The health cost of living in a city: The case of France at the end of the 19th century.

Lionel Kesztenbaum
Jean-Laurent Rosenthal*

This version: July 2009

Very preliminary, do not quote
References are incomplete and conclusions are preliminary

Abstract

Despite a long standing debate over urban living conditions during industrialization, the impact of rural-urban migrations on health and mortality remains an open question. In particular how did rural migrants converge on urban mortality rates? We observe both mortality and geographical mobility in a large longitudinal dataset of French males born in 1860. We compute the marginal effect on health of living in an urban area. We show that this effect is negative and remains significant in most part of the life cycle. At the same time, rural-urban migrants benefited from clear advantages in terms of health over those who already lived in the city. However, this benefit fades in a few years: after staying in an urban area, migrants rapidly face the same mortality rates as those who grew up there. Further we find no evidence of a spike in mortality among rural migrants as they encountered the more severe disease environment of cities, instead it seems their initially superior physical human capital was depleted over time.

Keywords: migration, health, differential mortality, rural-urban gap.

JEL codes: I18, N33, N93

* Respectively: INED, 133 bd Davout 75940 Paris Cedex 20. lionel.kesztenbaum@ined.fr and CalTech, jlr@hss.caltech.edu.

Introduction: Standards of living and local health conditions

Among 19th century social scientists, the higher mortality of urban over rural areas was a commonplace (Vedrenne-Villeneuve 1961). As early as 1778, scholars showed, by exploring monks' mortality and comparing it with that of 'ordinary' people that mortality may differ according to social and economic conditions of living (Moheau 1778). Later on, they also pointed out the awful living conditions of cities and the very high mortality that prevailed there (Villermé 1828 and 1830). However, the causes of the urban rural mortality gap and their precise contribution to it were unknown. Then, the causes the most frequently put forward were overpopulation, housing conditions, water supply, slope of the land and, of course, poverty. More recent studies have also pointed out the higher mortality of cities throughout the industrialization process (Landers 1993, especially chapter 7). Not only did cities offer worse living conditions --quality of housing, of food or the disease environment-- but working conditions were much harder than in the countryside (Neven 1997), Szreter and Mooney 1998).

More broadly we can contrast two views about the sources of high initial urban mortality. On the one hand, there are those who view it as the consequence of a very low stock of urban infrastructure combined with a high influx of poor migrants. The lack of clean water, healthy food and transport infrastructure, meant that cities were very crowded and hazardous. Part of the growth process would be to improve these living conditions but initially at least, they may have worsened (Williamson 1982, (Chamla 1962), Steckel 1992, Steckel 1995, Komlos 1994, (Komlos 1998). Most scholars now agree that there was indeed a sharp decline in health during industrialisation but long-term health conditions increased much more. In other words, the initial cost of industrialisation on health was, on a middle and long run basis, more than compensated (Galloway 1985, ; Fogel 2004).

The arguments about the low quality of the urban infrastructure extend directly to the health stock of the population. There is a growing literature dealing with the consequences of early life conditions on future mortality (Fogel 1986, (Bengtsson, Campbell et al. 2004). It shows that bad living conditions during childhood have negative effect on latter life: the worst are living conditions during childhood, the higher his adulthood mortality, especially at older age (Ferrie 2003). What remains in question here is how this specific effect does interact with later mortality differentials. In other words, we may wonder what part of urban-rural mortality differential is due to bad early-life conditions. As a significant part of the urban population was born there, higher urban mortality may be due, at least in part to bad living conditions during childhood. But in the opposite case, we may also think that those who grow up in the city are heavily selected and thus more able to resist bad city conditions during adulthood.

The other approach focuses more directly on the disease environment, some scholars argue for an immunization process: cities were the place where people from different regions and areas went together, all carrying different germs of various diseases (Elo and Preston 1992, ; Costa 2003). As prior exposure to infectious disease reduces the chances of dying later in life, the higher urban mortality in cities may result from the high migration rate to and within cities.

Migrants appear to be the key to these puzzles: they acquired their health stock in one location but experienced the living conditions of another. Now consider an age cohort in a population, if the disease environment was the primary source of urban over-mortality then we should observe a spike in the mortality of recent arrivals to cities relative to those who were born there. In particular rural migrants would encounter a variety of health risk for which they had acquired no prior immunity. Then after this curing phase migrant mortality should converge *from above* to that of permanent urban dwellers. Urban to rural migrants should not experience a spike (because there are moving to a less virulent disease environment) and their mortality should match that of their new rural neighbours. Conversely, the disease and infrastructure environment weight relatively equally on everyone, but rural migrants arrive in the city healthier than residents of the same age, then migrant mortality should converge *from below,* to that of permanent urban dwellers. This process would arise as rural migrants depreciate their health stock in the harsh working and living conditions of cities. In that case, urban to rural migrants should converge from above as their health stock improve, but consistent with the literature on health we would expect this convergence to be less complete than of rural to urban migrants. Indeed it is possible to depreciate one's health stock rapidly but certain negative experiences at young ages leave permanent marks (Ferrie and Troesken 2008). Evaluating these hypotheses not only helps us understand the role of urbanisation on health but also contributes to explaining the urban mortality transition. Due to the difficulty to take into account selection bias as well as the complex patterns induced by migration, analyses of the comparative mortality of migrants and stayers are very limited. In fact, few studies, if any, have tackled these issues despite the importance of migration flows and of rural-urban mortality differentials. Without high urbanisation rates the history of the industrial revolution would have been very different and so would the history of health transition.

In a recent study, Farcy and Faure compare the mortality of migrants and stayers depending on their places of origin based on a large sample of military conscripts for whom detailed information was collected up to age 46. However, they don't analyse the life-cycle patterns of mortality and perform only a static analyses of death rates (Farcy and Faure 2003, chapter 13 p 461). Yet their data are ideal for resolving our questions because they provided detailed data on the mortality experience of a cohort of individuals. The paper proceeds as follows, in the next section we detail the unusual data set compiled by Farcy and Faure, and lay out the basic characteristics of the sample. In Section III irstly, we compute mortality rates for both rural and urban areas (either as place of birth or place of residence). In section IV we focus on migrants so as to compare their living conditions to stayers in both departure and arrival areas. To tackle the problem of endogeneity section V examines mortality of experiences for given completed length of residence in urban and rural areas. We thus ask the question what was the mortality experience individuals who reached age 25 and distinguish between migrants, non-migrants, as we argue below this is a good first step in distinguishing between the two hypotheses. Overall, we find little evidence in favour of the disease environment hypothesis, instead it seems that health stocks were far more important.

II Data

Just after the defeat against Germany in the 1870 war, the French army was transformed into a full conscription army: military duty applied to everyone except for medical reasons. In particular, the replacement and exemptions options that had favoured the rich were eliminated. Then in 1872, military service was divided into a short portion of active service and a longer portion in the reserve army. Thus, conscripts stayed under the army's thumb for twenty-six years in a combination of active and reserve services. While in the reserve, individual were required to participate in training periods as they could be recalled for active service in case of war. Thus the army had to be able to locate its men over a very long period of time. To do so it required men to register any change in residence, or risk penalties or even jail sentences. Addresses were then inscribed in individual files in military registers (*registres matricules*) by location of original conscription. The files were closed only by discharge from service. Apart from migrations, they also recorded death: each time an adult male die, the mayor of his place of residence had to notify the local military authorities. There can be little doubt that the files were accurate. The recording system was the key feature of the French military organisation at a time of constant tension with German. Further, conscripts were assigned regular training sessions, during which their residence, as well as the simple fact that they are still alive, could be controlled. Finally thirdly, the data prove to be very detailed, mentioning even changes of street in the same city.

For their work on migration Farcy and Faure collected data on almost 50,000 young French men. They were all born in 1860 and are first observed in 1880 as they were being examined by the army (by the *conseil de revision*). Some of them were then discharged for medical reasons and disappear from the army's rolls. The rest were liable for a twenty-six years long military service with an active part of one to four years, from 1881 to 1882 or 1885, and a reserve part until 1906. All of these young men are observed from the end of their active military service until either their death, the end of their military service for medical reasons before 1906 or the complete discharge of their military duty in 1906. The use of failure-time data analysis allows us to consider the complexity of this dataset by analysing when the conscripts are at risk.

