Physica A 486 (2017) 542–553

Contents lists available at ScienceDirect

Physica A

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

Impact of marital status on health

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

- Census disability data are a possible substitute for death data.
- A change in living place results in a temporary disability spike.
- The post marriage disability rate decreases with time.

ARTICLE INFO

Article history:
Received 14 February 2017
Received in revised form 8 May 2017
Available online 12 June 2017

Keywords:
Marital status
Migration
Death rate
Disability

ABSTRACT

The Farr–Bertillon law states that the mortality rate of single and widowed persons is about three times the rate of married people of same age. This excess mortality can be measured with good accuracy for all ages except for young widowers. The reason is that, at least nowadays, very few people become widowed under the age of 30. Here we show that disability data from census records can also be used as a reliable substitute for mortality rates. In fact excess-disability and excess-mortality go hand in hand. Moreover, as there are about ten times more cases of disability than deaths, the disability variable is able to offer more accurate measurements in all cases where the number of deaths is small. This allows a more accurate investigation of the young widower effect; it confirms that, as already suspected from death rate data, there is a huge spike between the ages of 20 and 30.

By using disability rates we can also study additional features not accessible using death rate data. For example we can examine the health impact of a change in living place. The observed temporary inflated disability rate confirms what could be expected by invoking the "Transient Shock" conjecture formuladted by the authors in a previous paper. Finally, in another observation it is shown that the disability rate of newly married persons is higher than for those who have been married for more than one year, a result which comes in confirmation of the "newly married couple" effect reported in an earlier paper.

1. Introduction

1.1. The Farr–Bertillon law on the mortality effect of marital status

The Farr–Bertillon law [1–3] states that married persons have a lower death rate than non-married persons be they single, divorced or widowed. In its simplest form this law has been known for over one century. In a previous paper [3] the present authors added several new features to our knowledge of this effect.

* 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.2017.05.079

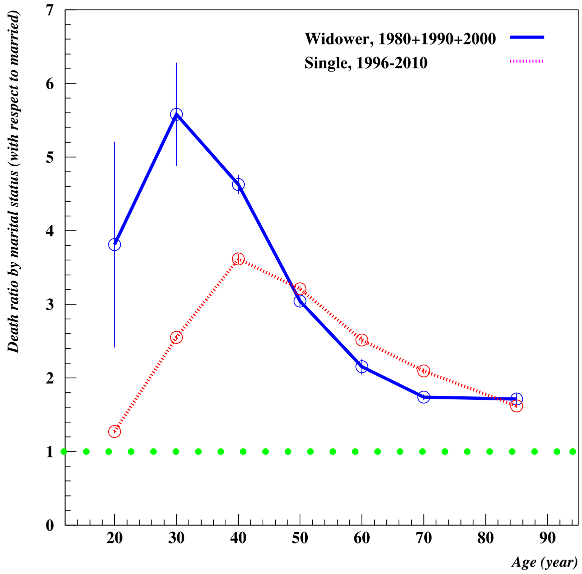

Fig. 1. Age-specific death ratio (for males) with respect to married persons in the United States. The death ratio is the death rate of widowers (or single persons) divided by the death rate of married persons. In accordance with Gompertz's law all these death rates increase exponentially with age. Dividing by the death rate of married persons has the advantage of removing the exponential behavior; thus, it can be seen as a kind of renormalization procedure. These curves should be compared with the curves of Fig. 2(b) which, instead of death, are based on disability. The data points give the death ratio for 10-year age intervals. The error bars represent $\pm\sigma$ where σ is the standard deviation of the average. This corresponds to a confidence level of 0.68. The curve for "single" is the average of 15 annual series which is why the error bars are so small that they are hardly visible. In contrast, for young widowers, because their number is so small and must therefore be measured with high accuracy, we can only use census data.
Source: Richmond et al. [3,4]: widowed: p. 757, single: p. 755.

• As a function of age the death ratio, namely the ratio (rate of non-married)/(rate of married), has the same shape today as it had 130 years ago.

• The law holds not only for the global death rate but also separately for all major classes of diseases, e.g. heart disease, cerebrovascular diseases, pulmonary diseases. The law is also valid for suicide and accidental causes of death.

• Chinese data for 1990 show a pattern that is very similar to that observed in western countries. These data provide a more accurate picture of the young widower effect than that provided by western data due to: (i) the sheer size of the Chinese population and (ii) the tradition of early marriage which was still common even in 1990 for some rural provinces. It turns out that, around the age of 20, for men as well as for women, the widowed/married death ratio displays a huge spike with amplitude about 20.

As a reminder and for the purpose of comparison with subsequent disability-based graphs, Fig. 1 shows the shape of the death rate ratio for widowed and single persons. Mortality data are based on death certificates which record basic information about the deceased and the circumstances and causes of the death.

1.2. The challenge of getting appropriate data

Since the pioneering papers of Farr and Bertillon there have been a number of subsequent studies but one can safely say that all of them were hampered by the paucity of data sources. As an illustration one can consider the following studies.

(1) Gove [5] used data collected and published by the US Department of Health [6].

(2) Ben-Shlomo et al. [7] used data from a study of a total of 18,403 men aged 40–64 years who were followed over a period of 18 years.

(3) Williams et al. [8] used *self-assessed* physical health data for a group of men and women.

(4) Robards et al. [9] used data from a longitudinal study done by the British Office of National Statistics (ONS).

