

A non-parametric approach to decompose the young adult mortality hump by causes of death

Adrien Remund*, Tim Riffe†, and Carlo Giovanni Camarda‡

March, 2016

Abstract

We propose a cause-deletion non-parametric approach to decompose the young adult mortality hump by cause of death. This novel method isolates the specific causes that contribute to the temporary deviation of the age-specific mortality rates in early adulthood, irrespective of their share in the absolute number of deaths. We apply our method to US males and females from 1959 to 2010. Preliminary results show that life expectancy is typically reduced by about 1 year for males and 0.2 year for females due to the presence of the hump. For males, accidents used to account for the vast majority of the hump in the 1960s but have been now overtaken by homicides, suicides, and poisoning. For females, traffic accidents have remained the main driver of the variations until now. Moreover, each cause peaks at different ages, suggesting that they may be generated by different underlying processes.

1 Background

The young adult mortality hump has been known since at least the second part of the nineteenth century (Thiele, 1871), although it received little attention until the publication of Heligman Pollard's model of mortality (Heligman and Pollard, 1980). In comparison with senescence, which drives the exponential increase of the force of mortality during the second part of adulthood, and ontogenescence, which explains the decrease of the risk of death in the first years of life, young adult "excess" mortality has never been properly defined, measured, or explained.

We suggest using a broad definition for this phenomenon, namely as a "temporary deviation of the aggregate age-specific death rate from the exponential trend due to senescence, during adolescence and/or early adulthood" (Remund, 2015). This feature is most visible in logged mortality rates, rather than absolute rates, which means it is as much a feature of the rate of change over age of mortality rates as it is of absolute mortality rates. In fact the hump may or may not be present and visible in a log mortality curve under any absolute level of mortality. To compare humps between curves, it is thus necessary to adjust somehow for the overall mortality level, i.e. relate it to the full age pattern of mortality. Common age-cause decompositions of mortality differences miss this key aspect of the hump.

The study of the leading causes of death in early adulthood using raw age-specific death rates has already been attempted (Heuveline, 2002; Blum, 2009), but this approach suffers from two drawbacks. First, the contribution to the hump must depend on the shape of the cause-specific forces of mortality and not on the absolute levels of age-specific death rates. Second, the age range considered should not be arbitrary fixed but instead adjusted to the period of life that is affected by excess mortality in each population.

*Institut national d'études démographiques & University of Geneva

†Max Planck Institute for Demographic Research

‡Institut national d'études démographiques

In this paper we present a non-parametric method to decompose young adult excess mortality into age- and cause-specific components. This method has the key advantages to be independent of the general mortality level, adapt to a wide range of possible shapes of mortality schedules, and to assure a coherence between the all-cause and the cause-specific hump. We then use it to study how causes of death have shaped the evolution of the young adult mortality hump amongst US males from 1960 to 2010.

2 Methods

Parametric models that include a hump, such as (Thiele, 1871), Heligman and Pollard (1980) or Kostaki (1992), are limited because they fail to adapt to the diversity of mortality schedules. These limitations have motivated the development of non-parametric models, notably based on P -splines (Camarda, 2008), which are not designed to estimate young adult excess mortality separately from child and adult mortality. Alternatively, a non-parametric equivalent to the component models was proposed under the name of Sum of Smooth Exponentials (SSE) (Camarda et al., 2016).

Briefly, the SSE model describes the observed force of mortality over age $\boldsymbol{\mu}(\boldsymbol{x})$ as the sum of three vectors $[\gamma_1 : \gamma_2 : \gamma_3]$ over age. The superscripts denote which mortality component each γ_j refers to: infant, midlife and old-ages mortality, respectively. In other words it assumes that deaths are realizations from Poisson distributions with mean composed of three parts:

$$\boldsymbol{y} \sim \mathcal{P}(\boldsymbol{e}\boldsymbol{\mu} = \boldsymbol{C}\boldsymbol{\gamma}), \quad (1)$$

Given m age groups, the matrix \boldsymbol{C} is a m by 3 matrix repeating the exposures in each column, which, when multiplied with the γ_j , sums the components and transforms the result into death counts. In matrix notation, \boldsymbol{C} is given by

$$\boldsymbol{C} = \mathbf{1}_{1,3} \otimes \text{diag}(\boldsymbol{e}), \quad (2)$$

where $\mathbf{1}_{1,3}$ is a 1×3 matrix of ones, $\text{diag}(\boldsymbol{e})$ is the diagonal matrix of the exposure population and \otimes denotes the Kronecker product.

Unlike parametric models, in the SSE model there is no need to make strong assumptions about the functional form of each component. For each component we assume a discrete sequence and we apply the exponential function to ensure non-negative elements:

$$\gamma_j = \exp(\boldsymbol{X}_j\boldsymbol{\beta}_j), \quad j = 1, 2, 3. \quad (3)$$

In other words, each component has to be described by a linear combination of a model matrix \boldsymbol{X}_j and associated coefficients $\boldsymbol{\beta}_j$. The design matrices \boldsymbol{X}_j can represent parametric or non-parametric structures. In this way the composite force of mortality $\boldsymbol{\mu}$ can be viewed as sum of 3 exponential components, which potentially can be smooth. Further, the SSE model allows us to incorporate shape constraints to enforce senescent and young-adult components to be monotonically increasing and log-concave, respectively.

