Identification of neighbourhood clusters on data balanced by a poset-based approach

Margherita Silan, Pietro Belloni & Giovanna Boccuzzo

Statistical methods and models for complex data — Poster Session: 21 September 2022

Introduction

A **neighbourhood effect** is the independent causal effect of a neighbourhood (i.e., residential community) on any number of health and/or social outcomes.

Health outcomes are primarily influenced by **individual characteristics**, which also lead to a mechanism of **self-selection** and possible aggregation of individuals in neighbourhoods (Figure 1). At the same time, health outcomes can be influenced by characteristics of the surrounding environment, such as the presence of green areas, socialization spaces, walkability, and security.

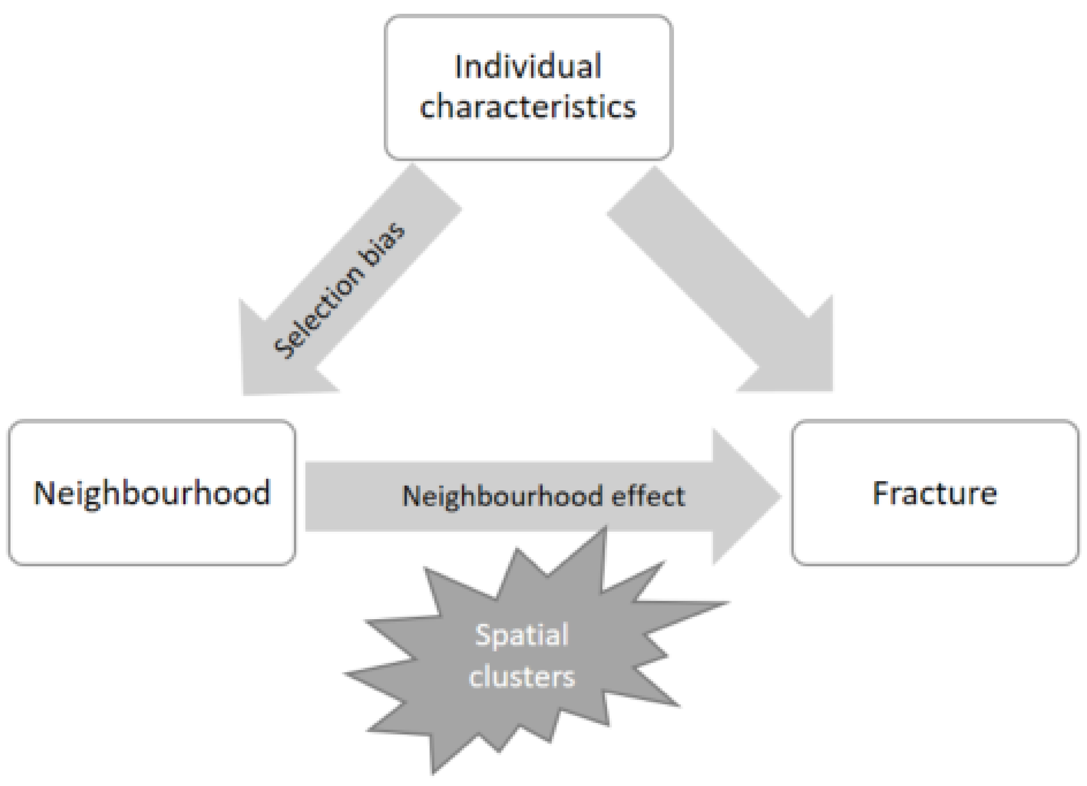

Figure 1. Fracture, neighbourhood effect and selection bias.

The goal of this work is to **identify and represent neighbourhood clusters at greatest health risk within a city, after balancing the distribution of confounders** among the neighbourhoods.

Data

We selected individuals living in the city of Turin, in north-western Italy, aged 60 or more from the 2001 census.

The considered **individual confounders** are: gender, age, region of birth, family composition, educational attainment, last known occupational condition, and home ownership.

