Statistical Office in 2012 [2]. Note that in space-time scan statistics, large clusters do not necessarily correspond to a high number of cases, but rather to an unusually high relative risk of COVID-19 cases detected in this area. For a similar number of cases observed in a defined area, the relative risk is per definition lower in high density populated areas. We differentiate clusters that had a p-value < 0.05 (red) or > 0.05 (gray) based on 999 Monte-Carlo permutations.



## A.5. Supplementary materials for Chapter 5

Figure A.5.2 – **Construction of minimum spanning tree from a network of SARS-CoV-2 genomes with their pairwise SNV distance.** i. 172 sequences represented as nodes are connected with vertices indicating SNV distance. ii. 20 “genomic groups” with identical sequences are identified. iii. Sequences in the same genomic groups are merged into a single node. iv. Minimum spanning tree was built from the previous network using Prim’s algorithm, which gives the shortest path to go through all nodes.

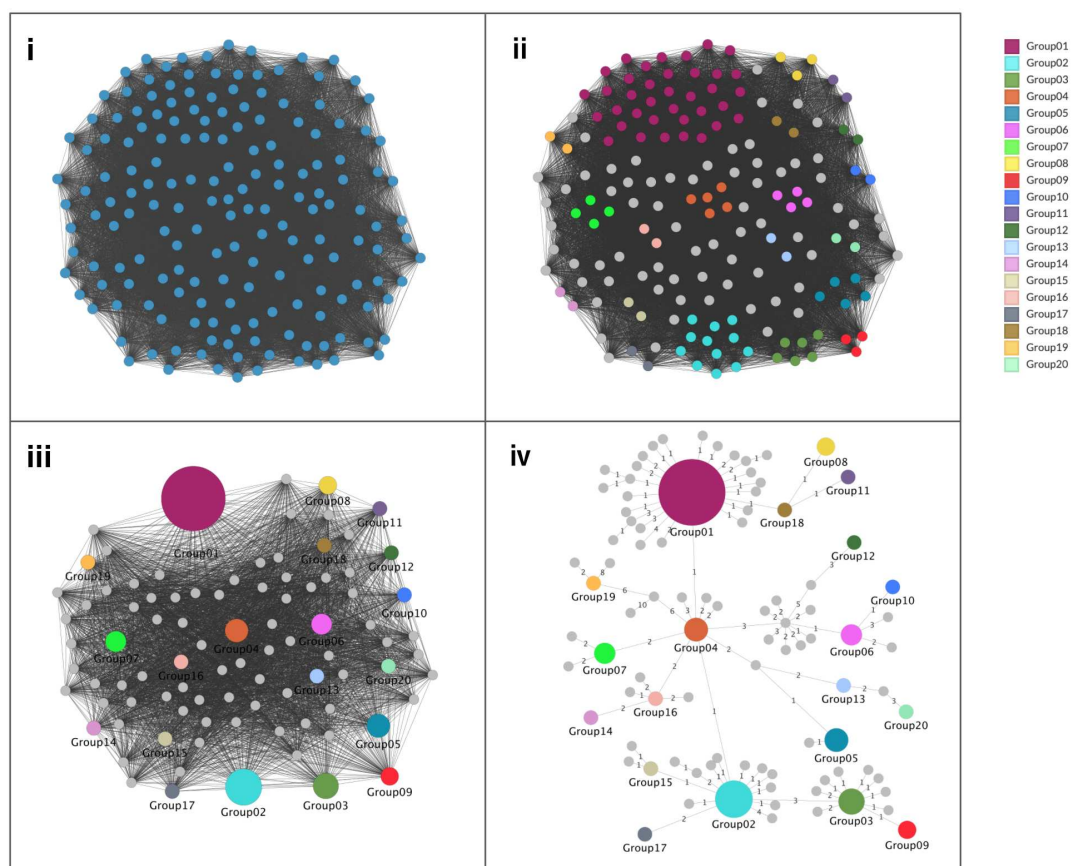


Figure A.5.3 – **Jaccard similarity index of 9 spatiotemporal clusters within the Lausanne region.** The overall median is indicated as a dotted line. Wilcoxon test was applied for cluster #2 and #09.

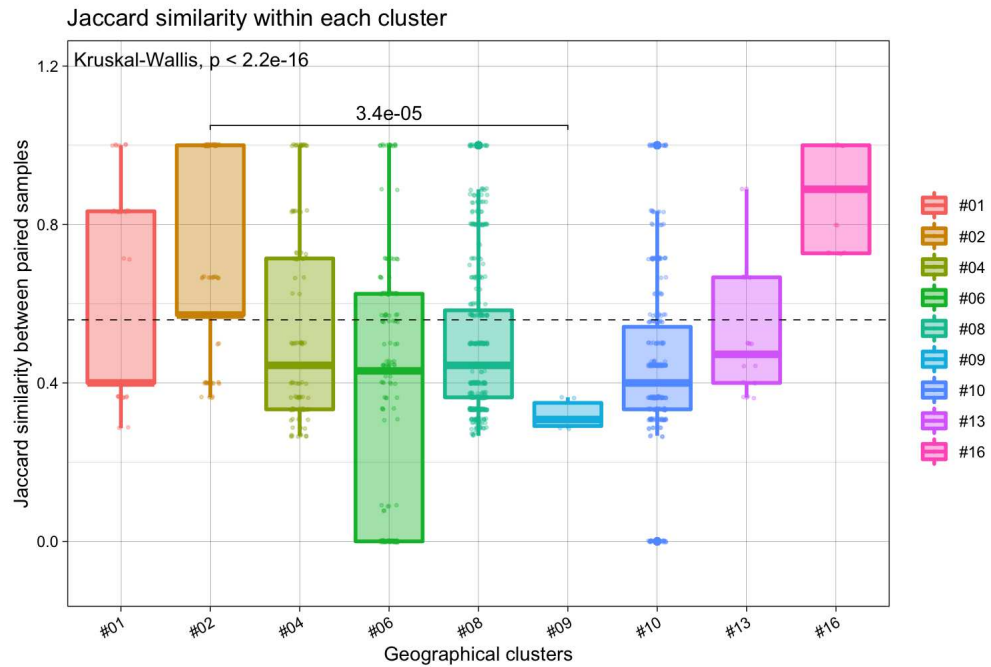
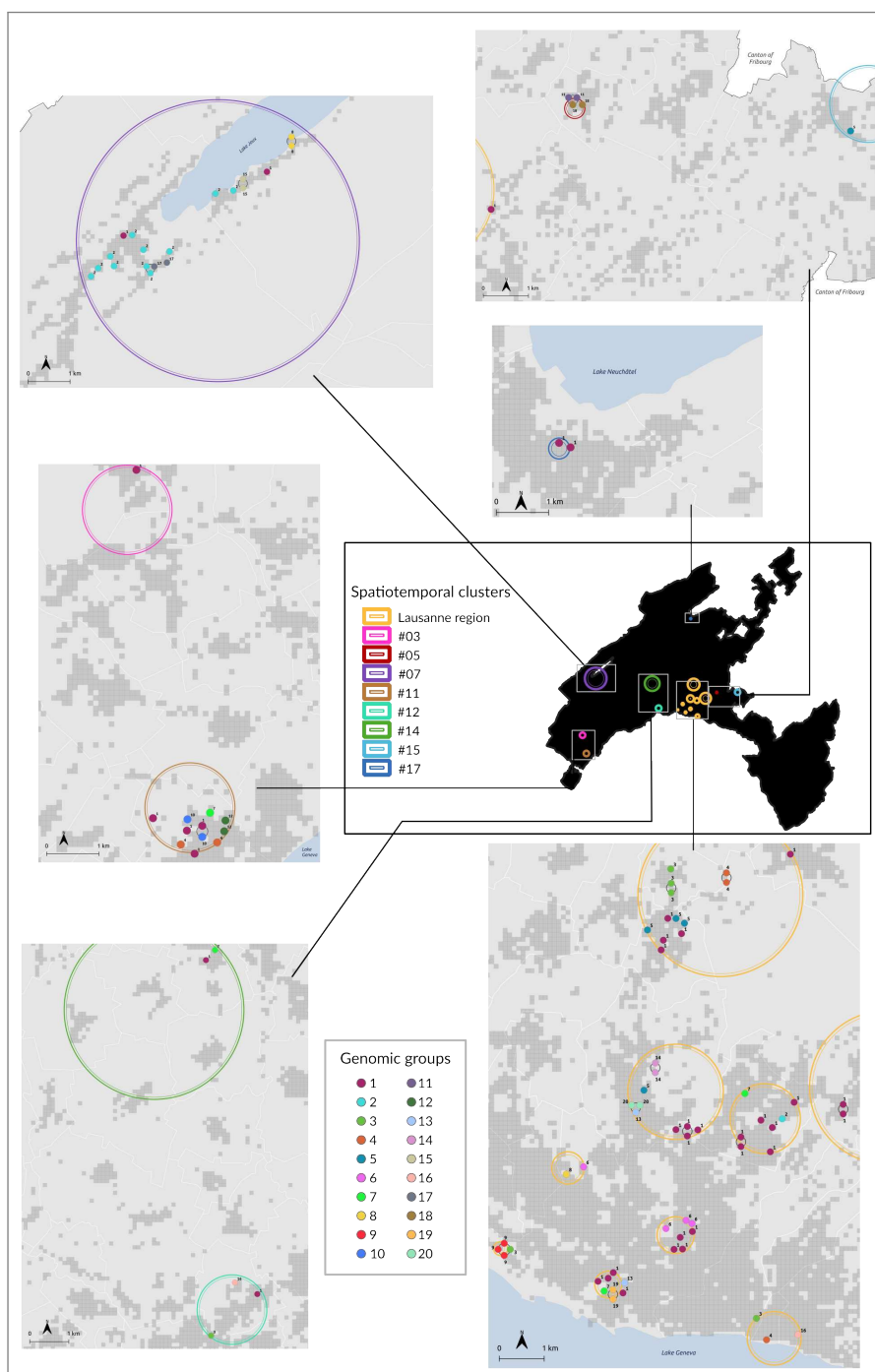


Figure A.5.4 – **Spatial distribution of genomic groups within spatiotemporal clusters.** Each point refers to a sequenced case geocoded to the patient's residential address, and is colored based on the genomic group to which it belongs. A small ring network represents cases that occurred in the same building. Colors of the large rings correspond to the color of the geographical clusters as summarized in the upper middle panel. Note that we randomly permuted the point's location within clusters to maintain confidentiality.