Of these sources the only one which is both publicly available and performed in a fairly systematic way is the first one. The three others are based on small samples (2), data of dubious quality (3) or studies that are not publicly available (4). These cases suggest that in the field of biodemography the data are the main challenge because they will permit the discovery of phenomenological laws from which appropriate models can then be derived.

1.3. Census data

In this paper we use census data. This is a completely different type of statistics in the sense that mortality data rely on continuous monitoring through the vital statistics network whereas censuses are taken every 10 years and provide a static picture of the whole population. It is difficult to decide which one of the two sources is more reliable but for our present study what is important is the fact that they are independent and very different from one another.

One might suspect the young widower effect observed from death rate data is a statistical artifact because under the age of 25 the number of cases is very small which makes these data very sensitive to any under-recording of the population of widowers. It is therefore important to observe this effect through a different data source. If census data display the same effect it will make us more confident that it is indeed genuine.

The census data that we use here are disability variables. Thus, the death ratio will be replaced by a disability ratio. A distinct advantage of the census data is that it is possible to compute the disability ratio for every year of age. This contrasts with mortality data which are given for only 5-year or 10-year age intervals. This difference becomes particularly crucial for the age interval around 20 where the young widower effect is expected.

1.4. Procedure

The paper proceeds as follows.

In Section 2 we describe the disability data and explain our methodology. Section 3 shows the results of our investigation. Next we explore the effect of a housing relocation on the disability rate from which we see that there is a distinctive increase albeit of a much smaller magnitude than for the effect of marital status. Finally, we examine the adverse short-term effect of marriage.

Before we start note that the present paper follows exactly the methodology associated with physics. Thus in physics a newly identified phenomenon is accepted as being real only after it has been observed in different experiments and under diverse conditions. For example, the speed of light was measured numerous times using various methods and with ever increasing accuracy. As a result it has become an unshakable pillar of modern physics. By way of contrast, what is striking in social science is that conflicting observations are cited in review papers and apparently accepted without any real attempt being made to discriminate between fact and artifact. How can something solid be built on such shaky foundations?

Our present analysis, based on disability data, of the incidence of marital status is akin to setting up a new experiment. If it confirms and supplements previous observations that strengthens our confidence in the Farr–Bertillon law and the young widower effect.

2. Data and methodology

2.1. Disability data

2.1.1. Census question

The US censuses of 1980 and 1990 asked the following question:

> "Does this person have a physical, mental, or other health condition which has lasted for 6 or more months and which prevents this person from working at a job?"

At first sight it might seem that the question concerned only persons who are not retired and it is true that in the census of 1970 the question concerned only persons under the age of 65. But in the censuses of 1980 and 1990 the question was asked of every person above the age of 15. In these cases "prevents this person from working at a job" must be understood as "a condition serious enough to prevent you from working at a *possible* job".

2.1.2. Increase as a function of age

As a function of age, the proportion of the persons afflicted with complete work disability (which corresponds to the code number 3) has the same behavior as the mortality rate. 0.95.

• It increases exponentially with age: the correlation between age and the logarithm of the proportion of work disabled is 0.995 with a confidence interval (0.993, 0.997) for a confidence level of 0.95.

• Its doubling time is $T = 11$ years which is about the same time as for the mortality rate.

At first sight this increase profile may seem surprising for the proportion of disabilities is bounded by 1 whereas the mortality rate is unbounded. However the upper limit plays no role for at the age of 85 the proportion reaches only about 0.5 (see Fig. A.1).

2.1.3. Alternative disability variable

The census of 1980 (but not the one of 1990) contained another question about disability which was the following:

> "Does this person have a physical, mental, or other health condition which has lasted for 6 or more months and which limits or prevents this person from using public transportation?"

Naturally one would expect the two disability variables to be strongly correlated and this is indeed the case. In what follows we use only the work disability variable because the other one is not available for 1990.

2.1.4. Relationship between disability proportion and death rate

A consistency test consists in checking whether the age-specific disability variable is correlated with the age-specific death rate (see Appendix).

Actually this is true not only for the whole population but also for its subsets. For instance, for widowed persons, if one denotes the number of persons with a disability in the 5% sample of the 1980 census by d and the number of deaths in the whole population by D one gets[1]:

$$D = 1.18d^{\alpha}, \quad \alpha = 1.06 \pm 0.07. \tag{1}$$

This relationship as well as the results given in Fig. A.1 show two things:

(1) The two variables are almost proportional to one another. However, as death numbers are available only for 10-year age intervals Eq. (1) indicates a global rather than a year-by-year proportionality. Actually, on a yearly scale one would not expect a close connection for a fairly simple reason. The death numbers are annual variables whereas the disability numbers are cumulative variables in the sense that disabilities which last more than one year will be added together. Thus, in a given year, in addition to the current number the disability level will also reflect extant past disabilities.

(2) D and d concern two different populations: for D it is the total population whereas for d it is the 5% sample. Nevertheless, D and d are of same magnitude. Thus we see that for the entire population the number of persons with a disability would be about 20 times larger than the number of deaths. Therefore, if one could get disability data for the whole population one would be in an excellent position to study the young widower effect. Unfortunately, as explained in the next subsection, only 1% and 5% samples are so far available.