A Penalized Composite Link Model (PCLM) has been proposed to estimate the SSE model, and more details on the model and its estimation procedure can be found in Camarda et al. (2016). In the following analyses, we focus on the accident hump, therefore estimations start from the onset of adolescence (age 10) disregarding the child mortality component.

Building on this non-parametric method, we propose a constrained approach to decompose the estimated hump into cause- and age-specific contributions. We proceed in four steps.

1. Based on cause-specific age trajectories, we identify those causes of death with an evident young adult hump component (cf. Data section).

2. We estimate an SSE model on the overall mortality in order to separate the senescent and young adult hump components.
3. We construct cause-deleted datasets by removing separately deaths from each cause that was identified in step 1.
4. We simultaneously estimate SSE models for each of these cause-deleted datasets, interpreting the diminution of each component as the contribution of this cause to that component, and constraining the sum of all these contributions to be equal to the components estimated in step 2.

To better grasp these steps, we render this approach in formula and apply it to a simplified simulation. For ease of presentation we use only three causes of death (A , B and C) with only A and C clearly affecting young-adult mortality (Figure 2). Cause A displays a strong hump between about age 15 and 40, and then decreases to a very low level. Cause C combines an exponential trend (Gompertz, 1825) and a clear hump, which reaches a higher absolute level than for cause A . Finally, cause B does not display any hump and follows a gompertz trend from about age 10 onward, although it always reaches a higher absolute death rate than cause A , even at the peak of the hump. If we were only focusing on the absolute cause-specific levels of mortality between 15 and 40 years of age, we would conclude to the following cause-specific contributions: C (52% of the deaths), B (40%) and A (8%). It is however quite obvious from the age- and cause-specific death rates that cause B should not contribute at all to the hump since it itself does not display any form of deviation around these ages.

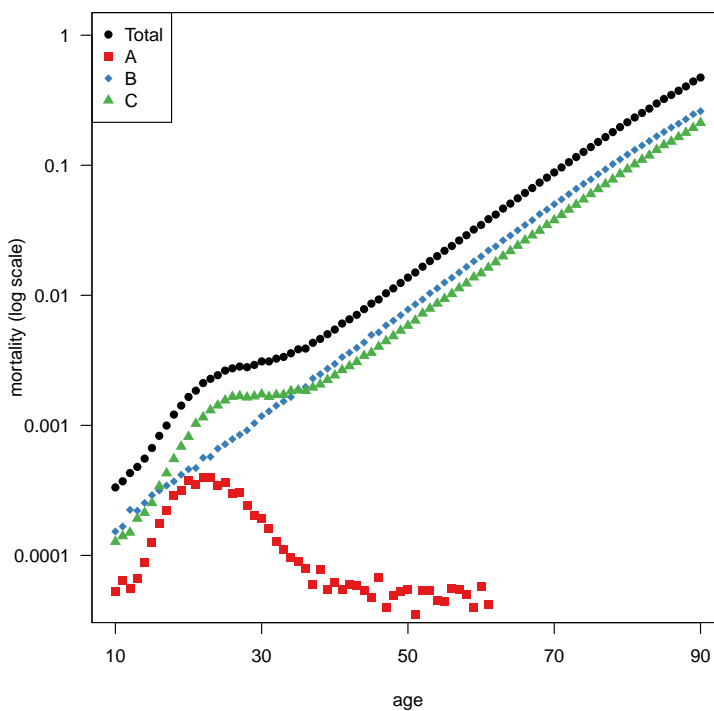


Figure 1: Cause- and age-specific death rates for simulated example

By fitting an SSE model on the overall force of mortality (Figure 2), we are able to estimate the expected deaths due to young adult excess mortality ($\hat{y}_1 = e \hat{\gamma}_1$) and senescence-driven mortality ($\hat{y}_2 = e \hat{\gamma}_2$). This provides a reference that the cause-specific contributions have to match in order to preserve the coherence between all-cause and cause-specific mortality.