## Appendix . Appendix

Table A.5.2 – Pangolin lineage and nucleotide sequence changes of 20 genomic groups.

Genomic groups	Pangolin lineage	Nucleotide sequence changes
Group01	B.1	C241T, C3037T, C14408T, C15324T, A23403G
Group02	B.1	C241T, C3037T, C14408T, A23403G, A26530G
Group03	B.1.128	C241T, C3037T, C8293T, C14408T, T18488C, A23403G, T24739C, A26530G
Group04	B.1	C241T, C3037T, C14408T, A23403G
Group05	B.1.93	C241T, C3037T, C13862T, C14408T, A23403G, C28045T, C28836T
Group06	B.1.1	C241T, C3037T, C14408T, T19839C, A23403G, G28881A, G28882A, G28883C
Group07	B.1	C241T, C3037T, C14408T, A20268G, A23403G, C28854T
Group08	B.1	C241T, C3037T, T14208C, C14408T, C15324T, A23403G, T24022C
Group09	B.1.128	C241T, C3037T, G7675T, C8293T, C14408T, T18488C, A23403G, T24739C, A26530G
Group10	B.1.1	C241T, C3037T, C14408T, T19839C, A23403G, G23426T, G28881A, G28882A, G28883C
Group11	B.1.1	C241T, C3037T, C4320T, T14208C, C14408T, C15324T, A23403G
Group12	B.1.1.1	C241T, C3037T, C4002T, G10097A, C13536T, C14408T, C20104T, C20946T, A23403G, C23731T, G28881A, G28882A, G28883C
Group13	B.1.93	C241T, C3037T, C13862T, C14408T, C15352T, A23403G, C28836T, G29229A
Group14	B.1	C241T, C1059T, C3037T, G11083T, C14408T, C16616T, A23403G, G25563T
Group15	B.1	C241T, C3037T, C12020T, C14408T, A23403G, A26530G
Group16	B.1	C241T, C1059T, C3037T, C14408T, A23403G, G25563T
Group17	B.1	C241T, G677A, C3037T, C11511T, C14408T, A23403G, A26530G
Group18	B.1	C241T, C3037T, T14208C, C14408T, C15324T, A23403G
Group19	B	G11083T, C14805T, T17247C, G26144T
Group20	B.1.93	C241T, C3037T, T3466C, C6738T, C6781T, C13862T, C14408T, C15352T, A23403G, C28836T, G29229A

## A.5. Supplementary materials for Chapter 5

Table A.5.3 – Advantages and disadvantages of spatiotemporal clustering and genomics in surveillance.

	Spatiotemporal Clustering	Genomics
<b>Advantages</b>	Early detection of case clusters; Indicate where to prioritize interventions and resources; Rely on already collected data; Low processing cost	Identification of known and novel strains; Tracking local and intercontinental transmission; International genome databases; Numerous user-friendly tools
<b>Disadvantages</b>	Do not guarantee epidemiological linkage within clusters; Centered around the residential setting	Considerable cost; Long turnaround time

## **A.6 Manuscript - Geographic footprints of life expectancy inequalities in the state of Geneva, Switzerland**

This appendix features an auxiliary study conducted in the canton of Geneva, Switzerland. The research presented in this manuscript employs Local Moran's I spatial autocorrelation statistics to analyze the spatial variation of Years of Potential Life Lost or Gained (YPLL<sub>G</sub>), thereby offering valuable insights into the geographic pattern of life expectancy inequalities in the canton of Geneva.

The manuscript, published in the journal *Scientific Reports*, is reproduced in its entirety and without modifications, in compliance with the Creative Commons Attribution 4.0 International License.

Citation: Ladoy, A., Vallarta-Robledo, J. R., De Ridder, D., Sandoval, J. L., Stringhini, S., Da Costa, H., Guessous, I., & Joost, S. (2021). Geographic footprints of life expectancy inequalities in the state of Geneva, Switzerland. *Scientific Reports*, 11(1), 23326. <https://doi.org/10.1038/s41598-021-02733-x>

Supplementary materials related to this manuscript can be accessed at the following URL: <https://doi.org/10.1038/s41598-021-02733-x>.





OPEN

# Geographic footprints of life expectancy inequalities in the state of Geneva, Switzerland

Anaïs Ladoy<sup>1,2,10</sup>, Juan R. Vallarta-Robledo<sup>2,3,4,10</sup>, David De Ridder<sup>1,2,3,4</sup>, José Luis Sandoval<sup>1,2,5,6</sup>, Silvia Stringhini<sup>2,3,4,7</sup>, Henrique Da Costa<sup>8</sup>, Idris Guessous<sup>2,3,6,10</sup> & Stéphane Joost<sup>1,2,4,9,10</sup>✉

Though Switzerland has one of the highest life expectancies in the world, this global indicator may mask significant disparities at a local level. The present study used a spatial cluster detection approach based on individual death records to investigate the geographical footprint of life expectancy inequalities in the state of Geneva, Switzerland. Individual-level mortality data ( $n = 22,751$ ) were obtained from Geneva's official death notices (2009–2016). We measured life expectancy inequalities using the years of potential life lost or gained (YPLL) metric, defined as the difference between an individual's age at death and their life expectancy at birth. We assessed the spatial dependence of YPLL across the state of Geneva using spatial autocorrelation statistics (Local Moran's  $I$ ). To ensure the robustness of the patterns discovered, we ran the analyses for ten random subsets of 10,000 individuals taken from the 22,751 deceased. We also repeated the spatial analysis for YPLL before and after controlling for individual-level and neighborhood-level covariates. The results showed that YPLL was not randomly distributed across the state of Geneva. The ten random subsets revealed no significant difference with the geographic footprint of YPLL and the population characteristics within Local Moran cluster types, suggesting robustness for the observed spatial structure. The proportion of women, the proportion of Swiss, the neighborhood median income, and the neighborhood median age were all significantly lower for populations in low YPLL clusters when compared to populations in high YPLL clusters. After controlling for individual-level and neighborhood-level covariates, we observed a reduction of 43% and 39% in the size of low and high YPLL clusters, respectively. To our knowledge, this is the first study in Switzerland using spatial cluster detection methods to investigate inequalities in life expectancy at a local scale and based on individual data. We identified clear geographic footprints of YPLL, which may support further investigations and guide future public health interventions at the local level.

Health inequalities, defined as differences in the population's health status<sup>1</sup>, remain a major challenge in public health<sup>2,3</sup>. Individuals more socioeconomically deprived usually face poorer health conditions and are at higher risk of presenting a premature death<sup>4–7</sup>. Premature death, evaluated through Years of Potential Life Lost (YPLL), which estimates the years a person did not live compared to an arbitrary age, is one of the most widely used mortality-based indicators to measure the population's health<sup>8–10</sup>. This indicator was first introduced by Mary Dempsey<sup>11</sup> to contrast the results obtained with mortality rate measurements in tuberculosis control.

Compared to other countries, Switzerland's health inequalities for all-cause mortality are below average<sup>9,12</sup>. Nevertheless, these statistics constitute a global estimation of the health situation at the country level, and

<sup>1</sup>Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. <sup>2</sup>Group of Geographic Information Research and Analysis in Population Health (GIRAPH), Geneva, Switzerland. <sup>3</sup>Faculty of Medicine, University of Geneva, Geneva, Switzerland. <sup>4</sup>Unit of Population Epidemiology, Department of Primary Care, Geneva University Hospitals, Geneva, Switzerland. <sup>5</sup>Department of Oncology, Geneva University Hospitals, Geneva, Switzerland. <sup>6</sup>Division of Primary Care Medicine, Department of Primary Care, Geneva University Hospitals, Geneva, Switzerland. <sup>7</sup>University Centre for General Medicine and Public Health, University of Lausanne, Lausanne, Switzerland. <sup>8</sup>Réseau Delta, HMO, Geneva, Switzerland. <sup>9</sup>La Source School of Nursing, University of Applied Sciences and Arts Western Switzerland (HES-SO), Lausanne, Switzerland. <sup>10</sup>These authors contributed equally: Anaïs Ladoy, Juan R. Vallarta-Robledo, Idris Guessous and Stéphane Joost. ✉email: stephane.joost@epfl.ch

substantial regional differences of mortality and socioeconomic status have been reported in Switzerland at the neighborhood level<sup>13,14</sup>.

Evidence shows that neighborhood conditions (i.e., social, economic, and physical) have an influence on the health of individuals, independent of personal characteristics<sup>15–17</sup>. Spatial analyses are therefore valuable for revealing geographic patterns in health inequalities, identifying populations at risk, and guiding public health interventions at local scales<sup>18,19</sup>. In the last decade, increasing evidence suggests strong patterns in mortality indicators across geographical space, both at large<sup>20–22</sup> and small geographic scales<sup>23–26</sup>. In addition, the spatial structure of premature mortality has been found to be significantly associated with deprivation status<sup>20–26</sup>, immigrant population size<sup>23</sup>, and multiple environmental features (e.g., pollution, greenspace, walkability)<sup>24–26</sup>.

These past analyses have been performed at aggregated levels, such as at county- or neighborhood-levels, making these results sensitive to variations in scale, which is commonly referred to as the modifiable areal unit problem (MAUP)<sup>27</sup>.

Furthermore, studies assessing the spatial distribution of mortality inequalities tend to focus only on premature mortality. Unfortunately, this indicator may smooth out mortality inequalities as it considers only individuals in the population that die before a certain age and ignores individuals living longer than expected.

We therefore sought to investigate mortality inequalities at a fine scale in the state of Geneva, using death records for 22,751 individuals from 2009 to 2016, with data georeferenced at the residential address. The outcome variable used here is an adaptation of the YPLL indicator, “years of potential life lost or gained” (YPLLG), and is defined as the difference (positive or negative) between age at death and life expectancy at birth (LEB). We believe that this indicator better captures mortality inequalities across geographical space. We assessed the spatial structure of YPLLG using spatial autocorrelation statistics, following which we also investigated the influence of individual-level and neighborhood-level covariates on the geographic footprint of YPLLG.

