August 05, 2025

Longevity Contests

Many 'longevity contests' pit performers against their professional peers. But how well do a specific group of people fare in a longevity contest against their *general population* peers?

In 2003, faced with the challenge of locating a natural comparison group for the Titanic survivors, we selected population comparators and exploited readily available vital statistics data from the USA and Sweden. The analyses required considerable care, just to come up with survival curves.[1]

I addressed the construction of these curves, as well as some of the formal statistical inference issues, in an abstract I submitted to, and in a presentation I gave at, the 2006 Conference of Applied Statisticians of Ireland (CASI).

In the two decades since we studied the longevity of the Titanic 'survivors', I have assembled a large and scattered collection of what I call 'longevity contests', and, with others, have been raising awareness of some of the special statistical pitfalls involved in them. Here is a portion of a talk I gave at U. Laval (Québec) in 2007, and another at U. Freiburg (Germany) in 2016.

Of course, I had been keenly aware of some of these pitfalls ever since the early 1970s, when statisticians reminded the heart-transplant surgeons that not all of the post-getting-on-the-waiting-list survival of the patients who were lucky enough to get a transplant could be attributed to receiving it.

Most recently, in 2024, just after I retired, I was invited to re-tell the story of the biggest such 'statistical blooper' that I have seen in my 50 years in the

[1] On the 26th slide in this talk in 2018, I told some of the 'after-math' to the publication and I mentioned some of the behind-the-scenes negotiations in the audio version, starting at about 25m40s. In class I would recount how, when I had not reported the results of a formal test, the Editor replied 'I must insist on a p-value.' When I then included the result of a simple t-one-sample test, the young statistical reviewer asked how could one do a t-test when lifetimes were censored. I pointed out that only 3 of the 435 were, and that I gave the 3 passengers an additional 20 years between them as a best case scenario, since they were already in their 90s. I mainly went by 'my eye' test, and did not want to make too much of a p-value. In the end I was able to avoid the phrase 'statistically significant' and didn't give any p-value. Moreover, I was able to convince the editor to drop BMJ's usual insistence on using a subtitle to classify the study design (the BMJ is still being silly about these 'RCT' vs. 'cohort' v.s. 'case-control' vs. 'cross-sectional' labels) and just put the subtitle 'A *Titanic* study'.

'longevity analysis' trade. The two Copenhagen statisticians who first noticed it in 2013, Theis Lange and Niels Keiding, also noted that it was accompanied by what may well be the smallest p-value ever published, surpassing by far the one published by John Arbuthnot in 1710.

Below, I first provide the abstract and presentation of my (shared only at CASI) 2006 work on effectively-infinite-size comparison groups.

Following that you will find a very long and winding piece entitled 'longevity contests'. It was my attempt, in 2022 as I was contemplating retirement, to pull together and share my large and scattered collection of what I call 'longevity contests' – and to use one (involving the top professional baseball players) to illustrate some design principles and data-analysis techniques that can help keep these contests fair, and statistically transparent.

I had not planned to also re-visit the one involving professional actors and actresses. However, the new analyses published in 2022 did not seem all that transparent to me, especially as I was not able to readily check them out using the dataset that was used. So, I now share a fresh dataset that I made myself: the data at issue are all from public sources – indeed, in both the 2001 and 2022 articles on this topic, some individual subjects were even identified by name. I apply to this new public dataset the same design principles and data-analysis techniques that I had applied to the – also public – data about the baseball players.

I had had several in-person discussions with Niels Keiding at the Freiburg workshop in 2016, and had also admired his writings on the Lexis diagram and its uses. So, as a tribute to him, I included his sadly-no-longer-censored lifeline (1944-2022) as the go-to example to explain (especially to North American readers!) the Lexis diagram I had included in Figure 1.

Here (belatedly in 2025) I thank three colleagues (SH and NG at McGill, and MC at UCC) who, in 2022 went through my draft, and offered detailed suggestions. One was to split it into two, which I did, partitioning the file into two 'manuscripts', I: study design, and II: Data Analysis.

I then emailed a journal editor to (pre-) enquire whether my (still) very long and winding piece might suit their journal.

I receives detailed responses and suggestions from the editor and their advisors. I (also belatedly) thank them for the considerable time they spent on it, and I am sorry that I did not find the time/energy to repay them by following their advice.