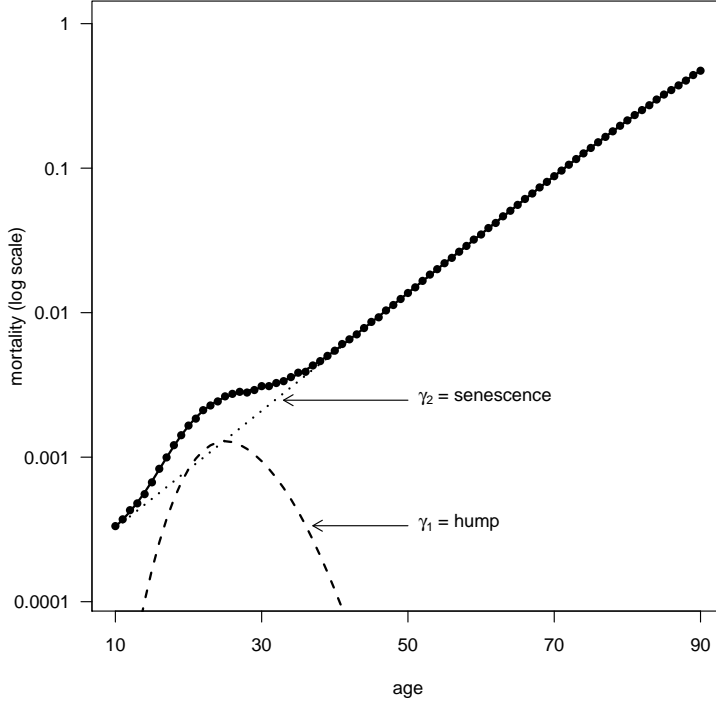


Figure 2: All-cause mortality for simulated example

The cause-specific contributions (δ_1^κ and δ_2^κ) to each of the two components (γ_1 and γ_2) are then estimated by refitting simultaneously the same SSE model on cause-deleted datasets. More specifically, we generate as many cause-deleted life tables as there are causes of death that are assumed to contribute to the hump. In our simulated example, only two such datasets must be computed, for causes A and C. This constrained model can be written as a system of SSE models and constraints such as

$$\left\{ \begin{array}{l} \mathbf{y}^{-A} \sim \mathcal{P}(\mathbf{C} \boldsymbol{\gamma}^{-A}) \\ \mathbf{y}^{-C} \sim \mathcal{P}(\mathbf{C} \boldsymbol{\gamma}^{-C}) \\ \hat{\mathbf{y}}_1 = \mathbf{e} \cdot (\delta_1^A + \delta_1^C) \\ \hat{\mathbf{y}}_2 - \hat{\mathbf{y}}^B = \mathbf{e} \cdot (\delta_2^A + \delta_2^C) \end{array} \right. \quad (4)$$

where the structures of matrices \mathbf{C} are given in (2), $\delta_j^\kappa = \gamma_j - \gamma_j^{-\kappa}$ (Figure 2) is the contribution of cause κ to the j -th component. By concatenating, the left-hand factors of (4) can be defined as $\check{\mathbf{y}}$, and the right-hand factors as $\check{\boldsymbol{\gamma}}$.

$$\check{\mathbf{y}} = \begin{bmatrix} \mathbf{y}^{-A} \\ \mathbf{y}^{-C} \\ \hat{\mathbf{y}}_1 \\ \hat{\mathbf{y}}_2 - \hat{\mathbf{y}}^B \end{bmatrix} \quad (5)$$

$$\check{\boldsymbol{\gamma}} = \begin{bmatrix} \boldsymbol{\gamma}_1^{-A} \\ \boldsymbol{\gamma}_2^{-C} \\ \boldsymbol{\gamma}_2^{-A} \\ \boldsymbol{\gamma}_2^{-C} \end{bmatrix} \quad (6)$$

We can therefore re-write the suggested approach as a single model with a composed mean as in (1)

$$\mathbf{y} \sim \mathcal{P}(\mathbf{C}' \boldsymbol{\gamma}'),$$

where the composite matrix takes the following form:

$$\check{\mathbf{C}} = \begin{bmatrix} \mathbf{I}_2 & \otimes & \mathbf{C} \\ \mathbf{1}_{1,2} & \otimes & \text{diag}(\mathbf{e} : \mathbf{e}) \end{bmatrix} \quad (7)$$

where \mathbf{I}_2 is an identity matrix of dimension 2, i.e. the number of components, and $\mathbf{1}_{1,2}$ is a matrix of ones of dimension 1×2 , i.e. the number of accident-hump related causes.

In this way, by augmenting both \mathbf{C} and $\boldsymbol{\gamma}$, a PCLM can be used for simultaneously estimating cause-specific contributions to each components which are constrained to the overall components. Obviously more complex settings with more hump-related causes can be accommodated by a small augmentation of the model elements.

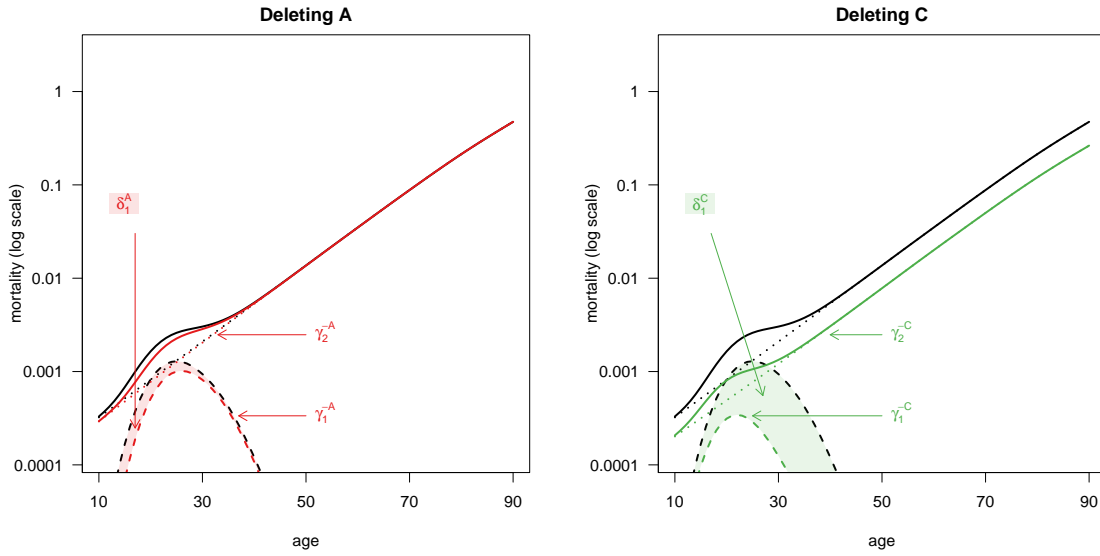


Figure 3: Cause-deleted mortality for simulated example (overall components from Fig. 2 in black)