## Methods and materials

**Death notice data.** Raw data included death records from the state of Geneva of 49,628 individuals between 1908 and 2017, with personal information of the deceased including name, date of birth, civil status, nationality (Swiss/non-Swiss), date of death, and residential address. Data were collected through web scraping of the official and publicly available death notices published in the *Feuille d'avis officielle* (FAO) until 2017 (Republic and Canton of Geneva, <https://fao.ge.ch>). Their use did not imply a request for authorization from the ethics committee of the canton of Geneva as the research protocol does not study disease, neither the structure nor the functioning of the human body (Federal Act on Research involving Human Beings, article 2). Moreover, the research protocol respects the Swiss Federal Act on Data Protection (Art. 22) specifying specifically that personal data can be processed for purposes not related to specific persons, and in particular for research, planning and statistics if the data is rendered anonymous as soon as the purpose of the processing permits, and if the results are published in such a manner that the data subjects may not be identified. For our analyses we only retained deaths from 2009 to 2016 ( $n = 27,889$ ; 56.2% of raw data), years where we have information consistent with the number of deaths reported by the Cantonal Statistical Office, OCSTAT ([www.ge.ch/statistique](http://www.ge.ch/statistique)). From this subset, we further removed duplicated entries, individuals with missing date of birth, date of death, or nationality ( $n = 408$ ; 1.5%), as well as individuals living outside the state of Geneva ( $n = 2094$ ; 7.5%), and individuals that could not be georeferenced ( $n = 1025$ ; 3.7%).

Due to the lack of specific gender information in the database, we used the genderize.io API for name-to-gender inference as it shows a correct performance rate compared to other web services<sup>28</sup>. This approach is commonly used in gender inequality research, such as investigating women's representation in academic literature<sup>29,30</sup>. The API returns the gender most commonly associated with a given first name, along with confidence parameters. With this process, we recovered gender for 94.2% of individuals in our dataset (1611 observations were removed).

After data filtering, we were left with 22,751 individuals (81.6% of 2009–2016 deceased) for further analysis. The study was carried out in accordance with the relevant guidelines and regulation.

**Mortality indicators.** (Cohort) life expectancy at birth (LEB) represents the average lifespan of a group of individuals born in any given year, considering the observed and forecasted evolution of death rates through their lifetime<sup>31</sup>. Official estimates of cohort life expectancy were extracted from the 1900–2030 Swiss cohort life tables (FSO, <https://www.bfs.admin.ch/>), calculated from a model developed by Jacques Menthonnex<sup>32</sup>. Cohort LEB was preferred over the traditional period LEB as we believe that it better captures changes in mortality conditions across a lifetime (see Supplementary Fig. S4). LEB was attributed to the deceased based on year of birth and gender.

Years of potential life lost (YPLL) estimates the years a person did not live compared to an arbitrary age (usually 75)<sup>8–10</sup> or compared to the individual's LEB<sup>11</sup>. Note that this indicator does not consider the years a person may live beyond this age cut-off.

Years of potential life lost or gained (YPLLG) is defined as the difference in years between the age at death and the individual's LEB. Positive values of YPLLG capture the years of life ‘gained’, while negative YPLLG values reflect the potential years of life ‘lost’, as defined by Dempsey<sup>11</sup>.

**Neighborhood-level characteristics.** To assess the influence of neighborhood characteristics on the spatial distribution of YPLL, we included yearly data of neighborhood socioeconomic status and median population age for the period between 2009 and 2016. Both indicators were available at the statistical subsector level ( $n = 475$ ), a geographic unit smaller than that of municipality, which is used by the state of Geneva for the diffusion of local aggregated statistical data<sup>33</sup>.

The neighborhood socioeconomic status was measured using the median annual neighborhood household income, which was obtained via a request from the Cantonal Statistical Office, OCSTAT (C. Stohr, personal communication, 2020). The transmitted data excluded unmarried individuals (i.e., single, divorced, widowed) from the calculation of the annual neighborhood income. The reasons given by the Statistical office were that: (1) their taxable income is not a good indicator of their quality of life, and (2) the information of non-taxable income such as social assistance, which constitutes a significant part of unmarried taxpayers' income, was not available before 2014.

The neighborhood population median age was estimated from the resident population by 5-year age groups, with a final open class of 100+. This information was obtained from the Cantonal Statistical Office website ([https://www.ge.ch/statistique/domaines/01/01\\_01/tableaux.asp#4](https://www.ge.ch/statistique/domaines/01/01_01/tableaux.asp#4)). As the calculation of median from grouped data requires classes with equal sizes, we assumed that individuals were not living longer than 105 years.

We assigned neighborhood-level characteristics to each individual based on the registered residential address at the date of the death. If data were not available for a specific neighborhood, we used the nearest neighborhood value.

**Statistical analysis.** We investigated the spatial structure of YPLL across the state of Geneva using the Local Moran statistic<sup>34</sup>. The statistic relies on a measure of spatial dependence (or spatial autocorrelation), i.e., how similar observations tend to be within a specific neighborhood (spatial lag), and identifies local clusters of low and high YPLL values. By differentiating the relationships between individuals and their surroundings into five categories, the Local Moran approach allows for precise interpretation of the spatial structure of a given phenomenon.

We decided to analyze the YPLL variable within a 1200-m buffer (spatial lag) around each individual's residential address. This methodological choice was supported by similar epidemiological studies conducted in the state of Geneva<sup>35,36</sup>.

For each residential address, the correlation between the observed variable and the mean of this variable in a given neighborhood (spatial lag) was calculated. The standardized scatterplot of this relationship allows to identify four distinct types of spatial association: (1) High–High clusters (dark green dots in the maps) represent individuals with high YPLL values (i.e., that live longer than expected) surrounded by individuals with high YPLL values; (2) Low–Low clusters (dark purple dots in the maps) represent individuals with low YPLL values (i.e., that live shorter than expected) surrounded by individuals with low YPLL values; (3) Low–High spatial outliers (light purple dots in the maps) represent individuals with high YPLL surrounded by individuals with low YPLL; and (4) High–Low spatial outliers (light green dots in the maps) represent individuals with low YPLL surrounded by individuals with high YPLL.

To assess whether or not the null hypothesis of no spatial association can be rejected, we performed a significance test using 99,999 Monte-Carlo permutations where the value  $y_i$  at a specific location  $i$  is held fixed for each step and the location of its neighboring values are randomly permuted<sup>34</sup>. Pseudo p-values were then calculated as the probability of obtaining a local Moran's  $I$  larger than observed<sup>37</sup>. To consider the effects of simultaneous multiple comparisons<sup>38</sup>, we applied a Bonferroni correction for an overall alpha level of 0.1, resulting in an individual significance level of  $1e-5$ . Non-significant locations (i.e., with pseudo p-value  $> 1e-5$ ) are shown in white on the maps.

To evaluate the degree to which neighborhood-level characteristics, such as socioeconomic status or population age, explain the spatial structure of YPLL, we performed the same analysis on adjusted YPLL values obtained with a median regression. This regression model is preferred to the traditional Ordinary Least Square model when the outcome variable does not follow a normal distribution<sup>39</sup>, which is the case of YPLL (Supplementary Fig. S1). We also included the nationality (Swiss/Non-Swiss) in the regression model to control for potential confounders. As gender and age were used to calculate an individual's LEB, they were not included as dependent variables. More detailed information about the regression model is provided in the Supplementary Materials (Eq. S1).

Methodological and computational issues may arise from conducting spatial statistics on such large datasets. First, the Bonferroni bound (defined as  $\alpha/n$ , where  $\alpha = 0.1$ , and  $n$  is the number of observations) requires  $9e+5$  permutations to be applied to the dataset. Second, as we are using point data, we cannot guarantee that the spatial structure of mortality discovered with Local Moran's statistic are not entirely due to the configuration of these specific data points for the period between 2009 and 2016. Therefore, we replicated the analysis (both for the raw and adjusted YPLL models) on ten random subsets, each containing 10,000 observations drawn from the 22,751 deceased. With this method, we could perform enough permutations to apply a Bonferroni correction while ensuring the robustness of the discovered spatial structure. Description of samples, characteristics of the spatial weights, and regression results for the ten subsets can be found in Supplementary Tables S1–S3. Since the spatial

structure of YPLLG was similar across subsets, only the maps of subset 8 are shown in the paper for descriptive purposes. However, the results for the other subsets can be found in the Supplementary Figs. S2 and S3.

For both the raw and adjusted YPLLG models, we summarized the results of the ten random subsets by calculating the range, mean, and standard deviation of population characteristics within each cluster type (i.e., Not significant, High–High, Low–Low, High–Low, Low–High). These population characteristics include the number of individuals within each cluster type, gender, nationality, neighborhood household income and population age, YPLLG value, and individual's YPLL. The YPLL was calculated using a 75-year cut-off<sup>40,41</sup>. We used Tukey's HSD test to compare all the possible pairs of means between each of the Local Moran cluster types to identify significant differences in population characteristics.

Spatial analyses were performed in R using the rgeoda package<sup>42</sup>.

**Ethics approval, consent to participate and consent for publication.** Data used for analysis were publicly available through the *Feuille d'avis officielle* (FAO) until July 2017 (Republic and Canton of Geneva, <https://fao.ge.ch>). The use of these data did not imply a request for authorization from the ethics committee of the canton of Geneva. Indeed, the research protocol used does not study disease, neither the structure nor the functioning of the human body (Federal Act on Research Involving Human Beings, article 2). Moreover, the research protocol respects the Swiss Federal Act on Data Protection (Art. 22) specifying that personal data can be processed for purposes not related to specific persons, and in particular for research, planning and statistics, if: (a) the data is rendered anonymous, as soon as the purpose of the processing permits; (b) the recipient only communicates the data to third parties with the consent of the body that transmitted them, and (c) the results are published in such a manner that the data subjects may not be identified.