The dataset, described in more details in (Farcy and Faure 2003), has been constructed to observe migrations to and within Paris, from a sample of ten *departement* (French counties) and Paris. Thus, it is representative of the rural and small city areas who fed Paris with migrants. It also includes a large sample of Parisians: among a total size of 48,136 conscripts, 36,429 come from various areas in France, 8,311 from Paris and 3,396 from *Banlieue*, the Seine county without Paris (Farcy and Faure 2003, 30). These two features are unquestionably a major advantage to study rural-urban differential mortality as we accurately observe both rural and urban areas and all migrations between the two.

The whole dataset is composed of 48,136 individuals, among which 6,030 (12.5%) are directly discharged of all military service for medical reasons. Some others are not observed due to particular reasons (professional soldiers, naval assignment, teachers or clergymen). The remaining dataset includes 37,423 individuals. They are observed during 758,618 years which represents on average twenty years per capita with a median of twenty-one years of observation for each individual of our sample (Table 1). For our purposes we are interested in where they live after they complete military service. Conscripts are released into the reserve on average at almost 23 years old and they are observed, on average, until 43 years old. Some enter at twenty years old, they have no active military service while some volunteer and have very long period of activity in the army, being at risk for the first time only at 44 years old. Overall, 6300 conscripts die under observation, which represents almost 17% of those observed for at least one year.

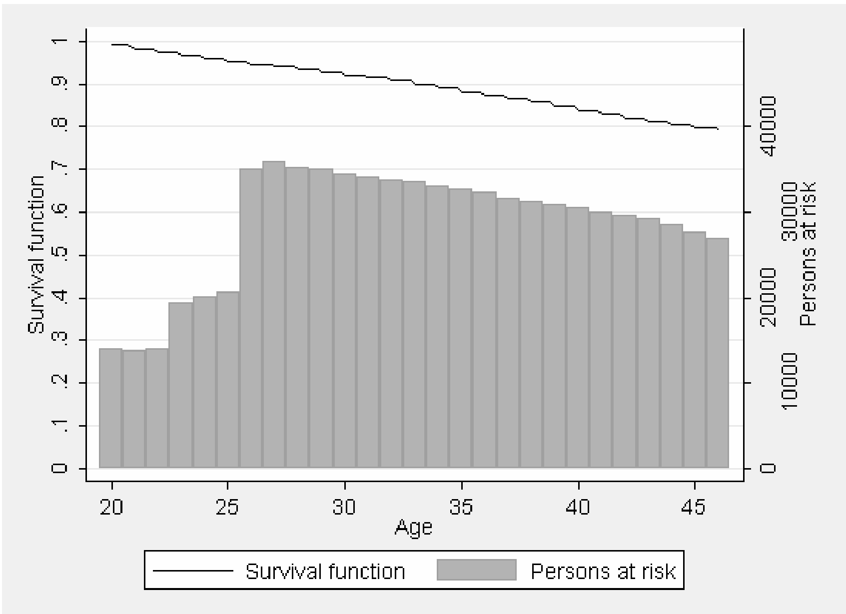
From this sample, we use failure analysis to study the duration before death. This enables us to take into account censoring, both *ex ante* --different conscript end their active military service at different moments of time-- and *ex post* --observation may end as conscripts are discharged from military service before their death. How we construct our analysis regarding to time is critical. The analysis starts when people are observed (after they complete active duty) but they may be at risk of dying before that.

In particular our case, we face left censoring, those who die before the age of twenty are never observed. It doesn't cause any damage to our analysis as long as we restrain ourselves to post-twenty years old mortality differences, which is what we do. We face a second problem: conscripts may die during active military service. Although we know if that occurred, such an event is not really comparable to mortality that occurs later. In particular it doesn't occur in a rural or urban place. Conscripts are mostly in barracks, some being in more hazardous places, if we may say so, like colonial Algeria or Indo-China. And most of the deaths occurring in the meantime are not linked to bad health due to living conditions but more to the difficult conditions of military life. So we decide to exclude that period from observation. To put it simply, people are at risk of dying from twenty years old but we consider that their deaths are only observed from the end of their active military service. The reason is that the length of military service varies between conscripts. Without this exclusion, we would have compared at the age of twenty-two two conscripts, one being at the army and the other not. In other words, the active military service is simply treated as censure. The length of this censure varies from one conscript to another but we take this effect into account through duration analysis. We'll also include this length in regression models later so as to control for the consequences of a longer active service. An alternative would have been to constrain the sample so that the origin of time is the end of the military service and not twenty years old. In that case, two conscripts with different ages, for instance twenty-one and twenty-five years old, would have entered at the same time. Thus, conscripts would have been compared according to time elapsed since the end of active military service. What we assume here is that age is the most important determinant of mortality, and not time under observation.

< Table 1 > *around here*

We thus consider a data set of young men observed during their most productive years. And we also collect their potential death during this period. Overall, the survival function in the sample is 79.84: one fifth of the conscripts die while under observation. As we study a rather smooth period of life and of history (there were hardly any war or epidemic during the period we consider) the survival curve decline monotonously (Figure 1). The sample grows in size as people complete active service and enter under observation. The maximum size of the sample is at 27 years old when a little more than 30,000 conscripts are at risk of dying --they are both observed and still alive at that age. From this moment on, the size of the sample shrinks slowly as conscripts die or are discharged from the army for medical reasons.

Figure 1 Survival function and size of the sample



Note: Associated with left axis, the line denotes the survival function: the probability for a given conscript to be both alive and under observation at a given age. The bins, related to right axis, refer to the number of conscripts under observation at a given age ('risk pool').

In a nutshell, here are the key features of our sample: we observe only males, fit for military service; aged twenty or more--we have all French young males who survived this age and whose parents where living in one of the sampled area-- but we record and analyse their deaths only for those who entered the reserve after their active service; we observe them until they die, are dismiss for medical reasons, or finish their time in the army at forty-six years-old; all conscripts are observed wherever places they move to[1]. So all mortality rates we compute are conditional on being still alive and fit for the reserve at the end of military duties. Overall, this selection bias is not a problem for us: if we observe a higher mortality for urban areas, we may suppose this gap would have been even higher without the selection bias. In fact, discharge of military duties for medical reasons is more frequent in urban areas than rural one: 7.5% against 5.8% of the conscripts observed at least once (Table 2). And, even more significant, the diseases observed there are much frequently potentially lethal than in rural areas. For instance, the number and proportion of conscripts being discharged for tuberculosis is twice as high in urban as in rural area.

We compare people that share many features: they were born the same year, they did their military service at the same moment and they are all in good health, enough to be accepted in the army during the whole part of their life cycle. Therefore, the selection process allows us to compare the mortality of healthy adult men. In these conditions, finding differential mortality between areas is even more convincing.

< Table 2 > *around here*

III The Urban Rural Gap in Life Expectancy

We begin our investigation to life expectancy by place of residence of the conscripts during the active part of their life—in effect this first effort neglects the endogeneity of location

[1] We also record international migrants.. However as our focus is on rural-urban differences, we have little interest in such moves. Therefore, we consider migration out of France as censure. Only a very small number of conscripts are concerned by such mobility.

decisions. We consider conscripts between the end of their active military service and their dismissal from military duties (due to death, disability or reaching the age of 46). During this period, we have longitudinal data that allow us to observe any occurring death and consequently to estimate mortality rates. Moreover, a conscript can be localized at each moment, either in a city or in the countryside. Consequently, each individual's residence can change over time. Nevertheless for each year between 1880 and 1906, we have the number of individuals living in rural or urban areas and we know how many die in each group. We can then precisely calculate mortality risk according to place of residence.

