

# **Lexis Fields**

**(authors redacted)**

## **Abstract**

### **BACKGROUND**

Lexis surfaces are an established visualization technique to show how a given value changes over age and time. Vector fields are a two-dimensional representation of two variables: direction and speed (or force).

### **OBJECTIVE**

We aim to increase the dimensionality of patterns shown on the Lexis surface by placing a vector field on the Lexis surface.

### **RESULTS**

We show Lexis fields of the relationship between life expectancy and the standard deviation of remaining lifespan over age and time. These instruments enable information layering on standard Lexis surfaces that is not common practice.

### **CONCLUSIONS**

Lexis fields extend the descriptive and analytic power of the Lexis surface, and these can be designed to display information at higher densities than commonly made Lexis surfaces.

TODO TR: add area examples.

## 1. Introduction

Lexis surfaces are a graphical tool used to display data on the Lexis coordinate plane, a Cartesian plane that is also a simplex relationship between age, period, and cohort. Surfaces are often displayed as heat maps, contour maps, perspective plots, or variants of these things (Vaupel, Gambill, and Yashin 1987). Various kinds of quantities, such as raw magnitudes, differences (Minton et al. 2017), excesses (Remund, Camarda, and Riffe 2018, Acosta and van Raalte (2019)) ratios (Canudas-Romo and Schoen 2005), intensities, proportions, derivatives (Rau et al. 2017), and even compositions (Schöley and Willekens 2017) can be displayed on Lexis surfaces in order to put age, period, cohort, or other patterns in relief.

Vector fields are a graphical form generally used to display variation in speed, direction, or force over a plane (Weiskopf 2007). Point estimates of these quantities on the plane are often represented with segments or arrows, where length may be proportional to a function of magnitude (force, speed), and angle indicates direction, potentially disambiguated with an arrowhead or articulated as a curve.

Maps in general combine layers of categorical, continuous, and symbolic information on a common spatial projection. Lexis surfaces in contrast almost exclusively display one visual layer at a time. We propose to enrich Lexis surfaces by adding a visual layer of quantitative information coded symbolically as a vector field, and we liken this to cartographic information layering. We propose a fusion of Lexis surfaces and vector fields, *Lexis fields*, as a tool to display variation in relationships between variables over age and time. A Lexis field may either be shown atop on a Lexis surface, representing two map layers — a true Lexis map — or as a single-layer stand-alone visualization. Visual patterns in a Lexis field reveal changes in complex relationships over age and time.

We give an overview of constructing a Lexis field with an application. Our example explores the relationship between remaining life expectancy and the standard deviation of remaining lifespan over age and time based on all available populations in the Human Mortality Database (2018) from 1950 onward. This example and others discussed motivate how Lexis fields could be used in other areas demographic research.

## 2. Lexis field construction

It makes sense to plot a Lexis field if data contain a relationship that can be summarized with a line, a simple curve, or similar, that varies by age and/or time. Constructing a Lexis field involves several design choices, which can however be codified into four basic

steps, which we summarize in Fig. 1. The steps to do so are outlined in the following description, referenced to regions of Fig. 1.

