Migrants' under-mortality: using causes of death to disentangle healthy migrant effect and salmon bias in early twentieth century Madrid

Michel Oris, Stanislao Mazzoni, Aitor Garcia Aguirre, Mélanie Bourguignon, Diego Ramiro Fariñas

Extended abstract

Migrants' under-mortality: using causes of death to disentangle healthy migrant effect and salmon bias in early twentieth century Madrid

Michel Oris, Stanislao Mazzoni, Aitor Garcia Aguirre, Mélanie Bourguignon, Diego Ramiro Fariñas

Topic

In debates on the causes of under-mortality among migrants, one interpretation emphasises **selection** at the point of origin, with the healthiest leaving while the others remained. Another interpretation is that this result is an artefact, as migrants fell ill in unsanitary cities and then returned home, where their deaths were recorded. This is known as the **salmon bias** (Oris and Alter, 2001).

To disentangle these two causalities, we propose an original approach based on the use of **causes of death of migrants and non-migrants**, while controlling for length of residence in the city and various social and environmental characteristics. Indeed, returning home was possible in the case of latent (though fatal) diseases such as tuberculosis, but not in the case of sudden death (such as cerebral haemorrhage, cardiac arrest, heart failure, etc.). Untangling the mortality advantage of migrants based on causes of death offers new perspectives.

Theoretical focus

A number of studies in historical demography have highlighted a mortality advantage among migrants in rapidly expanding cities in the 19th and early 20th centuries. This healthy migrant effect is a surprising phenomenon, as it compensates, to some extent, for the disadvantages experienced in cities by most new arrivals. It has also been observed in contemporary populations.

While the promotors of the "fundamental cause" theory assume that differences in access to both preventive and curative resources are related to socioeconomic status and explain a social gradient in mortality, whatever the epidemiological regime (Link and Phelan 1995; 2010; Clouston et al., 2016), they recognize that the healthy migrant paradox (poor socioeconomic conditions but better health) contradicts their theory (Link and Phelan 2010: 15; Bakhtiari 2018: 140).

Specifically, about tuberculosis, Hinde (2015) considers that its high prevalence in some British rural regions could be explained firstly by the selective (and gendered) migration of healthy rural people to the cities, and then by the salmon effect, i.e. the return of those who fell ill in unhealthy urban environments with, in particular, a high prevalence of tuberculosis (see also Reid and Garrett, 2018). Suffering from a fatal but generally latent disease, they had time to return home for treatment, especially since the representations of the time mythologized the countryside and its health virtues (Molero Mesas, 1989).

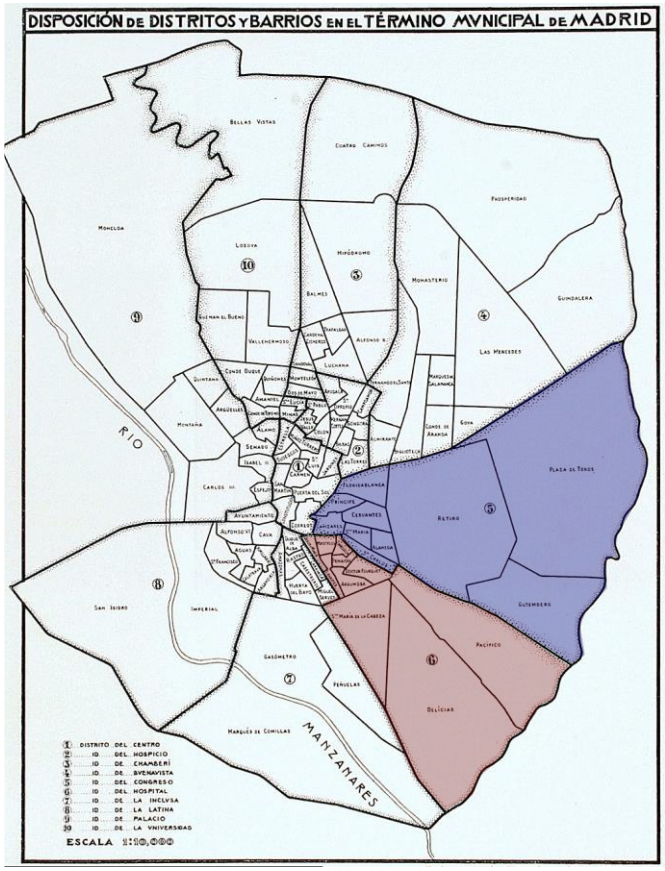
To our knowledge, the difference in mortality between migrants and natives for other causes of death has not yet been studied by historical demographers.