Once the cause-specific contributions to the hump δ_1^κ have been computed, they can be interpreted as densities (Figure 2), from which many different indicators of magnitude, timing and spread can be extracted for our simulated data (see box). By working in a smooth setting, we are able to evaluate the components with fine age-granularity. This allows us to consider components as continuous functions ($\delta_j^\kappa(x) \approx \delta_j^\kappa$).

$Area$	$= \int_0^\omega \delta_1^\kappa(x) dx$
Δ^κ	$= e_{10}^{-\delta_1^\kappa} - e_{10}$
$\bar{\kappa}_1$	$= \frac{\int_0^\omega \delta_1^\kappa \cdot x dx}{\int_0^\omega \delta_1^\kappa dx}$
M^κ	$= \arg \max_x \delta_1^\kappa$
σ^κ	$= \frac{\int_0^\omega \delta_1^\kappa \cdot \bar{x}_1^\kappa dx}{\int_0^\omega \delta_1^\kappa dx}$

Measures of magnitude include the total area of the contributions, which indicates in our example the following relative contributions : 77% (C) and 23% (A), which differ drastically from the initial proportions derived from the absolute death counts. Another measure of magnitude consists in estimating the potential gain of life expectancy at age 10 that would result from the deletion of the

cause-specific contribution to the hump. In our example, cause A induces a loss of 0.18 years of life expectancy, against 0.54 for cause C.

Measures of timing include the mean age at death and the mode of the cause-specific hump. The latter is preferable as the cause-specific contributions to the hump, and the overall hump itself, are rarely symmetrical. In our example, the peak of the overall hump is located at age 25, while it is reached at age 22 for cause A and 26 for cause C.

Measures of spread include the standard deviation of the age at death for the individuals who die from the cause-specific contribution to the hump. In our case, the standard deviation reaches 5.98 years for the total hump, against 5.69 for cause A and 5.78 for cause C, meaning expectably that the total spread exceeds that of each cause.

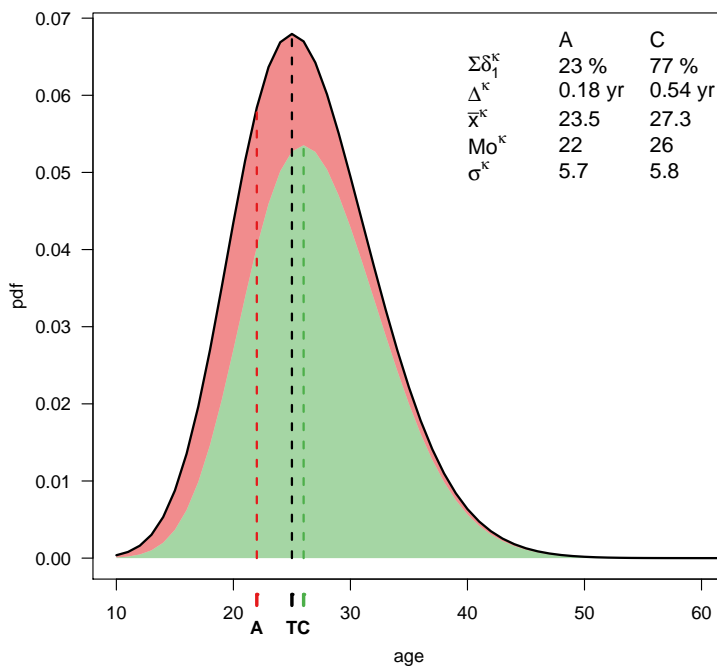


Figure 4: Cause-specific density of the hump and associated measures for simulated example

3 Data

We use an early release of data produced by the Human Mortality Database on cause- and age-specific death rates for the USA between 1959 and 2010¹. Working from the 2010 data for males that cover 92 causes of death, we look for the typology that best isolates those causes that are relevant to young adult excess mortality. To achieve this goal we cluster the 92 causes of death by the shape of their rate of ageing (relative first derivative of the force of mortality) between the age of 10 and 29. Then we compute the euclidean distances between each cause. Finally we group them into clusters using a hierarchical clustering algorithm. This technique, similar to the one used on mortality above age 30 by Camarda (2014), leads to the identification of six clusters with between one and 77 of the original

¹These data are not treated to insure smooth transitions between revisions of the ICD.

92 HMD standardized causes of death.

Three of the clusters are composed of a single cause, and the last one is the sum of eight causes. The single-cause clusters are traffic accidents, suicides, and homicides. Accidental poisonings also emerge as a single cluster, but we grouped it with alcohol abuse, accidental poisoning by alcohol and drug dependence for coherence. A fifth group that emerges is composed of several causes that share a similar age-shape and fall broadly in the category of other accidents. Moreover, we manually isolate HIV-AIDS, in order to take into account this cause of death that was much more present in the late 1980s and early 1990s. Finally, all the 77 other causes of death are grouped into a large cluster that does not display any hump (for details, see table 1 in Appendix). This list of six groups of causes of death that contribute to the hump, plus a seventh that has no hump, proved remarkably valid for all the periods considered and is thus applied throughout our analyses.