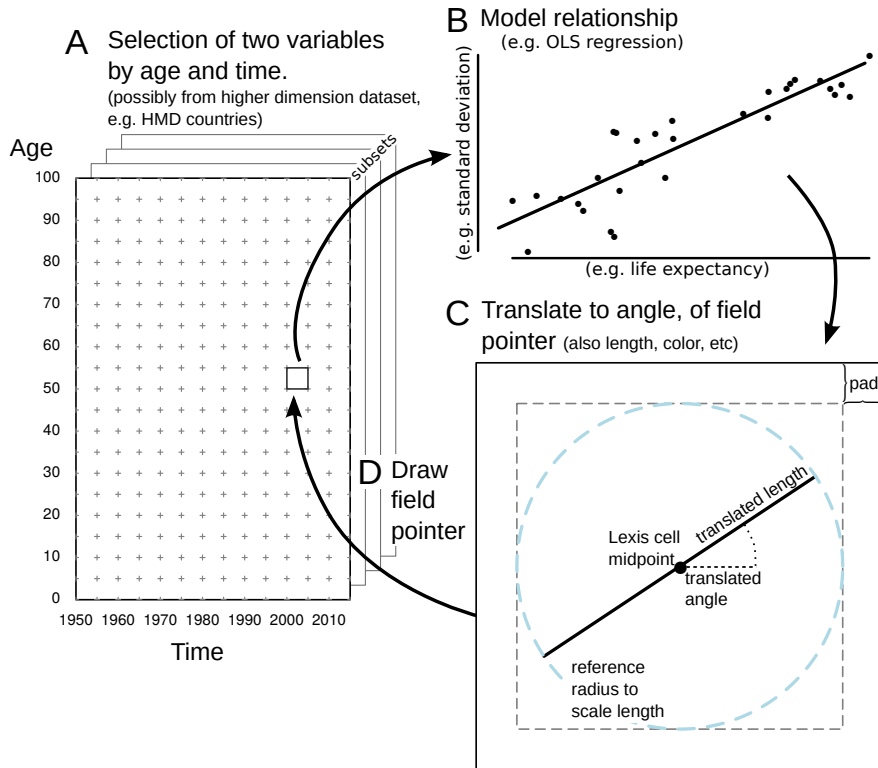
**A** Determine a grid over age and time on which to condition data selection. For example, a five-year grid implies  $5 \times 5$  Lexis cells. Data may be selected from multiple populations on the same reference grid or may consist in a selection of attributes from a single underlying Lexis surface. Presumably two variables are required.

**B** Abstract a model from the data, such as a linear, parabolic, or similarly simple relationship.

**C** Translate the model to the characteristics of a line segment, which we refer to as a field *pointer*. For example, the pointer angle can relate to the model relationship, and length and color may be determined by other aspects of the model, such as  $r^2$  or the Pearson's correlation coefficient.

**D** Draw the segment in the corresponding Lexis grid cell, repeating all steps for each cell in the diagram.

**Figure 1:** A diagram depicting the translation of functional relationships in data conditioned on age and time to visual encoding on a Lexis field. **A:** Condition data selection on age and time. **B:** Fit a model (e.g. linear) to the data subset. **C:** Translate the model to field elements, ‘pointers’, using angle, and possibly also length, curvature, thickness, etc. to represent model qualities. **D:** Populate the Lexis plane with field pointers to create a Lexis field.



### 3. Application

We select all HMD data available for females after 1950. For each life table we calculate an extra column for the standard deviation of remaining lifespan  $sd(x)$ . Each Lexis field element is based on the relationship between  $sd(x)$  and remaining life expectancy  $e(x)$

in  $1 \times 5$  Lexis cells <sup>1</sup>, as summarized by bivariate linear regression over the data points in each cell. The regression results used for each Lexis cell in resulting Lexis fields are identical, but the translation of regressions to field pointers varies between designs. We offer four examples of Lexis field designs, each drawn on  $5 \times 5$  Lexis subplots.

The first of these, Fig. 2, is a bare-bones Lexis field that serves to illustrate the underlying concept. This display draws each regression slope as a line segment of equal length (4 “years” long) and centered on each Lexis cell. The slope of each pointer is identical to regression slopes, which may be justified in this case because  $e(x)$  and  $sd(x)$  are both in year units. This is the truest and most literal depiction of how these regression slopes vary over age and time among females. From this figure we can see, for example, that there is some age where the relationship turns from negative to positive, which increased slightly over time. Slopes dampened in younger ages around the 1980s, but have since increased (except infants).

The second version, Fig. 3, draws the same slopes multiplied by two, with pointer lengths scaled proportional to the between-population interquartile range of the  $e(x)$  and  $sd(x)$  values used in regressions<sup>2</sup>, and line weight and grayscale proportional to the  $r^2$  of the regression fit. Segment lengths are therefore indicative of the spread in the data, while higher  $r^2$  results in darker and thicker pointers.

