Physica A 450 (2016) 748–767

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Effect of marital status on death rates. Part 1: High accuracy exploration of the Farr–Bertillon effect

Peter Richmond [a], Bertrand M. Roehner [b,*]

[a] *School of Physics, Trinity College Dublin, Ireland*

[b] *Institute for Theoretical and High Energy Physics (LPTHE), University Pierre and Marie Curie, Paris, France*

HIGHLIGHTS

- Non-married persons have higher death rates than married persons.
- This holds for all age groups and all diseases.
- Death rate by heart attack is 2.2 times higher for non-married.
- For young widowers the death rate is up to 20 times higher.

ARTICLE INFO

Article history:
Available online 21 January 2016

Keywords:
Death rate
Marital status
Widowhood
Young widowers
Social interaction
Apoptosis

ABSTRACT

The Farr–Bertillon law says that for all age-groups the death rate of married people is lower than the death rate of people who are not married (i.e. single, widowed or divorced). Although this law has been known for over 150 years, it has never been established with well-controlled accuracy (e.g. error bars). This even let some authors argue that it was a statistical artifact. It is true that the data must be selected with great care, especially for age groups of small size (e.g. widowers under 25).

The observations reported in this paper were selected in the way experiments are designed in physics, that is to say with the objective of minimizing error bars. Data appropriate for mid-age groups may be unsuitable for young age groups and vice versa.

The investigation led to the following results. (1) The FB effect is very similar for men and women, except that (at least in western countries) its amplitude is 20% higher for men. (2) There is a marked difference between single/divorced persons on the one hand, for whom the effect is largest around the age of 40, and widowed persons on the other hand, for whom the effect is largest around the age of 25. (3) When different causes of death are distinguished, the effect is largest for suicide and smallest for cancer. For heart disease and cerebrovascular accidents, the fact of being married divides the death rate by 2.2 compared to non-married persons. (4) For young widowers the death rates are up to 10 times higher than for married persons of same age. This extreme form of the FB effect will be referred to as the "young widower effect". Chinese data are used to explore this effect more closely.

A possible connection between the FB effect and Martin Raff's "Stay alive" effect for the cells in an organism is discussed in the last section.

* Corresponding author.
E-mail addresses: peter_richmond@ymail.com (P. Richmond), roehner@lpthe.jussieu.fr (B.M. Roehner).

http://dx.doi.org/10.1016/j.physa.2015.12.136

1. Introduction

Let us first define several terms which will be used throughout this article.

- The marital status of a person refers to one of the following situations: single, married, widowed, divorced. Needless to say, "single" means that the person has *never* been married for otherwise he (or she) would be widowed or divorced. These groups will be designated by the letters s, m, w, d respectively. The case of people who are married but separated or not married but cohabiting will also be considered later on albeit fairly shortly.
- For each of these groups of persons one can define a death rate in the standard way, that is to say by dividing the number of persons who die annually by the size of the group. In addition to the marital status distinction, one can order people by age group. For instance, $d_m(15:24)$ will be the death rate of married persons who are between 15 and 24 year old.
- Finally, we introduce the notion of *death rate ratio* which is the death rate of a given group divided by the death rate of married persons of same age. For instance, the death rate ratio of widowed persons in the age group $15:24$ will be:

$$\text{Death rate ratio of widowed persons: } r_w(15:24) = d_w(15:24)/d_m(15:24).$$

The expression *death rate ratio distribution of widowed persons* will refer to the curve of r_w as a function of age. Sometimes, death rate ratio distributions will also be named Farr–Bertillon distributions.

1.1. The Farr–Bertillon law

In the social sciences there are very few laws which are valid at any time and in any country. The Farr–Bertillon law[1] which states that for all age-groups married persons have a lower death rate than unmarried persons is one of them. More precisely, in all cases for which reliable data are available this law holds with error bars which are not broader than $\pm 10\%$.

At first sight, our assertion that there are few laws of this kind may seem surprising. For instance, is it not true that the frequency distribution of high incomes follows a Pareto law? Compared with the Farr–Bertillon law there are two major differences, however.

- The Pareto law contains a free parameter, namely the exponent of the power law. The Farr–Bertillon effect contains no free parameter.
- The Pareto law describes a frequency distribution whereas the Farr–Bertillon law is a relationship between two "physical" variables. In short, the Pareto law is of the same kind as the Maxwell–Boltzmann (MB) law which gives the velocity distribution of the molecules of a gas whereas the Farr–Bertillon law is similar (for instance) to Einstein's law which gives the relationship between specific heat and temperature. Needless to say, a relationship between physical variables tells us more about the system than a probability distribution.[2]

The Farr–Bertillon law [1] is named after William Farr (1807–1883) and Louis-Adolphe Bertillon (1821–1883). In 1859 Farr observed the effect on French data. Both Farr and Bertillon were among the main founders of medical demography. Bertillon's strong focus on comparative international investigations led him to recognize the existence of this effect in a broad range of countries [2]. As a matter of fact, in the one and a half century since its discovery, the Farr–Bertillon effect has been observed in *all* countries for which reliable data are available.

1.2. Measurement issues

Why did we stress the need for reliable data? The reason is that many death ratio curves (e.g. in Figs. 2c, 3c, 4, 6a,c,d,f) display huge random fluctuations which for young persons are commonly of the order of 100%. Thus, it can hardly be denied that it is a real challenge to keep the error bars under control. This can only be done by increasing the size, n, of the sample. As in physics it is the $1/\sqrt{n}$ factor which will dampen random fluctuations. However, in contrast to physics, here we cannot increase at will the number of experiments. Instead, we need to select the data with great care. For instance, Table 3b shows that for young people the data from US "Current Population Reports" (CPR) involve random fluctuations of the order of 100%. Thus, CPR data should be discarded, at least for young people. Further discussion will show that even nowadays in advanced countries it remains a real challenge to produce the kind of data that are needed to observe the FB law. It can even be said that present-day data are probably less accurate than those of 50 or 100 years ago for, as will be seen later, present-day statistics rely more and more on surveys based on population samples, that is to say on smaller values of n.

---

[1] So far, in the literature the FB law was variously referred to as the "marriage effect", the "widower effect" or the "bereavement effect". Adding to the confusion, some of these expressions were meant to describe special facets; for instance the term "bereavement effect" focuses on short-term rather than permanent effects. Here, as is standard in physics, this law will be designated by the name of its discoverers. We hope that following this usage will clarify its significance.

[2] The MB distribution for the speed of molecules is a consequence of the fact that each velocity component follows a centered Gaussian distribution. Because of the central limit theorem, Gaussian distributions are very common in the natural sciences which means that the exponential shape of the MB distribution only tells us that it belongs to this broad class rather than to the power law class. Actually, all significant physical information (about molecules masses and temperature) is contained in the width of the MB distribution.

Why is it a difficult task to measure the death rate of widowed persons and particularly of young widowers? As all death rates, the death rate $d(G)$ of widowers in age-group G is defined as a ratio:

$$d_w(G) = \frac{\text{Number of widowers who died during the year: } D(G)}{\text{Population of living widowers at beginning (or mid-year) in given age group: } P(G)}.$$

The numerator is fairly easy to measure because in all countries age and marital status are two characteristics recorded on death certificates. In contrast, it is difficult to get reliable estimates for $P(G)$. There are (at least) 4 difficulties in measuring $d(G)$ for young widowers.

(1) In the age groups under 30 there are few widowers. For instance in 1980 in the United States the 15–19 age group had only 6448 widows and 2081 widowers.[3] In the 15–24 age group there were 31,100 widows and 8050 widowers. In any census the task of identifying and counting accurately such small populations is not easy. Naturally, if instead of real censuses one relies on sample surveys, the task becomes even more difficult or altogether impossible if the samples are too small.
(2) Even though in a general way it is easier to measure death numbers than to count living persons, the fact that there are very few deaths of young widowers creates huge statistical fluctuations. Thus, in the United States in 1980 for the age-group 15–24 there were only 38 deaths of widows and 29 deaths of widowers. In the following age group 25–34 the numbers were about ten times larger, 323 and 219 respectively, but these are still small numbers. As a matter of fact, the deaths of widowed persons become "substantial" only over 65 years of age. Thus, in the age group 65–74 there were 35,630 deaths of widowers.
(3) $P(G)$ must be measured through a census but the problem with censuses is that they are based on the answers provided by the respondents. Even in countries such as the United States where censuses have been organized with much care, the enumerator relies entirely on the answers provided by the head of the household.[4] Yet, it is well known that the answers provided by the respondents are not always accurate. For instance, even such basic variables as age or the number of years spent in widowhood are not well remembered especially by elderly people. It appears that often such variables are rounded up to the nearest multiples of 10 or 5. That is why most census forms ask both the age and the year of birth. Moreover, the answers may be affected by other forms of bias. Thus, people may prefer to say that they are widowed rather than separated or divorced.
(4) In the interval between census years, most national statistical institutes carry out surveys based on population samples. The quality of such surveys greatly depends upon how well the samples are selected. In the United States, the annual "Current Population Reports" (CPR) are based on samples of some 60,000 persons, that is to say one per 20,000. Thus, even if the sample was selected properly, there will be substantial sampling errors. For a population of the order of 100,000 the sampling error of the CPR of 1960 was 37,000 that is to say nearly 40%; for a population of one million the standard error was still 12% [3, p. 6, Table D]. As a result, good measurements of $d(G)$ for young widowers can be obtained only in census years. A more detailed analysis is given below in Table 3b.

1.3. Farr–Bertillon effect in the 19th century

In this subsection and in the next we show some of the results due to Louis-Alphonse Bertillon. As already observed, contrary to William Farr, Bertillon was a "comparativist". After having identified this effect in France, his main concern was to see if it was also present in other countries.

As the author does not give the data for separate calendar years, we cannot compute the error bars. Around 1850 the populations of France, Belgium and the Netherlands were 36, 4.5 and 3.2 millions respectively. Thus, because of the difference in size one would expect larger random fluctuations in Belgium and the Netherlands than in France, especially for young and very old widowers. The most conspicuous feature displayed by these graphs is the great difference between the widower curves on the one hand and the never-married curves on the other hand. The widower ratios are higher for almost all ages and particularly for ages under 40. This feature will be confirmed for the 20th century by the graphs given subsequently (this time with the benefit of error bars).

1.4. Bonds between parents and children

If the bond between husband and wife plays a role in the FB effect it seems plausible to expect the ties between parents and their children to have a similar effect. This conjecture is confirmed by the data in Table 1.

The conceptual framework introduced by Durkheim in 1897 (that is to say some 18 years after Bertillon's statistical observations) postulated a close connection between suicide rates and the strength of interactions. In his work of 1897 Durkheim improved the dichotomic observation of Table 1 by showing that suicide rates decreased when the number of children increased.

3 There are less widowers than widows because at the same age there are less married men than married women.
4 In fact, after 1990 the census forms were mailed to the persons; visits by enumerators were limited to a few households.

Table 1
Effect of husband–wife ties and parents–children ties on suicide rates, 1861–1868.
Source: Bertillon [5, p. 474].