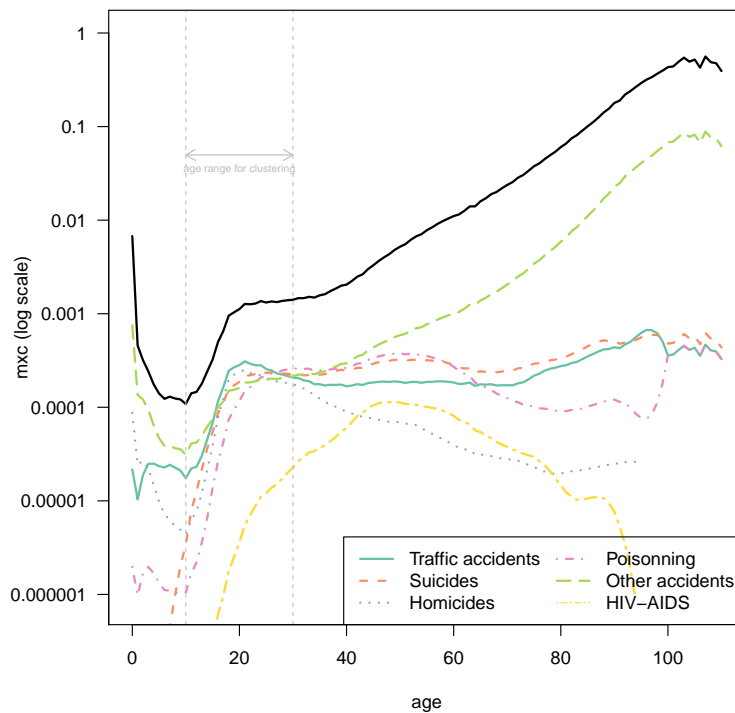


Figure 5: A typology of causes of death contributing to the hump, US males 2010

4 Results

We apply our model to the mortality of US males and females from 1959 to 2010 using the cause-of-death typology described above. The results are first presented in the form of Lexis surfaces of the cause-specific contributions to the hump (figures 4 and 4). The comparison of genders shows a more instable evolution of the females series, which is due to the smaller magnitude of the female hump that makes it more difficult to capture. This is especially true for the 1960s, when young adult excess mortality was almost absent for females, but is present for all cause-specific series over the whole period. The comparison however also highlights similarities between genders, such as the progressive widening of the overall hump from 1980, until its brutal shrinkage in 1995, and its re-widening in the last years. This 3-step evolution is the main characteristic of the american young adult excess mortality hump and deserves more attention.

For both genders, the overall hump remained relatively concentrated in the twenties from 1959 to about 1980, but later quickly spread to the thirties and even the forties. This evolution can be explained by the combined forces of the widening and the shift of different cause-specific contributions. Indeed, starting in the late 1970s, the contribution of suicides and poisoning to the hump tended to spread more and more to older ages. This could be due either to the fact that in this period individuals in their twenties and thirties became more and more susceptible to these causes of death, i.e. that the specific form of vulnerability associated with early adulthood spread to older ages. This would correspond to the process of destandardization that has been identified in the literature on the transition to adulthood (Billari and Liefbroer, 2010, e.g.), as well as the negative consequences of globalization on young adults that face more insecurity at a time of long-term decision-making (Blossfeld et al., 2005). Alternatively, this pattern could also be the result of a cohort effect associated with the individuals born during the baby-boom. According to this interpretation, the cohorts born just after the Second World War were negatively affected by their larger size in a so-called Easterlin's hypothesis (?).

Both interpretations differ in the fact that the increase of contributions to older ages is due either to period effects that extend the ages that are affected by early adulthood vulnerability while younger ages are not affected, or to cohort effects that progressively shift the contributions to older ages while the new cohorts are not affected when they enter adulthood. In practice, the former explanation seems to apply better to suicides, whereas the latter may be more suited to deaths by poisoning. Indeed, the contribution of suicides to the hump tends to remain relatively stable at age 20 while it expands in the thirties during the 1980s. Poisonings, on the contrary, tend to disappear in the early twenties during the same period. This difference might be explained by the fact that deaths by poisonings are often the consequence of long-term addictions to alcohol and/or drugs that start in adolescence and are transferred later into adulthood, in a process not dissimilar to smoking. Suicides, however, can be seen more as a reaction to distress caused by the current context and are thus more a mark of period than cohorts.

A third process explains the widening of the overall hump to older ages in the 1980s, namely the apparition of HIV-AIDS mortality. Indeed, this new cause of death pushes the force of mortality especially in the thirties and thus participates to the flattening of the hump after its peak. HIV-AIDS deaths appear in the data in 1988 and brutally decrease in 1996 with the introduction of antiretroviral therapies. This decrease coincides with the concentration of the overall hump around 20 years of age.