## Results

**Profile of the deceased.** Of the 22,751 deceased included in our analysis, 12,125 (53.3%) were women, and 18,101 (79.6%) were Swiss. Individuals lived in neighborhoods with a median household income of 128,012 ± 41,014 CHF per year and a median population age of 42.96 ± 9.70. The YPLLG distribution among individuals was negatively skewed, with a mean value of 5.19 ± 20.12 (min: −93.6, max: 49.9, median: 9.3) (Supplementary Fig. S1). The mean lifespan for the dataset was 79.42 ± 15.34 years.

Median YPLLG was significantly lower for men (8.2) than for women (9.9,  $p < 0.001$ ), and lower for non-Swiss (3.4) than for Swiss (10.6,  $p < 0.001$ ). Men tended to live in younger neighborhoods (i.e., based on the median age of the neighborhood population) than women (men: 40.4 vs. women: 40.8 years,  $p < 0.001$ ). Similar was found for non-Swiss compared with Swiss (non-Swiss: 40.0 vs. Swiss: 40.8 years,  $p < 0.001$ ). Non-Swiss were also located in more deprived neighborhoods than Swiss (non-Swiss: 111,076 vs. Swiss: 115,733 CHF per year,  $p < 0.001$ ). No significant differences were found in neighborhood income between the genders (men: 114,208 vs. women: 115,226 CHF per year,  $p = 0.55$ ).

**Characteristics of individuals within clusters.** To assess whether differences existed between individuals within and outside the clusters discovered by the Local Moran analysis, we compared summary statistics for individuals' gender, nationality, median neighborhood income and age, YPLLG, and YPLL taken from the ten random replications.

For the raw YPLLG model (Table 1), the Local Moran analysis detected a mean of 3502 (35%) individuals showing spatial dependence (i.e., that have a pseudo  $p$ -value  $< 1e-5$ ) for the ten random subsets, including 1,445 individuals belonging to high YPLLG clusters, 779 to low YPLLG clusters, 525 to Low–High spatial outliers, and 751 to High–Low spatial outliers. On average, individuals in high YPLLG clusters lived 19.21 years longer than expected (i.e., according to their Life Expectancy at Birth), individuals in low YPLLG clusters lived 15.06 years less than expected, and individuals showing no spatial dependence lived 4.40 years longer than expected. We observed similar trends when comparing the mean YPLL value between High–High (mean YPLL: 0) and Low–Low clusters (mean YPLL: 11.70). Other population characteristics significantly differed between clusters of low or high YPLLG values and locations showing no spatial dependence. In particular, the median annual neighborhood household income was significantly higher for individuals in high YPLLG clusters (122,870 CHF) than those in low YPLLG clusters (111,146 CHF,  $p < 0.001$ ), as well as the neighborhood population median age (High–High: 54.0, Low–Low: 38.9,  $p < 0.001$ ), the prevalence of women (High–High: 61.5%, Low–Low: 45.7%,  $p < 0.001$ ), and the prevalence of Swiss (High–High: 86.0%, Low–Low: 70.4%,  $p < 0.001$ ). We observed no significant differences in population characteristics between the ten random subsets.

Table 2 shows the summary results for the adjusted YPLLG model, in which we explicitly accounted for the effect of individual-level and neighborhood-level covariates with a median regression model. We observed fewer significant locations (1993; 20%) compared to the raw YPLLG model, but we found similar trends in population characteristics within each cluster type. For the ten random subsets, the analysis detected, on average, 752 individuals belonging to high YPLLG clusters, 375 to low YPLLG clusters, 406 to Low–High spatial outliers, and 457 to High–Low spatial outliers. Individuals in high YPLLG clusters lived, on average, 9.09 years longer than expected, while individuals in low YPLLG clusters lived 23.38 years shorter than expected, and individuals showing no spatial dependence lived 3.76 years shorter than expected. Median annual neighborhood household

		Not significant	High-High	Low-Low	Low-High	High-Low
N		5981–7018	1161–1715	627–883	385–615	599–835
		6497.9 (65.0%)	1445.4 (14.5%)	778.5 (7.8%)	524.9 (5.2%)	751.1 (7.5%)
		± 270.24	± 143.16	± 86.84	± 74.84	± 75.73
Mean YPLL		3.94–4.86	18.74–19.51	– 15.91 to – 14.22	– 11.77 to – 9.90	16.54–17.30
		4.40	19.21	– 15.06	– 10.56	16.95
		± 0.28	± 0.29	± 0.51	± 0.66	± 0.26
Gender	Women	3084–3664	722–1059	295–418	224–355	304–466
		3374.0 (51.9%)	889.3 (61.5%)	355.1 (45.7%)	298.9 (57.1%)	398.0 (52.9%)
		± 159.29	± 90.48	± 39.10	± 39.91	± 47.27
	Men	2897–3354	439–656	332–491	161–275	295–390
		3123.9 (48.1%)	556.1 (38.5%)	423.4 (54.3%)	226.0 (42.9%)	353.1 (47.1%)
		± 121.71	± 56.51	± 52.58	± 38.86	± 31.46
Nationality	Non swiss	1260–1422	174–239	184–269	94–149	124–156
		1343.5 (20.7%)	202.2 (14.0%)	230.5 (29.6%)	126.1 (24.1%)	139.2 (18.6%)
		± 48.01	± 18.78	± 30.28	± 17.51	± 10.06
	Swiss	4721–5619	987–1476	443–632	291–476	475–702
		5154.4 (79.3%)	1243.2 (86.0%)	548.0 (70.4%)	398.8 (75.9%)	611.9 (81.4%)
		± 229.96	± 125.47	± 58.42	± 59.53	± 68.64
Mean YPLL		4.27–4.64	0.00–0.00	11.01–12.32	7.60–9.22	0.00–0.00
		4.43	0.00	11.70	8.18	0.00
		± 0.13	± 0.00	± 0.38	± 0.56	± 0.00
Mean neighborhood household income		131,050–133,722	118,996–126,142	107,923–113,317	128,348–137,893	112,682–119,932
		132,267.2	122,870.5	111,145.6	133,403.8	116,913.6
		± 833.92	± 2677.39	± 1851.32	± 3121.58	± 2445.51
Mean neighborhood population age		40.81–41.42	52.97–55.70	38.70–39.01	45.72–48.06	39.03–39.43
		41.17	53.99	38.87	46.56	39.16
		± 0.22	± 0.78	± 0.11	± 0.75	± 0.12

**Table 1.** Characteristics of individuals within each Local Moran cluster type for the raw YPLL model. For each variable, the table shows range, mean, and standard deviation, calculated across the ten subsets. The totals per category may not be equal to 10,000 due to neighborless individuals who were excluded from the Local Moran's analysis.

income was significantly lower for individuals in low YPLL clusters (111,818 CHF) than in high YPLL clusters (131,410 CHF,  $p < 0.001$ ). Spatial clusters consisting of individuals living shorter than expected (i.e., low YPLL clusters) were also found in younger neighborhoods (Low-Low: 38.97, High-High: 46.68,  $p < 0.001$ ). Significant differences between low and high clusters were also observed for gender (58.7% of women in High-High clusters, and 46.5% of women in Low-Low clusters,  $p < 0.001$ ), and nationality (83.6% of Swiss in High-High clusters, and 76.1% of Swiss in Low-Low clusters,  $p < 0.001$ ).

**Geographic footprint of YPLL.** We identified the type of spatial association undergoing at each location by comparing the YPLL value of an individual with the mean YPLL value of its neighbors, illustrated on the Local Moran cluster maps for the raw and the adjusted YPLL models (Figs. 1A, 2A). On the maps, white dots represent locations showing no spatial dependence, dark green dots (High-High) and dark purple dots (Low-Low) represent individuals with high YPLL, respectively low YPLL, surrounded by individuals with similar values. Light purple dots (Low-High) and light green dots (High-Low) represent discordant behaviors (spatial outliers), where an individual's YPLL value differs considerably from the mean YPLL value of its neighbors.

Because similar geographic footprints were obtained for the ten subsets, in this section, we only refer to the Local Moran cluster map from subset 8. Results from the other subsets are available in the Supplementary Materials (Figs. S2, S3).

Analysis revealed a clear spatial structure for the raw YPLL model in the state of Geneva (Fig. 1A). Low-Low clusters were mostly concentrated downtown and in the west areas (landmarks #1–#3, #5, #8), while High-High clusters were located in the south and north areas (landmarks #4, #6, #7). The distribution of YPLL values within each cluster type (Fig. 1B) showed significant differences between clusters of low and high YPLL values (mean difference of 34.93 years,  $p < 0.001$ ).



		Not significant	High–High	Low–Low	Low–High	High–Low
N		7271–8866	466–1097	179–500	224–550	263–579
		8007.1 (80.1%)	752.3 (7.5%)	375.3 (3.8%)	405.8 (4.1%)	457.3 (4.6%)
		± 431.16	± 172.92	± 101.42	± 94.11	± 106.13
Mean adjusted YPLL		– 4.10 to – 3.47	8.46–9.74	– 24.32 to – 21.70	– 20.84 to – 18.71	8.58–9.55
		– 3.76	9.09	– 23.38	– 19.75	9.03
		± 0.22	± 0.39	± 0.90	± 0.76	± 0.36
Gender	Women	3824–4692	279–664	88–237	134–301	140–301
		4226.9 (52.8%)	443.3 (58.7%)	173.5 (46.5%)	229.3 (56.9%)	242.3 (53.1%)
		± 232.73	± 113.66	± 46.01	± 49.00	± 54.87
	Men	3447–4174	187–433	91–263	90–249	123–291
		3780.2 (47.2%)	309.0 (41.3%)	201.8 (53.5%)	176.5 (43.1%)	215.0 (46.9%)
		± 209.30	± 62.83	± 57.28	± 47.78	± 52.78
Nationality	Non swiss	1528–1788	80–169	41–121	42–98	60–143
		1645.5 (20.6%)	122.0 (16.4%)	90.3 (23.9%)	73.6 (18.1%)	110.1 (24.1%)
		± 79.34	± 23.84	± 26.72	± 18.30	± 26.39
	Swiss	5743–7078	386–928	138–389	182–452	203–440
		6361.6 (79.4%)	630.3 (83.6%)	285.0 (76.1%)	332.2 (81.9%)	347.2 (75.9%)
		± 355.62	± 150.50	± 76.78	± 76.84	± 81.00
Mean YPLL		4.02–4.41	0.00–0.02	11.56–13.40	7.83–9.72	0.02–0.05
		4.23	0.01	12.59	8.65	0.04
		± 0.13	± 0.00	± 0.61	± 0.62	± 0.01
Mean neighborhood household income		127,443–130,671	119,384–148,294	108,454–116,478	123,051–141,971	118,187–126,765
		129,059.2	131,410.1	111,817.5	131,932.4	120,792.6
		± 1151.89	± 9419.46	± 2222.63	± 6406.92	± 2495.34
Mean neighborhood population age		42.51–43.38	44.19–48.86	38.56–39.40	42.47–44.43	39.61–40.71
		42.93	46.68	38.97	43.59	39.93
		± 0.28	± 1.68	± 0.22	± 0.68	± 0.38

**Table 2.** Characteristics of individuals within each Local Moran cluster type for the adjusted YPLL model. For each variable, the table shows range, mean, and standard deviation, calculated across the ten subsets. The totals per category may not be equal to 10,000 due to neighborless individuals who were excluded from the Local Moran's analysis.