Situation	Male	Female	Male	Female
Married with children	20	4.5	1	1
Married without children	47	16	2.4	3.6
Widowed with children	52	10	2.6	2.2
Widowed without children	100	23	5.0	5.1

Notes: The table gives average suicide rates (per 100,000 people) in France over the 8-year time interval 1861–1868. The two columns on the right-hand size repeat the same data with a normalization based on the situation "married with children". If one accepts the explanation introduced by Emile Durkheim [6] that it is the severance (or lack) of bonds and especially of family bonds which is the main factor in the phenomenon of suicide, then these data allow us to compare the respective strengths of the bonds between husband and wife on the one hand and between parents and children on the other hand. The fact that the suicide rate is almost the same for married persons without children as for widowed persons with children suggests that the parents–children and husband–wife bonds are of same strength.

1.5. Systems science perspective

In answer to a question raised by one of the referees, we would like to explain why this paper is published in a physics journal rather than in a journal specialized in demography. This is indeed a natural question and is in close connection with the very rationale of econophysics.

How is our approach related to physics? It does not use the mathematical formalism of theoretical physics but it uses the methodology of *experimental physics*. What made physics a very successful field was its ability to focus on a phenomenon and to explore it through appropriate experiments until it is well understood.

This is what we do in the present paper for the FB effect.

It is usual in physics to study any given phenomenon in *several* of its manifestations. For instance, taking the example of gravity, physics does not just consider the fall of apples, it seeks to explain the fall of all objects, including the "fall" of the Moon towards the Earth or that of Mars towards the Sun. So, in this paper, while studying the effect of social contacts on the length of life, we do not limit ourselves to human populations, we also consider drosophila, ants and the Raff effect for microorganisms. Such a broad systems perspective is not usually found (nor accepted) in the more specialized journals.

In short, the systems science perspective is not only a key component of physics, it can also be seen as an essential and fruitful contribution of econophysicists to the diverse fields in which they work.

1.6. "Explanation" of the Farr–Bertillon law

What explanation of the Farr–Bertillon law can be offered at this point?

- In this paper we did not try to provide a model. We think one should first set out the facts as accurately as possible.
- True, our observations are guided by a conjecture according to which it is the strength of the inter-individual interactions which is the key-factor. The fact that, as shown in Table 1, the suicide rate decreases not only through the presence of a spouse but also through the presence of children suggests that male–female interaction is not the only ingredient. Martin Raff's "Stay alive" effect as well as the experiments on ants described in Ref. [7] also support such an interaction-based conjecture.

2. Toward accurate observations of the Farr–Bertillon law

2.1. Mid-age groups versus young age groups

In the previous section we emphasized the fact that the Farr–Bertillon effect holds with a level of precision akin to what one is used to in the natural sciences. However, in order to reduce the error bars as much as possible an appropriate methodology must be used. In this respect age-groups over 35 and age-groups under 35 will require different techniques.

- To estimate the sizes of the age groups over 35 one does not necessarily need to use censuses. Estimates from surveys based on population samples may be sufficient at least if the samples are "not too small". This will allow observations over time intervals containing a substantial number, k, of inter-census years. For the averages computed over such time intervals, the error bars will be reduced by the standard $1/\sqrt{k}$ factor.
- On the contrary, in the investigation of the young widower effect one needs to focus on age groups under 35 and, as already mentioned, this requires to rely on population data from decennial censuses. To some extent, the procedure based on census data is also needed for elderly age groups over 75 because of their small size.

Table 2
Summary of the observations of the Farr–Bertillon effect.

Fig.	Country	Period	Population estimates	Error bars	Shape of w/m	Quality stars
1a	France	1856–1865	?	No	\	*
1b	Belgium	1851–1860	?	No	\	*
1c	Netherlands	1850–1859	?	No	⋀	*
2a–c	USA	1996–2010	?	Yes	?	**
3a–c	USA	1940, 1950, 1960	Census	Yes	⋀	***
4	USA	1980, 1990, 2000	Census	Yes	⋀	***
5	USA	2005–2010	ACS	Yes	⋀	***
6	USA	1980, 1990	Census	Yes	\ ⋀	**
7a	France	1968–1993	?	Yes	\	**
7b	France	1981–1993	?	No	\	*

Notes: The column "Population estimates" indicates how the populations by age and marital status have been estimated. ACS means "American Community Survey". An interrogation mark in this column means that the technical notes of the publication failed to explain how the estimates were computed. This concerns especially inter-census years. The column "Error bar" indicates if it was possible to estimate the standard deviation of the death rate ratios. The column "Shape of w/m" indicates whether the curve for the death rate ratio of widowers is steadily decreasing (\) or shows a maximum for the second youngest age group (⋀). It is the last case which prevails in the observations of highest quality. The column "Quality stars" gives an estimated quality index for each observation: 3 stars (highest quality) = census based + error bars, 2 stars = uncertainty about the origin of the data + error bars, 1 star = same uncertainty and no error bars. It can be noted that similar death statistics by marital status and age are also available for England (see Ref. [9], review of the Registrar General, National Center for Health Statistics [10], Registrar General [11]) and Germany (see Statistisches Jahrbuch [12]).

In short, there will be two phases in our investigation. In a first phase we will focus on the central part of the age interval (30–60) and use as many years as possible to get the smallest error bars.

In the second phase, we will use accurate population data available for only a few years. This will give the death ratio for young age-groups. Though this procedure will of course also provide results for central age-groups, they will be less precise than those computed in the first phase.

2.2. *Methodological options*

In the present paper we perform repeated observations. At first sight one may think that they should be aggregated. However, as these observations are not performed under identical conditions (see Table 2) lumping them together would lead to unpredictable and uncontrollable results. As explained below, this is a widespread difficulty in the social sciences.

This difficulty is best seen in review papers. Most often, the authors of such papers report conflicting results obtained by different researchers but without describing the conditions under which the observations were made and how the data were analyzed. Inevitably, this makes readers uncomfortable. One gets the feeling of being confronted with a soft, multiform, shapeless and labile world about which no clear, univocal statement can ever be made. This is a great source of concern because reproducibility is a crucial requirement in any science. Let us illustrate this point through an example. A paper by Kposowa et al. [8] found that no additional risk of suicide is *significantly* associated with the marital status of widowed or never-married persons. Such a conclusion is at variance with the results reported consistently by numerous former and subsequent studies, including the present one. If presented without specific explanations about its methodology, this study would give the impression that even the most unlikely claim can be made and sustained. A closer look reveals that, in contrast with most other investigations, this one does not rely on aggregated data but on a multivariate analysis of individual data. The sample contains only 216 suicide cases. As such a small sample implies broad confidence intervals it is hardly surprising that the study could not find any significant connection between marital status and suicide rates. This does not mean that the connection does not exist but rather that the data used in this study were dominated by background noise.

Only observations of same nature and quality can possibly be lumped together. Thus, in the observations listed in Table 2 it would be possible to lump together the observations 3 and 4. However, from 1940 to 2000 they would span a time interval of 60 years during which important population changes took place in the United States. By keeping these observations separate one can control whether or not there was a possible shift.

Before carrying out the program outlined in Table 2, some preliminary tests are required. In the previous discussion we said that population estimates based on surveys may be acceptable provided that the samples are "not too small". Obviously, one needs to clarify what is meant by this expression. This will be done in the next section.

3. Sampling errors for population estimates

The expression "sampling errors" corresponds to measurement errors due to purely random fluctuations. However, we will see that for some sample estimates there are also non-sampling errors which refer to more or less systematic biases. As an example, one can mention the response rate. Nowadays, once the sample has been selected the forms are mailed to the respondents. Not all of them will reply, however. In the United States, response rates usually range between 80% and 95%. The persons who do not respond most likely are "unstable" households who move frequently and for that reason may not have received the form and also elderly persons who are in hospital or nursing homes.

Table 3a
Statistical sources for US population by marital status and age.

Year	Source	Size of sample
Census years		
1900–1940	Historical statistics of the US (p. 20–21)	Whole population
1950, 1960	Historical statistics of the US (p. 20–21)	25% sample
1970	Historical statistics of the US (p. 20–21)	5% sample
1980	Census volume PC80-1-D1-A	Whole population
1990	Census volume CP-1-1	Whole population
2000	Census table PCT007 on "FactFinder"	20% sample
2010	*Not recorded*, replaced by ACS (see below)	
Inter-census years		
1901–1959	*No data are available*	
1961–2004	Current Population Reports (CPR)	33,000–57,000
2005–	American Community Survey (ACS) on "FactFinder"	2.5 million

Notes: The population samples used in the CPR are much too small to reflect widower populations under age 40 (see Table 3b). This means that before 2005 there are in fact no appropriate data for inter-census years. Incidentally, it can be observed that the data for marital status by age which are published in the annual volumes of the "Statistical Abstract of the United States" are identical to those published in the CPR (P20 Series). The only difference is that the age-group 17–18 is omitted. This omission is probably motivated by the fact that for this age-group the estimates would be fairly poor. However, the comparisons performed in Table 3b show substantial discrepancies even for older age groups up to 35–44. "FactFinder" which is mentioned in the table refers to a search engine for statistical tables which is available on the website of the US Census Bureau.

Table 3b
Percentage errors in various estimates of widower population.
Source: Census 1980: Table 264 in the following census publication volume: US summary, Ch. D, Section A (available on the website of the US Census Bureau); CPR 1980: Series P-20, No 365, survey of March 1980 (issued in October 1981); Census 2000: Table PCT007 available on the FactFinder website of the US Census Bureau; CPR 2000: Series P-20, No 537 (issued in June 2001). ACS: Table B12002 available on the FactFinder website of the US Census Bureau; ASES 2010: "America's Families and Living Arrangements, Supplement", 2010. Many thanks to Dr. Rose Kreider from the US Census Bureau for her help.

Year	Source	Sample size						
1980								
Age			15–17	18–19	20–24	25–29	30–34	35–39
	Census	100%	992	1,089	5,970	11,759	16,531	22,337
	CPR	0.02%	0	0	2,000	8,000	11,000	19,000
	$\frac{\text{Census-CPR}}{\text{Census}}$		100%	100%	66%	32%	33%	15%
2000								
Age				15–19	20–24	25–29	30–34	35–44
	Census	100%		n.a.	n.a.	n.a.	n.a.	n.a.
	Census	20%		13,814	19,376	19,604	26,939	106,135
	CPR	0.02%		3,000	0	9,000	15,000	96,000
	$\frac{\text{Census-CPR}}{\text{Census}}$			78%	100%	54%	44%	10%
2010								
Age			15–17	18–19	20–24	25–29	30–34	35–39
	Census	100%	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Census	20%	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	ACS	1%	825	1,372	4,572	9,199	15,876	28,757
	ASES	0.05%	5,000	3,000	3,000	21,000	28,000	29,000
	$\frac{\text{ACS-ASES}}{\text{ACS}}$		−506%	−119%	34%	−128%	−76%	−1%

Notes: CPR means "Current Population Reports"; ACS means "American Community Survey"; ASES means "Annual Social and Economic Survey"; n.a. means "not available". The ratios (Census-CPR)/Census represent the errors in CPR estimates. The ratios (ACS-ASES)/ACS can also be seen as roughly representing the errors in ASES estimates. The sample size is given as a percentage of the total US population. In the 2000 census the question of marital status was not asked on the short form sent to all people but only on the long form filled by about 20% of the population. In the census of 2010 the marital status question was not asked at all. It was replaced by the ACS, yet with lower accuracy due to a sample size which is only about 1% of the US population. Thus, surprisingly, over the past two decades census data about marital status by age have become less and less accurate. Incidentally, it can be observed that the CPR data are *systematically* below the census data which shows that the differences cannot solely be explained as being due to random sampling errors; there must also be a non-sampling error component.