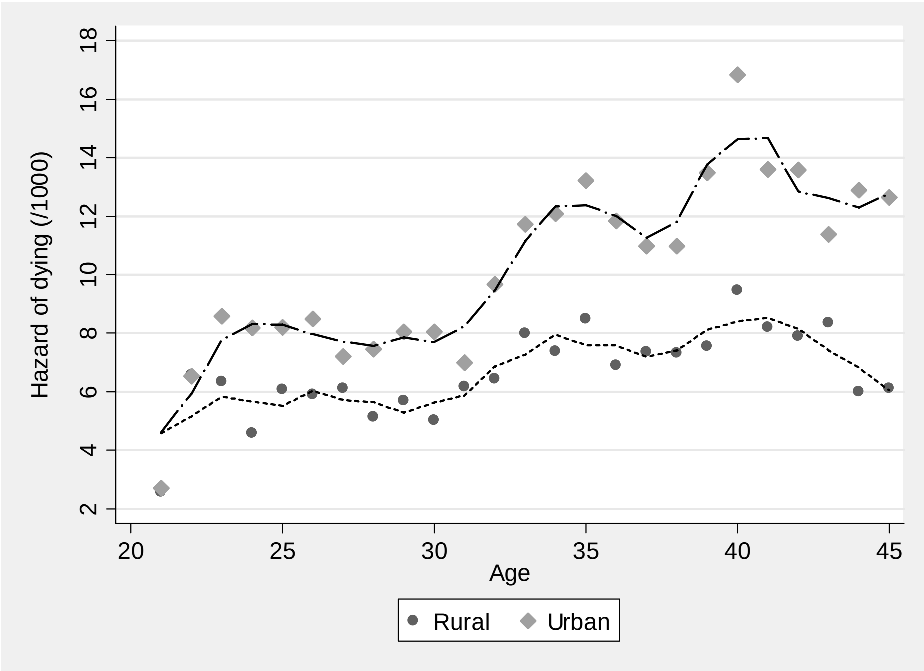
For our purposes we will group locations into four groups, "rural" refers to municipalities with a population smaller than 2500; 'small cities' have between 2,500 and 10,000 inhabitants while 'large cities' have more than 10,000 inhabitants, Paris is in its own category. When we write of urban rural differences we aggregate the last three categories into one.

There is a sharp urban penalty: the mortality rate is almost double in cities than in rural areas (Table 1). The disadvantage of urban living grows with the size of the city. Even in small cities, mortality was significantly higher than in rural area. And large cities were not as deadly as Paris, which was by far the largest city in France. Living there was twice as deadly as leaving in a rural area. If we assume that cities are deadlier than rural places then urban inhabitants may have a higher mortality either because most of them were born in a city or because of their current exposure to cities' risks, or both. Later on, we will try to disentangle these two effects by comparing migrants and stayers.

< Table 3 > *around here*

Beyond these blunt aggregate figures we can examine how the mortality differential between town and country evolves with age. Not only is this evolution crucial to understand mortality patterns during industrialization, it can help us determine the causes of this very large gap. In the case of a higher prevalence of infectious diseases in cities, we expect age-related patterns to be identical in the city and the countryside, with the difference between the two being roughly constant. If, instead, higher urban mortality is related to bad living--and working--conditions, we should observe a growing gap between the two, as mortality in urban area increases with time. Indeed in this case urban dwellers' health capital depreciates with age at a faster than rate than that of rural folks. To observe age-shaped mortality differentials we take advantage of the longitudinal feature of our dataset and compile hazard of dying depending on the type of place of residence (Figure 2). In both rural and urban areas the hazard starts low and then rises. This is hardly surprising: mortality inevitably rises with age. However, there is an important difference between the two areas in the magnitude of the increase. When in their 20's, conscripts face a yearly risk of dying: around 6 per 1000 in the countryside and 8 per 1000 in cities. In the countryside, the hazard then increases to a ceiling around 8 per 1000 for individuals in their 40s. In cities, however, picture is dramatically different. Past age 20 men's mortality risk jumps; in their 30s they face a risk of 12 per 1000, and later, in the 40s one of 14 per 1000. Thus the age effect is critical to understanding the differences between cities and countryside. Moreover the pattern is the reverse of what one might have expected if infectious diseases were the critical issue: at the beginning, urban conscripts die a lot but as time passes, they become immunized and their mortality declines towards that of rural areas. Instead, the sharply rising curve is more consistent with the argument that urban dwellers degradated the physical capital. People in cities suffered from their harsh living conditions of their environment: at the beginning, their mortality is quite comparable to that of their rural counterparts but after ten years, the toll is such that they die at higher rates.

Figure 2 Hazard of dying according to the place of residence



Note: hazard and 3-years moving average.

There is one obvious limitation to this first analysis: the structure of the two populations is very different. Even if we neglect the migration problem, we cannot assume that people living in rural and urban are identical. They have different incomes, education levels and family conditions. In other words, they not only differ in the environment they live in but also in their personal characteristics. To account for those differences, we use a regression model introducing some observed characteristics of the conscripts at the age of twenty: height, education, orphanage and occupation at that age; and type of military service. We also consider markers of early-life conditions as they may have later consequences on mortality. To do so, we interact the rural-urban character of the place of birth with that of the place of residence at twenty years old.

We estimate three Cox relative risk models to measure the effect of these covariates on the time until death. For our simplest model we assume there is a baseline hazard shared by all individuals and the covariates act multiplicatively on this hazard[2]. To put it differently, the risk of dying of two conscripts with different values of the same covariates are proportional one to another. This implies that the effect of a given covariate is independent of time--and we assume that the effect of living in an urban area is time-invariant. We then relax this assumption in a second frame where we assume this interaction is purely linear: the mortality differential between rural and urban places monotonously increase or decreases over time. This seems consistent with the idea that pernicious consequences of living in a city have cumulative effects. If city life leads to a more rapid degradation of health than in the countryside, we expect that mortality will rise with age faster than in cities. Finally, following Figure 2, we may also think that effects of living in a city are mostly non-linear: in a third model, we assume that they change non-monotonously over time. We look at rural-urban mortality differentials over five-year time-spans, under the implicit assumption that these effects are constant within five years.

More specifically, we model the hazard ratio as follow: $\lambda(t,x) = \lambda_0(t)\phi(X\beta)$ where $\lambda(t,x)$ is the hazard function. It depends on time and on covariates, x whose influence goes through a function $\phi()$, and $\lambda_0(t)$ is the baseline hazard depending solely on time. To model covariates influence we use the simplest functional form: $\lambda(t,x) = \lambda_0(t)\exp(X\beta)$[3].

The first model assumes that the effect of living in a city is independent of time. Thus it can be written as: $X\beta = \beta_1 urban + \alpha Z$ where αZ is a vector of control variables. The second model assumes a linear relationship between the hazard of death and the rural-urban differential of mortality or $X\beta = \beta_1 urban + \beta_2 urban \times time + \alpha Z$. The third model

[2] This assumption is not quite true for the type of military service. Using a stratified Cox model allows to correct this flaw but it doesn't alter the results for the other variables.

[3] For details, refer to (Kalbfleisch and Prentice 2002, 42-45) and (Cameron and Trivedi 2005).

decomposes the influence of time on the difference in six dummies indicating each five-year period since the beginning of observation. The model is then:

$$X\beta = \beta_1 urban + \beta_2 \delta(5 < t \leq 10)urban + \beta_3 \delta(10 < t \leq 15)urban + \beta_4 \delta(5 < t \leq 10)urban$$
$$+ \beta_5 \delta(15 < t \leq 20)urban + \beta_6 \delta(20 < t \leq 26)urban + \alpha Z$$

where $\delta(5 < t \leq 10)$ is a dummy variable taking for value 1 when an individual has resided the majority of time in cities between 5 and 10 years after the beginning of observation time. In that formulation, β_1, the coefficient on urban, measure the average effect of mortality of living in an urban place, whereas β_2 to β_5, estimate the non-monotonic part of rural urban mortality differentials. For instance, β_4 expresses the additional disadvantage (or the advantage) in terms of mortality of living in urban places between 35 and 40 years old, relatively to that disadvantage between 20 and 25 years old.