We considered **72 neighborhoods** with a population size that varies between 672 and 7758 individuals (3136.5 on average).

The considered variables at neighborhood level are: deprivation, availability of green and pedestrian areas, walkability, and rate of violent crimes.

The **health outcome** considered is **hospitalized fractures**, observed in 2002 from helthcare dataflows collected by the Turin Longitudinal Study. Its occurrence is less than 1% in the population, with some differences in the distribution among neighbourhoods (Figure 2). However, the crude distribution of fractures in the territory is affected by the unbalanced distribution of individual characteristics.

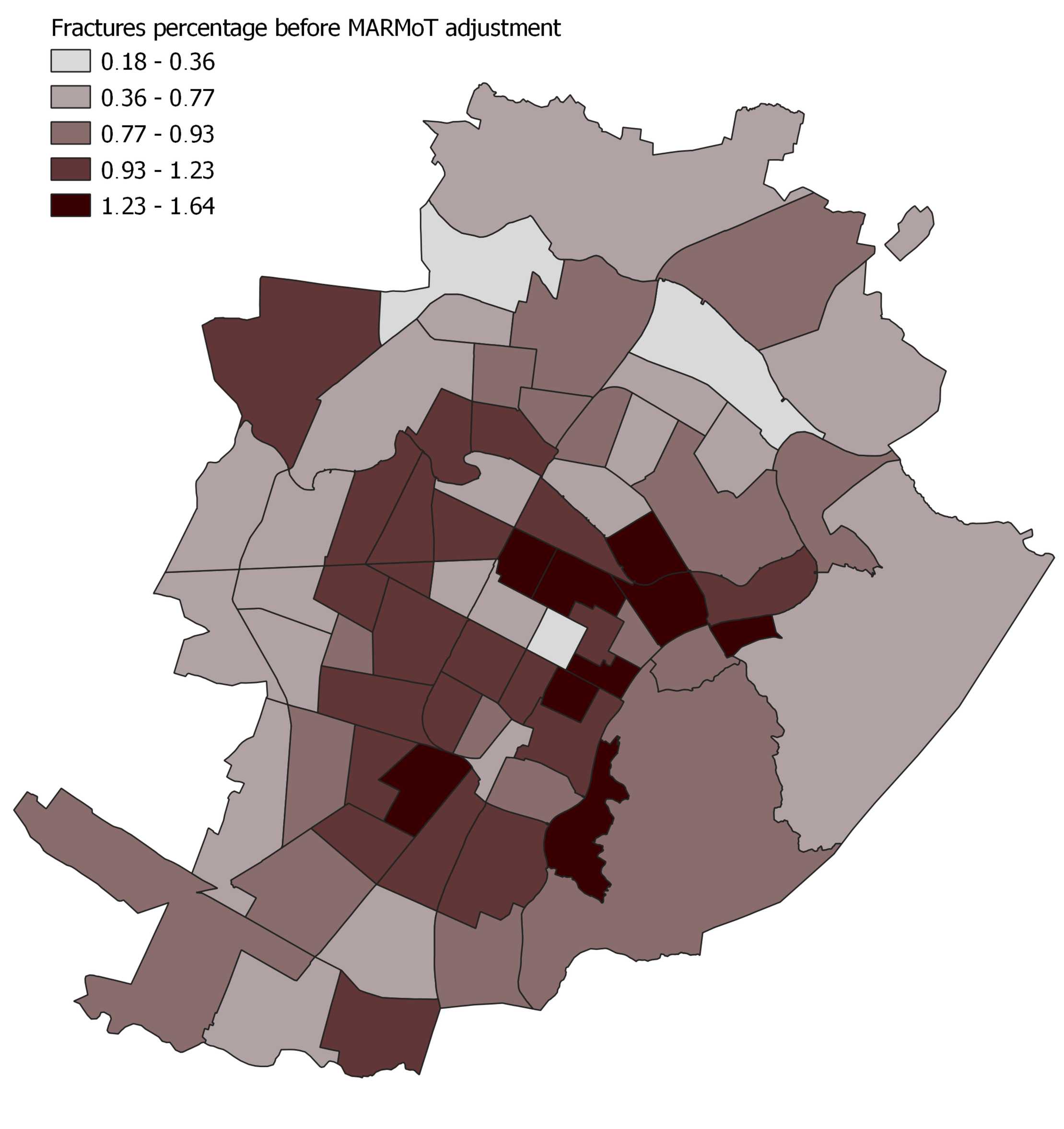

Figure 2. Fractures ratio before the MARMoT adjustment

Balancing procedure

One of the most relevant methodological problems to face when estimating neighbourhood effects is the selection bias. Indeed, having observational data, it is difficult to **establish whether differences with respect to the outcome between treatment groups can be attributed to the treatment itself, rather than to differences between subjects' characteristics in the groups**. Thus, in order to solve this issue, the focus is about balancing the distribution of confounders among the treatment groups. In a multiple-treatment framework, a promising technique to do so is the **Matching on Poset based Average Rank for Multiple Treatments (MARMoT)** proposal.

MARMoT