Table 3a describes US statistical sources. Their accuracy will then be ascertained through a number of tests performed in Table 3b.

Table 3a shows a great range of variations in the size of the samples: from CPR to census the sample size is multiplied by 10,000. In terms of the $1/\sqrt{n}$ factor, this will result in a division by one hundred of random background fluctuations.

The main conclusion that should be drawn from Table 3b is that the errors due to limited sample size are huge and that they have not only a (random) sampling component but also a non-sampling (i.e. non-random) component.

3.1. Discussion of computational methods for estimating populations

Before we close this section about population estimates an additional observation is in order concerning computational methods.

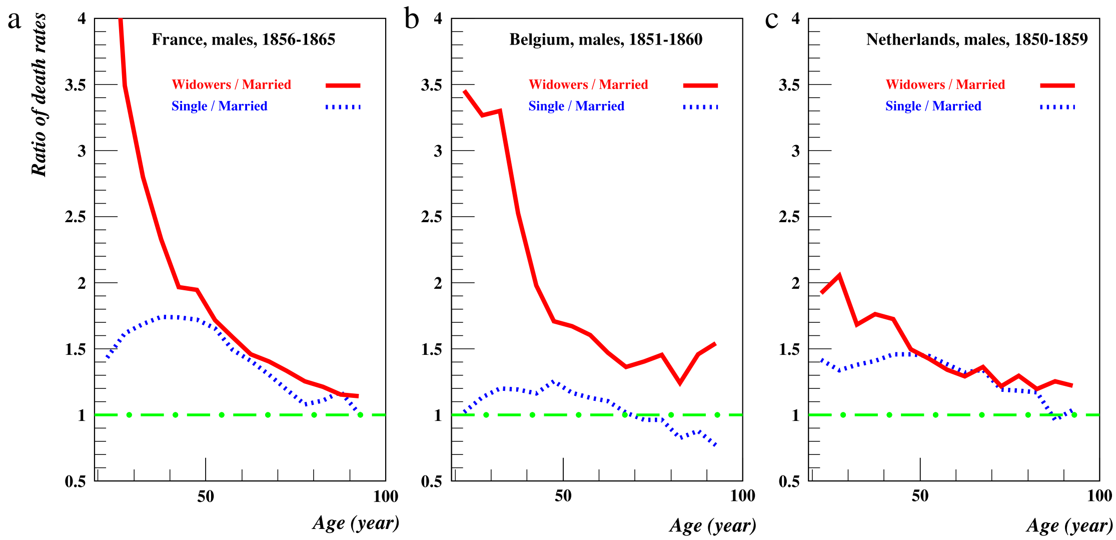

Fig. 1. Death rate ratio according to marital status. The data points correspond to 16 age groups ranging from 20–25 to 95–100. It can be seen that the ratio widowers/married differs from the ratio single/married both in shape and in magnitude. Incidentally, Bertillon was suspecting a possible statistical bias for the death rates of single persons in Belgium because they are two thirds the size of those in France and the Netherlands. In this graph as in the rest of the paper "single" means "never married".
Source: Bertillon [2]; the graph for France appears also in Ref. [4, p. 781].

At first sight it might seem that in the intervals between census years it is easy to *compute* the sizes of age-groups. Indeed, based on the numbers of deaths, marriages, divorces in each age-group, one should be able to predict the sizes of relevant age-groups. Such a procedure which would permit to follow each age group year after year until the next census may work in some countries, but in the United States it does not. There are three main obstacles.

(1) One does not know the flows of illegal immigrants. Although this difficulty exists in all countries it is more or less serious depending on the magnitude of illegal immigration.
(2) The annual data about marriages and divorces are known to be fairly incomplete in some US states. Until 1996 total divorces were reported by the Federal Government. Subsequently, it ceased to publish national divorce data.
(3) Deaths which occur overseas are not included in the death numbers published by the US Census Bureau. In other words, the deaths of US soldiers in Europe, Korea, Vietnam, Afghanistan or Iraq were not included in annual death statistics.

It is true that fatality data are published by the Pentagon. However, such data are incomplete in two respects.

- Firstly, the Department of Defense does not publish official data for the fatalities among civilian contractors working for the armed forces. Whereas, during the Vietnam War the proportion of military personnel to civilian personnel was 6:1, during the occupation of Iraq it was almost 1:1 [13]. In addition to the personnel under contract there are also persons who are not considered as contractors. For instance, one can mention news correspondents, businessmen, embassy personnel, Peace Corps affiliates, members of the Young Men's Christian Association (YMCA), and so on.[5] Young age groups will be particularly affected by the omission of overseas fatalities in death data.
- Prior to 1980 the US Department of Defense did not publish worldwide fatality data. This point was made clear in 1993 when a data revision was announced. Previously it had been said that 54,246 soldiers had died in the Korean War. According to the revision, there had been 36,516 deaths in Korea and 17,730 worldwide outside of the war theater.

Globally the total of omitted overseas deaths is certainly much smaller than the deaths of illegal immigrants, however, as most of those deaths are due to accidents, their omission will considerably affect the death rate by accident of young age-groups.

In spite of these difficulties, computational methods are commonly used. For instance, in France population numbers by marital status and age were computed for every year from 1901 to 1993 [14]. The main problem with such estimates is that it is impossible to control their quality. Usually, in such calculations one needs to make some assumptions. If for some reason (e.g. omitted overseas deaths) these assumptions are not correct then the results of the calculation will be biased. This is a non-sampling error which will not be removed by taking averages over several years (at least if the bias persists).

On the contrary, for population estimates based on samples, the statistical uncertainty is well known and in addition it can be reduced by averaging over several years. Most of the data shown in this paper have been obtained in that way. The only exceptions are Figs. 1, 2, 7.

5 However, in recent years the US State Department has published the names and cause of death of Americans who die abroad.

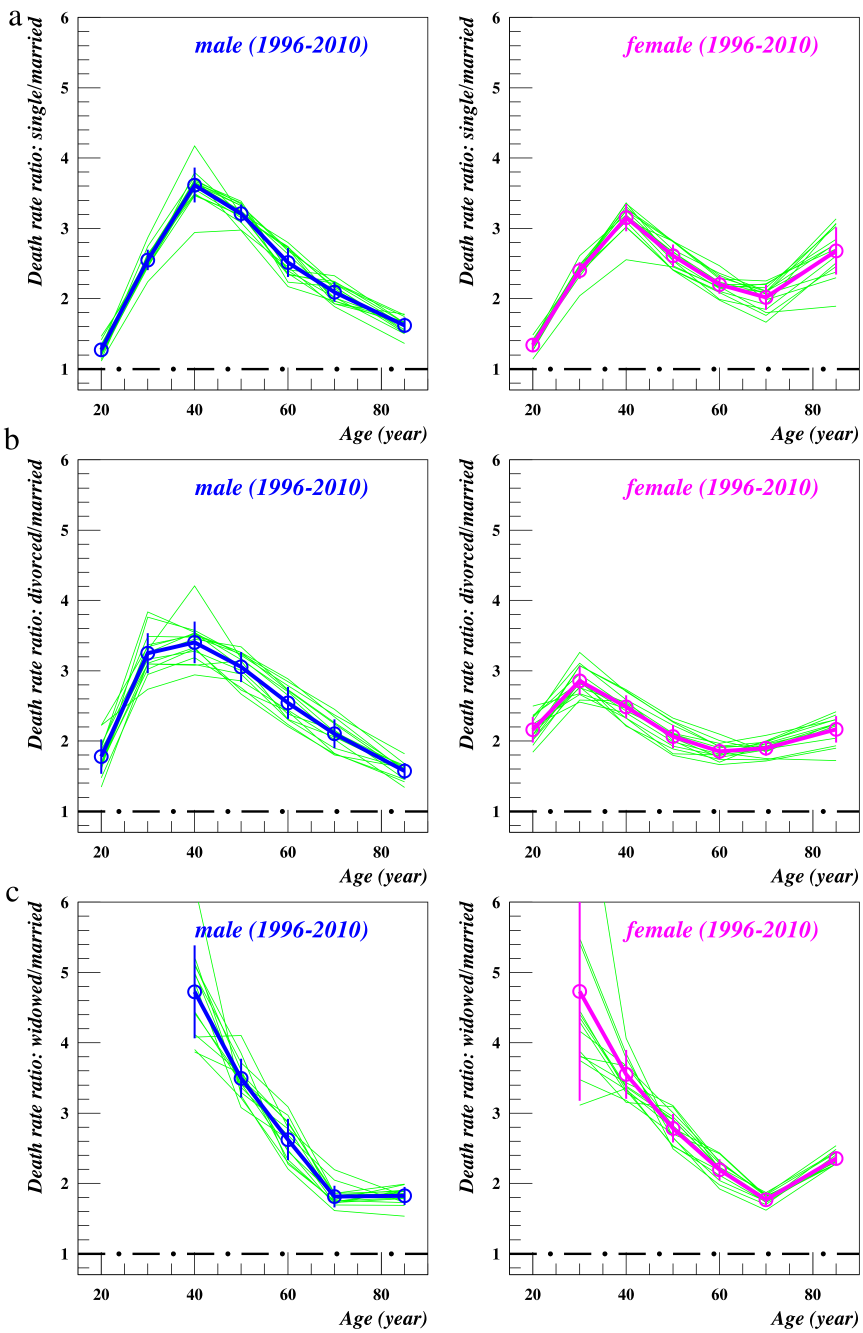

Fig. 2. Death rate ratio according to marital status in the United States. Intercensus years. From top to bottom: single/married, divorced/married, widowed/married. The thin (green lines) are the yearly curves for the 15 successive years. The thick lines show averages over the 15 years. During this period the ratios did not display any trend, there were only random fluctuations. There are 7 age groups ranging from 14–24 to 64–75, >75 but in the source no data are given for the youngest age groups of the w/m case. In this graph as well as in all subsequent graphs the length of the error bars is $\pm 1.96\sigma$ (where σ is the standard deviation of the average) which corresponds to a probability confidence level of 0.95. For widowed persons there are no data points for young age groups because for intercensus years there are no reliable estimates of the corresponding populations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Source: National Vital Statistics Reports. Deaths: Final Data. Successive years from 1996 to 2010 [15]. National Center for Health Statistics. The publication gives the death numbers and the rates. How were these rates computed? The "Technical notes" attached to the table do not give any specific information. They say only that the population data "were produced under a collaborative arrangement with the US Census Bureau".