After controlling for nationality, neighborhood income, and median neighborhood age, we obtained a similar geographic footprint of YPLL values (Fig. 2A), with a moderate reduction in the size of low (43%) and high (39%) YPLL cluster footprints. The adjustment thinned down the low YPLL clusters in landmarks #2, #3, and #5, while the Low–Low cluster in landmark #3 has disappeared. For high YPLL clusters, the reduction mainly affects clusters located in landmarks #4 and #7. The difference in YPLL values between individuals in clusters of low and high YPLL (Fig. 2B) was also considerable (mean YPLL difference of 32.93 years,  $p < 0.001$ ).

## Discussion

In using spatial cluster detection methods on individual georeferenced mortality data, our analyses revealed a clear spatial structure of YPLL across the state of Geneva. We found that individuals living shorter than expected (i.e., with low YPLL values) were clustered in downtown and western areas (landmarks #1–#3, #5, #8 in Figs. 1A, 2A), while individuals living longer than expected (i.e., with high YPLL values) were clustered in the southern and in the northern areas of the state (landmarks #4, #6, #7 in Figs. 1A, 2A). In addition, we observed demographic and socioeconomic differences between low YPLL clusters (Low–Low) and high YPLL clusters (High–High), where the latter population was characterized by a higher proportion of non-Swiss and men who were living in more poorer and younger neighborhoods. Adjusting for individual-level and neighborhood-level covariates moderately mitigated the size of YPLL clusters (43% reduction for Low–Low clusters and 39% reduction for High–High clusters). While this thinned down most of the spatial clusters detected in Fig. 1A, the geographic footprint of YPLL remained, suggesting that population demographics and socioeconomic status might not fully explain patterns in mortality across the state of Geneva.

The significant association between YPLL and neighborhood median income is consistent with results from other studies, where a relationship between mortality and the socioeconomic position in small geographic areas

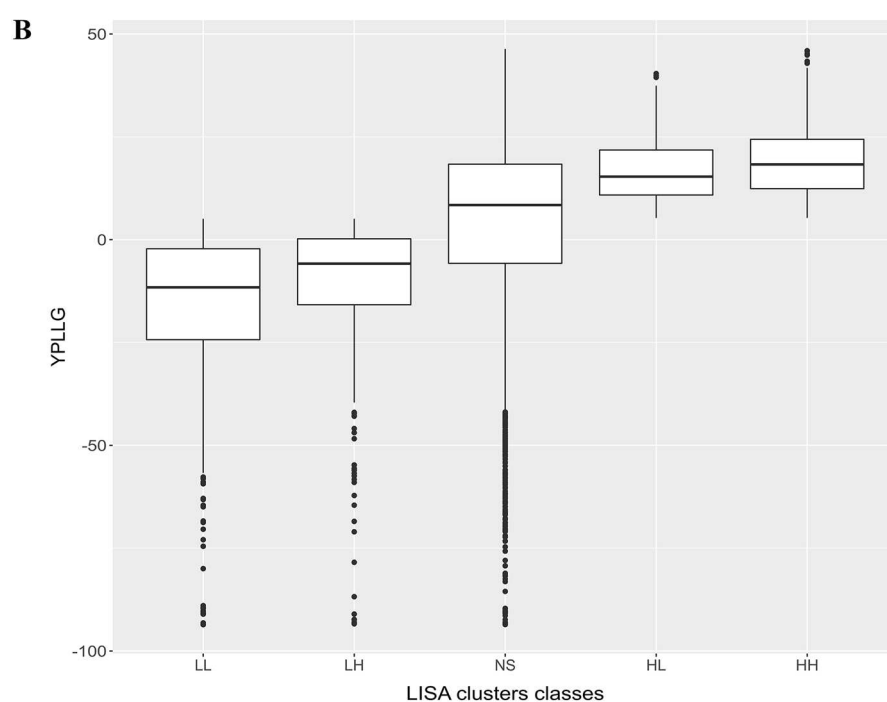
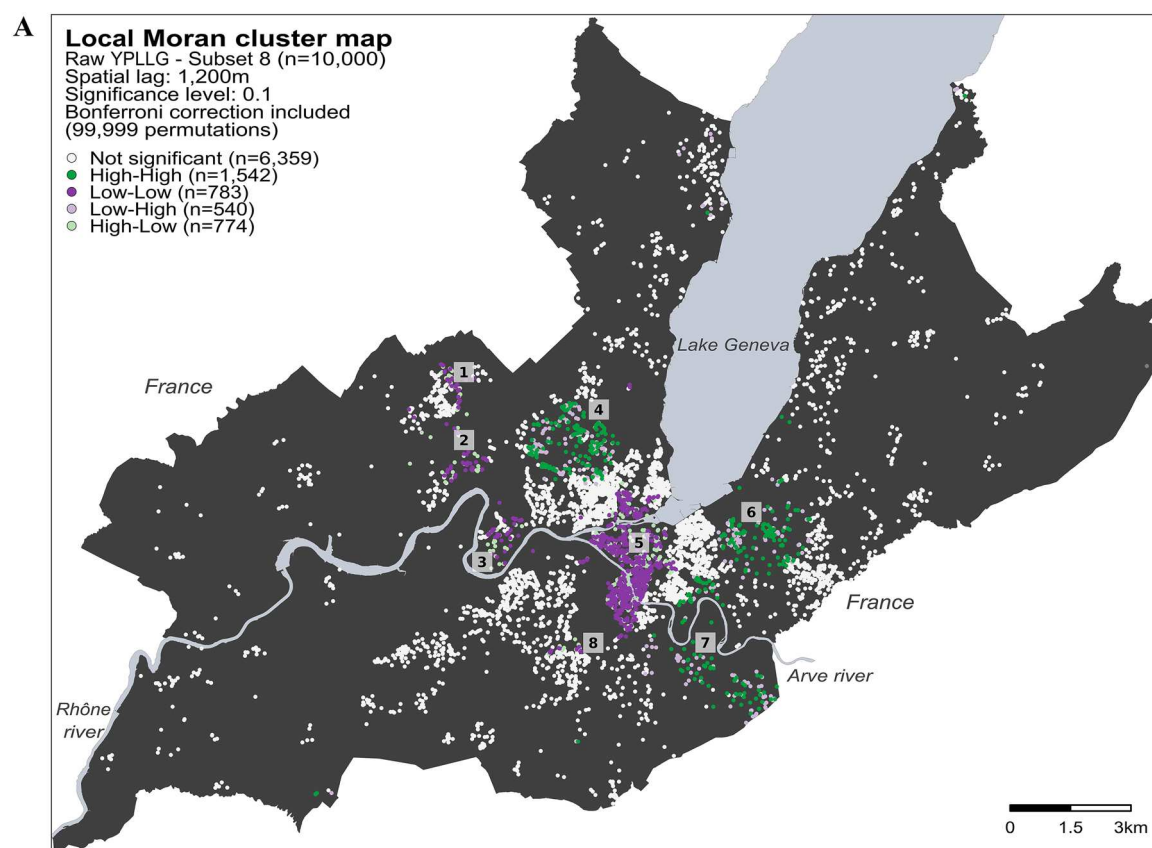
are also observed<sup>24–26,43</sup>. These studies suggest interconnected pathways between social and health inequalities. Unlike Buajitti et al.<sup>23</sup>, we did not find a higher proportion of foreigners in areas of lower premature mortality, which may contradict the immigrant health advantage described elsewhere<sup>44</sup>. However, the literature also states that this advantage is mitigated over time<sup>45</sup>. Therefore, indicators using a general age cutoff (e.g., 75 years) may not reflect the entire lifespan of individuals and overestimate the immigrant health advantage. Furthermore, our analysis was conducted on individual-level data rather than at an aggregated scale which might exacerbate these differences. The higher proportion of foreigners found in low YPLL clusters could partially be explained by the fact that we assigned to individuals the Life Expectancy at Birth from Swiss life tables, which is among the highest in the world<sup>12</sup>. Hence, it could overestimate the LEB of non-Swiss people and, thus, the absolute value of YPLL. It may also indicate unequal situations between Swiss nationals and migrants as foreigners face worse living conditions and quality of life in Switzerland<sup>46</sup>. This is consistent with other studies highlighting associations between deprived life conditions and higher mortality rate<sup>3,47–49</sup>. Further epidemiological studies are necessary to disentangle the underlying factors leading to the geographic footprints of life expectancy inequalities discovered in this study.

Interestingly, we identified a few similarities between the spatial patterns of health inequalities revealed by YPLL and those detected in another study assessing the spatial dependence of body mass index (BMI) in Geneva<sup>36</sup>. Indeed, some clusters of elevated BMI overlap clusters of low YPLL and conversely (data not shown). Thus, we may presume that both outcomes are spatially interlinked and that some of these premature deaths may be related to conditions associated with a high BMI, such as cardiovascular disease and diabetes.

When comparing YPLL and YPLL values among cluster types, we observed similar trends between both indicators, showing that they measure comparable mortality inequalities. For instance, High–High clusters present a YPLL value of 0, while Low–Low clusters have a positive YPLL mean value of 12 in both raw and adjusted models. However, due to the nature of YPLL that only identifies subjects that faced premature death, YPLL may constitute a more appropriate indicator to measure health inequalities.