Estimations results are presented in Table 4. First, it should be noticed that coefficients for the control variables are almost identical in the three models. Hence the primary value of the models lies with the time interactions. The control variables have mostly expected effects on the hazard of death. For instance, conscripts with one parent still alive when reaching twenty years of age have a higher mortality risk than those with both parents alive. However this effect is not gendered: the sex of the remaining parent does not matter. But having lost both is an even greater handicap. The coefficient is strong and highly significant; orphans at the age of twenty have a quarter chances less to survive after that age than conscripts with both parents alive. This turns out to be one of the largest effects considering that we only assess the consequences of parental loss on mortality at adult ages. Education matters less with small differences within groups. All coefficients have the expected sign: illiterate and those who can only read and write have a higher mortality than those who are also numerate, the vast majority of conscripts, whereas those with secondary education do have a lower mortality. The difference between those are not numerate and those who are, which isn't very surprising considering the small educational differential between the two groups[4]. The only strong coefficient is the one between conscript with secondary education and the rest--the former having at least a 15% advantage over the latter-- but it is not significant, a fact that is certainly explained by the small number of individual with secondary education, less than 10% of our sample, a proportion quite comparable to France at that time.

On the other side, there are only small differences by occupation, except for farmers who have a substantially lower mortality than all other occupational groups. This may be connected either to their way of living --better access to food for instance-- or to a higher wealth, as being a farmer means here owning some land . To be sure, however, we only record occupation at the age of twenty, which may explain why there are so little differences between occupations.

Military service shows the complex effects of selection: people doing a shorter active military service are more exposed to mortality risk in later life. This is likely due to two complementary effects. First, when they were conscripted those fit for a long service were in better health than those slotted into other duties. Secondly, those who survive the army may be in better conditions after their service.

We take into account place of birth and place of residence at the age of twenty and the interaction between the two in a single variable so as to estimate the consequences of early-life living conditions on later mortality. There are four different situations: One could be

[4] Another difficulty is that these mentions are declarative and it is thus hard to understand what the educational differences between the two groups precisely are.

born in cities or in the countryside, and one could continue to reside in the same area or have migrated. The analysis clearly shows that conscripts both born in and living at the age of twenty years old in a rural area have a substantially lower mortality. Those born in cities and still there at twenty years old where at highest risk closely followed by those who were born in the countryside and migrated to the city before twenty years old. In contrast, , migrants from the city to the countryside during their childhood appear in better health than urban dwellers. This confirms that being born in a city is not by itself damaging; what hurts is growing up in a city. We don't have detailed information on migration between birth and twenty years old so we cannot say much more about the way early-life conditions influence adulthood mortality. But what about simply living in one, no matter where you were born?

Let's now turn to our main interests, rural-urban mortality differentials during the life-cycle. Because the control variables have almost identical effects across a variety of statistical models, the rest of this discussion neglects them. At heart we want to estimate both an average penalty for urban residence, and the consequence of longer residence in cities. If we take the whole sample, the average effect is statistically significant and large; mortality in urban areas is 1.27 times higher than in rural area (first model). The time effects have the right signs reasonable magnitudes, but they are not statistically significant. Under the linear hypothesis rural-urban mortality differential increases by 6% over the baseline (model 2). Model 3 suggests, there is no significant mortality differential in the two first periods with a very small coefficient close to zero. For the last two age spans the coefficient is much higher, though it appears not significant for the third period and only slightly significant for the last one.

< Table 4 > *around here*

We can investigate these effects in more detail, by considering first a sample of stayers (individual who always resided either in the countryside or in cities) and collapsing the residence dummies into decades. In this case both the time coefficient in the linear model and the dummy for residence in cities past 35 in the non-linear model are positive and significant. Thus we find strong support for the idea that an individual's physical capital or health stock degraded over time in cities.

We run the same estimation contrasting migrants to cities versus rural stayers, coefficients are small, not statistically significant and can take the wrong sign. These results point to the fact that migrants are heavily selected (including on the fact they must survive until they migrate) and must be treated with special care in the analysis.

< Table 5 > *around here*

This section has established that urban residence was costly in terms of life expectancy in nineteenth century France, and the differential in risk of death between rural and urban dweller seems to rise with age. While that is consistent with the health stock hypothesis, it does not tackle the questions of what degraded the health stock, or of migrants' need to acquire immunities from diseases prevalent in urban areas. The regressions that break out migrants are not consistent with the 'curing' hypothesis but they are so subject to selection problems that we prefer to seek better evidence.

IV Migrants and stayers

Migration plays an important role in demography, even in a country as stable as 19th century France (Pinol 1996, Farcy and Faure 2003 ; Moch 1992). Therefore we have to study more carefully migrants' and stayers' mortality patterns. They are the key to understand both the causes and consequences of this urban-rural gap. Such a task is being complicated by the specific situation of migrants, usually considered as highly self-selected. In other words, migrants are healthier than the rural population they left. Various arguments can be put forward to explain this hypothesis. The main economic point is that the decision to migrate, as an investment, relies on the expected net value of migration (Sjaastad 1962). So if wages are higher at the place of destination the more time spent there, the higher is migration return and the higher its net value. This explains why young people migrate more. The same holds true for healthier individuals: as their life expectancy is higher, so is their migration return and net value. Thus they have more incentives to migrate and will, on average, migrate more. Another economic argument is related to the other side of the value equation: costs. As migration is costly, only those with some money may undertake one which explains that, on average, migrants belong to higher social groups. Other arguments can be put forward and in fact are mentioned in the literature: migrants have better networks, so more people to take care of them in case of trouble; they have better habits and cultural differences that favour them against natives; and so on.

In fact, there is a huge and increasing literature focusing on what is now called "the Latino mortality paradox", in economics but also in epidemiology or sociology. Latino because this literature mostly focus on Mexican and Hispanic international migration to the US nowadays and paradox because these immigrants do have a much lower social status than natives American combined with a lower mortality. Besides the fact that they focus on a very specific migration situation, this literature does have some important shortcomings. First, the formulation of the selection argument, sometimes called the 'healthy migrant effect', is a little confused: if migrants are positively selected it is towards stayers from the departure area, and not stayers from the arrival area. For instance, Mexican immigrants may be in better health than the average Mexican but it does not explain why they would be in better health than the average American. Second, they mostly rely on data at the place of arrival, which imply they can neither compare immigrants to those who stayed in the country of departure nor consider specific characteristics of immigrants. An interesting exception is (Rubalcava, Teruel et al. 2008) who, using panel data from the Mexican Family Life Survey compare various health measure of Mexican migrants and stayers. They found no evidence of better health for those who migrate. This illustrates the main trouble with the actual literature: there is no clear conclusion to whether there is a healthy migrant effect or not and to what explain International migrants' lower mortality. Some authors argue that Mexican lower mortality is related to cultural (Abraido-Lanza et al. 1999); others conclude that migrants are selected on socio-economic basis (Akresh and Frank 2008) while, on the other side, scholars challenge the existence of such effect, arguing either that they come from statistical errors (ref) or that they are an artefact due to return migration: the 'salmon bias effect' postulate that immigrants go back to their country of origin when they are sick and so are more prone to die outside of all statistical observation (Palloni and Arias 2004). Interestingly enough, one of the few studies focused on other situation than Hispanic in the US does contest the existence of such effect: looking at various foreign nationality in Belgium, (Deboosere and Gadeyne 2005) demonstrate that the magnitude of the return migration must be so high in order to explain immigrants-native mortality differential that such a pattern is clearly impossible to imagine. After a careful review of mortality differential by cause, they argue that immigrants advantage come mostly from cultural differences, way of life and less risky behaviours.

Overall, it would be difficult to survey in more details the whole literature on the healthy migrant effect. Here, we will simply describe the specificity of the situation we consider: rural-urban migration in France at the end of the 19th century. We will review and discuss more formally out assumptions *infra*, in the estimation section. Contrary to the contemporary situation of migration from developing to developed countries, our case is a little simpler: firstly, living conditions are better at the place of origin than at the place of destination. Therefore, even without any selection effect, we may expect migrants to be in better health than natives of the place of destination. Secondly, rural migrants are not much different, culturally speaking, than urban stayers. The difference in way of life, food habits or networks, shouldn't be underestimated but it is clearly not that of today's Mexican immigrants in the US. Thirdly, thanks to the structure of our data, we don't have any under-reporting problem, which means that no artefact effect may bias rural-urban mortality differential. Fourthly, similarly in the contemporary situation, most, if not all, migrations are labour-related and there is both network linked mobility and a high rate of return migrations. Overall, we must take into account migrants' selection effect even though we do not especially wish to explain it.