4. Phase 1: Mid-age part of the Farr–Bertillon distribution

Table 3b shows that, except for middle age-groups, the accuracy provided by the "Current Population Reports" is fairly low especially for widowed persons. Another concern is the existence of a systematic non-sampling error component. In

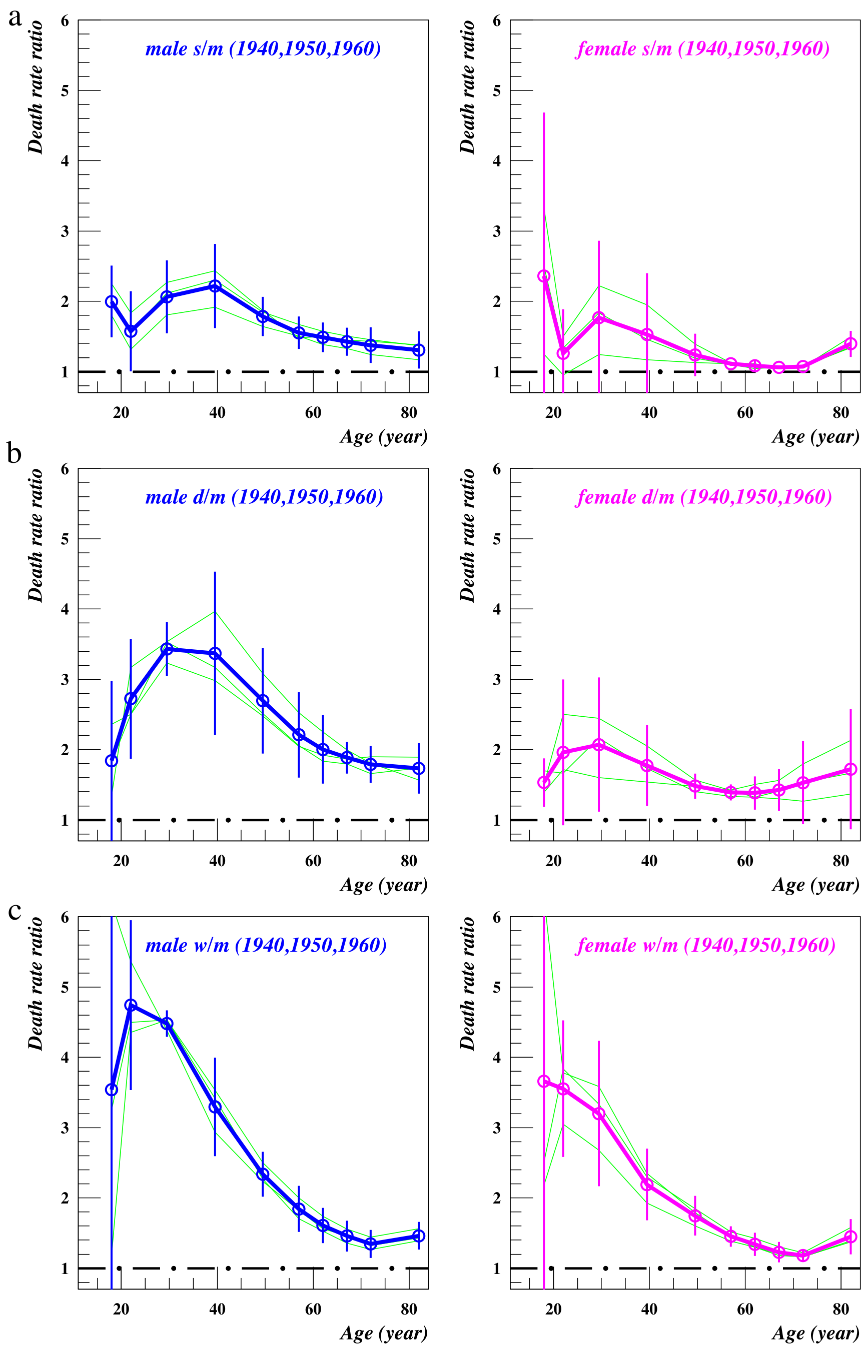

Fig. 3. Death rate ratio according to marital status in the United States. Census years: 1940, 1950, 1960. From top to bottom: single/married, divorced/married, widowed/married. The thin (and green) curves are for each of the 3 years while the thick lines with the round circles show their average. There are 10 age groups: <20, 20–24, 25–34, 35–44, 45–54, 55–50, 60–64, 65–69, 70, 74, >75. As expected, the error bars become fairly large for young age groups, particularly for young widowers and widows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Source: Grove and Hetzel [16, p. 334]. The publication gives directly the rates.

fact, we do not really know what population estimates were used to compute the death rates given in the "National Vital Statistics Reports" (NVSR) that were used in Fig. 2. The technical notes of the reports say only that "the populations used to calculate death rates were produced under a collaborative arrangement with the US Census Bureau". At least, this sentence suggests that the population estimates were not merely drawn from the CPR. Probably the CPR were used as a starting point and, in some (unspecified) way were corrected for small age-groups. The omission in NVSR data of the youngest age groups of widowed persons is a cautious and sensible step.

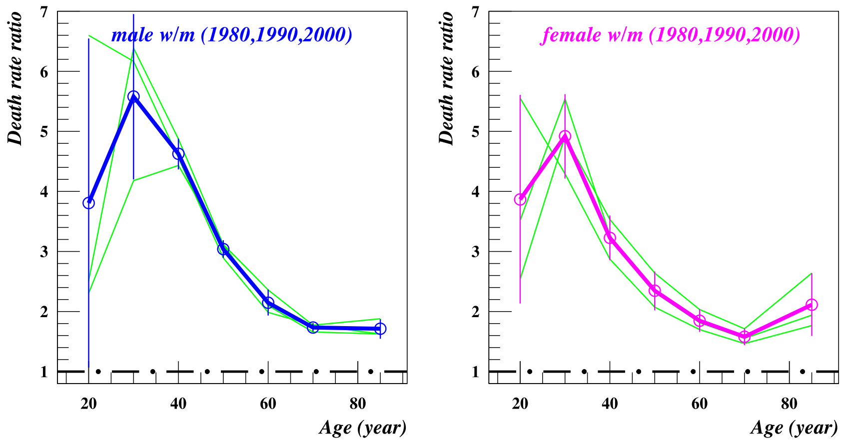

Fig. 4. Death rate ratio for widowed persons in the United States. Census years: 1980, 1990, 2000. The thin (and green) curves are for each of the 3 years while the thick lines with the round dots show their average. The age groups are the same as in Fig. 2. As expected, the error bars become fairly large for young widowers and widows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Source: Census population data: 1980: Vol. PC80-1-D1-A (Table 264, p. 1-67); 1990: General population characteristics, Vol. CP-1-1, (Table 34, p. 45); 2000: "American FactFinder" website of the US Census Bureau, Table PCT007. Mortality data: 1980: Vital Statistics of the United States 1980, Vol. 2, part A, Table 1-31 (p. 315); 1990: Vital Statistics of the United States 1990, Vol. 2, part A, Table 1-34 (p. 387); 2000: National Vital Statistics Report, Vol. 50, No. 15, 16 Sep. 2002.

5. Phase 2: The young widower effect

In this section we present three sets of results.

- The first series of graphs (Fig. 3) is based on the censuses of 1940, 1950 and 1960. For these data the rates were computed by the US Department of Health and published in the data compilation done by Robert Grove and Alice Hetzel [16].
- The second series of graphs (Fig. 4) is based on the censuses of 1980, 1990 and 2000. Actually in the census of 2000, marital status data were asked only on the so-called "long-form" which was distributed to 20% of the households.
- In Figs. 3 and 4 the error bars for widowed persons in the young age groups remain very large. In an attempt to reduce them, we use data for 6 successive years. These data are based on the "American Community Survey" which is answered by about 1% of the households. Thanks to the 6-year interval, the error bars are notably reduced.

It should be noted that in all the 12 graphs of Figs. 2 and 3 the scales of the x and y axis are exactly the same, which allows easy comparison of the magnitude of the FB effect. In the 4 graphs of Figs. 4 and 5 the scales are also exactly the same. Moreover, in Figs. 2, 3 and 4, 5 the scales are almost the same: the only slight difference being that the vertical axis $(1, 6)$ was replaced by $(1, 7)$.

We see that of the three effects the widower effect is always the strongest whereas the effects for divorced and single persons are more or less of same amplitude.

5.1. Error bars

As stated in the caption of Fig. 2, the lengths of the error bars are $\pm 1.96\sigma$ (average). The standard deviation of the average was computed by dividing the standard deviation of k annual curves $Y_j,\ j = 1, \ldots, k$ by the standard $1/\sqrt{k}$ factor. However, this factor is correct only when the Y_j are not correlated. While there is indeed a low correlation for young age groups, for older age groups there is an average correlation $r_m = 0.90$. For these data points the factor $1/\sqrt{k}$ should be replaced by $f = \sqrt{1 + (k-1)r_m}/\sqrt{k}$ [17, p. 45]. With $k = 6$ and $r_m = 0.90$ the factor f is almost equal to 1. In other words, except for young age groups, the error bars shown in the graphs underestimate the actual confidence intervals. On the other hand, using the factor f everywhere would result in overestimating the confidence interval for young age groups, the only ones which really matter in this respect.

5.2. What is the influence of cohabitation and separated couples?

In recent decades the traditional picture of family life has become more complicated due to the following trends.

(1) In 1960, 72% of all American adults were married; in 2012 just 50% were.
(2) During the same time interval, the number of cohabiting non-married couples of opposite sex jumped from 1.1% to 11%. Note that because different states may not have the same definition of cohabitation the last percentage may be subject to an error margin of about $\pm 10\%$.

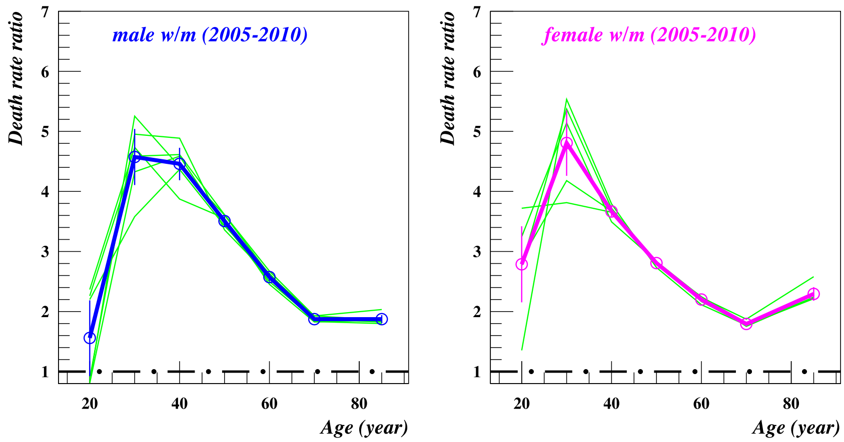

Fig. 5. Death rate ratio of widowed persons, United States, 2005–2010. The thin (and green) curves are for the 6 successive years while the thick lines with the round dots show their average. The age groups are the same as in Fig. 2. The length of the error bars is $\pm 1.96\sigma$ which corresponds to a probability confidence level of 0.95. The data used in this graph for the populations are not census data but are based on samples of about 2.5 million respondents. This "experiment" confirms the existence of a dip for the youngest age group 15–24. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Source: Populations: Starting in 2005 data by marital status and age are provided by the "American Community Survey" and are made available on the "American FactFinder" website set up by the US Census Bureau (Table B12002). Mortality data: National Vital Statistics Reports entitled "Deaths: Final Data". In 2015 the most recent year available was 2010.