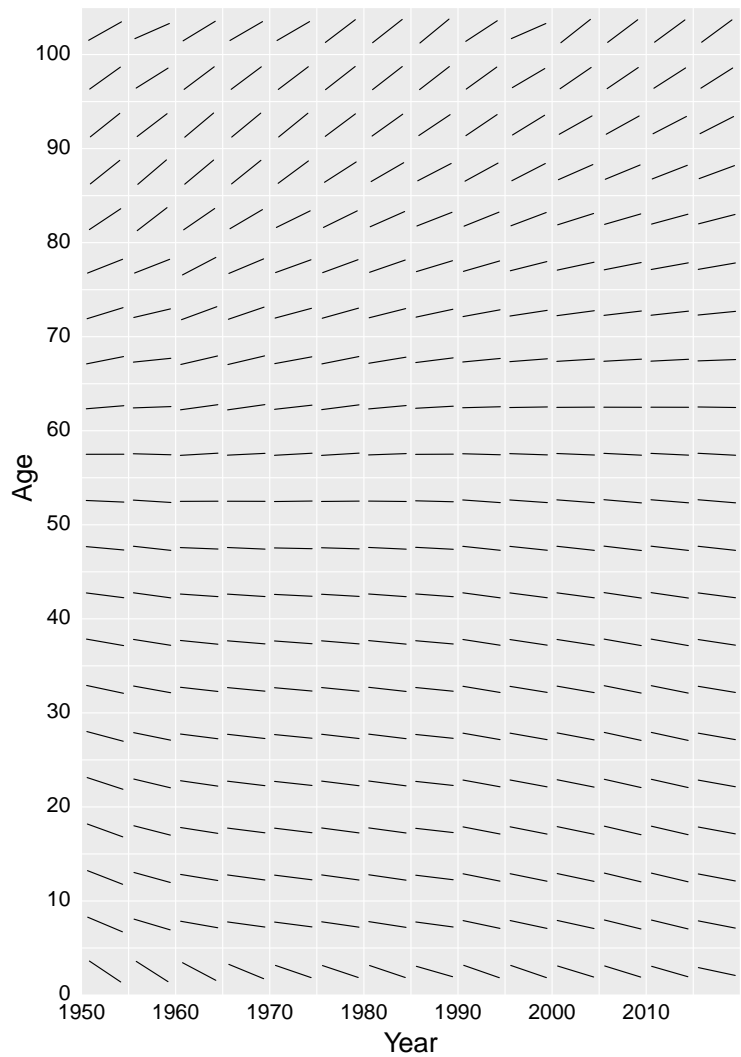
The final example, Fig. 4 is a true Lexis map. The base of the map is a filled contour plot of the mean (over HMD populations) of the lifetable survivorship column  $\ell(x)$ , converted to proportions. This quantity is interpreted as the probability of surviving from birth until at least age  $x$ , which means that contours in the Lexis map are interpreted as survivorship quantiles. The survivorship surface is redundantly drawn with a sequential color palette and labelled contours. The same field from Fig. 4 is layered atop the survivorship quantile surface, achieving a true layered map.