2.2. The IPUMS database

Now that we know the 1980 and 1990 censuses contain the data we need, how can we access it? Over the past decades the University of Minnesota has developed a database containing *individual* records of all US censuses except that for 1890 which was destroyed in a fire. "Individual" means that the database will deliver files in which each line corresponds to one person and contains as many coded variables as the user selects. Access is free and the data are provided in several formats. For our research we have used the "text only" format, formerly called the ASCII (American Standard Code for Information Interchange) format. However, there are two limitations.

- The data are available only in the form of random samples, either 1% or 5% samples.[2] The results given below are based on the 5% samples of the 1980 and 1990 censuses. As the disability variable exists only above the age of 15 and under 90 we limited our samples to the age interval (16, 89). As a result, the files of the 5% sample of 1980 and 1990 contained 8,746,006 and 9,529,970 lines respectively.
- For reasons of confidentiality the variables we use do not allow precise localization of the individuals. Information about their place of residence is limited to county level or even to a cluster of several counties when the counties are small. As the present investigation does not use residence location variables this limitation is of no concern.

2.3. The small n difficulty

As already explained one of our main objectives is to explore the Farr–Bertillon effect for widowed persons in the age interval from 20 to 30. The main difficulty comes from the fact that even for a large country like the United states there are only few young widowers and among them only a small percentage with a disability. This difficulty is illustrated in Table 1 which gives the number of widowed persons and the numbers within this group with a disability.

Two measures were taken to limit the impact of statistical fluctuations arising from such small numbers.

(1) The analysis used both the samples of 1980 and 1990 which together comprise 18 million individual records.

(2) A centered moving window averaging was applied. We tried widths of 5 and 11 years and the latter proved the most satisfactory,

2.4. Uniformity test

In order to make sure that the shape of the age-specific disability ratio is not brought about by a sub-sample of outliers, we tested some 10 sub-samples each comprising 2 million records. They all led to curves of same shape peaking in the 20–30 age interval and decreasing toward 1.2 at old ages.

---

1 The death data by age and marital status are from the 1980 volume of "Vital Statistics of the United States", table 1–31.

2 Full scale 100% data are available only for a few censuses, for instance in 1940.

Table 1
Number of widowed persons in the 5% sample of the US census of 1980.
Source: 5% sample of the census of 1980: Ruggles et al. [10] (IPUMS).

Age	Men	Men with disability	Women	Women with disability	Men + women	Men + women with disability
16	10	0	28	6	38	6
20	42	4	128	9	170	13
25	93	5	464	21	557	26
30	139	4	699	30	838	34
35	209	17	1005	64	1214	81

Notes: All numbers are for widowed persons. "Disability" refers to a condition which prevents people from working. This "work disability" variable was given in the censuses of 1980 and 1990. Needless to say, the small number of cases in the 16–30 range gives rise to strong inter-age statistical fluctuations.

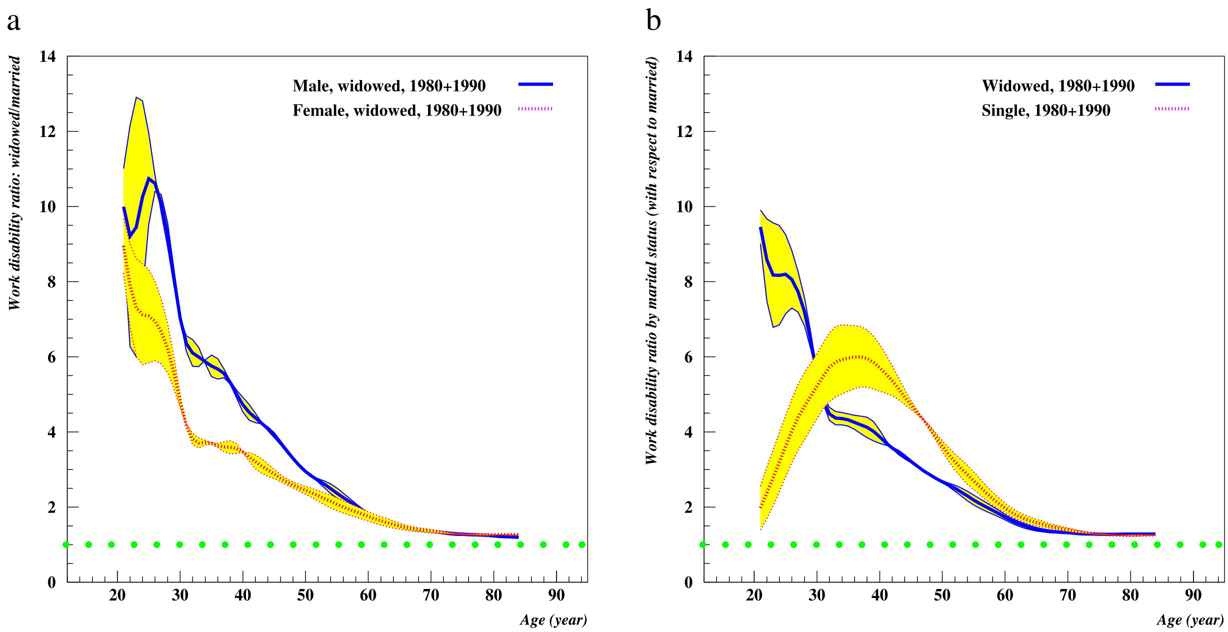

Fig. 2. Age-specific disability ratio with respect to married persons in the United States. (a) The disability ratio is the disability rate of widowed persons divided by the disability rate of married persons. (b) The curve for widowed persons is the average of the male and female curves of (2a); "single" refers to persons who never got married. Fig. 2(b) should be compared with Fig. 1 which is similar except that disability is replaced by death. Because there is a data point for each year, the error bars are not drawn as bars but are shown as a (yellow) error band. It represents $\pm\sigma$ where σ is the standard deviation of the average respective to subsamples, e.g. 1980,1990. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Source: 5% samples of the US censuses of 1980 and 1990; available from [10] (IPUMS).