(3) Finally around 2005, in about 8% of married couples one of the spouses was not present. In 1980 this proportion was about 6%.

Needless to say, such trends are by no means special to the United States; they are shared by many other western countries. However, the trends are perhaps more surprising in the United States because traditionally this country has put a strong social emphasis on family life. In this respect it can be recalled that in 2014 three states, Mississippi, Michigan, and Florida, still had laws against cohabitation by opposite-sex couples.

How do the previous trends affect the interpretation of our results?

We will successively consider the effects of cohabitation and separation.

5.3. Implication of cohabitation

In itself point 1 will not affect our results but in fact it is strongly connected with the second point: those people who do not get married are doing so because they are living together without being married.

To make the argument clearer let us make the following simple assumptions. We assume that the *real* death rates in the married, single and widowed classes are 1–3 per 1000. In addition we assume that 50% of the persons registered as single or as widowed are in fact cohabiting with a partner. For the sake of simplicity we assume that there are 2000 single and 2000 widowed persons.

Under these assumptions, what are the death rates, d_{me}, and death rate ratios, r_{me}, that will be measured and how do they compare to the real death rate ratios r_{re} (the subscripts *me* and *re* mean "measured" and "real" respectively)?

- **Death rate ratio of single persons**

$$d_{me}(s) = (2+1)/2 = 1.5 \rightarrow r_{me}(s) = 1.5/1 = 1.5, \quad \text{whereas } r_{re}(s) = 2/1 = 2.$$

- **Death rate ratio of widowed persons**

$$d_{me}(w) = (3+1)/2 = 2.0 \rightarrow r_{me}(w) = 2.0/1 = 2.0, \quad \text{whereas: } r_{re}(w) = 3/1 = 3.$$

In other words, due to cohabitation our measurements will underestimate the actual death rate ratios of single, and widowed persons. The same argument applies of course to divorced persons.

5.4. Implication of separation

In the present argument we suppose that there is no cohabitation which means that the death rates of single and widowed persons are correct. In addition, in the same way as above, we assume that 50% of the married persons are in fact separated. With the same real death rates as above what will be the measured death rate ratios? Whereas previously, the numerators of the death rate ratios were affected, this time the denominators are affected.

Table 4
Ranking of causes of death according to their average death ratio.
Source: Same sources as for Fig. 6.

Marital status						Average (all 5 causes)
Never married						
	Suicide	Heart	Cerebrovasc.	Motor veh.	Cancer	
Death ratio s/m	1.8 ± 0.4	1.8 ± 0.3	1.76 ± 0.4	1.5 ± 0.1	1.3 ± 0.1	1.6 ± 0.1
Widowed						
	Suicide	Motor veh.	Heart	Cerebrovasc.	Cancer	
Death ratio w/m	4.1 ± 1.7	4.0 ± 2	3.1 ± 1.4	2.4 ± 0.7	1.6 ± 0.4	3.0 ± 0.4

Notes: The figures given in this table are averages over the 7 age groups considered in Fig. 6. For heart disease and cerebrovascular accidents taken together the average ratio is 2.2. Overall the death ratio for w/m is about twice the death ratio of s/m. The error bars are for a probability confidence level of 0.95. The "pulmonary disease" cause of death has not been included in this ranking because it has very large fluctuations: it coefficients of variation is 40% for s/m and 102% for w/m respectively. The 6 causes of death under consideration correspond to the following code numbers in the 9th Revision of the International Classification of Diseases of 1975: heart diseases: 390–398 + 404–429, cancer: 140–208, cerebrovascular diseases: 430–438, pulmonary diseases: 490–496, motor vehicle accidents: $E810$–$E825$, suicide: $E950$–$E959$.

- **Death rate ratio of single persons**

$$d_{me}(m) = (1+2)/2 = 1.5 \rightarrow r_{me}(s) = 2/1.5 = 1.33, \quad \text{whereas: } r_{re}(s) = 2/1 = 2.$$

- **Death rate ratio of widowed persons**

$$d_{me}(m) = (1+2)/2 = 1.5 \rightarrow r_{me}(w) = 3.0/1.5 = 2.0, \quad \text{whereas: } r_{re}(w) = 3/1 = 3.$$

In other words, separation will also make our measurements underestimate the actual death rate ratios.

Because the two effects go in the same direction, their combination should also result in underestimating the real death rate ratios. An additional conclusion is that if we see a weakening of the Farr–Bertillon effect in the coming decades it may well be a statistical artifact due to persistent cohabitation and separation trends.

6. Death ratios by marital status and age for selected causes

In this section we consider death rate ratios according to selected causes of death. Previously we have seen that the death ratios for never-married and divorced persons are somewhat similar. Therefore, the present investigation will be restricted to never-married and widowed persons.

First, we must decide what causes of death we are going to select. The source offers 9 causes. The three that we left out were: "All accidents", "Other accidents except motor vehicle", "Homicide". We discarded the first two because of their vagueness and the third because it implies an exogenous factor and constitutes a different phenomenon.

Regarding the 6 causes that we retained, their full specifications are as follows:

(i) Diseases of heart (ii) Malignant neoplasms, including neoplasms of lymphatic and hematopoietic tissues (iii) Cerebrovascular diseases (iv) Chronic obstructive pulmonary diseases (v) Motor vehicle accidents (vi) Suicide. To what extent do the 6 selected causes taken together represent all causes? For married persons, in the age-group 15–24 the cumulative death rates (per 100,000) of the 6 selected causes total 51; that is 30% less than the total for all causes which is 74.

Once again, we see that the effect is strongest for widowers. At the other end of the spectrum one is not surprised to see that the effect is smallest for cancer because this is a disease whose evolution (most often) is fairly slow; therefore, except for persons in an advanced stage of the disease, their health will not be much affected by a widowhood shock.

The main difficulty with data by cause of death is the fact that the death numbers are fairly small which creates large fluctuations. In an attempt to smooth them out as far as possible we lumped together not only successive years (as was already done in previous graphs) but also the two genders.

Incidentally, it can be noted that death numbers by cause of death, marital status and age exist from 1979 to 1993, but only 1980 and 1990 can be used because for the populations we must rely on census data.

The graphs in Fig. 6 can help us to better understand the origin of fluctuations observed earlier in young age-groups of the curves for widowed persons. If we denote the first two data points of these curves by r_1 and r_2, we see that in some graphs $r_1 > r_2$ whereas in others it is the opposite. In the age-group 15–24 the three leading causes of death are motor vehicle accidents with a rate of 31 per 100,000 followed by suicide with a rate of 11 and cancer with a rate of 5.1. The fact that vehicle accident fatalities are a fairly volatile variable (for instance seat belts regulations may reduce death numbers) may explain the fluctuations of r_1 and r_2.

6.1. Ranking of causes of death according to death ratios

Because almost all death ratios documented in Fig. 6 are larger than 1 it makes sense to consider averages over all age groups. This will allow a ranking of the causes of death according to their death ratios (Table 4).

Because the w/m death ratios are based on smaller population numbers than the s/m ratios they have higher volatility.

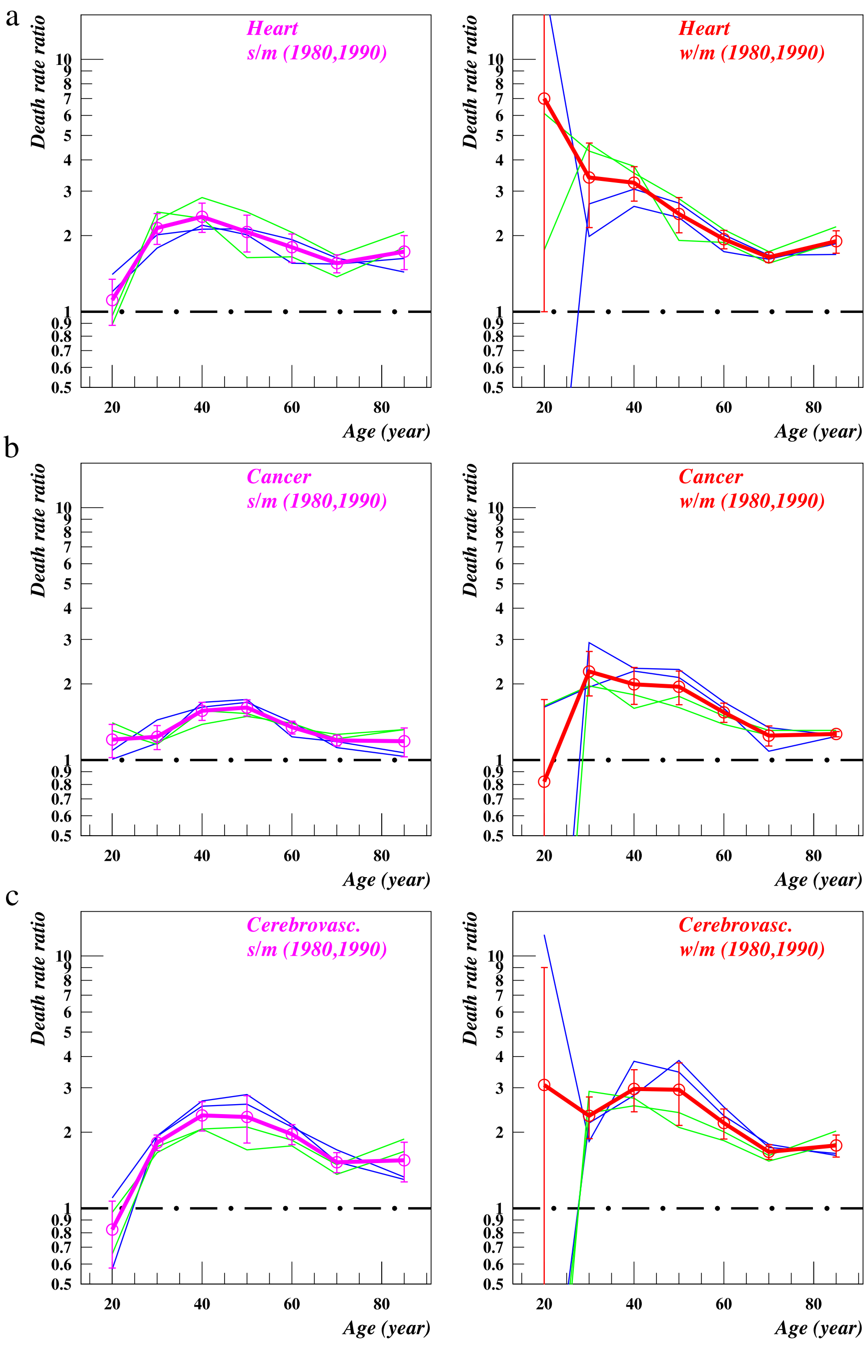

Fig. 6a. Death rate ratio for different causes of death, United States. The s/m graphs on the left-hand side are for single/married, whereas the w/m graphs are for widowed/married. The graphs show 1980 and 1990 data for males (thin blue lines) and for females (thin green lines). The thick lines represent the averages of the 4 series. The age groups are the same as in Fig. 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Source: 1980: Vital Statistics of the United States, 1980, Vol. 2, Part A, table 1-31, p. 315–324. 1990: Vital Statistics of the United States, 1990, Vol. 2, Part A, table 1-34, p. 387–400. In order to compute the rates we used census population data which is why the analysis is restricted to census years.