But I do hope that some younger/energetic people will consider doing so, and I would be delighted to share *all* of my material with them, and have them turn it into a proper piece. In any case, if people wish to, they are free to use it in teaching.

Sincerely,

James Hanley
webpage: https://jhanley.biostat.mcgill.ca | email: james.hanley@mcgill.ca

A finely stratified log-rank test with effectively-infinite-size comparison groups

James A Hanley[1]

[1] Dept. of Epidemiology, Biostatistics & Occupational Health, McGill University, Montreal, Quebec, H3A 1A2, Canada. E-mail: James.Hanley@McGill.CA

Abstract

This presentation will elaborate on a previously published longevity analysis of ours. It will discuss some technical statistical issues related to how we created synthetic comparison groups, and how we avoided the errors made by other authors, e.g., those who have studied the longevity of jazz musicians. The differences between 'current lifetables' – commonly used, and frequently misunderstood by the public – and cohort lifetables will be illustrated. The synthetic nature of the comparison groups we created is used to illustrate the limiting behaviour of the stratified log rank test when each stratum consists of one index person and an 'effectively infinite' number of comparison persons.

Keywords: Survival Analysis, Lexis Diagram, Current and Cohort Life-Tables, Log-rank Test.

1 Introduction

Failure to recognize what is now termed "immortal time bias" has, over the years, lead to a large number of invalid survival comparisons in the medical and epidemiologic literature (Hanley et al. 2006). In teaching how to perform more valid analyses, we have used as an example the longevity of the passengers who survived the sinking of the Titanic (Hanley et al. 2003). We compared the post-1912 longevity of each passenger with the remaining life-course of an age- and sex-matched group of peers alive in 1912, using each (passenger, peer-group) as a separate 'stratum.' Since each comparison lifetable was reconstructed from national vital statistics data, the comparison group for each passenger was effectively infinite in size.

2 Methods

Unable to find a comparison group with the same mix of backgrounds and selection factors, we created two 'next best' comparison groups from available national data. We calculated what proportions of an age and sex matched group of white Americans alive in 1912 would be alive at each anniversary of the sinking. To do so, we converted current (cross sectional) life tables for the years 1912-20002 into cohort life tables. The Lexis Diagram helped guide the calculations. We created a second comparison group from life table data for Sweden, which were already in cohort form. Longevity differences were assessed by several methods, including a stratified log rank test.

3 Results

Substantively, the survival of the 435 passengers who had been traced was slightly, but not significantly, longer than that of the two comparison groups. Despite their social advantage, first class passengers did not fare particularly well.

Methodologically, if we denote by t the age at death of a passenger, and by $S[t]$ the corresponding proportion in the comparison population still alive at that age, then the stratified Log-Rank statistic (with $1df$) has the simple form $\{\sum(1+\log S[t])\}^2/\{-\sum \log S[t]\}$ where $\sum$ denotes summation over the n pasengers. Alternatively, if we combine the passenger-specific P-values, we obtain the Chi-squared ($2n$ df) statistic $-2\sum \log S[t]$.

4 Discussion

Given our inability to find a comparison group with the same mix of backgrounds and selection factors, the inaccuracies in the data, and the fact that some 17% have not been traced, the substantive results should not be over-interpreted. However, the special nature of this particular stratified log-rank test, when each stratum consists of one person in the index category, and an infinite number in the reference category, does lead to some insights into the structure of the limiting case of the log-rank test, and into the name of the test itself.

References

Hanley JA, Carrieri MP and Serraino D (2006). *International Journal of Epidemiology* Advance Access published March 16, 2006

Hanley JA, et al. (2003). *British Medical Journal* 327 (7429):1457-8.

A finely stratified log-rank test with effectively-infinite-size comparison groups

[How long did their hearts go on? Survival analysis of the Titanic Survivors]

Background

Erroneous analyses in longevity comparisons [Jazz Musicians, Oscar winners]
Beyond "who survived": longterm effects

Data

Passengers ; Comparison Groups

Methods

Passengers: K-M curves
Comparison Groups: "Cohort from Current" (U.S.) & Cohort(Sweden) Lifetables

Results

Overall; By Gender and Class

Methodological

Stratified log-rank test: each passenger versus effectively infinite comparison group

Peer-review and beyond

BMJ ; Media

James.Hanley@McGill.CA http://www.epi.mcgill.ca/hanley CASI, May 17-19, 2006

Natural Sciences & Engineering
Research Council of Canada

Québec

Fonds Québécois de la recherche
sur la nature et les technologies.