3. Results

3.1. Observations

Fig. 2 represents disability ratios; how are they defined?
First, we introduce the disability rates which, for any marital status σ, are defined as follows.

$$f(t;\sigma) = \frac{\text{Number of persons of marital status } \sigma \text{ with a disability}}{\text{Total number of persons of marital status } \sigma}$$

where t represents the age.

The case when σ is the married status is of special importance because this rate is used as a reference for the definition of the disability ratios. For instance the disability ratio of widowed persons shown in Fig. 2(b) will be defined as: $r(t;w) = f(t;w)/f(t;m)$. More generally, the disability ratio for any marital status σ will be defined as: $r(t;\sigma) = f(t;\sigma)/f(t;m)$.

Fig. 2(a) shows the disability ratios for σ = widowed males and σ = widowed females respectively.

Fig. 2(b) compares the disability ratios for σ = widowed persons (male or female) and σ = single (i.e. never married) persons respectively.

These figures lead to the following conclusions.

• Between the ages of 20 and 35 the curves for widowed and single are very different: decreasing with age for widowed and increasing for single persons. This observation confirms what was found with Chinese death-ratio data. For US death ratio data (Fig. 1) the difference was less striking in the sense that in the age interval (20, 30) the two curves are fairly parallel. This is probably due to the large error bars for young widowers.

• Fig. 2(a) does not show any fundamental difference between men and women except that the disability ratio of widows is slightly lower than for widowers.

• The peak for widowed persons reaches a level of about 10 which is intermediate between the value of 6 observed in the US (Fig. 1) and the value of 20 observed in China.

Incidentally, it should not come as a surprise that the shape of the death ratio for young people is country-dependent. This is due to the fact that between 16 and 30 the main causes of death are not diseases but external causes such as traffic accidents, homicide or suicide. The frequency of traffic accidents is of course conditioned by the number of young people who drive cars or motorbikes.

3.2. Other groups of non-married persons

So far we have considered only disability ratios for widowed and single persons. But the marital status variable of the census defines 6 different situations. Their definitions and respective fractions in 1980 are as follows (in 1990 the percentages are almost identical).

(1) married, spouse present: 56%,
(2) married, spouse absent: 1.2%,
(3) separated: 2.2%,
(4) divorced: 6.2%,
(5) widowed: 7.6%,
(6) single: 26%.

Groups 2 and 3 are too small to be analyzed in a meaningful way. Group 4, divorced persons, leads to a disability ratio curve which is intermediate between “widowed” and “single”. This contrasts with “widowed” which has a rising part which peaks around the age of 32 which is slightly earlier than the curve for “single”. The peak reaches a level of about 3, about 2/3 the level reached by the curve for “single” and only 1/3 the level reached by the curve for “widowed”.

4. Health impact of a change in living place

4.1. A testable prediction of the “Transient Shock” conjecture

For death occurrences the only data available about the deceased are those recorded on the death certificate. This includes only basic information such as age, cause of death, marital status. For the persons enumerated in a census much more information is available which can be linked to the data about disability. Here this kind of linkage is illustrated by a particular interesting case which yields a test of the transient shock conjecture proposed in one of our earlier papers. The “Transient Shock” conjecture introduced in [4] posits that:

> “Any abrupt change in living conditions generates a mortality spike which acts as a kind of selection process”.

Moving from one place to another is a fairly sudden change although hardly as dramatic as moving from home into a nursing home.[3] This conjecture leads us to expect that a change of residence will have an adverse effect on health; although because such changes are fairly common one expects a small impact. Yet thanks to the disability variable it is possible to test this prediction.

4.2. Data and methodology

In the censuses of 1980 and 1990 the question about the living place was as follows.

“When did the person move into this house or apartment?”

Because it is well known that recollection of the date of past events is fairly unreliable, the questionnaire proposed a number of fairly broad time intervals. In 1980 the first interval was 1979–1980, the second was 1975–1978. Altogether there were 6 code numbers with the final one standing for “moved in more than 31 years ago”.

Based on the analysis conducted in [4] we expect adverse effects to appear fairly quickly after the change. This led us to consider only two cases:

(i) Short stay: 1979–1980 (ii) Long stay: all years before 1979.

If we admit that the census question was asked in mid-1980, the first case corresponds to a stay of between 0 month (for a change occurring just prior to the census interview) and 18 months (for a change occurring on 1 January 1979). Thus, for this short-stay case the average length is 9 months.

Because we expect the long-stay case to have a lower disability rate than the short-stay case, it will be the analog of the married status considered in the previous section. Thus, in the same way as we computed disability ratios with respect to the married status, here we will compute disability ratios with respect to the long-stay case. As before we compute this ratios for all ages between 16 and 89. In addition we repeat this calculation for different marital statuses.

[3] An effect of that kind for elderly persons was studied in [4].