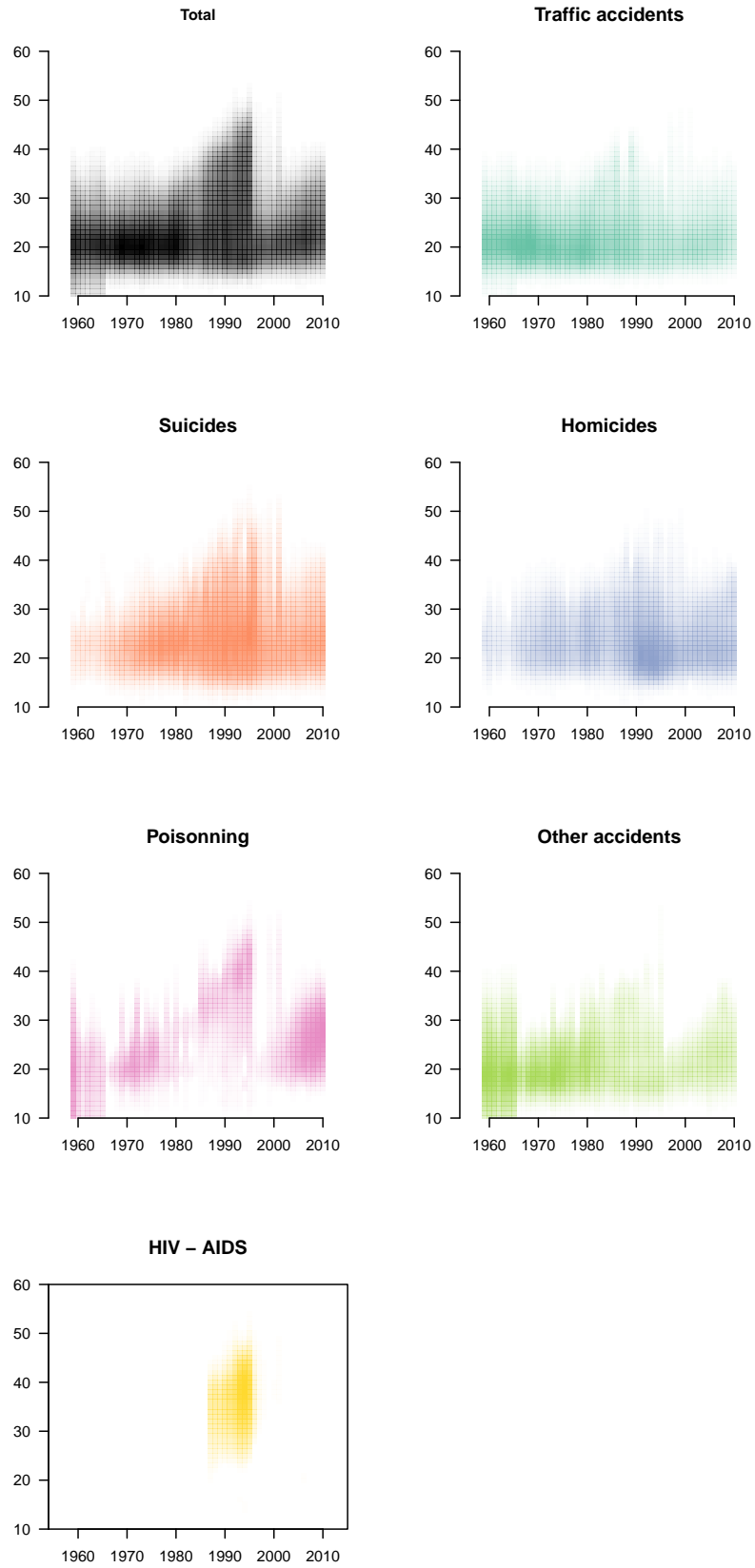


Figure 6: Lexis surfaces of cause-specific contributions to the hump, US males 1959-2010