Data

Our analyses are based on two data sources. The first one is the *padron*, the list of inhabitants that was established by the municipal authorities every five years. For 1905, we collected the individual notices,

organized by household, for all the inhabitants of two about the ten districts of Madrid: Hospital and Congreso, which were two contrasted areas, the former accumulating the penalties, the latter being relatively favored (Mazzoni, 2022). 112'296 inhabitants constitute the population at risk. Based on a linkage between the death certificates and the padron, we will analyze the risk of dying: deaths that happened during the period 1905-1908 were attributed to the individuals present at the time of the census (which was done late November, early December 1905).

Figure. Map of Madrid and its ten districts



Source: Núñez Granés (1910)

In a first step, analyses were made within the boarder of Hospital and Congreso districts and linked 4061 death certificates with the individual notices in the padron. In a second step, 1173 additional deaths were found, from individuals living in Hospital or Congreso in 1905 and who died elsewhere, but in Madrid city, between the census and the end of 1908. So, it corresponds to a total of 5234 deaths. A comparison with the death rates published by the municipality of Madrid suggests that we missed less than 200 cases. Those low values can be explained by the fact that even if urban populations are known to be highly mobile, moreover in a city of the size of Madrid, our death certificates database covers all the city and not only the two districts we study. It proved to be very helpful[1]. Deaths in the other eight districts could not only be a consequence of normal internal mobility, but also of the recommendations to change living environment that medical doctors gave, especially to their tuberculous patients[2].

[1] Anecdotally, although not so anecdotical at the end, the Spanish custom of mentioning two family names, was positive to increase both the number of linkages and their reliability.

[2] This is reported, although some 30 years after the period we study, in 1935, by Dr. Carlos Díez Fernández, who was the head of the Madrid dispensary. We take this information from Miralles Buil, 2014, 437.

For the remaining gap, for the deaths it seems we miss, they can be due to imperfect data or to errors during the data collection preventing the linkage of the individual notices in our two data sources. Furthermore, we can suspect the salmon bias previously evoked. When our models reveal an under-mortality from the migrants, it is not a proof, strictly speaking, but a support for such an interpretation.

Methods

We use a statistical approach that is still rarely employed in epidemiological research, and quite innovative in historical demography. In a multinomial logistic design, we consider "staying alive until 1908" as the reference event, and we contrast it with four possible other outcomes: "dying from pulmonary tuberculosis", "dying from respiratory diseases", "sudden deaths", and "dying from other causes of death". We will also test competitive risk models (Fine-Gray).

These outcomes require a few explanations. In 1905-08, we are 13/16 years after Koch identified *mycobacterium tuberculosis* which stabilized the disease etiology and was rapidly diffused among the Spanish physicians. There was however a dissimulation of tuberculosis as cause of death at that time, for explicit medical reasons, but also due to pressure exerted by families (probably the wealthy ones) on physicians. Indeed, they considered pulmonary tuberculosis, bronchitis and pneumonia as more susceptible of prevention that curation, and also that the two latter were often the ultimate cause of the death after a more or less long period of tuberculosis morbidity (Miralles Buil, 2014, 87-88). That is why, in our analysis, we compare the deaths from pulmonary tuberculosis with those from respiratory diseases.

Then comes a central point in this communication. Pulmonary tuberculosis resulted rarely in a rapid death. In most of the cases, as previously noted, it was most often a latent disease, so that leaving the unhealthy city and returning home was possible. In that case, the death was recorded in the place of origin and not in the place of destination, where the migrant was contaminated. This salmon bias, however, does not apply to the cases of sudden deaths, due to diseases like cerebral haemorrhage, cardiac arrest, heart failure, etc.

Our analyses rely on regressions using characteristics observed in 1905 to explain the risk of dying during the following three years. The explanatory variables are organized according to a central hypothesis rooted in the literature, i.e. the opposition between social and environmental (or urban) factors. Starting with sex, age, and matrimonial status as both control and explanatory variables, several nested models test the impact of social, then environmental variables.

Among the social factors, our proxy variables of the inequalities in both resistance and exposure are the individual-household social status, the neighborhood dominant social character, and the individual occupation. For the occupations, we used the HISCO classification (Van Leeuwen, Maas, Miles, 2002), more exactly the synthetic typology in seven modalities proposed by Trigueros, de Miguel Salanova and Simón (2014, 119). Then HISCLASS was used for the social status (Van Leeuwen, Maas, 2011). Each individual was classified in the socioeconomic stratification according to his/her own status, and then people without any recorded position were given the social status of their household head. So, in a bourgeois family with servants, the latter were coded as servants, but wife and children were given the HISCLASS code of the *pater familia*. For the sake of parsimony, we have grouped the 13 classes of HISCLASS into 3 groups (elite, middle classes, popular classes).