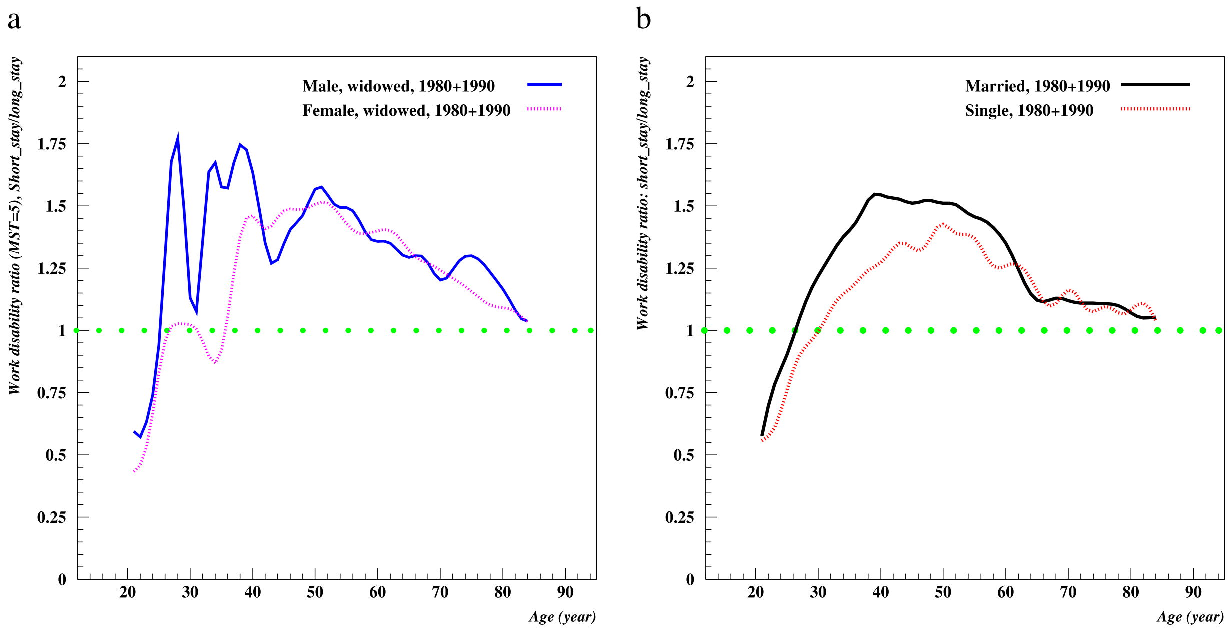

Fig. 3. Age-specific disability ratio according to time of residence in one and the same place. The disability ratio is the disability rate of persons who have been moving to another place in the 18 months preceding the census interview divided by the disability rate of persons who have remained in the same place for more than 18 months. For the sake of clarity the error bands have been omitted; they have basically the same shape as in Fig. 2.
Source: The results are based on the 18,037,222 records of the 5% samples of the US censuses of 1980 and 1990; available from [10] (IPUMS).

4.3. Results

With respect to the analysis conducted in the previous section adding a new variable namely the length of stay will further reduce the number of persons who qualify. For that reason we need big samples and this led us to analyze the merged data of both the 1980 and 1990 censuses.

The curves presented in Fig. 3 were computed from the eighteen million data lines of this merged file. They correspond to the following disability ratio:

$r(t; s) = f(t; \text{short stay}; s)/f(t; \text{long stay}; s)$, where t, s and f have the same meaning as previously. For ages over 25 the curves show indeed disability ratios which are larger than one, thus confirming the prediction based on the "Transient Shock" effect.

4.4. Discussion of the shock effect

When a group of persons experiences a transition from a state A to a state B its disability (or death) rate $d(t)$ may display two effects.

• A smooth transition from the rate d_A to the rate d_B. This effect exists always but, of course, when $d_A = d_B$ it vanishes.

• A shock effect during which $d(t)$ displays a peak that is higher than d_A and d_B. This effect is predicted by the "Transient Shock" conjecture (stated in [4]); it exists whenever this conjecture applies.

In the present experiment $d_A = d_B$ for neither of the two places is assumed to be better than the other. Thus, the excess disability revealed by Fig. 3 must be attributed entirely to the shock effect.

It can be summarized by saying:

> For persons who have been in a new place for an average length of time of 9 months, the disability rate is some 1.4 times higher than for persons who did not move during the past 18 months.

4.5. Discussion of some surprising features

The curves have also some surprising features.

• The maximum level which is reached between the ages of 40 and 60 is almost independent of the marital status.

• After the age of 60 the curves fall until converging towards a stationary level of 1.07.

What is surprising in these observations can be summarized by saying that one would expect the impact of relocation to be stronger for more "fragile" groups. It would be reasonable to think that widowed persons are more fragile with respect to relocation than are married persons. However, the results show basically the same effect for widowed and married people.