6.2. Marital ties as an effective drug

Can marriage be considered as an effective multipurpose drug? Yes and no.

"No" for a very obvious reason: it can only reduce the death rates of persons who are not already married. What proportion do non-married persons (to be distinguished from "never-married") represent in the age-group 65–74? In 2005

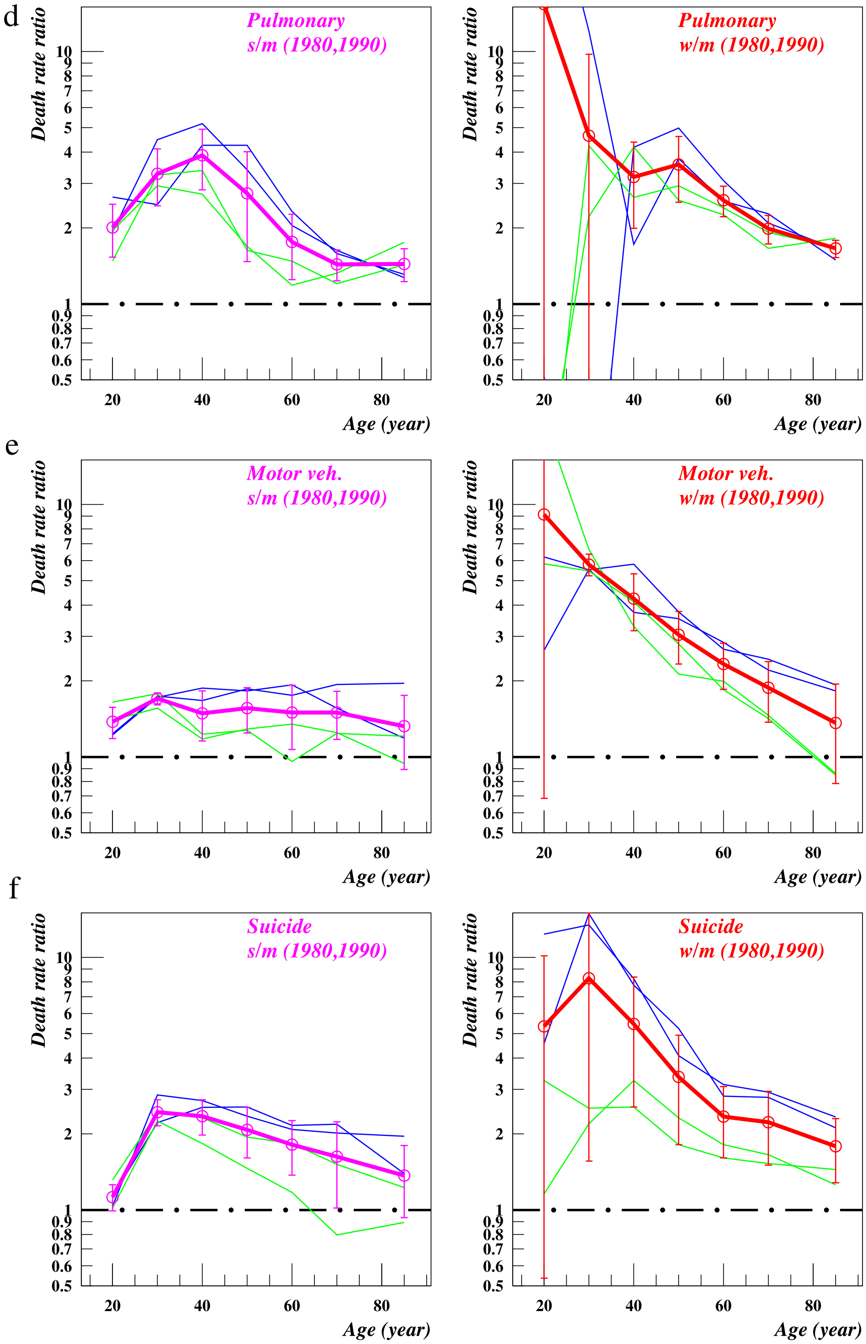

Fig. 6b. Death rate ratio for different causes of death, United States. The comments made in Fig. 6a apply as well to the present graphs. *Source:* Same as for Fig. 6a.

for instance, according to the data provided by the "American Community Survey" the non-married were 21% for men and 43% for women.

The previous question can also be answered affirmatively because for single, divorced or widowed persons, marriage makes really a big difference. In this respect, one should remember that in clinical test trials most pharmaceutical drugs, for instance those for heart disease, provide at most a 20%–30% benefit (more details can be found in Ref. [18]). On the contrary, Table 4 shows that for heart disease and cerebrovascular accidents the average death rate is divided by 2.2.

7. Suicide

7.1. Why is suicide of special interest?

Among the causes of death considered previously, suicide has a special significance for (at least) four reasons.

- Historically, since the mid 19th century, the phenomenon of suicide arose considerable interest among sociologists. The work of Emile Durkheim [6] is probably the most well known but there were many other studies, for instance by Louis-Adolphe Bertillon and his son Jacques Bertillon.
- Durkheim showed that persons with many family links have smaller suicide rates. For married persons with respect to never-married or widowed persons, this connection was already well-known before Durkheim. Although the influence of children presence had also been observed (particularly by Bertillon, see above), the relationship was not known accurately because of the fact that death certificates do not contain information about the number of children. Durkheim was able to establish the existence of a negative correlation between number of children and suicide rates by taking advantage of regional differences. This observation (largely forgotten nowadays) was a strong argument in favor of Durkheim's thesis of a connection between the strength of family ties and low suicide rates.
- In biology a phenomenon called apoptosis is often referred to as "cell suicide". Apoptosis (which should be distinguished from necrosis) occurs when cells no longer receive "stay alive" signals from their neighbors. Thus, a lack of communication with neighbors means that no "stay alive" signals will reach the cell under consideration. As a result, it will die by apoptosis. By contrast such a regulation mechanism does not exist for cancer cells, thus allowing their endless multiplication even when they are in a wrong location (e.g. breast cancer cells on the liver). More details can be found in Ref. [19] and in chapter 12 of Roehner [17]. This parallel with apoptosis gives at least a plausible mechanism for the observed connection between suicide and interaction with neighbors: like apoptosis, suicide will be more frequent when there is no communication with the neighbors.

  For other causes of death we do not yet have the beginning of an understanding.
- Finally, a look at Table 4 shows that suicide is the cause of death for which the death ratios are highest. For suicide the average of "never married" and "widowed" is 2.95. It is followed by "motor vehicle" whose average is 2.75.

7.2. Suicide ratios in France (1968–1993)

In Fig. 6b we have fairly large statistical fluctuations. As always in such a situation, we wish to reduce them. There is only one way to do that: one must increase the numbers of the events. This means either increasing the size of the country or increasing the number of years. Here we adopt the second approach. We consider France, which is smaller than the United States, but for which data over a period of 26 years can be obtained. The population of France is about 5 times smaller than the US population but with respect to the graphs displayed in Fig. 6b we will gain a factor: $(26/2)(1/5) = 2.6$. In addition, shifting from the US to another country will tell us something about the robustness of the Farr–Bertillon effect.

Regarding the accuracy of the present data set we must also ask ourselves how the populations of the age-groups by marital status have been estimated. For inter-census years they were estimated through a computational procedure whose results were published in Ref. [14]. As already observed, the correctness of such a procedure can hardly be checked.

A comparison of Figs. 1a and 7a shows that the shapes are very much the same but in Fig. 7a the magnitude of the effect is about twice as large. This observation is consistent with similar evidence from US data given in Table 5.

8. Conclusions and perspectives

The Farr–Bertillon effect is not yet well recognized as a major determinant of death rate. This is particularly true for medical doctors and is attested by the fact that the characteristics of patients taking part in medical trials include many parameters, yet their marital status is usually omitted in spite of the fact that it will substantially affect the outcome of the trial. More details on this point can be found in Ref. [18]. Moreover additional evidence and references can be found in [20–23].

8.1. Conclusions

As observed by Ausloos [26] who reminds us of the work of Elliot Montroll, the investigation of social forces has been on the agenda of econophysicists from the very beginning of econophysics. In the present paper, our investigation of a powerful social force, namely the Farr–Bertillon effect, led us to the following conclusions.

- In spite of the important demographic and sociological changes that took place over the past century the Farr–Bertillon effect remained fairly unchanged.
- The death rate ratios computed from different data sets (based on censuses or surveys) are well consistent with one another.

  The results summarized in Table 5 are restricted to the age interval 40–60 because Fig. 2 is limited to this interval.

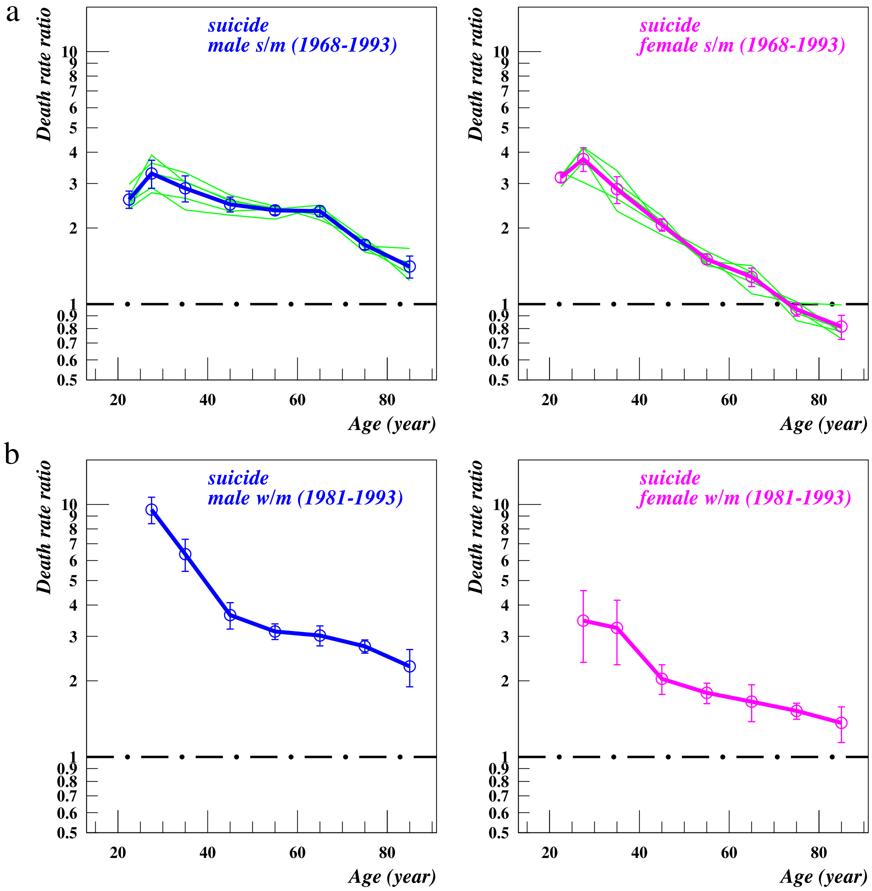

Fig. 7. Death rate ratio for suicide in France. The age groups are: 20–24, 25–29, 30–39, 40–49, 50–59, 60–69, 70–79, >80. In Fig. 7a the thin, green curves are not for single years but instead for groups of years: 1968–1973, 1974–1978, 1979–1983, 1984–1988, 1989–1993. For widowers the source gave only the average. Thus, the error bars were estimated from Fig. 7a and taking into account the reduced number of years. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Source: Besnard [25, p. 744, 752].