The two questions we ask are straightforward. First, how much does a year in a city cost for migrants? Another formulation could be: how longer would migrants' life have been, had they not migrated? Be there no selection effect, the answer would be immediate, by comparing mortality of migrants from the countryside and mortality of those who stay there. But if one assumes that migrants are different from stayers in terms of health --and previous results strongly support this assumption-- then the answer is a little trickier.

Second, which part of the mortality differential is due --or may be attributed to-- migrants? In other words, be there no migrant at all, would the rural-urban mortality differential be higher or lower? Both questions refer to the underlying problem we focus on: the causes of urban-rural differentials. To put it short, we need to examine if urban higher mortality is linked with worse living conditions, which will induce mortality to be converging between migrants and stayers; or if it is linked with different epidemiological context, which should induce mortality of migrants to be higher than that of stayers, especially in the first years of arrival in the city.

To sum up, we assume that migrants are healthier than the rest of the population but we try to distinguish migrants from both stayers in the place of departure and stayers in the place of arrival. In addition, we also try to disentangle how these effects evolve over time. Finally, we also take into account the long term consequences of early-life conditions. We want, firstly, to quantify the diverse differences existing between migrants and stayers -- rural to rural, urban to urban and so on; and, secondly, to use the differences to explain the rural-urban penalty. To start, we focus on early-life conditions.

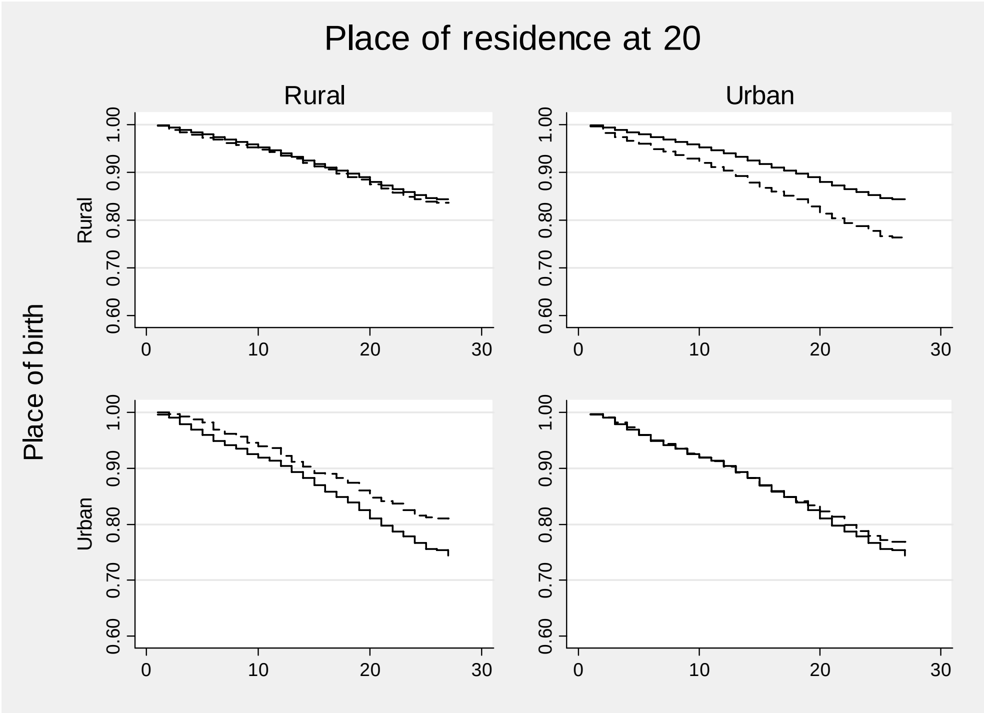
The privilege of birth?

In this preliminary section, we explore mortality after twenty years old according to geographical mobility before that age. It is both demonstrated and acknowledge that early-life conditions do have long-term consequences. The worst are childhood conditions the higher is latter mortality (Bourdieu and Kesztenbaum 2004). Here, we cannot directly look at infant mortality as we observe only survivor at age twenty. But we do have some clues on later effects of childhood living conditions. To do so, all we need to know is where people grew up. To identify this, we use the situation of the parents and we assume that, had they moved, they had taken the children with them during childhood. This is what we want to explore in more details. We compare conscripts whose parents change municipality to those who stay in place during their childhood. Doing so, we consider directly migrants' mortality without having trouble with selection effects. Contrary to the active part of the life cycle, we cannot assess precisely conscripts parents' migration behaviour as we do not have a continuous following of their residences. However, with some basic assumptions, we can examine part of their mobility decisions. For each conscript, we compare his place of birth and the place of residence of his parents when he is twenty years old. We assume that the place of birth is a good proxy for the residence of the parents at the time of birth. And we also assume that this mobility occurs while the conscript was still living with his parents.

The global rural urban gap we shown *supra* is still present here as we compare the graphics on the anti-diagonal (Figure 3). On the one hand (top right corner), conscripts who migrate to a city as a child experience a much higher mortality during their adulthood than their counterparts who stayed in the countryside. On the other hand, those who move to the countryside after being born in a city have a lower mortality than those who stayed in the city during their childhood. These are two ways of considering the urban penalty: a person who grow up in a city has a higher mortality; it does not matter whether he is born there and stays there (compared to those who left) or he is born in the countryside but moves to a city as a child with his parents. This is coherent with previous finding on child mobility (Kasakoff and Adams 2000). On the other hand, there are small if any differences between migrants within the same area, rural-rural or urban-urban. In both cases, however, migrants appear to have lesser mortality risk which is coherent with a selection effect --

very small in that case, as the selection certainly affects much less children when considering their parents migration.

Figure 3 Survival curve after 20 years old according to rural urban migration before 20 -- parental migration



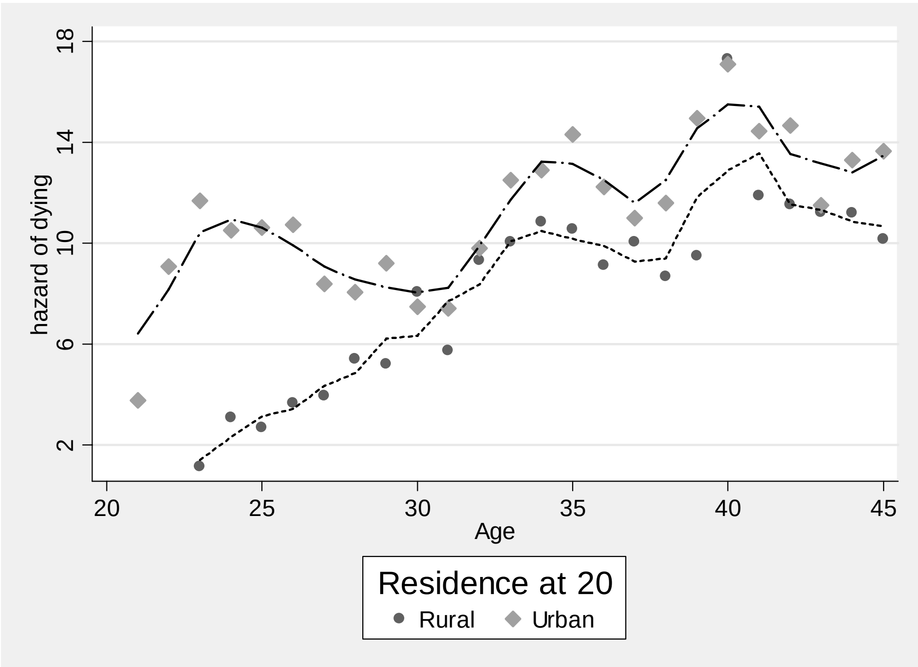
Note: in all graphs, the straight line stands for stayers while the dashed line represents movers. By construction, the straight lines are identical horizontally. Each graph compares people born in the same place according to their parents' mobility status. For instance, the graph in the top-right corner shows survival curves for those born in a rural area depending on whether they moved to an urban place or not. Finally, all graphs have the same axis scales and definitions, x is time since twenty years old and y is survivor curve (the probability to be alive and under observation at a given age).