The MARMoT technique summarizes individual characteristics that need to be balanced among treatment groups with the help of **partially ordered set (poset) theory**. Each subject is identified by a **profile** corresponding to the set of confounders. An approximation of the **average rank** is associated with each profile, used as a balancing tool. Once the average rank is computed, it is involved in the matching procedure that produces a synthetic population with similar composition in all treatment groups (Silan et al., 2021). We rely on the **absolute standardized bias (ASB)** to check if the balance is reached after the algorithm. The ASB is defined as:

$$\text{ASB} = \frac{|\bar{X}_t - \bar{X}|}{\sqrt{\frac{S_t^2}{2} + \frac{S^2}{2}}}$$

where $\bar{X}$ and $\bar{X}_t$ are the means of the variable X of individuals living respectively in the whole city, and in the neighbourhood t; and S and S_t are the standard deviations of the variable X of individuals living respectively in the whole city, and in the neighbourhood t.

Balance evaluation after MARMoT procedure

The **distribution of confounders** among neighbourhoods before the balancing procedure is highly **unbalanced**. For every variable level and every treatment group, a value of ASB is computed (Table 1). The most unbalanced variables are educational attainment (with a mean ABS equal to 25.88%), region of birth (with mean ASB 16.13%) and home ownership (with mean ASB 13.48%). All the ASBs are reduced after MARMoT procedure.

ASB	1^{st} Qu.	Med.	Mean	3^{rd} Qu.	Max.	$> 5\%$	$> 10\%$
Before	2.56	5.54	9.53	11.41	78.75	742	402
After	1.31	2.87	4.09	5.44	46.39	377	100

Table 1. Distribution of ASBs before and after MARMoT adjustment.

After MARMoT balancing procedure, the differences observed in the distribution of hospitalized fractures among neighbourhoods should no longer depend on individual confounders, under the unconfoundedness assumption.