**Strengths.** As far as we know, this is the first study that analyzes spatial dependence of a life expectancy indicator using a large sample size of individual-level mortality data ( $n = 22,751$ ), making it possible to identify small areas inequalities in health. The fact that no significant differences were detected in the geographic footprint of YPLL between the ten random subsets demonstrates that our results are not specific to our dataset, and that pattern of life expectancy inequalities are deeply embedded in the territory of the Geneva state.

**Limitations.** Several important variables were not available in the original raw dataset, including socioeconomic status, cause of death, and prevalence of comorbidities. In addition, we removed 1611 individuals (5.8% of the original dataset) for whom gender could not be assigned using the name-to-gender inference, which could





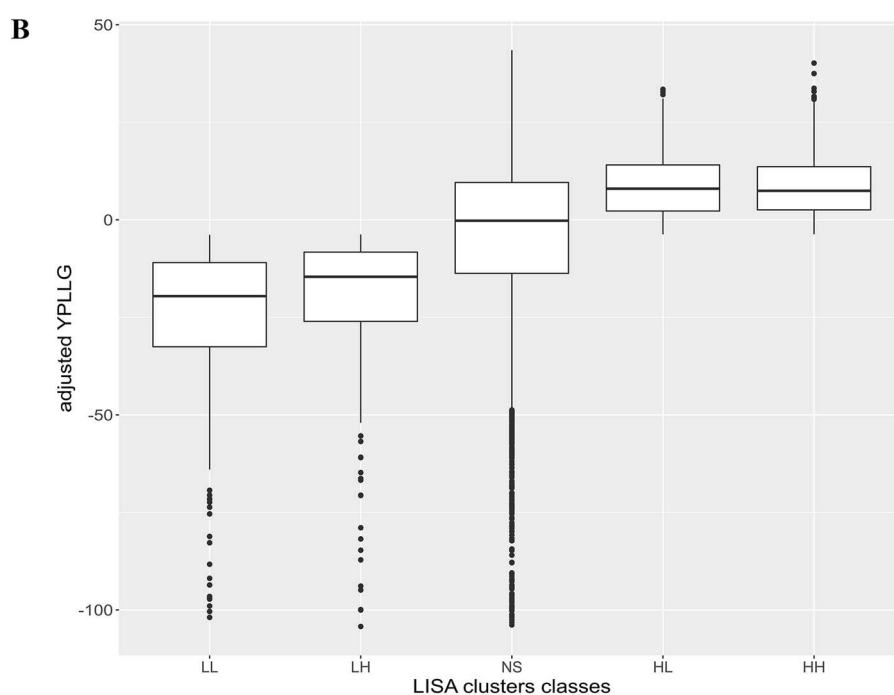
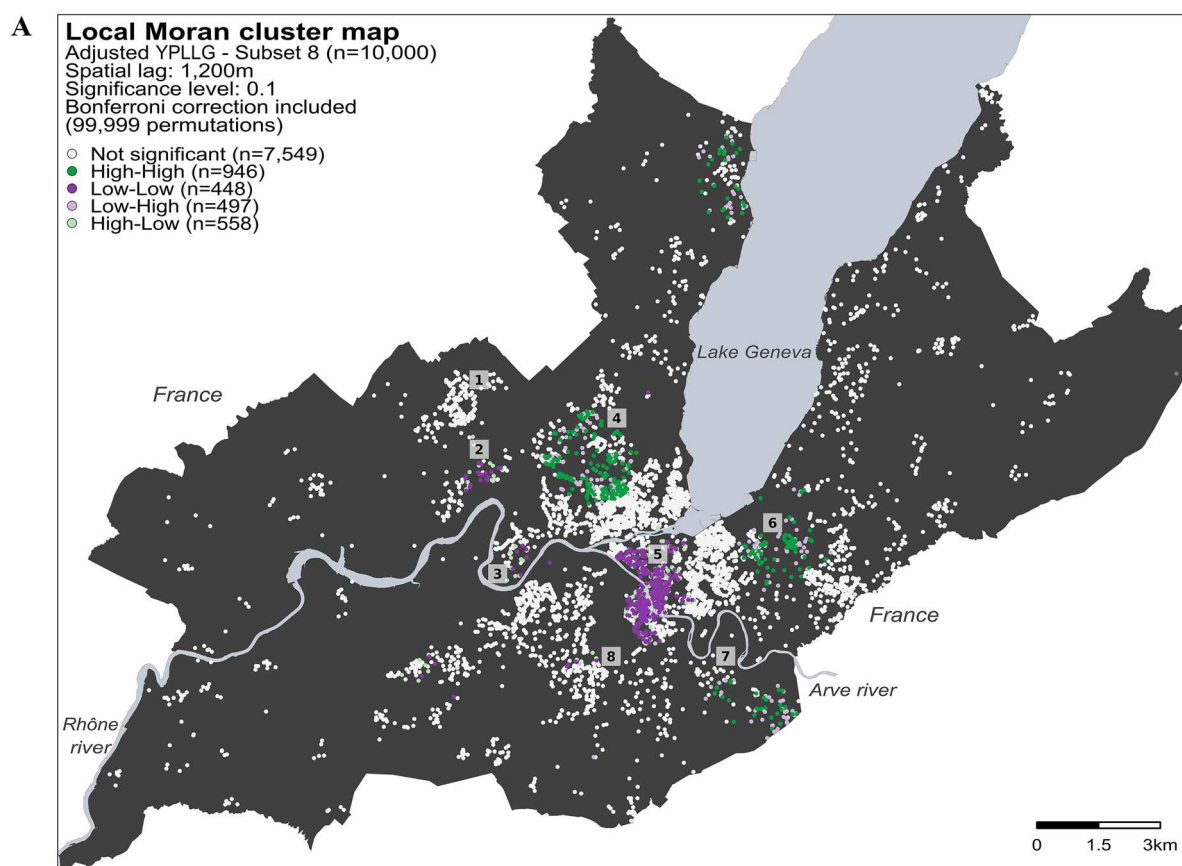
**Figure 1.** Spatial footprint of YPLL for the raw model. **(A)** Local Moran cluster map was calculated for a random subset of 10,000 individuals taken from the 22,751 individuals in the state of Geneva for the 2009–2016 period (Subset 8). White dots represent individuals without spatial dependence (i.e., those whose Local Moran's  $I$  p-values adjusted with Bonferroni are not significant). Dark green dots (High–High cluster) represent individuals with high YPLL values (i.e., those that lived longer than expected) surrounded by individuals with similar YPLL values within a distance of 1200 m. Dark purple dots (Low–Low cluster) represent individuals with low YPLL values (i.e., those that lived shorter than expected) surrounded by individuals with similar YPLL values. Light purple dots (Low–High spatial outliers) represent individuals with high YPLL values surrounded by individuals with low YPLL values, and light green dots (High–Low spatial outliers) represent individuals with low YPLL values surrounded by individuals with high YPLL values. Indicative landmarks are shown on the map to facilitate the interpretation of the results (#1–#8). [Source (administrative boundaries): <https://www.swisstopo.admin.ch/>, 2020; the map was produced using R, package ggplot version 3.3.5.]. **(B)** Distribution of YPLL values for each Local Moran cluster type (*LL* Low–Low, *LH* Low–High, *NS* not significant, *HL* High–Low, *HH* High–High).

lead to underestimating the proportion of foreigners in the final dataset<sup>30</sup>. However, we did not notice any trend in the missing cases when we compared the number of deaths included in our study (stratified by gender and nationality) with those published by the Cantonal Statistical Office at the municipality level (Supplementary Table S4). Additional limitations may also originate from the fact that only married couples could be included in the calculation of household income, and we only included the last place of residence of individuals, which may not represent where they spent most of their lives. Furthermore, our results may only be representative of the period included in the analysis (2009–2016).

**Policy implications.** Switzerland, and in particular the state of Geneva, has one of the highest LEB and quality of life worldwide. However, this does not prevent the state from presenting considerable health inequalities. Contrasts in socioeconomic conditions like nationality and neighborhood income result in profound geographic disparities and reveal deprived living environment areas. The same areas were by the way recently shown to be exposed to SARS-CoV-2 clusters that persisted significantly longer than elsewhere in the state<sup>50</sup>. However, one can also consider the existence of such spatial structures as opportunities for intervention. The present findings should encourage authorities to acknowledge geographic areas facing health inequalities and to favor, in these zones, the development of adequate public health policies to create conditions of more equitable living environments. Such policies should consider the social component rather than focusing exclusively on treating risk factors<sup>6,51</sup>. For instance, decision-makers could: (i) favor the development of urban districts socioeconomically mixed; (ii) improve living conditions of neighborhoods with a high rate of foreigners; (iii) favor access to health in local areas where exist higher prevalence of chronic diseases or; (iv) allocating economic assistance in elderly.

## Conclusion

Individual-based spatial patterns of life expectancy translate health inequalities footprints on a territory. Our study revealed specific spatial patterns of YPLL in the state of Geneva using spatial cluster detection methods on individual georeferenced mortality data. The proportion of women, the proportion of Swiss, and the median neighborhood income were significantly lower for populations in low YPLL clusters than for populations in high YPLL clusters. Adjustment for nationality and neighborhood income slightly reduced the footprint of low YPLL clusters but did not modify the population characteristics within clusters. Results highlight the worth of precision public health relying on spatial methods to assess health inequalities at a local level and target vulnerable populations.



**Figure 2.** Spatial footprint of YPLL for the adjusted model. **(A)** Local Moran cluster map was calculated for a random subset of 10,000 individuals taken from the 22,751 individuals in the state of Geneva for the 2009–2016 period (Subset 8). White dots represent individuals without spatial dependence (i.e., those whose Local Moran's I p-values adjusted with Bonferroni are not significant). Dark green dots (High–High cluster) represent individuals with high adjusted YPLL values (i.e., those that lived longer than expected) surrounded by individuals with similar values within a distance of 1200 m. Dark purple dots (Low–Low cluster) represent individuals with low adjusted YPLL values (i.e., those that lived shorter than expected) surrounded by individuals with similar values. Light purple dots (Low–High spatial outliers) represent individuals with high adjusted YPLL values surrounded by individuals with low adjusted YPLL values, and light green dots (High–Low spatial outliers) represent individuals with low adjusted YPLL values surrounded by individuals with high adjusted YPLL values. Indicative landmarks are shown on the map to facilitate the interpretation of the results (#1–#8). [Source (administrative boundaries): <https://www.swisstopo.admin.ch/>, 2020; the map was produced using R, package ggplot version 3.3.5.]. **(B)** Distribution of adjusted YPLL values for each Local Moran cluster type (LL Low–Low, LH Low–High, NS not significant, HL High–Low, HH High–High).