---

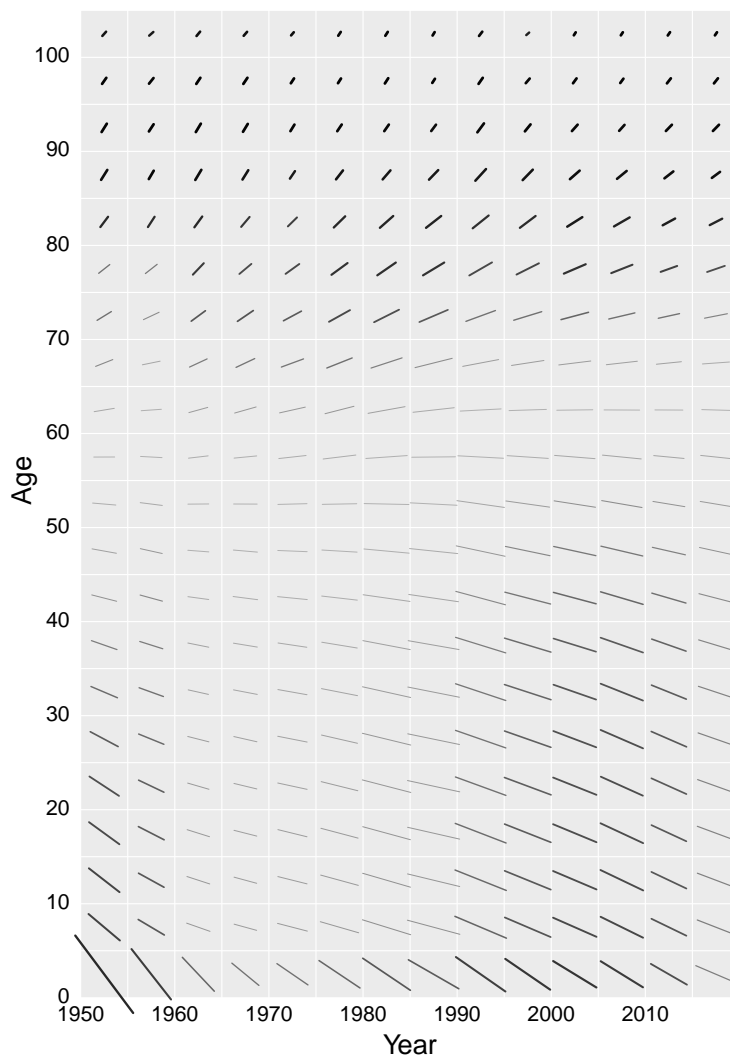
<sup>1</sup>Data points included in regressions are single ages evenly divisible by five for each of the five years included in a Lexis reference cell.

<sup>2</sup>Specifically, pointer length is proportional to the central spread of the relationship between  $e(x)$  and  $sd(x)$ , approximated as  $\sqrt{IQR(e(x))^2 + IQR(sd(x))^2}$

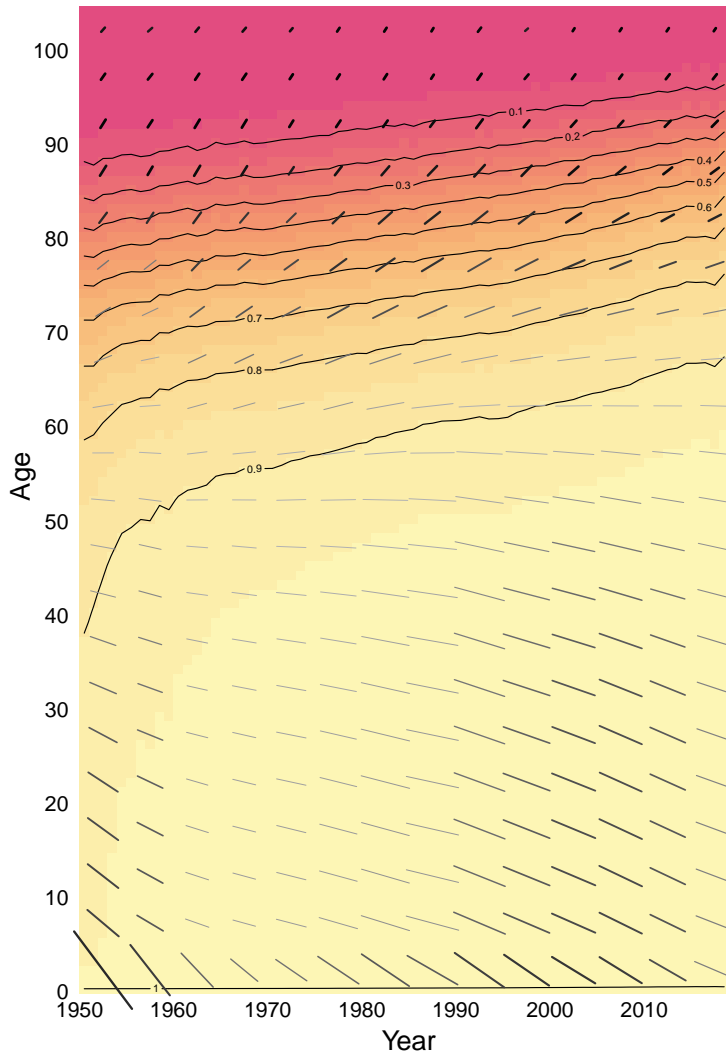
**Figure 2:** Lexis field:  $sd(x)$  by  $e(x)$  linear fits, with slopes drawn equal to regression line slope coefficients. Pointer lengths are all equal.