Premature Death in Jazz Musicians: Fact or Fiction?

commonly held view: More liable than other professions to die early from drink, drugs, women, or overwork.

Statistical Study: 70 (82%) of 85 US-born jazz musicians listed in university syllabus exceeded their life expectancy

Spencer FJ. Am J Public Health. 1991 81(6):804-5: Am J Public Health. 1992 82(5):761.

Longevity of popes and artists between 13th & 19th century

Likely, in past centuries, to be better fed, clothed & sheltered, and to had better medical care & to survive longer than most of their contemporary people.

Longevity significantly longer than that of artists (P = 0.02); ... artists had 1.5-fold higher risk of death before age 70 years than Popes (95% CI: 1.08–2.16)

Serraino D, Carrieri M-P: International Journal of Epidemiology 2005; 34: 1435–1436

Survival in Academy Award–Winning Actors and Actresses

Social status is an important predictor of poor health. Most studies of this issue have focused on lower echelons of society

Life expectancy 3.9 years longer for Academy Award winners than for other, less recognized performers (79.7 vs. 75.8 years; P = 0.003).

Donald A. Redelmeier, MD, and Sheldon M. Singh, BSc *Ann Intern Med.* 2001;134:955-962.

Male / Female

Adult / Child

Socio-Economic Class
[1 /2 /3 / unclassified]

Survived?

titanic.dat
titanic.txt

JSE ARCHIVE
http://www.amstat.org/publications/jse/

NAME: Population at Risk and Death Rates for an Unusual Episode
TYPE: Complete record for all of population at risk
SIZE: 2201 observations, 4 variables

The article associated with this dataset appears in the *Journal of Statistics Education*, Volume 3, Number 3 (November 1995).

SUBMITTED BY:
Robert Dawson
Dept.of Mathematics and Computing Science
Saint Mary's University
Halifax, Nova Scotia
Canada B3H 3C3
email: rdawson@husky1.stmarys.ca