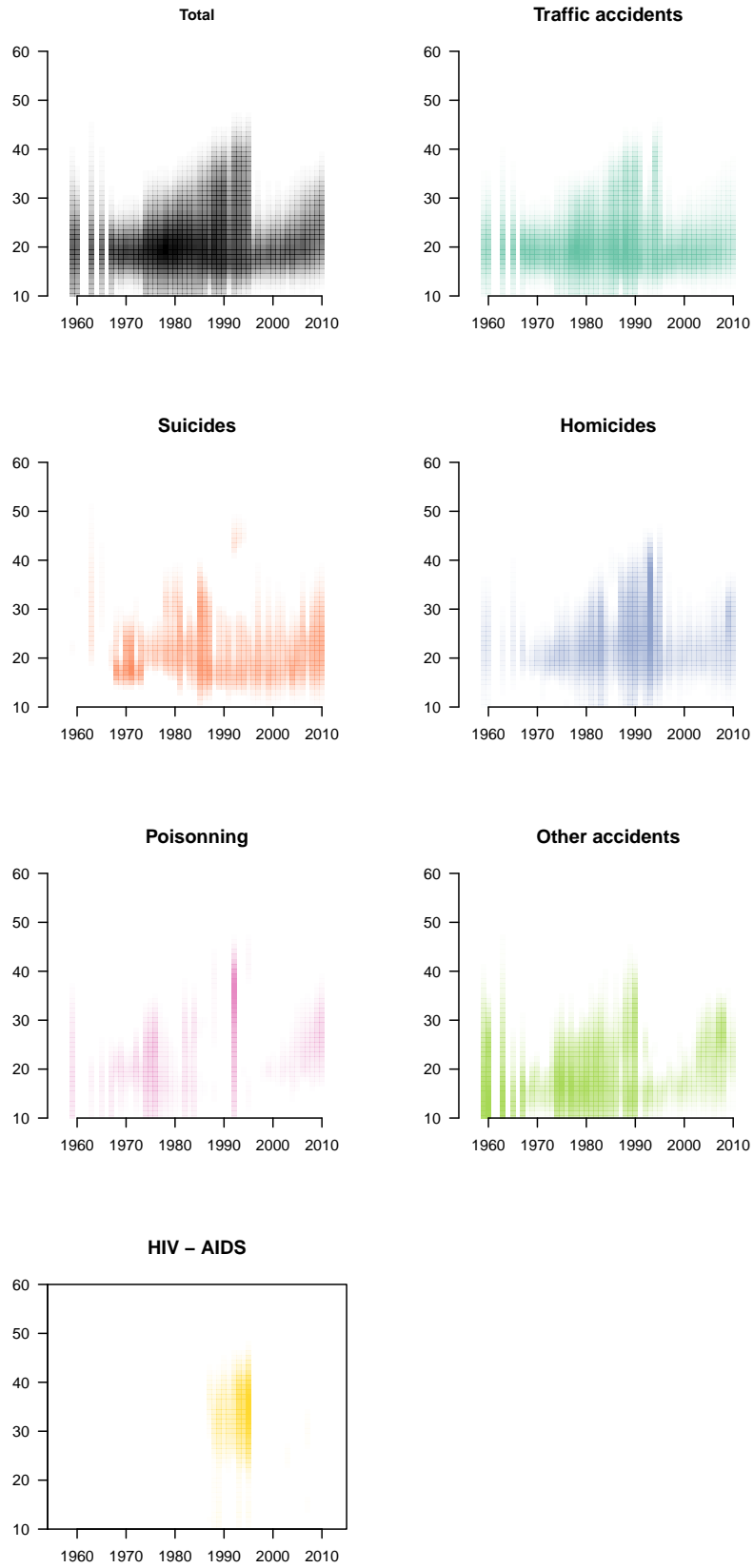


Figure 7: Lexis surfaces of cause-specific contributions to the hump, US females 1959-2010

It is when focusing on the magnitude of the hump that gender differences become more obvious. Overall, young adult excess mortality costs slightly less than one year of life expectancy for males, while it only peaks at a third of this for females in 1990. Although both genders experience a drop of the hump in the mid-1990s, the similarities stop here both in the overall hump and in the cause-of-death contributions. Throughout the whole period from 1959 to 2010, males typically lost about 0.8 year of life due to the hump, with a 0.3 increase from 1988 to 1995 due to HIV-AIDS. Females, however, experienced virtually no young adult excess mortality in the early 1960s, but it gradually increased until 1990, before plummeting to a minimum in 1997, with a slow regain during the last years.

In terms of cause-of-death contributions, it appears that in 1960 the often-used denomination of "accident hump" (Heligman and Pollard, 1980; Goldstein, 2011) was justified for males as traffic and other accidents made up about 80% of the hump. This situation slowly evolved however over time and nowadays these two causes only account for a third of the hump. Meanwhile, suicides and homicides have grown from less than 10% to about half of the hump. As for poisonings, they generally did not contribute more than 5%, except in the last years when their share grew to 17%. The story for males is thus very much about suicides and homicides slowly replacing accidents in an otherwise stable overall hump.

The pattern is completely different for females, as most of the evolution of the overall hump can be explained by traffic accidents. Indeed, not only do they make up most of the hump at any point in time (from 40% to 80%), their evolution is also highly correlated with the overall hump ($r = 0.84$). The other causes of death never really significantly contributed to the hump for females, not even HIV-AIDS in the early 1990s. Perhaps quite surprisingly, although males are often pictured as the main victims of traffic accidents, this cause of death is actually much more influential on female young adult excess mortality.