**Figure 3:** Lexis field:  $sd(x)$  by  $e(x)$  linear fits, with slopes exaggerated by 2. Pointer length is proportional to the diagonal of the IQR box around  $sd(x)$  and  $e(x)$ , while grayscale and segment width are proportional to  $r^2$  (darker and thicker = higher  $r^2$ ).



**Figure 4:** A Lexis map: Lexis surface of mean survivorship quantiles as a filled contour plot plus a Lexis field of  $sd(x)$  by  $e(x)$  linear fits, slopes exaggerated by 2. Pointer length is proportional to the diagonal of the IQR box around  $sd(x)$  and  $e(x)$ , while grayscale and segment width are proportional to  $r^2$ .





## 4. Discussion

We suggest the use of vector fields on the Lexis surface, introducing the notion of a Lexis field, which is a standard vector field on a regular Lexis grid over age and time. This instrument allows researchers to overlay Lexis surfaces and display relationships in complex ways. We demonstrate some of the designer freedom in translating data into the elements of a Lexis field, as well as an instance of Lexis map layering. These examples serve to illustrate the construction of Lexis fields, but do not pretend to be *best practice* Lexis surfaces in terms of visual design or legibility. It is our sense that the patterns revealed in Figures 2-4 are accessible to the viewer and lend themselves to substantive interpretation. The field pointers we use represent linear models. Although patterns in data may be much more complex than can be expressed with a single linear model, in fields each model fit can be thought of as a local *zoom-in* on a complex macro pattern — subtle changes between neighboring pointers reveal interpretable patterns as the eye *zooms out*.

The composite surfaces of Schöley and Willekens (2017) also display layered information as a single visual layer. Small multiples of Lexis surfaces (for example panel plots of Lexis surfaces) on the other hand constitute a de-layering (c.f. Remund, Camarda, and Riffe 2018, Kashnitsky and Aburto (2019)), as these are spatially disjoint. Comparisons between plots require an extra attentive step from the the viewer to cross-reference patterns or values at specific coordinates of age and time. More dimensions (e.g. causes of death) imply more such cross-referencing work from the viewer if displayed in this way. On the other hand, an overlaid Lexis field may imply a lower degree of age and time resolution for each of the variables overlaid — for example we used a  $5 \times 5$  grid in our application. We do not evaluate the trade-off between the potential clarity of individual but disjoint graphs versus comparisons using Lexis fields. Researchers may wish to experiment using the reproducible code we provide.

The idea to draw a Lexis field arose in practice in an attempt to investigate the apparently mechanical relationship between lifespan variation and average length of life with a macro view. Evidence suggests a negative relationship between lifespan variation and life expectancy when measured at young ages such as 0 or 15 (Wilmoth and Horiuchi 1999, Smits and Monden (2009), Vaupel, Zhang, and van Raalte (2011), Alvarez, Aburto, and Canudas-Romo (2019), Colchero et al. (2016)). This means that at the aggregate level, as life expectancy at birth increases in low mortality countries, length of life becomes more equal, and this relationship appears to hold up generally between human populations (Colchero et al. 2016). Lifespan variation has become an important topic in demographic research because larger lifespan variation implies greater uncertainty about the timing of death for individuals and, at the population level, it implies that health improvements are unevenly shared in a population (van Raalte, Sasson, and Martikainen 2018). More recently, several cases have been uncovered where an increase in life expectancy is not

followed by decrease in lifespan variation. For example, in Eastern Europe, in periods of slow improvements in life expectancy at birth, lifespan variation moved in the same direction, running against the usual narrative (Aburto and van Raalte 2018). Similarly, studies have pointed out that the age used for truncation is important to determine the strength and direction of the relationship between life expectancy and lifespan variation (Myers and Manton 1984, Nusselder and Mackenbach (1996), Robine (2001), Engelman, Canudas-Romo, and Agree (2010), Németh (2017)). Our visual approach reveals that the relationship starts as strongly negative, and somewhere near age 55 it reverses to become positive in a systematic way. A crossover was already documented by Myers and Manton (1984), but it has not been previously shown in such a systematic and comprehensive way, possibly because standard ways of looking at these data would have required dozens or hundreds of graphs. We aim to fill this gap by proposing a way to visualize these complex patterns in a single plot.