Overall, the general pattern means a high mortality penalty that goes on in later life: what matters is not really place of birth but place of growing up. To precise this point, we now consider mortality in the same space --rural or urban-- for individuals with different backgrounds.

The cost of migration

We consider the direct influence of the place of origin on mortality. In other words, we compare mortality hazards in the same place, for instance in cities, according to the place of residence at twenty. This is a key feature of our analysis as it enables us to assess precisely if the conditions of living have a direct effect on health and to quantify this effect. There is a clear gap between rural and urban, at least at the beginning (Figure 4). At the same time, rural migrants face an increasing mortality and, slowly but surely, they finally suffer the same mortality patterns as urban natives. The gap in the first years is huge with a mortality rate of less than 2 per 1000 against 9 per 1000 for natives. However, after 8 years of residence in the city, the mortality of rural dwellers is almost identical to that of their urban counterparts. It seems obvious that the initial advantage of rural dwellers fade away after less than ten years. Of course, the time we consider is not time spend in city but if we assume that there are only few returns after some years spend in the city (we will discuss return migration later on), it seems obvious that the initial advantage of rural dwellers fade away after less than ten years.

Figure 4 Hazard of dying in a city depending on the place of residence at age twenty

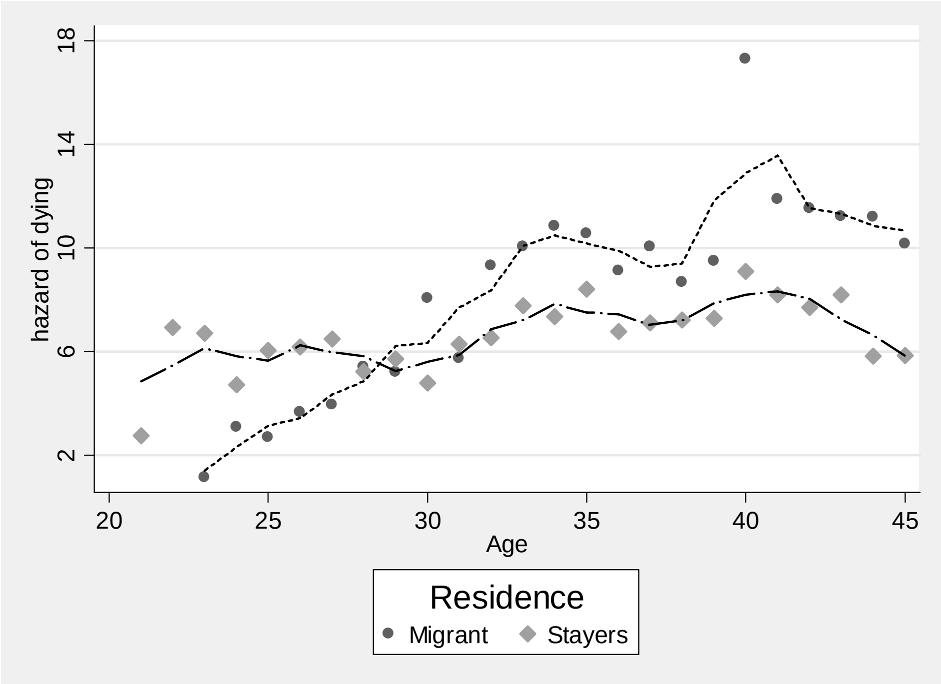


Note: hazard and 3-years moving average. There are no individual at risk before age 23 from rural origin -- ie there is no individual residing in a rural area before the military service who has achieved his service and migrated to an urban dwelling immediately after.

The previous figure does show a very important advantage for migrants compared to city natives. But it isn't only the result of selection effects. In fact, it does mix together all the effects we previously mentioned: mortality is lower in the countryside so even if migrants were not selected, there would have an advantage over urban stayers; they also have a lower mortality due to their early-life living conditions; and, finally, they may beneficiate from positive health selection. The difference between rural migrants and urban native is the sum (sort of) of the difference between rural migrants and urban stayers, on the one side, and that between rural migrants and rural stayers on the other side. To get an idea of the selection effect we can, on the other side, compute the mortality of migrants against those they left, rural stayers (Figure 5). The picture looks somehow different from the previous one and, to some extends, disappointing: the selection effect --expressed by the differences between the two curves-- seems rather small and quick to vanish. In fact, after eight years, rural migrants have the same mortality as their staying counterparts and, from that moment on, they will have a higher mortality.

These two figures put a visual picture on migrants' mortality advantage. They are particularly useful for acknowledging the evolution of both effects with time. But they mix two very distinct effects: duration of residence in the city and age at arrival. These two patterns are very difficult to disentangle but, using a Cox relative hazard model drawn from the previous section, we will try to take into account simultaneously all effects influencing migrants' mortality differential.

Figure 5 Hazard of dying for those living in a rural area at 20 years old depending on their mobility status



Note: hazard and 3-years moving average. There are no individual at risk before age 23 from rural origin -- ie there is no individual residing in a rural area before the military service who has achieved his service and migrated to an urban dwelling immediately after.

We wish to estimate the cost of migrating to the city. To do so we need to control for characteristic before the migration that may distinguish migrants and stayers. A straightforward way to do so is to compute mortality conditional to the previous place of residence. Take a conscript living in the countryside between 20 and 24 years old. Will his mortality between 25 and 29 years old be higher or lower than his city counterpart, no matter were they live at these ages? As Table 6 shows, his mortality is much lower. More important, the differences increase with age; and this is true no matter which benchmark age is considered. This means health is degrading very fast in cities, or at least much faster than in the countryside. As our previous results, these results also strongly support the hypothesis of a degradation of the health stock in the city.

< Table 6 > *around here*

Concluding remarks

The history of the Industrial Revolution is full of controversy about the rise, stagnation or fall of living standards. There is a general consensus on the existence of an urban-rural divide when speaking of living conditions and especially of mortality. However the way this differential affected people over their life cycle is less well known.

What is at stake here is then the extension of return migration and the duration of mobility to cities. As rural-urban migration become more and more permanent, the consequences of the urban penalty are harsher for the population as a whole. But at the same time, the extension of migration is produced at a time when the urban disadvantage starts to reduce.

We use the detailed analysis of one cohort to light up the consequences of rural-urban differential mortality. We compare people that are almost completely identical: they were born the same year, they did their military service at the same moment and they are all in good health, enough to be accepted in the army during the whole part of their life cycle. In these conditions, finding substantial difference in mortality between areas is even more convincing. Secondly, a surprising result is how fast migrants lose their initial advantage: migrants from the countryside are in good health when then arrived, both because they come from a safest place and are positively selected but their advantage fade away a few years upon their arrival.

Overall, we find no evidence of an increasing immunization in the cities. Our results strongly support the interpretation that urban penalty is largely due to bad cities conditions. This is coherent with recent works demonstrating how improving cities sanitary conditions reduced mortality.

Table 1 Summary of the failure data sample

```
                                   |-------------- per subject --------------|
Category                   total        mean         min      median        max
------------------------------------------------------------------------------
no. of subjects            37423
no. of records            115585    3.088609           1           2         27

(first) entry time                  2.756513           0           2         24
(final) exit time                   23.02795           1          26         27

subjects with gap              0
time at risk              758618    20.27144           1          21         27

failures                    6300    .1683457           0           0          1
------------------------------------------------------------------------------
```

Table 2 Medical discharges of the army by group of diseases and place of residence

	Not discharged		Reason unknown		Pulmonary tuberculosis		Non tuberculosis related bronchitis		Hernia		Physical trouble	
	N	%	N	%	N	%	N	%	N	%	N	%
rural	17,334	94.24	74	0.40	110	0.60	87	0.47	88	0.48	314	1.71
urban	17,269	92.55	202	1.08	226	1.21	128	0.69	94	0.50	284	1.52
Total	34,603	93.39	276	0.74	336	0.91	215	0.58	182	0.49	598	1.61

	Earing or visual trouble		Wounds or related		Mental trouble		Variscose vein and related		Others		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
rural	93	0.51	9	0.05	81	0.44	90	0.49	113	0.61	18,393	100.00
urban	110	0.59	12	0.06	111	0.59	105	0.56	118	0.63	18,659	100.00
Total	203	0.55	21	0.06	192	0.52	195	0.53	231	0.62	37,052	100.00

Note: Figures are computed on all conscripts observed at least once. Khi² is 114.95.