## Data availability

Datasets used in the current study are available at <https://zenodo.org/badge/latestdoi/418981695>.

## Code availability

All code for data cleaning and analysis associated with the current submission is available at <https://zenodo.org/badge/latestdoi/418981695>.

Received: 31 May 2021; Accepted: 22 November 2021

Published online: 02 December 2021

## References

- Murray, C. J., Salomon, J. A. & Mathers, C. A critical examination of summary measures of population health. *Bull. World Health Organ.* **78**, 981–994 (2000).
- European Public Health Alliance (EPHA). *Health Inequalities a Public Health Challenge for European Policy Makers* (European Public Health Alliance, 2019).
- Marmot, M. Health equity in England: The Marmot review 10 years on. *BMJ* **368**, m693 (2020).
- Lewer, D. *et al.* Premature mortality attributable to socioeconomic inequality in England between 2003 and 2018: An observational study. *Lancet Public Health* **5**, e33–e41 (2020).
- Marmot, M. Status syndrome. *Significance* **1**, 150–154 (2004).
- Stringhini, S. *et al.* Socioeconomic status and the 25 × 25 risk factors as determinants of premature mortality: A multicohort study and meta-analysis of 1.7 million men and women. *The Lancet* **389**, 1229–1237 (2017).
- Vineis, P. *et al.* Health inequalities: Embodied evidence across biological layers. *Soc. Sci. Med.* **246**, 112781 (2020).
- Etches, V., Frank, J., Ruggiero, E. D. & Manuel, D. Measuring population health: A review of indicators. *Annu. Rev. Public Health* **27**, 29–55 (2006).
- Holly, A. & Benkassmi, M. *Health and Health Care Inequalities in Switzerland* 39 (United Nations Research Institute for Social Development, 2003).
- Turnock, B. J. *Public Health: What It Is and How It Works* 5th edn. (Jones & Bartlett Learning, 2012).
- Dempsey, M. Decline in tuberculosis: The death rate fails to tell the entire story. *Am. Rev. Tuberc.* **56**, 157–164 (1947).
- WHO. *World Health Statistics 2019: Monitoring Health for the SDGs, Sustainable Development Goals* (World Health Organization, 2019).
- Moser, A. *et al.* What does your neighbourhood say about you? A study of life expectancy in 1.3 million Swiss neighbourhoods. *J. Epidemiol. Community Health* **68**, 1125–1132 (2014).
- Panczak, R. *et al.* A Swiss neighbourhood index of socioeconomic position: Development and association with mortality. *J. Epidemiol. Community Health* **66**, 1129–1136 (2012).
- Diez Roux, A. V. Investigating neighborhood and area effects on health. *Am. J. Public Health* **91**, 1783–1789 (2001).
- Diez Roux, A. V. Neighborhoods and health: What do we know? What should we do? *Am. J. Public Health* **106**, 430–431 (2016).
- Ribeiro, A. Public health: Why study neighborhoods? *Porto Biomed. J.* **3**, e16 (2018).
- Fei, L. & Chenghu, Z. Spatial autocorrelation analysis on regional economic disparity of northeast economic region in China. *Chin. J. Popul. Resour. Environ.* **7**, 27–31 (2009).
- Jourkavitch, J., Burgert-Brucker, C., Assaf, S. & Delgado, S. Using geographical analysis to identify child health inequality in sub-Saharan Africa. *PLoS ONE* **13**, e0201870 (2018).
- Cheng, E. R. Disparities in premature mortality between high- and low-income US counties. *Prev. Chronic Dis.* <https://doi.org/10.5888/pcd9.110120> (2012).
- Plümper, T., Laroze, D. & Neumayer, E. The limits to equivalent living conditions: Regional disparities in premature mortality in Germany. *Z. Gesundh.* **26**, 309–319 (2018).
- Plümper, T., Laroze, D. & Neumayer, E. Regional inequalities in premature mortality in Great Britain. *PLoS ONE* **13**, e0193488 (2018).
- Buajitti, E. *et al.* Regional variation of premature mortality in Ontario, Canada: A spatial analysis. *Popul. Health Metr.* **17**, 9 (2019).
- Awuor, L. & Melles, S. The influence of environmental and health indicators on premature mortality: An empirical analysis of the City of Toronto's 140 neighborhoods. *Health Place* **58**, 102155 (2019).
- Kihal-Talantikite, W., Legendre, P., Le Nouveau, P. & Deguen, S. Premature adult death and equity impact of a reduction of NO<sub>2</sub>, PM<sub>10</sub>, and PM<sub>2.5</sub> levels in Paris—A health impact assessment study conducted at the census block level. *Int. J. Environ. Res. Public Health* **16**, 38 (2018).
- Livingston, M., Whyte, B., Walsh, D. & Bailey, N. *Investigating the Impact of the Spatial Distribution of Deprivation on Health Outcomes* (Glasgow Centre for Population Health, 2013).
- Fotheringham, A. S. & Wong, D. W. S. The modifiable areal unit problem in multivariate statistical analysis. *Environ. Plan. Econ. Space* **23**, 1025–1044 (1991).
- Santamaría, L. & Mihaljević, H. Comparison and benchmark of name-to-gender inference services. *PeerJ Comput. Sci.* **4**, e156 (2018).

29. Nguyen, E., Robinson, R. & Hoover, R. Changes in gender representation in pharmacy research literature. *J. Health Disparities Res. Pract.* **12**, 18 (2019).
30. Qureshi, R., Lê, J., Li, T., Ibrahim, M. & Dickersin, K. Gender and editorial authorship in high-impact epidemiology journals. *Am. J. Epidemiol.* **188**, 2140 (2019).
31. Guillot, M. Period versus cohort life expectancy. In *International Handbook of Adult Mortality* (eds Rogers, R. G. & Crimmins, E. M.) 533–549 (Springer, 2011).
32. Menthonnex, J. *La Mortalité par Génération en Suisse: Évolution 1900–2150 et Tables par Génération 1900–2030: Rapport Technique* (Federal Statistical Office, 2009).
33. Office cantonal de la statistique (OCSTAT). Le découpage du canton de Genève en sous-secteurs statistiques. In *Révision 2005*, 44 (2006).
34. Anselin, L. Local indicators of spatial association-LISA. *Geogr. Anal.* **27**, 93–115 (1995).
35. Guessous, I. *et al.* A comparison of the spatial dependence of body mass index among adults and children in a Swiss general population. *Nutr. Diabetes* **4**, e111 (2014).
36. Joost, S. *et al.* Overlapping spatial clusters of sugar-sweetened beverage intake and body mass index in Geneva state, Switzerland. *Nutr. Diabetes* **9**, 35 (2019).
37. Ernst, M. D. Permutation methods: A basis for exact inference. *Stat. Sci.* **19**, 676–685 (2004).
38. Getis, A. & Ord, J. K. Local spatial statistics: An overview. *Spat. Anal. Model GIS Environ.* **374**, 261–277 (1996).
39. McGreevy, K. M., Lipsitz, S. R., Linder, A., Rimm, E. & Hoel, D. G. Using median regression to obtain adjusted estimates of central tendency for skewed laboratory and epidemiologic data. *Clin. Chem.* **55**, 165 (2008).
40. Beer, V., Bisig, B. & Gutzwiller, F. Social class gradients in years of potential life lost in Switzerland. *Soc. Sci. Med.* **37**, 1011–1018 (1993).
41. Statistics Canada. *Health Indicators. Report No.: 82-221-X*, 1, 28–31 (2008).
42. Li, X. *rgeoda: R Library for Spatial Data Analysis* (2020) (Accessed 20 October 2021).
43. Borrell, C., Mari-Delolmo, M., Serral, G., Martínez-Beneito, M. & Gotsens, M. Inequalities in mortality in small areas of eleven Spanish cities (the multicenter MEDEA project). *Health Place* **16**, 703–711 (2010).
44. Aldridge, R. W. *et al.* Global patterns of mortality in international migrants: A systematic review and meta-analysis. *The Lancet* **392**, 2553–2566 (2018).
45. McDonald, J. T. & Kennedy, S. Insights into the ‘healthy immigrant effect’: Health status and health service use of immigrants to Canada. *Soc. Sci. Med.* **59**, 1613–1627 (2004).
46. FSO. *Comment se Porte la Population Issue de la Migration en Suisse?* (Federal Statistical Office, 2019).
47. Ford, M. M. & Highfield, L. D. Exploring the spatial association between social deprivation and cardiovascular disease mortality at the neighborhood level. *PLoS ONE* **11**, e0146085 (2016).
48. Maguire, D. & Buck, D. *Inequalities in Life Expectancy* 76 (King's Fund, 2015).
49. Marmot, M. Closing the health gap. *Scand. J. Public Health* **45**, 723 (2017).
50. De Ridder, D. *et al.* Socioeconomically disadvantaged neighborhoods face increased persistence of SARS-CoV-2 clusters. *Front. Public Health* **8**, 626090 (2021).
51. Gostin, L. O. & Friedman, E. A. Health inequalities. *Hast. Cent. Rep.* **50**, 6–8 (2020).

## Author contributions

A.L., I.G. and S.J. conceived and designed the analysis. H.D.C. collected the data. A.L. and J.V. performed the analysis. A.L., J.V., I.G. and S.J. wrote the first version of the manuscript. D.D.R., J.S. and S.S. contributed to its improvement and in the writing of the final version. I.G. and S.J. supervised the project.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-02733-x>.

**Correspondence** and requests for materials should be addressed to S.J.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

## **A.7 Code availability**

This section provides the GitHub repositories containing the source code used for the analyses and applications discussed in each chapter of the thesis. Additionally, a repository for general functions applicable across multiple case studies is included.