Fig.4 serves to illustrate that Lexis fields can be layered with traditional Lexis surfaces that are color-coded, increasing the information and pattern density on the Lexis plane with little drawback in terms of legibility. This allows the viewer to interpret field patterns conditional on the underlying surface. For example, the increase in the (mean) 90% survival contour appears to have been faster than the increase in the age at which pointer slopes switch from negative to positive. Few people reach ages in which there is a strong positive relationship between lifespan uncertainty and average remaining lifespan, but the fraction who do so is increasing.

Fields could also be derived from a single underlying pattern rather than a series of regressions on different populations or variables. For example, Shang (2018) recommends the use of phase diagrams to represent the rate of change of the hypothetical life course implied by period fertility curves. This construct could be translated to a Lexis field in a straightforward way, with pointers mapping to the notions of acceleration and velocity. Certainly variants of vector fields could be used to intuitively display other demographic phenomena and components of demographic change, and these do not need to adhere to the Lexis grid. For example, changes in the relationship between male and female fertility rates over age and time could be shown with a field. Other well-known relationships, such as the *Preston curve* (Preston 1975) or *Taylor's Law* (Cohen, Bohk-Ewald, and Rau 2018), might also lend themselves to Lexis field representation. We also suggest wider use of fields based on demographic relationships in spatial settings.

## 5. Conclusions

We describe the construction and use of vector fields on the Lexis plane. We argue that this technique can increase the information density and scope displayed on the Lexis surface. We also argue that fields are easily made compatible with Lexis surfaces drawn

as filled contour plots, which lends them to map layering. We suggest alternative field designs and uses that could be easily applied for other demographic processes and research questions. In sum, displaying a larger variety of demographic quantities on the Lexis plane and increasing the information density on the Lexis plane using layering techniques such as fields should broaden the scope of demographic exploration and sharpen the instruments of demographic pattern detection.

## **6. Reproducibility**

Calculations and visualizations in this manuscript were all done in the R programming language (R Core Team 2016). Code used to produce the experimental visualizations here is available in a public repository:

<https://github.com/timriffe/MacroShape/DR>

## **7. Acknowledgements**

We wish to thank Alyson van Raalte for helpful comments received that improved this manuscript.

## References

- Aburto, J.M. and van Raalte, A. (2018). Lifespan dispersion in times of life expectancy fluctuation: the case of central and eastern europe. *Demography* 55(6): 2071–2096. doi:10.1007/s13524-018-0729-9.
- Acosta, E. and van Raalte, A. (2019). Apc curvature plots: Displaying nonlinear age-period-cohort patterns on lexis plots. *Demographic Research* 41.
- Alvarez, J.A., Aburto, J.M., and Canudas-Romo, V. (2019). Latin american convergence and divergence towards the mortality profiles of developed countries. *Population studies* 1–18. doi:10.1080/00324728.2019.1614651.
- Canudas-Romo, V. and Schoen, R. (2005). Age-specific contributions to changes in the period and cohort life expectancy. *Demographic Research* 13: 63–82. doi:10.4054/DemRes.2005.13.3.
- Cohen, J.E., Bohk-Ewald, C., and Rau, R. (2018). Gompertz, makeham, and siler models explain taylor's law in human mortality data. *Demographic Research* 38: 773–841. doi:10.4054/DemRes.2018.38.29.
- Colchero, F., Rau, R., Jones, O.R., Barthold, J.A., Conde, D.A., Lenart, A., Nemeth, L., Scheuerlein, A., Schoeley, J., Torres, C. et al. (2016). The emergence of longevous populations. *Proceedings of the National Academy of Sciences* 113(48): E7681–E7690. doi:10.1073/pnas.1612191113.
- Engelman, M., Canudas-Romo, V., and Agree, E.M. (2010). The implications of increased survivorship for mortality variation in aging populations. *Population and Development Review* 36(3): 511–539. doi:10.1111/j.1728-4457.2010.00344.x.
- Human Mortality Database (2018). University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 15 Dec., 2018).
- Kashnitsky, I. and Aburto, J.M. (2019). Geofaceting: Aligning small-multiples for regions in a spatially meaningful way. *Demographic Research* 41: 477–490. doi:10.4054/DemRes.2019.41.17.
- Minton, J., Shaw, R., Green, M.A., Vanderbloemen, L., Popham, F., and McCartney, G. (2017). Visualising and quantifying excess deaths in Scotland compared with the rest of the UK and the rest of Western Europe. *J Epidemiol Community Health* jech-2016. doi:10.1136/jech-2016-207379.
- Myers, G.C. and Manton, K.G. (1984). Compression of mortality: myth or reality? *The gerontologist* 24(4): 346–353. doi:10.1093/geront/24.4.346.