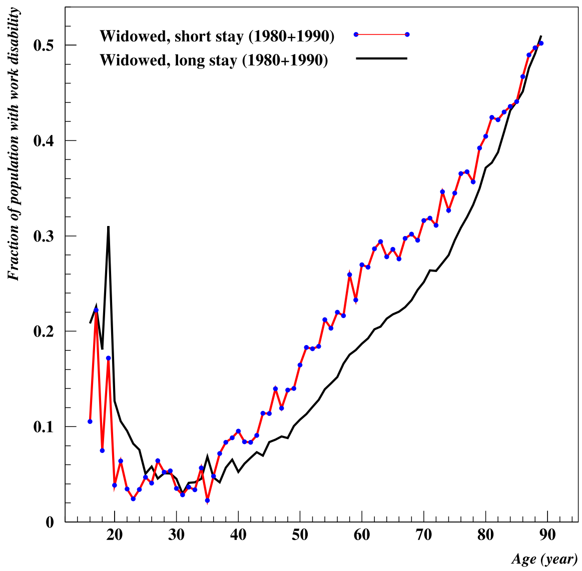

Fig. 4. Fractions of widowed groups with disability: short stay versus long stay. "Short stay" means same residence during a lapse of time comprised between 0 and 18 months. As both fractions increase along with age, the relocation effect becomes drowned in the other many disability causes; as the later are common to the two subgroups the impact of the length of stay is reduced.
Source: 5% samples of the US censuses of 1980 and 1990; available from [10] (IPUMS).

Similarly one would expect elderly people to be more fragile than people in their 40s. However, the results of Fig. 3(a), (b) show the opposite. They reveal a relocation effect that is smaller for elderly groups than for midlife people.

One can propose the following explanation.

With increasing age come more and more disability factors mostly related to health issues; relocation represents one of these factors but as the number of the other factors increases the relative weight of relocation decreases. This explanation is confirmed by comparing the numerator and denominator of the disability ratio (Fig. 4). As age increases, both terms increase but at the same time the gap between the two curves narrows.

In the next section we use the disability data to test (or rather re-test) a surprising effect already identified in [4].

5. Short-term disability increase following marriage

5.1. Method and data

Based on the "Transient Shock" conjecture it was predicted in [4] that after marriage there should be a temporary increase in the mortality rate (more will be said below about this prediction) This prediction was tested and confirmed by three different methods which all relied on *mortality* data.

The fact that disability rates can be used as a proxy for death rates opens a new possibility and it is therefore interesting to see whether our previous tests can be complemented.

This investigation is based on the answers to the following question which was asked in the 1% censuses conducted between 2008 and 2015.[4]

"In the past 12 months did this person get married?"

If there is a temporary surge in disability in the months following marriage, one should see an inflated rate for the persons who got married within the past 12 months with respect to those who have been married for a longer time. Fig. 5 shows that this is indeed the case except (for reasons as yet unknown) for the youngest and oldest age intervals, namely 16–25 and 76–85.

5.2. Selection of the disability variable

Note that for the ACS surveys the disability variable is not defined in the same way as in the censuses of 1980 and 1990. Here, there are 3 different disability variables corresponding to different aspects of the situation of the persons:

4 Actually as they are done on 1% samples they are not real full scale censuses. Such surveys are called "American Community Surveys" (ACS).

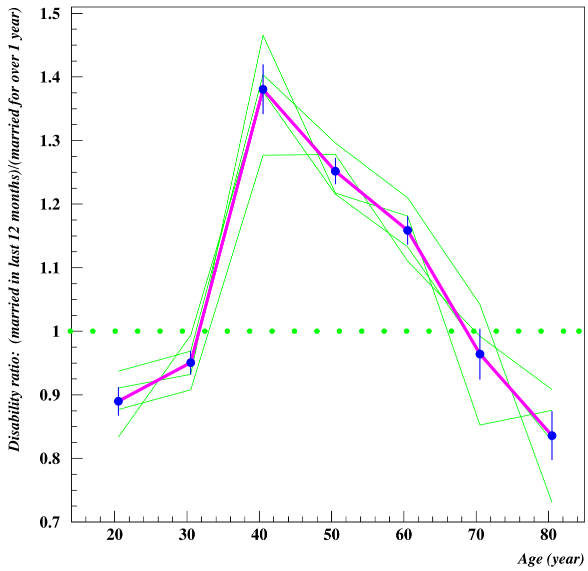

Fig. 5. Disability rate of persons who got married in past 18 months divided by the disability rate of persons who have been married for a longer time. The data points are for 10-year age intervals: (16–25), (26–35), ..., (75–85). The four thin lines in green correspond to: (i) 2008–2010 (7,203,967 persons), (ii) 2011–2012 (4,988,147 persons), (iii) 2013–2014 (5,035,986 persons), (iv) 2015 (2,542,244 persons). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Source: Data from the annual American Community Surveys from 2008 to 2015; available from [10] (IPUMS).

d_1: mobility at home, d_2: mobility outside home, d_3: difficulty in bathing, dressing. It turns out that as a function of age, $\log(d_1)$, $\log(d_2)$, $\log(d_3)$ are highly correlated: $r(1,2) = 0.93$, $r(1,3) = 0.97$, $r(2,3) = 0.99$.

With respect to the mortality rate D in seven 10-year age intervals from 16 to 85 they have the following relationships (the corresponding correlations are given within parentheses):

$$d_1 \sim D^{1.21} \quad (r = 0.975), \qquad d_2 \sim D^{1.93} \quad (r = 0.982), \qquad d_3 \sim D^{1.53} \quad (r = 0.995).$$

So we select the disability d_1 as being the condition most closely connected with the mortality rate.

In the phenomenon of post-marriage disability one should distinguish two effects: the transition effect on the one hand and the shock effect on the other hand. First we will explain the two effects and then we discuss which one we can see in Fig. 5.