### **General Functions**

- <https://github.com/aladoy/GIRAPH-functions>

### **Chapter 2: Supporting health promotion initiatives with geographic information systems**

- General analysis: <https://github.com/aladoy/commune-en-sante>
- Web application: <https://github.com/aladoy/GEOSAN>

### **Chapter 3: Spatial patterns and factors associated with cardiovascular risk factors**

- <https://github.com/aladoy/GEOSAN-geochronic>

### **Chapter 4: Detecting daily spatio-temporal COVID-19 case clusters**

- <https://github.com/aladoy/GEOCOVID-phase1>

### **Chapter 5: Combining spatio-temporal and genomic analyses to improve COVID-19 surveillance system**

- <https://github.com/aladoy/GEOCOVID-genomics>

### **Chapter 6: Prioritization of populations for mobile COVID-19 vaccination campaigns**

- <https://github.com/aladoy/GEOSAN-vaccination>



# CURRICULUM VITAE

## Publications

### SUBMITTED MANUSCRIPTS

Lengen, G., Hugli, O., De Ridder, D., Guessous, I., Ladoy, A., & Joost, S. (2023). Spatial dependence of non-traumatic out-of-hospital cardiac arrest in a Swiss region: A retrospective analysis. *Under review at Resuscitation*.

Ladoy, A., Marques-Vidal, P., Guessous, I., & Joost, S. (2023). Identifying hot spots of cardiometabolic risk factors in a Swiss city: Impact of individual and environmental factors. *Preprint Available in: <https://doi.org/10.21203/rs.3.rs-3359714/v1>*.

De Ridder, D., Ladoy, A., Choi, Y., Jacot, D., Vuilleumier, S., Guessous, I., Joost, S., & Greub, G. (2023). Environmental and geographical factors Influencing the Spread of SARS-CoV-2 Over Two Years: A Fine-Scale Spatiotemporal Analysis. *Under review at Frontiers in Public Health*.

Joost, S., Ladoy, A., Kathari, G., Renevey, R., Goldberg, M., Zins, M., Matta, J., & Czernichow, S. (2023). Spatio-temporal evolution of Body Mass Index and Obesity in France 1990-2015: The GAZEL cohort. *Under review at International Journal of Health Geographics*.

### PEER-REVIEWED JOURNAL PUBLICATIONS

Choi, Y., Ladoy, A., De Ridder, D., Jacot, D., Vuilleumier, S., Bertelli, C., Guessous, I., Pillonel, T., Joost, S., & Greub, G. (2022). Detection of SARS-CoV-2 infection clusters: The useful combination of spatiotemporal clustering and genomic analyses. *Frontiers in Public Health*, 10. <https://www.frontiersin.org/articles/10.3389/fpubh.2022.1016169>

Ladoy, A., Vallarta-Robledo, J. R., De Ridder, D., Sandoval, J. L., Stringhini, S., Da Costa, H., Guessous, I., & Joost, S. (2021). Geographic footprints of life expectancy inequalities in the state of Geneva, Switzerland. *Scientific Reports*, 11(1), 23326. <https://doi.org/10.1038/s41598-021-02733-x>

Vallarta-Robledo, J. R., Sandoval, J. L., De Ridder, D., Ladoy, A., Marques-Vidal, P., Humair, J.-P., Cornuz, J., Probst-Hensch, N., Schaffner, E., Stringhini, S., Joost, S., & Guessous, I. (2021). Spatial clusters of daily tobacco consumption before and after a smoke-free policy implementation. *Health & Place*, 70, 102616. <https://doi.org/10.1016/j.healthplace.2021.102616>

Ladoy, A., Opota, O., Carron, P.-N., Guessous, I., Vuilleumier, S., Joost, S., & Greub, G. (2021). Size and duration of COVID-19 clusters go along with a high SARS-CoV-2 viral load: A spatio-temporal investigation in Vaud state, Switzerland. *Science of The Total Environment*, 787. <https://doi.org/10.1016/j.scitotenv.2021.147483>

## **Presentations**

MAY 15, 2023 - BrownBag Seminar

*Organizers:* Institute of geography and sustainability (IGD), University of Lausanne (UNIL)

*Location:* University of Lausanne (UNIL), Lausanne, Switzerland

DECEMBER 15, 2022 - Workshop Health & Place

*Organizers:* Laboratory of Geographic Information Systems (LASIG), École Polytechnique Fédérale de Lausanne (EPFL) | Laboratory of Geographic Information Visualization and Analysis, University of Zurich (UZH)

*Location:* École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

JUNE 16, 2022 - Green Days of the Environmental Engineering Institute

*Organizers:* Environmental Engineering Institute (IEE), École Polytechnique Fédérale de Lausanne (EPFL)

*Location:* Hotel Zodiaque, Anzère, Switzerland

FEBRUARY 19, 2020 - Workshop HABITAT

*Organizers:* Habitat Research Center (HRC), École Polytechnique Fédérale de Lausanne (EPFL)

*Location:* École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

NOVEMBER 9, 2019 - 50ème journée des Alumni

*Organizers:* École Polytechnique Fédérale de Lausanne (EPFL)

*Location:* École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

MAY 23, 2019 - Workshop "Utilisation des données géographiques en santé publique"

*Organizers:* Direction Générale de la Santé Vaud (DGS) | Geographic Information Research and Analysis in Public Health (GIRAPH)

*Location:* Swiss Graduate School of Public Administration, University of Lausanne (UNIL), Lausanne, Switzerland

# ANAÏS LADOY

@ anais.ladoy@epfl.ch    +41 78 842 64 04    Avenue de Sévery 8, 1004 Lausanne  
https://github.com/aladoy    in https://www.linkedin.com/in/aladoy    ORCID 0000-0001-7036-4617



## EDUCATION

### PhD in spatial epidemiology

École Polytechnique Fédérale de Lausanne

2019 - 2023    Lausanne, CH

Doctoral thesis carried out in the GEOME group (LGB, EPFL) under the supervision of Stéphane Joost and Idris Guessous (SMPR, HUG). The thesis is about the *role of spatial epidemiology to support public health and prevention policies*, through case studies in the canton of Vaud (GEOSAN project).

### MSc in Environmental Sciences and Engineering

École Polytechnique Fédérale de Lausanne

2016 - 2018    Lausanne, CH

Specialization in Monitoring and Modeling of the Environment.

### BSc in Environmental Sciences and Engineering

École Polytechnique Fédérale de Lausanne

2012 - 2016    Lausanne, CH

## EXPERIENCE

### Teaching assistant

Laboratoire des Systèmes d'Information Géographique (LASIG), EPFL

Jan 2019 - Jun 2022    Lausanne, CH

Geographic Information Systems (GIS) course for EPFL undergraduates. Teaching QGIS software and Python programming for geospatial and mapping purposes.

### Web development internship

The Fair Traveller

Jul 2017 - Sept 2017    Lausanne, CH

Maintain the database and contribute to site development.

### Land-use planning internship

Communauté de Communes du Genevois

Oct 2015 - Feb 2016    Archamps, FR

Assessment of the energy and land use targets foreseen in the Schéma de Cohérence Territorial (SCOT).

## SKILLS

Programming    Python    R    SQL    Bash

Geospatial    PostgreSQL / PostGIS    QGIS    GeoDa    SaTScan

GDAL    GeoPandas / PySAL    Leaflet / Folium

Others    Office    LaTeX    git    Inkscape    Linux

## LANGUAGES

French    Mother tongue  
English    Fluent in written and spoken  
Spanish    Basic knowledge

## RESEARCH

### GEOCHRONIC project

EPFL

Jul 2022 - Aug 2023    Lausanne, CH

Stéphane Joost, Idris Guessous

Investigate the influence of physical and social environments on the spatial distribution of cardiometabolic risk factors in the Lausanne population.

### Vaud Mobile Vaccination Campaign

Protection Civile, DGS, EPFL

Apr 2021 - Aug 2021    Lausanne, CH

Louis-Henri Delarageaz, Jérôme Mouton, Stéphane Joost

Prioritize the location of vaccination sites using geospatial approaches.

### GEOCOVID project

CHUV, La Source, EPFL

Sept 2020 - Nov 2022    Lausanne, CH

Gilbert Greub, Séverine Vuilleumier, Stéphane Joost

Detection of geographical clusters of COVID-19 in the canton of Vaud during the first wave of the epidemic. Genomic analysis of the clusters and recommendations for improving surveillance systems.

Choi et al. (2022) Detection of SARS-CoV-2 infection clusters: The useful combination of spatiotemporal clustering and genomic analyses. *Frontiers in Public Health*.

Ladoy et al. (2021) Size and duration of COVID-19 clusters go along with a high SARS-CoV-2 viral load: A spatio-temporal investigation in Vaud state, Switzerland. *Science of The Total Environment*.

### "Commune en santé" label

Unisanté, EPFL

Jun 2019 - Aug 2023    Lausanne, CH

Oriana Villa, Stéphane Joost

Development of indicators that characterize the living environment of the population of Vaud to support labeling initiatives. Creation of a web application and training of public health practitioners on Geographic Information Systems (GIS).

### Chicago Array of Things

Center for Spatial Data Science (CSDS), University of Chicago

Feb 2018 - Sept 2018    Chicago, US

Luc Anselin, Julia Koschinsky

Design of a spatial data infrastructure that collects and analyzes daily air pollution data from a network of sensors distributed throughout the city of Chicago.

### Life expectancy inequalities

Geographic Information Research in Population Health (GIRAPH)

Sep 2017 - Dec 2017    Lausanne / Genève, CH

Stéphane Joost, Idris Guessous

Investigate spatial inequalities in life expectancy in the canton of Geneva.

Ladoy et al. (2021) Geographic footprints of life expectancy inequalities in the state of Geneva, Switzerland. *Scientific Reports*.

### Modeling ultrafine particles

Institut de Santé au Travail (IST)

Feb 2017 - Jun 2017    Lausanne, CH

Jean-Jacques Sauvain, Stéphane Joost

Sampling and spatial modeling of ultrafine particles in a manufacturing facility to assess potential worker exposure.