Table 3 Mortality rate depending on the place of residence (per 1000)

	Person-time	*Death*	*Rate*	95% CI *Lower*	95% CI *Upper*
All	759 301	6 560	8.640	8.433	8.851
Rural	381 182	2 625	6.887	6.628	7.155
Urban	360 531	3 813	10.576	10.246	10.917
Rural	381 182	2 625	6.887	6.628	7.155
Small cities	75 060	619	8.247	7.622	8.923
Large cities	120 404	1 256	10.432	9.870	11.025
Paris	165 067	1 938	11.741	11.229	12.725

Note: ‘small cities’ refers to between 2,500 and 10,000 inhabitants while ‘large cities’ means more than 10,000 inhabitants.

Table 4 Effects of individual characteristics on the hazard of dying -- Cox relative risk model

	Model 1		Model 2		Model 3	
	coef	*se*	coef	*se*	coef	*se*
Urban	0,351***	*0,036*	0,266***	*0,068*	0,155*	*0,093*
Urban x t			0,006	*0,004*		
Urban x t>5					0,016	*0,107*
Urban x t>10					0,046	*0,103*
Urban x t>15					0,135	*0,103*
Urban x t>20					0,185*	*0,103*
Birth and 20						
Rural then city	0,141***	*0,046*	0,143***	*0,046*	0,209***	*0,046*
City then rural	0,136*	*0,073*	0,135*	*0,073*	0,147**	*0,073*
City	0,159***	*0,040*	0,161***	*0,040*	0,237***	*0,040*
Missing	0,024	*0,093*	0,026	*0,093*	0,089	*0,093*
Education						
Illiterate	0,014	*0,043*	0,015	*0,043*	0,011	*0,043*
Read and/or write	0,055*	*0,031*	0,055*	*0,031*	0,054*	*0,031*
Secondary education	-0,164	*0,107*	-0,164	*0,107*	-0,165	*0,107*
Missing	0,110**	*0,054*	0,110**	*0,054*	0,114**	*0,054*
Orphanage						
No Mother	0,139***	*0,045*	0,139***	*0,045*	0,142***	*0,045*
No Father	0,145***	*0,034*	0,146***	*0,034*	0,148***	*0,034*
Orphan	0,214***	*0,050*	0,214***	*0,050*	0,215***	*0,050*
Occupation						
Unskilled	0,116***	*0,036*	0,115***	*0,036*	0,123***	*0,036*
Skilled	0,147***	*0,053*	0,146***	*0,053*	0,159***	*0,053*
White collar	0,183**	*0,076*	0,182**	*0,076*	0,193**	*0,076*
Other	0,176***	*0,056*	0,175***	*0,056*	0,186***	*0,056*
Height						
Small	-0,099***	*0,037*	-0,099***	*0,037*	-0,100***	*0,037*
Tall	0,010	*0,037*	0,010	*0,037*	0,010	*0,037*
Missing	0,051	*0,071*	0,052	*0,071*	0,052	*0,071*
Military service						
Short service	0,122***	*0,032*	0,122***	*0,032*	0,122***	*0,032*
Auxiliary service	0,161***	*0,042*	0,162***	*0,042*	0,162***	*0,042*
Conditional service	-0,155*	*0,083*	-0,154*	*0,083*	-0,153*	*0,083*
Volunteer	0,173***	*0,059*	0,172***	*0,059*	0,175***	*0,059*
campagne==1	0,055	*0,050*	0,055	*0,050*	0,056	*0,050*
Number of observations	761 633		761 633		761 633	
N_fail		6 183,000		6 183,000		6 183,000
N_sub		37 239,000		37 239,000		37 239,000
risk		743 848,908		743 848,908		743 848,908
Log-Likelihood	-63 426,49		-63 424,07		-63 445,43	

note: *** p<0.01, ** p<0.05, * p<0.1
references are as follow:
For Birth and 20: both rural
For Education: read, write and calculate
For Orphanage: both parents alive
For Occupation: farmer
For Height: average height
For Military service: regular active service

Table 5 Effects of individual characteristics on the hazard of dying: sub-samples.

	No time interactions		Linear interactions		5 years interactions		10 years interactions	
	coef	*se*	coef	*se*	coef	*se*	coef	*se*
Panel A: full sample								
Urban 1 year	0,349***	*0,036*	0,259***	*0,068*				
Urban x t			0,006	*0,004*				
Urban 5 years					0,382***	*0,093*		
Urban x 5<t<10					-0,046	*0,106*		
Urban x 10<t<15					-0,095	*0,103*		
Urban x 15<t<20					0,007	*0,102*		
Urban x 20<t<26					0,033	*0,103*		
Urban 10 years							0,273***	*0,057*
Urban x 10<t<20							0,049	*0,062*
Urban x 20<t<26							0,126*	*0,073*
Panel B: stayers only								
Urban 1 year	0,482***	*0,038*	0,288***	*0,077*				
Urban x t			0,014***	*0,005*				
Urban 5 years					0,408***	*0,102*		
Urban x 5<t<10					-0,015	*0,124*		
Urban x 10<t<15					-0,025	*0,121*		
Urban x 15<t<20					0,175	*0,121*		
Urban x 20<t<26					0,183	*0,121*		
Urban 10 years							0,398***	*0,062*
Urban x 10<t<20							0,085	*0,076*
Urban x 20<t<26							0,193**	*0,090*
Panel C: city migrants versus rural stayers								
Migrant	0,176***	*0,034*	-0,382***	*0,085*	-0,402***	*0,120*	-0,152**	*0,065*
Migrant x t			0,037***	*0,005*				
Urban x 5<t<10					0,354**	*0,142*		
Urban x 10<t<15					0,545***	*0,136*		
Urban x 15<t<20					0,784***	*0,136*		
Urban x 20<t<26					0,750***	*0,136*		
Urban x 10<t<20							0,416***	*0,079*
Urban x 20<t<26							0,501***	*0,091*

Panel A:

Number of observations	761 633
N_fail	6 183,000
N_sub	37 239,000
risk	743 848,908

Panel B:

Number of observations	469 569
N_fail	3 800,000
N_sub	23 162,000
risk	462 908,090

Panel C:

Number of observations	534 923
N_fail	3 866,000
N_sub	25 435,000
risk	525 537,000

note: *** p<0.01, ** p<0.05, * p<0.1
Stayers means no migration between birth and exit of the sample
City migrants are those who live in a city at least once between birth and exit of the sample
Urban 5 year is dummy variable indicating if the individual live in a city or in the countryside by five years period. It is constructed according to the mean of number of years spend in a city.
The same holds true for Urban 10 years

Table 6 Mortality rate at a given age according to the place of residence before that age

	Place of residence at age…					
	20-24		25-29		30-34	
Age	rural	urban	rural	urban	rural	urban
25-29	6,10	9,76				
30-34	7,77	11,39	6,95	11,02		
35-39	8,51	12,92	7,40	12,90	7,35	13,02
40-46	7,80	13,00	7,44	6,85	7,28	12,71

References

Akresh, I. R. and R. Frank (2008). "Health Selection Among New Immigrants." *American Journal of Public Health* 98(11): 2058-2064.

Bengtsson, T., C. Campbell and J. Lee (2004). *Life under pressure. Mortality and living standards in Europe and Asia, 1700-1900*. Cambridge (MA), MIT Press.