5.3. Explanation of the transition effect

Whilst the transition from being single to married in a legal sense occurs instantaneously, the physical and mental state does not change immediately but rather with a delay that we denote by θ; this is quite an important parameter for it gives an estimate of how quickly persons adapt to a new living way. The single and married states of mind we denote by S and M.

Within time θ, among a sample of legally married people, the proportion p_s of S-persons will fall from 1 to 0. What consequence does this have with regard to the group of persons who became married in the 12 months before the census was taken? For the sake of simplicity let us assume that the census was taken in December.[5] What will be the composition of the sample E_{12} of people who got married in the past 12 months?

(Case 1) If $\theta = 0$ then E_{12} will have no S persons: $p_s = 0$.

(Case 2) If $\theta = 60$ (i.e. 5 years) then E_{12} will have almost only S-persons: $p_s \simeq 0$ because for most of them one year will be too short a time to change to the M state.

(Case 3) Under an intermediate and more realistic assumption, e.g. $\theta = 6$ months, the persons who got married between January and June will be M-persons at the time of the census whereas those married between July and December will be partly M and partly S which leads to $p_s \sim 0.25$.

In other words, by measuring p_s we can infer θ.

In case 1 the disability ratio (married in last 12 months)/married would be equal to the ratio married/married which is of course identically equal to 1.

In case 2 the disability ratio would be equal to the ratio single/married shown in Fig. 2(b) (at least if we forget for a moment that Fig. 2(b) is for 1980 + 1990 whereas Fig. 5 is for 2008–2015). Although the two curves have the same shape

5 In fact in 1980, 1990 it was usually taken between April and June.

(with their maximum around 40 years), the amplitude of the curve in Fig. 5 is 4 times smaller than the amplitude $r_{s/m} = 6$ in Fig. 2(b).

In case 3, the disability ratio r at the top of the peak will be approximately:

$$r = 0.25r_{s/m} + 0.75 \simeq 0.25 \times 6 + 0.75 = 2.25.$$

As this result is higher than the value of 1.4 observed in Fig. 5 it leads us to the conclusion that the real value of θ is somewhere between 0 and 6.

By repeating the same calculation for $\theta = 3$ months one gets a peak-amplitude of 1.6 which is close to the actual one. In other words, θ can be considered as being of the order of 3 months, an estimate which is consistent with the one obtained in [4] by a method based on death rates.

5.4. *Explanation of the shock effect*

The previous effect is always present as soon as there is a transition from a state *A* to a state *B* but in addition there can be a shock effect. This expression refers to the fact that instead of decreasing steadily from the "single" to the "married" state, the death or disability rate may overshoot the "single" rate. One of the most spectacular illustrations of the shock effect is birth that is to say the transition from fetus to newborn. In this case, instead of falling steadily from the pre-natal (fetus) rate to the infant rate, the post-natal death rate shoots up to a level that is over 50 times higher than the late fetal mortality rate.

Regarding marriage, in [4] it was shown that there is a shock effect in the transition from "single" to "married" with a rate that peaks at 2.5 times the death rate in the "single" state.

Here, however, we cannot see this effect. In Fig. 5 the disability rate peaks at only 1.4 times the rate in the "married for over one year" state which is much less than the amplitude of the single/married rate shown in Fig. 2(b). The reason for that is explained in the previous subsection. It is due to the fact that the one year time interval corresponding to the IPUMS data is much too long compared with $\theta = 3$ months. As a result, in this group there are mostly persons whose transition has already been completed. One can predict that, if available, data for a 3-month time interval would lead to a much higher peak which would display the expected shock effect.

6. Conclusion

6.1. *Overall results*

In this paper we exploited the fact that disability rates can be used as near-substitutes for mortality rates. Whereas the second are based on death certificates, the first are recorded individually in some (but not all) censuses. This gives much more flexibility because censuses record more personal information than that given on death certificates.

As a result we were able to estimate the impact on health of three different conditions: (i) marital status (ii) moving from one living place to another (iii) getting married. For (i) and (iii) we had a prior knowledge of what to expect from a previous study based on mortality rates. Our observations led to effects similar to what was seen with mortality rates thus confirming that the disability rates are acceptable substitutes for mortality rates.

The effect (ii) of a living place change in the 18 months preceding the census interview could be predicted based on the "Transient Shock" conjecture. The results show that, on average over all ages, disability rates are inflated by a factor 1.5 with respect to the persons of same age and same marital status who did not move.

6.2. *Assets and promises of biodemography*

We finish with a word about the field of biodemography. A few physicists have recently begun to explore this field (see in particular [11]), but so far it has not attracted their attention to the same degree as has finance. When econophysics began some 20 years ago it mainly focused on the study of stock prices for a simple reason which was the availability of big databases giving high frequency transaction data (so-called "tick by tick" data) for thousands of stocks.

Throughout the development of biodemography, data availability has also been a permanent challenge. From John Graunt (1620–1674, who established the first life tables based on rudimentary weekly statistics of death in London), to Sébastien de Vauban (1633–1707, who correctly predicted the growth of the population of Canada), to Edmund Halley (1656–1742, who improved Graunt's tables thanks to accurate data for the German city of Breslau) further progress always came about thanks to better data. In this respect, the Internet has opened a wonderful new window of opportunity. We hope the present paper will encourage other physicists to take advantage of this opportunity.

Thanks to the Internet many data sets can now be readily interrogated in order to illuminate regularities whose level of noise is much lower than that of stock data.