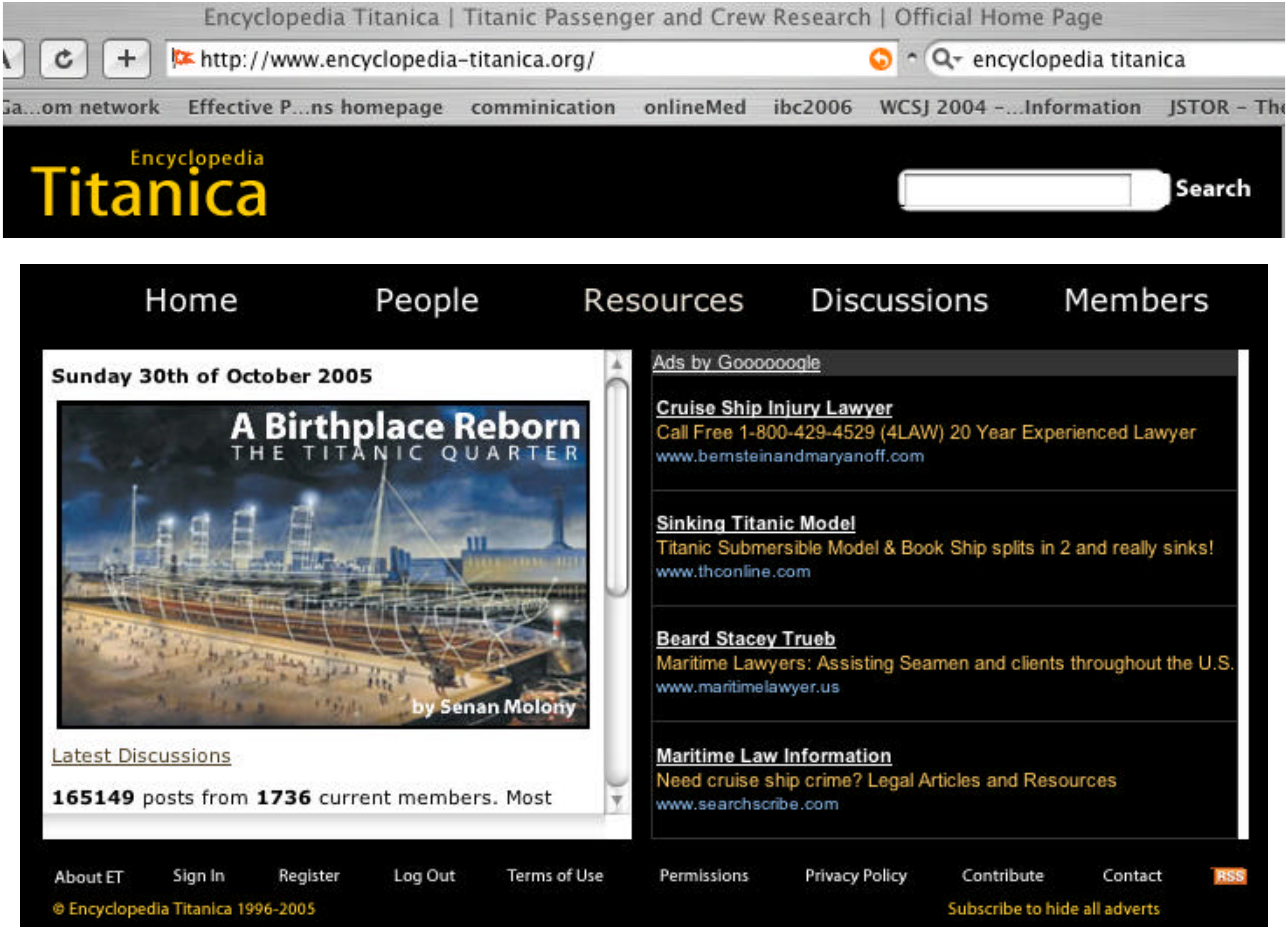
How long did their hearts go on? A *Titanic* study

James A Hanley, Elizabeth Turner, Carine Bellera, Dana Teltsch

Several studies have examined post-traumatic stress in people who survive disasters but few have looked at longevity. The 1997 film *Titanic* followed one character, apparently fictional, but the longevity of the actual survivors, as a group, has not been studied. Did the survivors of the sinking of the *Titanic* have shortened life spans? Or did they outlive those for whom 14-15 April 1912 was a less personal night to remember?

Subjects, methods, and results

We limited our study to passengers. We used data from biographies listed in Encyclopedia Titanica, a website that claims to have "among the most accurate passenger and crew lists ever compiled."[1] Of the 500 passengers listed as survivors, 435 have been traced. We calculated the proportion alive at each anniversary of the sinking.


Encyclopedia Titanica | Titanic Passenger and Crew Research | Official Home Page
http://www.encyclopedia-titanica.org/
encyclopedia titanica
Ga...om network
Effective P...ns homepage
comminication
onlineMed
ibc2006
WCSJ 2004 -...Information
JSTOR - The
Encyclopedia Titanica
Search
Home
People
Resources
Discussions
Members
Sunday 30th of October 2005
A Birthplace Reborn
THE TITANIC QUARTER
by Senan Molony
Latest Discussions
165149 posts from 1736 current members. Most
Ads by Goooooogle
Cruise Ship Injury Lawyer
Call Free 1-800-429-4529 (4LAW) 20 Year Experienced Lawyer
www.bernsteinandmaryanoff.com
Sinking Titanic Model
Titanic Submersible Model & Book Ship splits in 2 and really sinks!
www.thconline.com
Beard Stacey Trueb
Maritime Lawyers: Assisting Seamen and clients throughout the U.S.
www.maritimelawyer.us
Maritime Law Information
Need cruise ship crime? Legal Articles and Resources
www.searchscribe.com
About ET
Sign In
Register
Log Out
Terms of Use
Permissions
Privacy Policy
Contribute
Contact
RSS
© Encyclopedia Titanica 1996-2005
Subscribe to hide all adverts