Bourdieu, J. and L. Kesztenbaum (2004). "Vieux, riches bien portants. Une application de la base "TRA" aux liens entre mortalité richesse." *Annales de Démographie Historique* p. 79-105.

Cameron, A. and P. Trivedi (2005). *Microeconometrics. Methods and applications*. Cambridge University Press, Cambridge.

Chamla, M. C. (1962). "L'accroissement de la stature en France de 1880 à 1960; comparaison avec les pays d'Europe occidentale." *Bulletins Mémoires de la Société d'Anthropologie de Paris* 11e série, 6: 201-278.

Costa, D. (2003). *Health and labor force participation over the life cycle. Evidence from the past*. Chicago, University Chicago Press & NBER.

Deboosere, P. and S. Gadeyne (2005). "La sous-mortalité des immigrés adultes en Belgique : une réalité attestée par les recensements et les registres." *Population* 60(5-6): 765-811.

Elo, I. T. and S. H. Preston (1992). "Effects of early-life conditions on adult mortality: A review." *Population Index* 58: 186-212.

Farcy, J.-C. and A. Faure (2003). *La mobilité d'une génération de Français : recherche sur les migrations et les déménagements vers et dans Paris à la fin du XIXe siècle*. Paris, INED.

Fogel, R. W. (2004). *The escape from hunger and premature death*. 1700-2100 europe, america and the third world Cambridge New York Cambridge University Press.

Galloway, P. R. (1985). "Annual Variations in Deaths by Age, Deaths by Cause, Prices, and Weather in London 1670 to 1830." *Population studies* 39(3): 487-505.

Herscovici, S. (1998). "Migration and economic mobility: wealth accumulation and occupational change among antebellum migrants and persisters." *The Journal of Economic History* 58(4): 927-955.

Kalbfleisch, J. D. and R. L. Prentice (2002). *The statistical analysis of failure time data*. Hoboken (N.J.), J. Wiley & sons.

Kasakoff, A. and J. Adams (2000). "The Effects of Migration, Place and Occupation on adult mortality in the American North, 1740-1880." *Historical Methods* Volume 33, Number 2.

Komlos, J. (1998). "Shrinking in a Growing Economy? The Mystery of Physical Stature during the Industrial Revolution." *The Journal of Economic History. 58 (3)* p. 779-802.

Landers, J. (1993). *Death and the metropolis: studies in the demographic history of London, 1670-1830*. Cambridge, Cambridge University Press.

Neven, M. (1997). "Epidemiology of town and countryside: Mortality and causes of death in East Belgium, 1850–1910." *Revue Belge d'Histoire Contemporaine* 27: 39-82.

Palloni, A. and E. Arias (2004). "paradox lost: explaining the Hispanic adult mortality advantage." *demography* 41(3): 385-415.

Pinol, J. L. (1996). "Mesurer les mobilités urbaines. Strasbourg 1870-1940 : trajectoires individuelles espace urbain." *Enquête* 4, EHESS/Parenthèses, Marseille: 93-106.

Rubalcava, L. N., G. M. Teruel, D. Thomas and N. Goldman (2008). "The healthy migrant effect: new findings from the Mexican family life survey." *American Journal of Public Health* 98(1): 78–84.

Vedrenne-Villeneuve, E. (1961). "L'inégalité sociale devant la mort dans la première moitié du XIXe siècle." *Population* 4: 665-698.

Appendix

Appendix A. Descriptive statistics of the covariates used in the regressions

	N subjects	%	N observations	%	N failures	mortality rate
All conscripts	37,052		112,214		6,183	8.343
Place of residence						
Rural	18,369	49.58	38,933	34.70	2,512	6.594
Urban	18,683	50.42	73,281	65.30	3,671	10.194
Birth and 20						
Rural	20,931	56.49	50,883	45.34	2,945	6.796
Rural then city	4,337	11.71	14,974	13.34	870	10.381
City then rural	1,182	3.19	3,462	3.09	207	8.795
City	9,686	26.15	38,822	34.60	2,008	10.868
Missing	916	2.47	4,073	3.63	153	9.779
Education						
Read, write and calculate	4,304	11.62	12,168	10.84	681	7.760
Illiterate	9,044	24.41	25,950	23.13	1,494	8.200
Read and/or write	20,965	56.58	64,687	57.65	3,515	8.418
Secondary education	669	1.81	1,987	1.77	102	7.586
Missing	2,070	5.59	7,422	6.61	391	9.741
Orphanage						
Both alive	24,777	66.87	72,214	64.35	3,834	7.728
No Mother	3,253	8.78	10,182	9.07	578	9.091
No Father	6,653	17.96	21,655	19.30	1,297	9.555
Orphan	2,369	6.39	8,163	7.27	474	10.382
Occupation						
Farmer	8,340	22.51	18,219	16.24	1,119	6.412
Unskilled	20,861	56.30	66,872	59.59	3,571	8.591
Skilled	3,688	9.95	13,705	12.21	721	10.259
White collar	1,393	3.76	4,360	3.89	250	9.423
Other	2,770	7.48	9,058	8.07	522	9.654
Height						
Average height	24,692	66.64	74,475	66.37	4,141	8.400
Small	5,642	15.23	17,679	15.75	911	7.840
Tall	5,550	14.98	16,183	14.42	913	8.306
Missing	1,168	3.15	3,877	3.46	218	9.924
Military service						
Regular active service	20,556	55.48	61,956	55.21	3,079	7.910
Short service	9,526	25.71	29,113	25.94	1,795	8.636
Auxiliary service	3,741	10.10	10,348	9.22	756	8.797
Conditionnal service	1,257	3.39	3,637	3.24	184	7.259
Volunteer	1,972	5.32	7,160	6.38	369	11.293
Campagnaining						
No	34,084	91.99	101,745	90.67	5,705	8.276
Yes	2,968	8.01	10,469	9.33	478	9.236

Appendix B. Descriptive statistics of time in the sample

	Time at risk (year)						
	Total	25%	Median	75%	Mean	Survival function	Khi²
All conscripts	741,070	20	21	26	20.28	80.82	
Place of residence							
Rural	380,947	21	21	26	20.82	84.47	295***
Urban	360,123	19	21	24	19.75	77.07	
Birth and 20							
Rural	433,315	21	21	26	20.87	84.17	
Rural then city	83,808	18	21	24	19.65	76.34	
City then rural	23,537	20	21	24	20.17	80.53	319***
City	184,762	17	21	24	19.48	75.59	
Missing	15,645	16	20	23	18.54	78.39	
Education							
Read, write and calculate	87,755	21	21	26	20.59	82.06	
Illiterate	182,186	20	21	26	20.34	81.29	
Read and/or write	417,543	20	21	24	20.22	80.61	14.97***
Secondary education	13,445	21	24	25	20.81	82.49	
Missing	40,138	19	21	26	19.83	77.76	
Orphanage							
Both alive	496,098	20	21	24	20.29	82.24	
No Mother	63,580	20	21	24	19.84	79.17	83***
No Father	135,736	20	23	26	20.74	78.04	
Orphan	45,655	18	21	25	19.59	76.65	
Occupation							
Farmer	174,514	21	21	26	21.04	85.10	
Unskilled	415,671	20	21	24	20.18	80.31	
Skilled	70,281	17	21	25	19.62	76.69	130***
White collar	26,530	18	22	24	19.80	78.27	
Other	54,070	19	21	24	19.87	78.36	
Height							
Average height	492,979	20	21	24	20.22	80.69	
Small	116,197	21	21	26	20.83	81.97	10.72**
Tall	109,926	20	21	25	20.13	80.81	
Missing	21,966	16	21	26	19.74	77.70	
Military service							
Regular active service	389,265	20	21	21	19.15	82.93	
Short service	207,840	21	26	26	22.13	79.61	
Auxiliary service	85,941	26	26	26	23.24	79.32	66***
Conditionnal service	25,347	22	23	24	20.98	82.75	
Volunteer	32,674	15	20	21	17.13	75.19	
Campagnaining							
No	689,312	21	21	26	20.49	80.92	2.32
Yes	51,756	17	20	21	17.87	77.21	