Table 5
Change of the US death rate ratio w/m from 1940 to 2010.

Male or Female	Age	Average 1940–1960	Average 1990–2010	Coefficient of variation for 1990–2010
Male				
	40	3.2	4.5	2.6%
	50	2.3	3.3	7.9%
	60	1.7	2.5	12.0%
Female				
	40	2.2	3.5	5.7%
	50	1.8	2.7	8.6%
	60	1.4	2.5	11.1%

Notes: The tables summarizes the death rate ratios w/m computed in Fig. 3 for 1940–1960 and in Figs. 2, 4, 5 for 1990–2010. The coefficient of variation, that is to say σ/m, gives an estimate of the fluctuations due to background noise. The age column indicates the middle of the 10-year intervals of the age groups. A similar increase is also observed in Europe as shown by the death rate ratios given in Ref. [27, p. 318–321]. We do not yet know what are the factors which bring about such increases.

- Whereas for single and divorced persons the death rate ratios are bell-shaped with a maximum around the age of 40, for widowed persons it is a function which either decreases steadily from youngest to oldest age groups or which has a maximum at the second youngest age interval. It is this last shape which is found in the observations of highest quality. However, we have seen that the shape of the curve for young age groups is mainly determined by the traffic accidents, a component which may be modified by changes in traffic regulation rules.
- When different causes of death are investigated it appears that suicide leads to the highest average death rate ratio while cancer leads to the lowest.
- In recent decades, death rate ratios for widowhood have been increasing in Europe as well as in the United States.

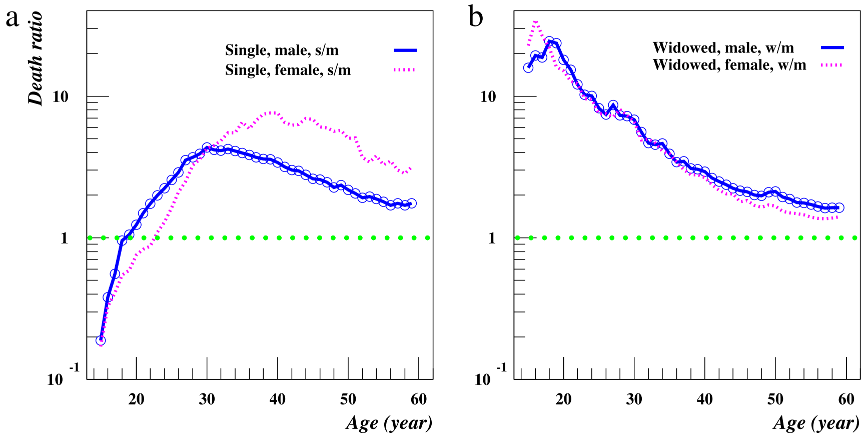

Fig. 8. The Farr–Bertillon effect in China, 1990. (a) Single/married. (b) Widowed/married. The male and female amplitudes of the FB effect are closer than they are in western countries; this is particularly clear for widowed persons. No error bars could be drawn because the data are for the whole country and for a single census year; however, the smoothness of the curves suggests that the fluctuations are small.
Source: The primary sources are of course official census and death records. These data are also available at the following address: http://bbs.pinggu.org/thread-1530030-1-1.html. (This Chinese website is a database which allows users to share their resources.) We are most grateful to Ms. Haokun Song, Xinhang Song, Shuyu Wang and Ximeng Wu from the "Beijing University of Post and Telecommunication" for bringing this resource to our attention.

8.2. Previous results supported by Chinese data

China is an interesting case for at least three reasons.

(i) It is important to see whether the FB effect also holds in a non-western country where family relations are different.
(ii) Because of the large population of China we can expect more accurate results than in the case of countries with smaller populations. Thus, instead of using 5- or 10-year age groups, here we can use 1-year age groups. This is of special importance for young widowers.
(iii) Several provinces (e.g. Fujian, Jiangsi, Sichuan, Xinjiang) have a tradition of early marriage which also contributes to magnify the young widower effect.

The population data used for calculating the death rates are those of the census carried out in June 1990. For instance, this dataset tells us that there were 237 male widowers of age 15; for females the number was 187.

The distributions in Fig. 8 are quite consistent with the most accurate results given earlier for western countries. This provides further confirmation of the widespread validity of the FB effect.

8.3. How can one observe the dynamical response?

In fact, demographic statistics of the kind considered in the present paper give little information about the dynamical aspect of this phenomenon. We learned that "on average" for 10-year age-groups, widowers have a death rate which is 3 times the death rate of married persons. However, this observation does not tell us anything about the transition from one state to another. How long does it take? Does the death rate of recent widowers increase steadily toward a steady state or is there a shock effect during which the death rate ratio first overshoots its steady limit?

One way to answer this question is to follow a sample of married persons over several years. This was done in a number of studies: Bojanovsky [28,29], Frisch and Simenson [30], Helsing et al. [31], Mellström et al. [32], Parkes et al. [33], Thierry [34,35], Young et al. [36]. Needless to say, in order to observe a substantial number of deaths in a sample of married people followed by a sizable number of deaths of widowers one should work on a sample of elderly persons. That is why most of the previous studies concern persons over 50 or 60. Basically, they found that the death rate of widowers peaked in the first 3 or 6 months after widowhood and then returned to the rate of married persons. More results will be given in Part 2 of this study.

How can one connect this observation to the death rate ratios measured in the present paper? Most of our observations concerned 10-year age groups? Such age groups will contain a mixture of widowers who differ both in age (a) and in the length of widowhood time (w).

Through the studies mentioned above, we know that the death rate of widowers (d) will be a function not only of a but also of $w : d = d(a, w)$. Similarly, the remarriage rate of widowers (m') will also be a function of a and $w : m' = m'(a, w)$. Because age is recorded on death and marriage certificates, one knows how, for fixed w, d and d' depend upon a. However, because w is *not* recorded on such certificates, there is no direct information about how d and m' depend upon w (for fixed a). The functions $d_a(w)$, $m'_a(w)$ summarize the dynamic responses of a group of widowers of age a.

The objective of the second part of this study will be to learn more about these response functions.

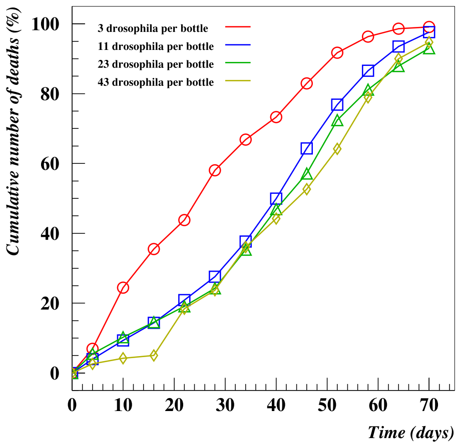

Fig. 9. Cumulative death number in groups of drosophila as a function of number of individuals per bottle. The bottles containing only 3 flies (circles) have a much higher mortality than the bottles with 11 (squares), 23 (triangles) and 43 (diamonds) drosophila respectively. The flies were put in the bottles on the day of their birth without any selection being done as to their sex. Thus, sex is not likely to play any role in the present experiment.
Source: Pearl and Parker [38].

Table 6
Life duration in segregated versus mixed populations of Drosophila.
Source: Hall [37, p. 211].

	Case: Population per vial	Mean life span male (days)	Mean life span female (days)
1a	10 females		59.0
1b	10 males	52.0	
2	5 females + 5 males	52.9	53.8
3	1 female + 1 male	50.4	41.9

Notes: This experiment did not measure the effect of the severance of the male–female interaction; instead it measured the effect of the non-existence of that interaction. The lower life expectancy of females in case 2 (with respect to 1a) is probably due to the burden of laying eggs because we do not observe a similar effect for the males. The author did not try a single male or female in each vial. However, case 3 is a proxy of this situation because after the death of the partner the survivor remains alone. This case shows that a situation close to solitary confinement leads to higher mortality, especially for females. This is confirmed by Wang et al. [7] and by Fig. 9.

8.4. Does the Farr–Bertillon law also apply to non human populations?

This question would deserve a detailed investigation.[6]

In a systems science perspective, the key factor is the inter-individual interaction. The precise nature of the individuals is largely irrelevant. Thus, if in a population there are strong interactions, one would expect their severance to lead to higher death rates just as seen in the FB law. Two experiments involving drosophila are described below.

- There is one interaction that exists for sure in many populations, namely the male–female interaction.[7] In an experiment performed by Jeffrey Hall [37] three groups of Drosophila were raised together: (i) virgin males (ii) virgin females (iii) mixed population with equal numbers of males and females. In each group there were 10 individuals in each of the 240 vials. In addition, (iv) there were 120 vials each containing only one male and one female.

  Three observations are in order about the results presented in Table 6.

  (i) For males there is a small (and probably not significant, although we cannot really know because the author does not give the standard deviation) benefit for being with females. For females it is the opposite. However, copulation also means laying eggs. As for all females, giving birth involves risks. Back in the 18th century, many women died as a result of child delivery. Similarly, among males there can be a competition for the females. Whether or not this can be life-threatening (or life-shortening) is of course species-dependent.

  (ii) The comparison of cases 2 and 3 shows that smaller groups result in shorter life spans. The experiment summarized in Fig. 9 leads to the same conclusion. It can be observed that their authors did not try the case of solitary confinement because the context of their experiment was in fact fairly different from our own questioning.

  (iii) When interpreting the results of such an experiment one should keep in mind that the drosophila were reared in conditions that were very different from natural living conditions of fruit flies. They were confined in small bottles (10 per bottle the size of which is not given) and every 3 days they were transferred to new bottles. Under such conditions what does one measure exactly? For humans it would correspond to observations made on inmates.

6 The short indications that follow are given at the request of one of our reviewers. We are planning to expand this aspect of our study in a subsequent paper.

7 However, this interaction also results in the production of offsprings which has implications that need to be considered separately.

Table A.1
Death rates by marital status and age, USA, census years: 1940, 1950, 1960.
Source: Grove and Hetzel [16, p. 334].

	<20	20–24	25–34	35–44	45–54	55–59	60–64	65–69	70–74	>75
Men, death rate										
Single, 1940	4.5	2.9	4.7	9.2	17.4	28.7	38.3	52.3	75.1	132
Single, 1950	3.3	2.2	3.6	8.3	17.2	29.6	40.8	55.0	79.5	137
Single, 1960	2.7	2.2	3.4	7.3	15.7	23.7	38.0	53.8	76.3	138
Married, 1940	2.5	2.2	2.6	4.8	10.6	19.1	27.6	39.4	60.3	113
Married, 1950	1.7	1.4	1.7	3.6	9.3	17.7	25.9	36.6	54.6	100
Married, 1960	1.2	1.2	1.5	3.0	8.4	16.0	25.3	37.2	53.4	100
Widowed, 1940	15.5	11.8	11.4	14.1	23.7	34.8	43.1	57.4	79.7	162
Widowed, 1950	2.0	6.1	7.7	12.3	21.1	30.2	39.6	49.7	69.1	139
Widowed, 1960	3.9	5.4	6.8	10.6	21.0	32.0	44.0	57.9	77.0	156

Notes: The rates are given in deaths per 1000 population of specified groups. It can be observed that, in accordance with Gompertz law [24], the death rates increase exponentially with age with a doubling time of about 10 years.