What is appealing in biodemography is the fact that there are numerous well-defined and intriguing questions. By way of illustration we note two: (i) Why are suicide rates highest in May and lowest in December? (ii) Why are the death rates of young widowers some 6 times higher than those for married persons of the same age? Although these are long-standing problems at the present moment the answers still elude us.

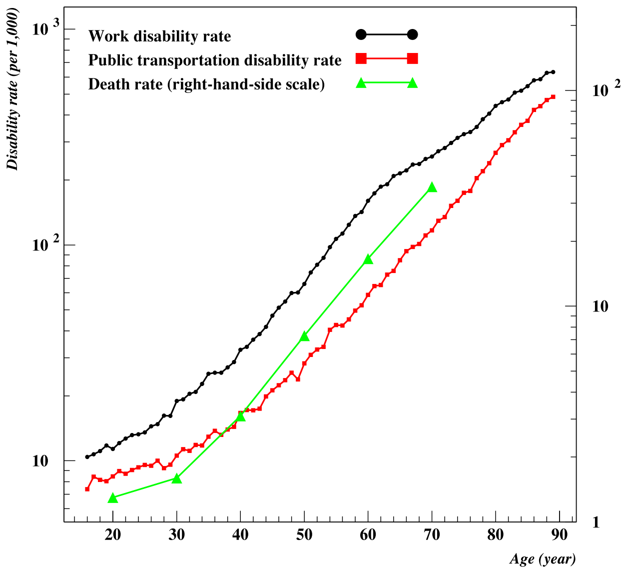

Fig. A.1. Comparison of two types of disability rates to death rates, US census of 1980 (1%). The fact that the three lines are roughly straight and parallel shows that they follow Gompertz's law according to which death rates increase exponentially with age. Note that the amplitudes of the two vertical axis have been taken identical for otherwise the comparison of the slopes would not be significant.
Source: IPUMS database, 1% sample of the census of 1980; available from [10] (IPUMS).

Appendix. Complementary data

In this appendix, on the request of one of our reviewers, we present some complementary data. We think this practical information can be useful for readers who wish to start their own research based on IPUMS data. In contrast with data for stocks that many econophysicists are using, the IPUMS data are structured, low-noise data. Apart from the US censuses used in this paper, IPUMS has also an international section which provides similar census records for a number of countries.

A.1. An example of data provided by IPUMS

The IPUMS database is very user-friendly and easy to use. It is free, one only needs to register. In the data files which are downloaded from IPUMS each line corresponds to one person. The variables given for this person are selected by the user from a menu.

The following line belongs to the file used to draw Fig. A.1. Altogether this file has 1,748,374 lines (14 M). Below we give one line as an example.

One individual → one line → 2061611 → 2 061 6 1 1 →
2 = female
061 = age
6 = Never married
1 = No disability that affects work
1 = No public transportation disability.

In contrast with 061 which represents a "real" data, all other numbers are code numbers which take their significance only through the conversion tables given in IPUMS.

A.2. Relationship between death rates and disability rates

The logarithms of the work- and transportation disability rates have a correlation of 0.984 and the relationship between them reads: $T \sim W^{0.96}$.

The public transportation disability data are available only in the census of 1980. That is why we have been using the work disability data.

References

[1] L.-A. Bertillon, Article "mariage", in: Dictionnaire Encyclopédique Des Sciences Médicales, in: Encyclopedic Dictionary of Medical Sciences, Vol. 5, 1872, pp. 7–52 2nd series. Available on "Gallica", the website of digitized publications of the "Bibliothèque Nationale de France" (French National Library) at: http://www.bnf.fr.

[2] W. Farr, The Influence of Marriage on the Mortality of the French People, Savill and Edwards, London, 1859.
[3] P. Richmond, B.M. Roehner, Effect of marital status on death rates. Part 1: High accuracy exploration of the Farr-Bertillon effect, Physica A 450 (2016) 748–767.
[4] P. Richmond, B.M. Roehner, Effect of marital status on death rates. Part 2: Transient mortality spikes, Physica A 450 (2016) 768–784.
[5] W.R. Gove, Sex, marital status, and mortality, Am. J. Sociol. 79 (1) (1973) 45–67 [This study relies on the data given in "US Public Health Service" 1970 cited below.].
[6] US Public Health Service. Mortality from selected causes by marital status. United States, Part A. Rockville (Maryland), 1970.
[7] Y. Ben-Shlomo, G.D. Smith, M. Shipley, M.G. Marmot, Magnitude and causes of mortality differences between married and unmarried men, J. Epidemiol. Community Health 47 (3) (1993) 200–205.
[8] K. Williams, D. Umberson, Marital status, marital transitions, and health; a gendered life course perspective, J. Health Soc. Behav. 45 (1) (2004) 81–98.
[9] J. Robards, M. Evandrou, J. Falkingham, A. Vlachantoni, Marital status, health and mortality, Maturitas 73 (4) (2012) 295–299.
[10] S. Ruggles, K. Genadek, R. Goeken, J. Grover, M. Sobek, Integrated Public Use Microdata Series, (IPUMS), University of Minnesota, Minneapolis (Minnesota), 2017.
[11] G.M. Viswanathan, M.G.E. da Luz, E.P.R. Raposo, H.E. Stanley, The Physics of Foraging. An Introduction to Random Searches and Biological Encounters, Cambridge University Press, Cambridge, 2011.