First Class Passengers

We found 346 people . Showing 1 to 346

Name v	Age	Class/Dept	Ticket	Fare	Group	Ship	Joined	Job	Boa	
ALLEN, Miss Elisabeth Walton	29	1st Class	24160	£211 6s 9d			Southampton		2	
ALLISON, Mr Hudson Joshua Creighton	30	1st Class	113781	£151 16s			Southampton	Businessman		
ALLISON, Mrs Bessie Waldo	25	1st Class	113781	£151 16s			Southampton			
ALLISON, Miss Helen Loraine	2	1st Class	113781	£151 16s			Southampton			
ALLISON, Master Hudson Trevor	11m	1st Class	113781	£151 16s			Southampton		11	
ANDERSON, Mr Harry	47	1st Class	19952	£26 11s			Southampton	Stockbroker	3	
ANDREWS, Miss Kornelia Theodosia	62	1st Class	13502	£77 19s 2d			Cherbourg		10	
ANDREWS, Mr Thomas	39	1st Class	112050		H&W Guarantee Group		Belfast	Shipbuilder		
APPLETON, Mrs Charlotte	53	1st Class	11769	£51 9s 7d			Southampton		D	
ARTAGAVEYTIA, Mr Ramon	71	1st Class	17609	£49 10s 1d			Cherbourg	Businessman		22
ASTOR, Colonel John Jacob	47	1st Class	17757	£247 10s 6d			Cherbourg	Property Developer / Real Estate		124
ASTOR, Mrs Madeleine Talmage	18	1st Class	17757	£247 10s 6d			Cherbourg		4	

Miss Elisabeth Walton Allen

Elisabeth Allen

SUMMARY

BORN: SUNDAY 1ST OCTOBER 1882 IN ST. LOUIS MISSOURI UNITED STATES
AGE: 29 YEARS 6 MONTHS AND 14 DAYS.
MARITAL STATUS: SINGLE.
LAST RESIDENCE: IN ST. LOUIS MISSOURI UNITED STATES
1ST CLASS PASSENGER
FIRST EMBARKED: SOUTHAMPTON ON WEDNESDAY 10TH APRIL 1912
TICKET NO. 24160 , £211 6S 9D
CABIN NO. B5
RESCUED (BOAT 2)
DISEMBARKED CARPATHIA: NEW YORK CITY ON THURSDAY 18TH APRIL 1912
DIED: FRIDAY 15TH DECEMBER 1967
CAUSE OF DEATH: HEART FAILURE / DISEASE

Miss Elisabeth Walton Allen, 29, was born in St. Louis, Missouri, USA, on 1 October 1882, the daughter of George W. Allen, a St. Louis judge, and Lydia McMillan. She was returning to her home in St. Louis with her aunt, Mrs Edward Scott Robert , and her cousin, fifteen-year-old Georgette Alexandra Madill . Miss Madill was the daughter of Mrs Robert from a former marriage.

Miss Allen was engaged in 1912 to a British physician, Dr. James B. Mennell, and was going home to St. Louis to collect her belongings in preparation for moving to England where she would live with her future husband. Miss Allen, Mrs Robert , Miss Madill , and Mrs Robert's maid Emilie Kreuchen all boarded the *Titanic* in Southampton. For the voyage, Miss Allen was in cabin B-5 , along with cousin Miss Madill , while Mrs Robert was across the hall in cabin B-3 . The entire party travelled under ticket number 24160 (£221 16s 9d). She escaped with her relatives in lifeboat 2 , one of the last boats to leave the *Titanic* , under the command of Fourth Officer Joseph G. Boxhall . After the sinking, Elisabeth filed a $2, 427.80 claim against the White Star Line for the loss of personal property in the disaster.

Master Hudson Trevor Allison

Grave of Hudson Trevor Allison
Courtesy of Jason D. Tiller
more pictures

SUMMARY
BORN: SUNDAY 7TH MAY 1911
AGE: 11 MONTHS AND 8 DAYS.
LAST RESIDENCE: IN MONTREAL QUÉBÉC CANADA
1ST CLASS PASSENGER
FIRST EMBARKED: SOUTHAMPTON ON WEDNESDAY 10TH APRIL 1912
TICKET NO. 113781 , £151 16S
CABIN NO. C22/26
RESCUED (BOAT 11)
DISEMBARKED CARPATHIA: NEW YORK CITY ON THURSDAY 18TH APRIL 1912
DIED: WEDNESDAY 7TH AUGUST 1929
CAUSE OF DEATH: PTOMAINE POISONING
BURIED: MAPLE RIDGE CEMETERY CHESTERVILLE ONTARIO CANADA

Master Hudson Trevor Allison, 11m, was born May 7, 1911 in Westmount, Quebec.