A second cluster of social variables is about the origin of the inhabitants of Madrid. The *padron* indicates the place and province of birth for each individual counted. Based on previous research (Oris, Mazzoni, Ramiro Fariñas, 2024), we defined six modalities: Natives of Madrid province, born in Castilla la Mancha, Castilla y Leon, North of Spain, Rest of Spain, and Others. The duration of residence (in

year) in Madrid is also mentioned, although this info is missing in 12% of cases. The groups are: Natives, <1 year, 1-4, 5-9, 10+, unknown.

Then, a first group of environmental variables focus on the household of every person at risk. For the Spanish medical doctors after Koch discovery, household size was crucial since reflecting the contamination potential in the closest, intimate, nest (Miralles Buil, 2017, 70). We also distinguish family households from pensions or hotel, since the literature review that the latter played an important role in the (male) urban excess mortality.

We will proceed by successive spatial levels. First, we will look at the number of inhabitants at ego address, second at the number of inhabitants divided by the number of buildings (addresses) at the street level, and third we calculate the same indicator at the neighborhood level. So here, to the household characteristics we add the house, street and neighborhood environments, with indicators of size and crowding. Once again, the doctors of the time inspired us, since some of them had the intuition that the narrower the spatial scale of observation, the more important was the role of these factors (Miralles Buil, 2014, 206). We can also suspect that in a patriarchal context, women were more affected in the private space (household, building), and men in the public space (street, neighborhood). Recently, we added two important indicators that we have not yet tested: the floor on which the accommodation was located, and the renting prices.

Expected findings

In first analyses, we have already observed that female service workers benefited from a strong under-mortality by pulmonary tuberculosis (O.R. 0,304). They were mainly young immigrant women working as servants in the middle and high classes' households. Although their work was hard, it was not necessarily as harder as in their village. Their income was often miserable, but they lived in the neighborhoods and homes of their wealthy masters, and were fed in their kitchens, even if it was with leftovers. When compared with the male day laborers, we see at which point the city of the poor immigrant women differed from the city of the poor male immigrants of the same age (Oris, Ritschard, Ryczkowska, 2006).

Precisely, when we study the effect of origin, the reference category are Madrid natives, which faced the higher risks of dying from pulmonary tuberculosis. Only the immigrants from Northwestern Spain (almost 7% of the population of Hospital and Congreso) did not differ significantly. This region showed a high prevalence of this disease, so that it is possible that migrants arrived already contaminated. Furthermore, going back implied a long uneasy travel, so that a higher proportion of them could have stayed, and ultimately died in Madrid. All the other migrant groups showed a much lower mortality from pulmonary tuberculosis in Madrid, especially the women from the two Castille, the regions that surrounded Madrid (and made almost 28% of the Hospital and Congreso inhabitants). In their case, the salmon bias is a possible explanation. Maybe those migrants fell sick in the Spanish capital and went back home, disappearing from Madrid mortality statistics.

However, the selection of migrants based on their health and the progressive impact of life in the city have also to be considered. These are tested in the following model testing the duration of residence in Madrid, only controlled by age and sex. "Since always", an expression identifying the natives, but also those who arrived babies, is the reference category, and definitively defines the population the most at risk of dying from pulmonary tuberculosis. More recent the immigrant arrival in the city, lower the risks, and inversely. Living in town was dangerous, especially for men.

Table. Logistic regressions on the risk of dying from pulmonary tuberculosis in the districts Hospital and Congreso. Age, sex and duration of residence in Madrid

Variables	**All**	**Men**	**Women**
Age groups			
0-4	0.333***	0.352***	0.304***
5-14	0.280***	0.142***	0.530***
15-24	1.206*	0,941	1.755***
25-49	(base)	(base)	(base)
50-59	0.632***	0,801	0.367***
60+	0.317***	0.358***	0.266***
Sex			
Male	(base)	(omitted)	
Female	0.607***		(omitted)
Residence duration			
< 1 year	0.284***	0.398**	0.144***
1-4	0.446***	0.475***	0.398***
5-9	0.532***	0.567**	0.484***
10+	0.754***	0,847	0.625***
« Since always »	(base)	(base)	(base)
Unknown	0.418***	0.340***	0.589**
_cons	0.012***	0.012***	0.007***
ll	-3338	-1909	-1408