Table A.2
Death and population by marital status and age, USA, census years: 1980, 1990, 2000.
Source: Same as for Fig. 4.

	15–24	25–34	35–44	45–54	55–64	65–74	>75
Men, deaths							
Single, 1980	30 799	13 739	6 463	10 154	16 251	20 804	25 702
Single, 1990	24 921	25 448	19 324	11 609	14 492	18 973	26 611
Single, 2000	21 247	16 615	21 731	20 464	15 203	18 119	28 299
Married, 1980	4 894	16 144	23 238	57 550	134 153	199 962	210 401
Married, 1990	2 386	12 868	24 234	43 802	105 347	195 495	257 631
Married, 2000	1 493	8 218	22 160	49 298	87 351	164 106	312 544
Widowed, 1980	76	223	539	2 704	12 656	38 900	130 987
Widowed, 1990	29	218	619	1 852	9 329	35 630	143 605
Widowed, 2000	49	148	656	2 125	7 278	30 800	182 316
Men, Population							
Single, 1980	17 723	4409	978	657	551	364	198
Single, 1990	16 516	7779	2493	838	555	392	225
Single, 2000	17 450	7791	4071	1 783	548	385	247
Married, 1980	3424	12 589	10 488	9 365	7 716	5 496	2 329
Married, 1990	2089	12 183	13 710	9 863	8 205	6 400	3 156
Married, 2000	2331	10 797	15 890	13 747	9 131	6 581	4 200
Widowed, 1980	8.05	28.2	53.3	152	366	602	891
Widowed, 1990	10.1	32.3	71.8	134	346	701	1 079
Widowed, 2000	28.3	120	304	665	1 776	3 587	5 637

Notes: The numbers of deaths are expressed in units while the populations are in thousands. The data given in the table are the primary data from which the death rate ratios s/m and w/m were computed. The data for females can be drawn from the same sources; they were omitted here in order to save space.

- Apart from the male–female interaction, there is one other case in which the existence of inter-individual interactions is certain. It is the case of social insects for indeed elaborate social organizations cannot exist without social interaction. The experiment done by Wang et al. [7] is by far the most extensive experiment that we know of that was destined to measure the effect of severing social ties. In this experiment two groups were formed: group A consisted of 40 boxes each containing 10 ants, whereas group B consisted of 40 boxes each containing only one ant. The same experiment was repeated for 3 different species of ants. On average the death rate in group B was consistently higher than in group A. In short, the evidence from this experiment supports the conjecture that more interaction increases life expectation.

Acknowledgments

We wish to express our gratitude to Drs. Jason Fields and Rose Kreider of the US Census Bureau, and to Dr. Betzaida Tejada Vera of the Center for Diseases Control. Their advice helped us to locate the most accurate data available. We also thank our reviewers whose observations allowed us to make the text more readable and comprehensive.

Appendix. Death rates by age, marital status and sex

This Appendix gives the death numbers and population data for US census years from 1940 to 2000.[8] These data should permit to test theoretical models. Although, most of these data are available on Internet, they are not easy to locate. For instance after 1970, census population data are buried among dozens of volumes and thousands of pages of census publications; this makes their identification and extraction fairly time consuming (see Tables A.1 and A.2).

References

[1] W. Farr, Influence of marriage on the mortality of the French people (12 p.), in: Transactions of the National Association for the Promotion of Social Science 1858–1859, 1859, pp. 504–520. The paper was republished in 1975 in Vital statistics, a memorial volume of selections from reports and writings of William Farr. Scarecrow Press, Methuen (New York).
[2] L.-A. Bertillon, Article Mariage, in: The Dictionnaire Encyclopédique des Sciences Médicales, in: Encyclopedic Dictionary of the Medical Sciences, vol. 5, 1872, pp. 7–52. 2nd series. [Available on "Gallica", the website of digitized publications of the French national library, at: http://www.bnf.fr].
[3] Accuracy of data on population characteristics as measured by CPS-census match, 1960. Series ER 60, No. 5. Bureau of the Census. US Government Printing Office, Washington DC.
[4] J. Bertillon, Les célibataires, les veufs et les divorcés, au point de vue du mariage. [Inclination to marriage respectively of bachelors, widowers and divorced persons.], Rev. Sci. Fr. Etranger 8 (33) (1879) 776–783.
[5] L.-A. Bertillon, Article France, in: The Dictionnaire Encyclopédique des Sciences Médicales, in: Encyclopedic Dictionary of the Medical Sciences, vol. 5, 1879, pp. 403–584. 4th series. [Available on "Gallica", the website of digitized publications of the French national library at: http://www.bnf.fr].
[6] E. Durkheim, Le suicide. Etude de sociologie, F. Alcan, Paris, 1897, [A recent English translation is: "On Suicide" (2006), Penguin Books, London.].
[7] L. Wang, Y. Xu, Z. Di, B.M. Roehner, How does group interaction and its severance affect life expectancy? 2013, arXiv Preprint arXiv:1304.2935. (9 April 2013).
[8] A.J. Kposowa, K.D. Breault, G.K. Singh, White man suicide in the United States. A multivariate individual-level analysis, Soc. Forces 74 (1) (1995) 315–323.
[9] Mortality statistics: review of the Registrar General on deaths in England and Wales. Series DHI, Number 16. Her Majesty's Stationary Office, London.
[10] National Center for Health Statistics (NCHS): Mortality from selected causes by marital status. Vital and Health Statistics, Series 20, Number 8, 1970.
[11] Registrar General. Statistical Review of England and Wales. Part III, Office of Population Censuses and Surveys, London, 1971.
[12] Statistisches Jahrbuch für die Bundesrepublik Deutschland: Stuttgart, 1978.
[13] S. Flounders, Why US occupation cannot transform Afghanistan or Iraq, 2009. [available on the website of the International Action Center].
[14] F. Daguet, Un siècle de démographie française. Structure et évolution de la population de 1901 à 1993. [Structure and evolution of the French population from 1901 to 1993], Institut National de la Statistique et des Etudes Economiques (INSEE), No 434-435, Paris, 1995.
[15] Deaths final data. National Vital Statistics Reports, Years 1996–2010. National Center for Health Statistics, Hyattsville, Maryland. [available on the website of the NCHS at: http://www.cdc.gov/nchs/products/nvsr.htm].
[16] R.D. Grove, A.M. Hetzel, Vital Statistics Rates in the United States, 1940–1960, United States Printing Office, Washington, DC, 1968.
[17] B.M. Roehner, Driving Forces in Physical, Biological and Socio-economic Phenomena, Cambridge University Press, Cambridge, 2007.
[18] B.M. Roehner, Incidence of the Bertillon and Gompertz effects on the outcome of clinical trials, Physica A 414 (2014) 300–307.
[19] M.C. Raff, Cell suicide for beginners, Nature 396 (1998) 119–122. 12 November.
[20] W.R. Gove, Sex, marital status, and suicide, J. Health Soc. Behav. 13 (1972) 204–213.
[21] F.E. Linder, R.D. Grove, Vital Statistics Rates in the United States, 1900–1940, United States Printing Office, Washington, DC, 1947.
[22] L. March, Some researches concerning the factors of mortality, J. R. Stat. Soc. 75 (1912) 505–538. [The paper includes a comparative analysis of mortality rates in France, Prussia and Sweden.].
[23] W. Stroebe, M.S. Stroebe, Bereavement and Health. The Psychological and Physical Consequences of Partner Loss, Cambridge University Press, Cambridge, 1987.
[24] B. Gompertz, On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies, Philos. Trans. R. Soc. 115 (1825) 513–585.
[25] P. Besnard, Mariage et suicide. la théorie durkheimienne de la régulation conjugale à l'épreuve d'un siècle. [The impact of marriage on suicide revisited one century after Durkheim's work], Rev. Fr. Sociol. 38 (1997) 735–758.
[26] M. Ausloos, Another analytic view about quantifying social forces, Adv. Complex Syst. 15 (8) (2012) 1250088-1–1250088-13.
[27] J. Vallin, F. Meslé, T. Valkonen, Tendances en Matière de Mortalité et Mortalité Différentielle, Editions du Conseil de l'Europe, Strasbourg, 2001. An English version was published under the title "Trends in mortality and differential mortality".
[28] J. Bojanovsky, Wann droht der Selbstmord bei Geschiedenen? [At what point after separation are divorced people most likely to commit suicide?], Schweiz. Arch. Neurol., Neurochirurgie Psychiatr. 125 (1) (1979) 73–78.
[29] J. Bojanovsky, Wann droht der Selbstmord bei Verwitweten? [After becoming a widower when is the likelihood of committing suicide largest?], Schweiz. Arch. Neurol., Neurochirurgie Psychiatr. 127 (1) (1980) 99–103.
[30] M. Frisch, J. Simonsen, Marriage, cohabitation and mortality in Denmark: national cohort study of 6.5 million persons followed for up to 3 decades, Int. J. Epidemiol. 1 (2013) 13.
[31] K.J. Helsing, M. Szklo, G.W. Comstock, Factors associated with mortality after widowhood, Am. J. Public Health 71 (1981) 802–809.
[32] D. Mellström, A. Nilsson, A. Odén, A. Rundgren, A. Svanborg, Mortality among the widowed in Sweden, Scand. J. Soc. Med. 10 (1982) 33–41.
[33] C.M. Parkes, B. Benjamin, R.G. Fitzerald, Broken heart. A statistical study of increased mortality among widowers, Br. Med. J. 1 (1969) 740–743. [This is the continuation of the Young et al. (1963) study. The same sample was observed over 4 more years i.e. a total of 9 years after the death of the spouse.].
[34] X. Thierry, Risques de mortalité et de surmortalité au cours des 10 premières années de veuvage. [Excess mortality during the first 10 years of widowhood.], Population 54 (2) (1999) 177–204.
[35] X. Thierry, Risques de mortalité et causes médicales des décès aux divers moments du veuvage, Gérontol. Soc. 95 (2000) 27–43.
[36] M. Young, B. Benjamin, C. Wallis, Mortality of widowers, Lancet 2 (1963) 254–256.
[37] J.C. Hall, Age-dependent enzyme changes in *Drosophila melanogaster*, Exp. Geront. 4 (1969) 207–222.
[38] R. Pearl, S.L. Parker, Data on the influence of density of population on duration of life in Drosophila. Paper no 4 in the series entitled: Experimental Studies on the duration of life, Am. Nat. 56 (645) (1922) 312–321.

[8] The regular publication of death data by age, marital status and sex began in 1979. However, because marital status was recorded on death certificates, Grove and Hetzel [16] from the US Census Bureau were able to compute and publish death rates for the census years 1940, 1950 and 1960. That is why no death numbers are available for these years. For 1970 there neither death rates nor death numbers.