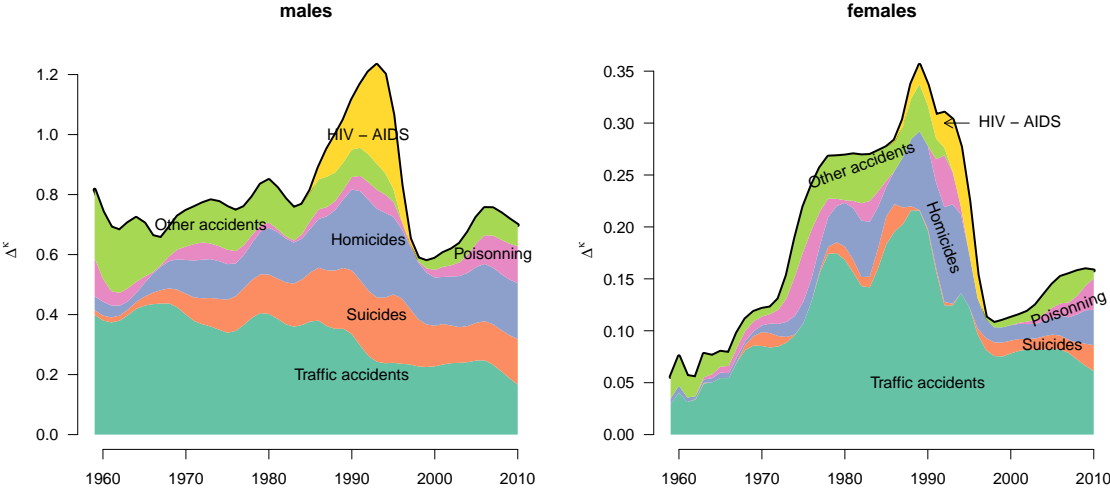


Figure 8: Cause-specific loss of life expectancy due to the hump, US 1959-2010

5 Conclusion

The young adult mortality hump is a well-known phenomenon, but its interpretations remain unsatisfactory. One reason for this gap is the lack of methodological tools to capture this deviation from senescence in the force of mortality, irrespectively from the absolute death rates. In this paper we offer

a new method that decomposes not only the force of mortality into a senescence and a hump components, but also the all-cause mortality hump into cause-specific contributions. In order to achieve this goal, we apply a sum of smooth exponentials on both the all-cause and cause-deleted forces of mortality within a constrained environment, and interpret the decrease of the hump after the deletion of a given cause as the contribution of this cause to the hump.

The application of this method to the American series from 1959 to 2010 shows both similarities and dissimilarities between males and females. Both genders experienced a widening of the young adult mortality hump during the 1980s, followed by a brutal concentration in the mid-1990s and a subsequent re-widening in the last years. This movement was probably driven by a spread of suicides to older ages, and an increase of poisoning due to the passage of the baby-boom cohort, as well as the apparition of HIV-AIDS mortality. Males and females differ however in the magnitude of the hump, that constantly costs to the former about 0.8 years of life expectancy, while the latter went from barely anything in the 1960s to 0.3 years lost in 1990, back to 0.1 in the past two decades. Moreover, while the men have experienced a continuous decrease of accidents in the hump, compensated by an increase of suicides and homicides, female young adult mortality has been almost solely driven by traffic accidents.

These results confirm the interest of developing more detailed measures of young adult excess mortality in order to better understand the forces that shape this peculiar demographic phenomenon. Further analyses should be run on a broader set of countries, and possibly on even older time periods, in order to see if the American case reflects the general evolution or if it is exceptional. In any case, our results already suggest that the term "accident hump" should be avoided when addressing young adult excess mortality, as other causes of death are involved in this phenomenon.

6 Appendix

Table 1: Typology of causes of death

Group	Contributing causes	Sum of deaths in 2010
Traffic accidents		25'286
Suicides		30'966
Homicides		12'926
Poisoning	Alcohol abuse	5'190
	Drug dependence	1'411
	Accidental poisoning by alcohol	1'652
	Other accidental poisoning	20'006
Other accidents	Accidental falls	12'913
	Events of undetermined intent	2'949
	Other diseases of the nervous system and the sense organs	16'630
	Other external causes	3'110
	Other forms of heart disease	74'840
	Other malignant neoplasms	38'842
	Unknown and unspecified causes	6'045
	Other accidents	15'952
HIV - AIDS		6'266
Other		996'289
total		1'232'432

References

- Francesco Billari and Aart Liefbroer. Towards a new pattern of transition to adulthood? *Advances in Life Course Research*, 15(2&€“3):59–75, 2010.
- Hans-Peter Blossfeld, Erik Klijzing, Melinda Mills, and Karin Kurz. *Globalization, uncertainty and youth in society*. Routledge, London, 2005.
- Robert W Blum. Young people: not as healthy as they seem. *The Lancet*, 374(9693):853–854, 2009.
- Carlo G. Camarda. *Smoothing methods for the analysis of mortality development*. PhD thesis, 2008.
- Carlo G. Camarda, Paul H. C. Eilers, and Jutta Gampe. Sums of smooth exponentials to decompose complex series of counts. *Statistical Modelling*, 2016.
- Carlo Giovanni Camarda. Reconstructing mortality series by cause of death: two alternative approaches, 2014.
- Joshua Goldstein. A secular trend toward earlier male sexual maturity: Evidence from shifting ages of male young adult mortality. *PLoS ONE*, 6(8), 2011.
- Benjamin 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. *Philosophical Transactions of the Royal Society of London*, 115:513–583, 1825.
- L. Heligman and J. H. Pollard. The age pattern of mortality. *Journal of the Institute of Actuaries*, 107, 1980.
- Patrick Heuveline. An international comparison of adolescent and young adult mortality. *The ANNALS of the American Academy of Political and Social Science*, 580(1):172–200, 2002.
- Anastasia Kostaki. A nine&€“parameter version of the heligman&€“pollard formula. *Mathematical Population Studies*, 3(4):277–288, 1992.
- Adrien Remund. *Jeunesses vuln“rables? Mesures, composantes et causes de la surmortalit“ des jeunes adultes*. PhD thesis, 2015.
- Thorvald Nicolai Thiele. On a mathematical formula to express the rate of mortality throughout the whole of life, tested by a series of observations made use of by the danish life insurance company of 1871. *Journal of the Institute of Actuaries and Assurance Magazine*, 16(5):313–329, 1871.