Shortly after Trevor was born, the Allison family travelled to England for business purposes, and it was in England that young Trevor was baptised.

He travelled on the *Titanic* with his father Hudson Allison his mother Bess Allison and sister Loraine . He was also accompanied by a nurse Alice Cleaver .

Of the Allison family, only baby Trevor was saved.

After the sinking, baby Trevor returned home to Canada, where he would be raised by his aunt and uncle, George and Lillian Allison.

Trevor died on 7 August 1929 at the age of 18 in Maine, USA of ptomaine poisoning and was buried beside his father in Chesterville, Ontario.

Third Class Passengers

We found 708 people . Showing 1 to 708

Name v	Age	Class/Dept	Ticket	Fare	Group	Ship	Joined	Job	Bo	
ABBING, Mr Anthony	42	3rd Class	5547	£7 11s			Southampton	Blacksmith		
ABBOTT, Mrs Rhoda Mary 'Rosa'	39	3rd Class	CA2673	£20 5s			Southampton		A	
ABBOTT, Mr Rossmore Edward	16	3rd Class	CA2673	£20 5s			Southampton	Jeweller		
ABBOTT, Mr Eugene Joseph	14	3rd Class	CA2673	£20 5s			Southampton	Scholar		
ABELSETH, Miss Karen Marie	16	3rd Class	348125	£7 13s			Southampton		16	
ABELSETH, Mr Olaus Jørgensen	25	3rd Class	348122	£7 13s			Southampton	Farmer	A	
ABRAHAMSSON, Mr Abraham August Johannes	20	3rd Class	3101284	£7 18s 6d			Southampton		15	
ABRAHIM, Mrs Mary Sophie Halaut	18	3rd Class	2657	£7 4s 7d			Cherbourg		C	
ADAMS, Mr John	26	3rd Class	341826	£8 1s			Southampton			103
AHLIN, Mrs Johanna	40	3rd Class	7546	£9 9s			Southampton			

Mr Abraham August Johannes Abrahamsson

Mr August Abrahamson, 20, a single man from Dalsbruk (Taalintehdas), Kimito Island, in southwest Finland boarded the *Titanic* at Southampton. He was travelling to Hoboken, New Jersey. He travelled with Eino Lindqvist and Helga Hirvonen. He shared a cabin with 5 other Finns.

At the time of the collision August was asleep, at first he had no intention to go up and investigate the cause, however, he changed his mind and went to the adjacent cabin to warn Eino Lindqvist, when he began to suspect something was wrong.

He went up to the Boat Deck and entered, most likely, lifeboat 15 he later reported hearing stifled explosions as the ship went down.

After his arrival in New York August was quartered at St. Vincent hospital in New York. He went back to Finland but, in 1914, got married and returned to America where he died in 1961.

References

Claes-Göran Wetterholm (1988, 1996, 1999) *Titanic*. Prisma, Stockholm. ISBN 91 518 3644 0

Acknowledgements

Claes-Göran Wetterholm, Sweden

Contributors

Leif Snellman, Finland

SUMMARY

AGE: 20 YEARS
LAST RESIDENCE: IN DAISBRUK FINLAND
3RD CLASS PASSENGER
FIRST EMBARKED: SOUTHAMPTON ON WEDNESDAY 10TH APRIL 1912
TICKET NO. 3101284 , £7 18S 6D
DESTINATION: HOBOKEN NEW JERSEY UNITED STATES
RESCUED (BOAT 15)
DISEMBARKED CARPATHIA: NEW YORK CITY ON THURSDAY 18TH APRIL 1912
DIED: 1961