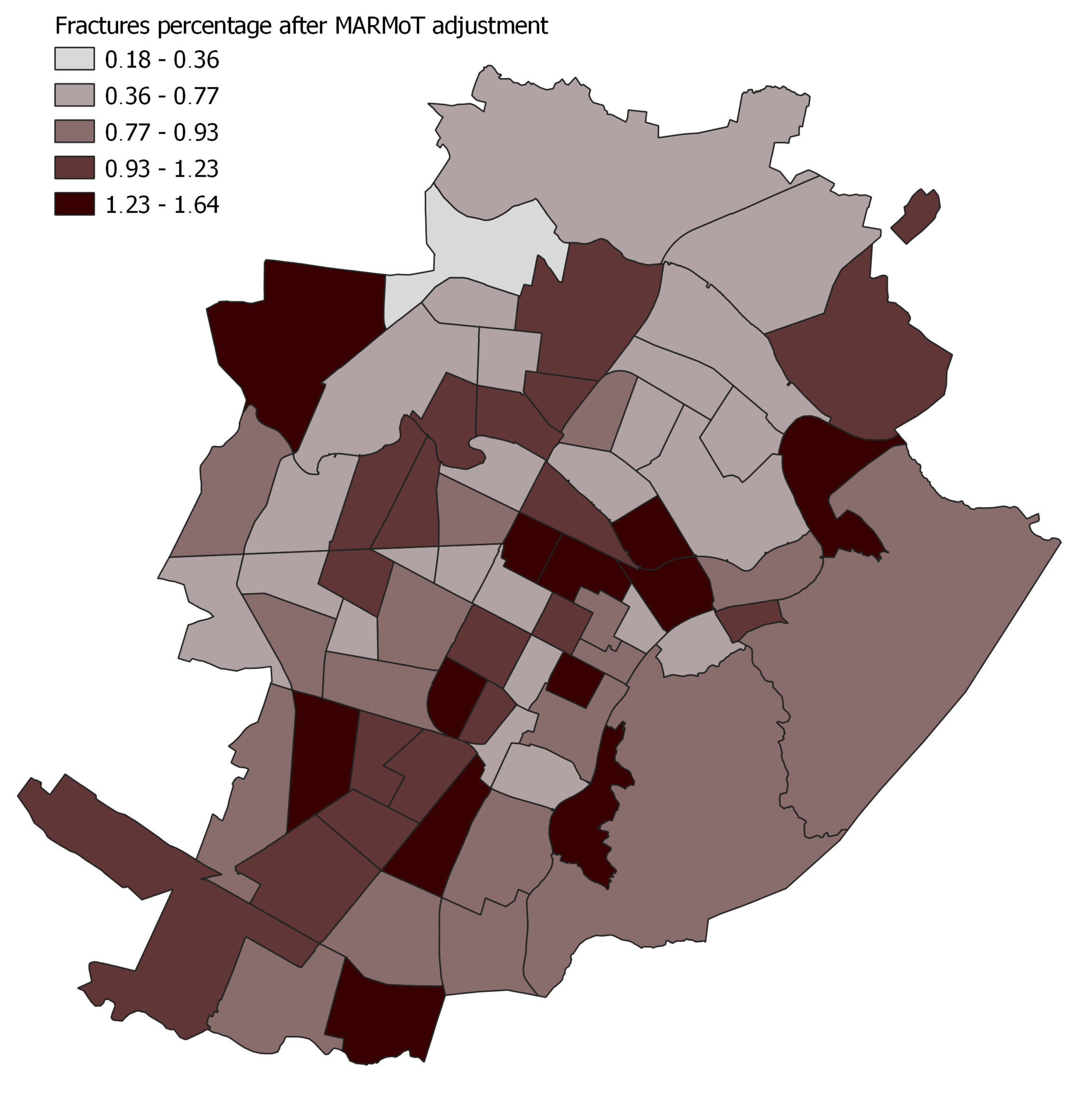

Figure 3. Fractures ratio after the MARMoT adjustment.

Clustering

A further step after estimating the neighborhood effects is to check **whether the risk of fracture** for the subjects analyzed **is constant** across the territory **or is significantly higher in some contiguous areas** (known as geographical clusters). Therefore, we used a **spatial scan** to **identify the presence of clusters** and quantifying the increased risk.

Disease clustering using a spatial scan

The idea behind a spatial scan is to conduct the test

$$H_0 : \gamma_i = \gamma_{\bar{i}} \qquad H_1 : \gamma_i > \gamma_{\bar{i}}$$

where i is a candidate cluster part of the geographical space, γ_i is the risk within the cluster and $\gamma_{\bar{i}}$ is the risk outside the cluster. Our analysis relies on the spatial scan introduced by Gómez-Rubio et al (2019) and is able to adjust for multiple covariates at area level. This method is based on a Poisson GLM and uses a distinct dummy variable for each cluster to perform the likelihood test shown in (??).