- Németh, L. (2017). Life expectancy versus lifespan inequality: A smudge or a clear relationship? *PloS one* 12(9): e0185702. doi:10.1371/journal.pone.0185702.
- Nusselder, W.J. and Mackenbach, J.P. (1996). Rectangularization of the survival curve in the netherlands, 1950-1992. *The Gerontologist* 36(6): 773–782. doi:10.1093/geront/36.6.773.
- Preston, S.H. (1975). The changing relation between mortality and level of economic development. *Population studies* 29(2): 231–248. doi:10.1080/00324728.1975.10410201.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rau, R., Bohk-Ewald, C., Muszyńska, M.M., and Vaupel, J.W. (2017). *Visualizing Mortality Dynamics in the Lexis Diagram*, vol. 44. Springer. doi:10.1007/978-3-319-64820-0.
- Remund, A., Camarda, C.G., and Riffe, T. (2018). A cause-of-death decomposition of young adult excess mortality. *Demography* 55(3): 957–978. doi:10.1007/s13524-018-0680-9.
- Robine, J.M. (2001). Redefining the stages of the epidemiological transition by a study of the dispersion of life spans: The case of france. *Population: An English Selection* 173–193. URL <https://www.jstor.org/stable/3030265>.
- Schöley, J. and Willekens, F. (2017). Visualizing compositional data on the Lexis surface. *Demographic Research* 36: 627–658. doi:10.4054/DemRes.2017.36.21.
- Shang, H.L. (2018). Visualizing rate of change: An application to age-specific fertility rates. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* doi:10.1111/rssa.12359.
- Smits, J. and Monden, C. (2009). Length of life inequality around the globe. *Social science & medicine* 68(6): 1114–1123. doi:10.1016/j.socscimed.2008.12.034.
- van Raalte, A.A., Sasson, I., and Martikainen, P. (2018). The case for monitoring life-span inequality. *Science* 362(6418): 1002–1004. doi:10.1126/science.aau5811.
- Vaupel, J.W., Gambill, B.A., and Yashin, A.I. (1987). Thousands of data at a glance: Shaded contour maps of demographic surfaces URL <http://pure.iiasa.ac.at/2905>.
- Vaupel, J.W., Zhang, Z., and van Raalte, A.A. (2011). Life expectancy and disparity: An international comparison of life table data. *BMJ open* 1(1): e000128. doi:10.1136/bmjopen-2011-000128.

- Weiskopf, D. (2007). *GPU-Based Interactive Visualization Techniques*, Springer, chap. Vector Field Visualization: pp. 81–159. doi:[10.1007/978-3-540-33263-3.pdf](https://doi.org/10.1007/978-3-540-33263-3.pdf).
- Wilmoth, J.R. and Horiuchi, S. (1999). Rectangularization revisited: Variability of age at death within human populations. *Demography* 36(4): 475–495. doi:[10.2307/2648085](https://doi.org/10.2307/2648085).