NOSWORTHY, Mr Richard Cater	21	3rd Class	39886	£7 16s			Southampton	Farm Labourer		
NYSTEN, Miss Anna Sofia	22	3rd Class	347081	£7 15s			Southampton		13	
NYSVEEN, Mr Johan Hansen	60	3rd Class	345364	£6 4s 9d			Southampton	Farmer		
O'BRIEN, Mr Timothy	21	3rd Class	330979	£7 16s 7d			Queenstown			
O'BRIEN, Mr Thomas	27	3rd Class	370365	£15 10s			Queenstown	Farm Labourer		
O'BRIEN, Mrs Johanna "Hannah"	26	3rd Class	370365	£15 10s			Queenstown	Housewife		
O'CONNELL, Mr Patrick Denis	17	3rd Class	334912	£7 14s 8d			Queenstown	General Labourer		
O'CONNOR, Mr Maurice	16	3rd Class	371060	£7 15s			Queenstown	General Labourer		
O'CONNOR, Mr Patrick	23	3rd Class	366713	£7 15s			Queenstown	Farmer		
O'DRISCOLL, Miss Bridget	27	3rd Class	14311	£7 15s			Queenstown		D	
O'DWYER, Miss Ellen "Nellie"	25	3rd Class	330959	£7 17s 7d			Queenstown			
O'KEEFE, Mr Patrick	21	3rd Class	368402	£7 15s			Queenstown	Farm Labourer	B	
O'LEARY, Miss				£7						

Miss Hanora "Nora" O'Leary

Miss Hanora (Nora) O'Leary, 16, was born in Glencollins, Kingwilliamstown, Co. Cork on June 10, 1895. She was the daughter of John O'Leary and Johanna Healy and had five brothers and two sisters. She was going to her sister Ms. Katie O'Leary at 137 W. 11th Street, New York City.

She boarded the *Titanic* at Queenstown (ticket number 330919, £7 16s 7d). She was travelling in a group from the Kingwilliamstown area led by Daniel Buckley, and consisting of Hannah Riordan, Bridget Bradley, Patrick Denis O'Connell, Patrick O'Connor, and Michael Linehan.

Nora was rescued, probably in lifeboat 13.

Nora became a domestic in New York City. Upon returning to Ireland for a visit a few years later, she married Thomas J. (Tim) Herlihy and then remained in Ireland where she raised her son and four daughters. She spent the remainder of her life in Ballydesmond where she died on 18 May 1975. She is buried in the parish churchyard just a few feet from fellow survivor, Daniel Buckley.