Results

First, the disease clustering procedure has been run **without neighbourhood covariates**, with the aim of identifying the actual presence of clusters at higher risk. Then, it has been run **including the 4 area-level covariates**.

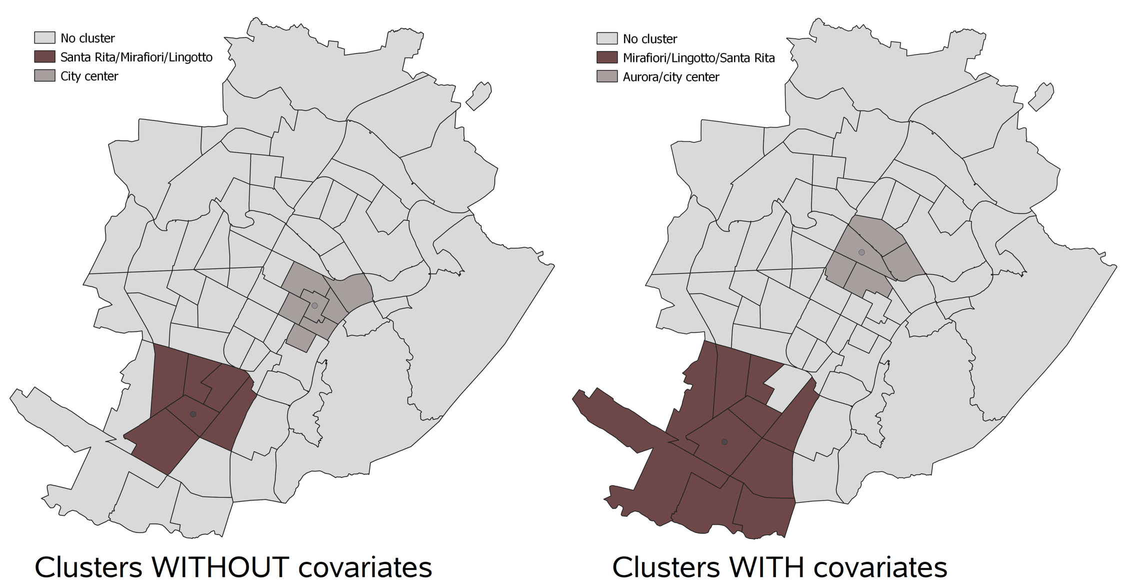

Figure 4. Clusters obtained with and without area-level covariates.

The reason given to differences between clusters identified with and without covariates is that the effect of the covariates is able to explain part of the clusters and, at the same time, their presence brings out new areas of high risk that were previously masked.

Clusters	Areas included	Increased risk
1 WO	Santa Rita, Mirafiori, Lingotto	25.19%
2 WO	City center	22.17%
1 W	Mirafiori, Lingotto, Santa Rita	22.87%
2 W	Aurora, City center	19.68%

Table 2. Details of the clusters obtained with (W) and without (WO) covariates.

Conclusions

In conclusion, thanks to the combination of both MARMoT procedure and a spatial scan, it was possible to highlight **two clusters of neighbourhoods** in Turin where there is **an increased risk of incurring hospitalized fractures for elderly people**. This is a starting point to implement focused **prevention policies** to reduce the occurrence of fractures among the elderly.

References

Gómez-Rubio, V., Moraga, P., Molitor, J., et al (2019). Dclusterm: model-based detection of disease clusters. *Journal of Statistical Software*, **90**, 1 – 26

Silan, M., Boccuzzo, G., Arpino, B. (2021). Matching on poset-based average rank for multiple treatments to compare many unbalanced groups. *Statistics in Medicine*, **40(28)**, 6443 – 6458

Silan, M., Belloni, P., Boccuzzo, G. Identification of neighbourhood clusters on data balanced by a poset-based approach. *Statistical Methods & Applications*, submitted

Contact information

Margherita Silan, research fellow
Department of Statistical Sciences, University of Padua
margherita.silan@unipd.it