Sources

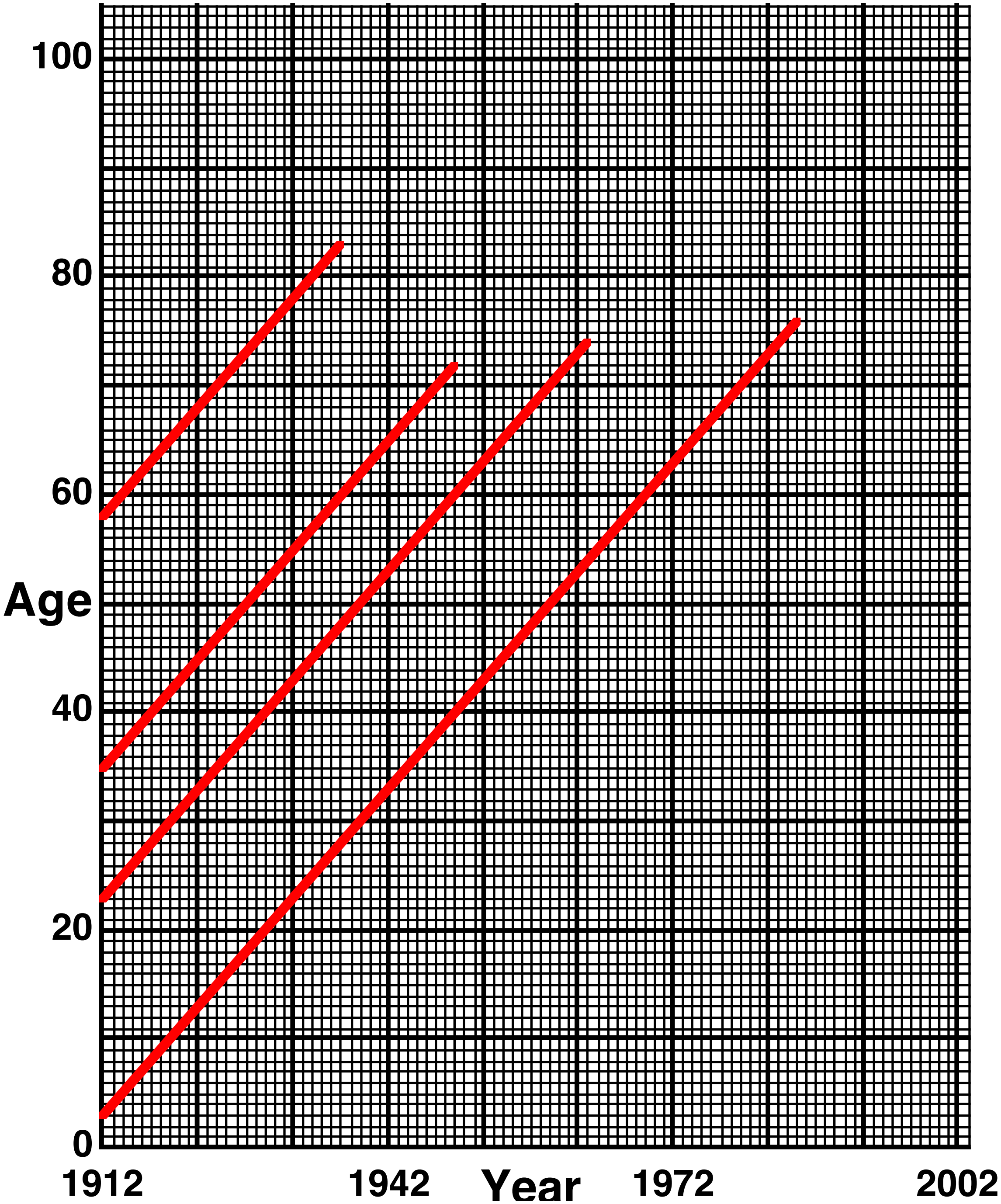
Contract Ticket List, White Star Line 1912 (National Archives, New York; NRAN-21-SDNYCIVCAS-55[279]).

Noel Ray (1999) *List of Passengers who Boarded RMS Titanic at Queenstown, April 11, 1912*. The Irish Titanic Historical Society

Contributors

Cameron Bell, Northern Ireland
Robert L. Bracken, USA
Michael A. Findlay, USA
Noel Ray, Ireland

The largest groups travelling in first and second class were North American or British; most of those in third class were emigrating from Europe to the United States. Unable to find a comparison group with the same mix of backgrounds and selection factors, we created two "next best" comparison groups from available data. We calculated what proportions of an age and sex matched group of white Americans alive in 1912 would be alive at each anniversary. To do so, we converted current (cross sectional) life tables for the years 1912-2000[2] into cohort life tables. We created a second comparison group from life table data for Sweden, which was already in cohort form.[3] Longevity differences were assessed by log rank tests.



Lexis Diagram

German statistician & actuary Wilhelm Lexis (1837–1914)

National Vital Statistics Reports NVSS
CDC SAFER·HEALTHIER·PEOPLE™
Volume 51, Number 3 December 19, 2002

United States Life Tables, 2000

by Elizabeth Arias, Ph.D., Division of Vital Statistics

Introduction

There are two types of life tables—the cohort (or generation) life table and the period (or current) life table. The cohort life table presents the mortality experience of a particular birth cohort, all persons born in the year 1900, for example, from the moment of birth through consecutive ages in successive calendar years. Based on age-specific death rates observed through consecutive calendar years, the cohort life table reflects the mortality experience of an actual cohort from birth until no lives remain in the group. To prepare just a single complete cohort life table requires data over many years.

Unlike the cohort life table, the period life table does not represent the mortality experience of an actual birth cohort. Rather, the period life table presents what would happen to a hypothetical (or synthetic) cohort if it experienced throughout its entire life the mortality conditions of a particular period in time. Thus, for example, a period life table for 2000 assumes a hypothetical cohort subject throughout its lifetime to the age-specific death rates prevailing for the actual population in 2000. The period life table may thus be characterized as rendering a "snap-shot" of current mortality experience, and shows the long-range implications of a set of age-specific death rates that prevailed in a given year. In this report the term "life table" refers only to the period life table and not to the cohort life table.

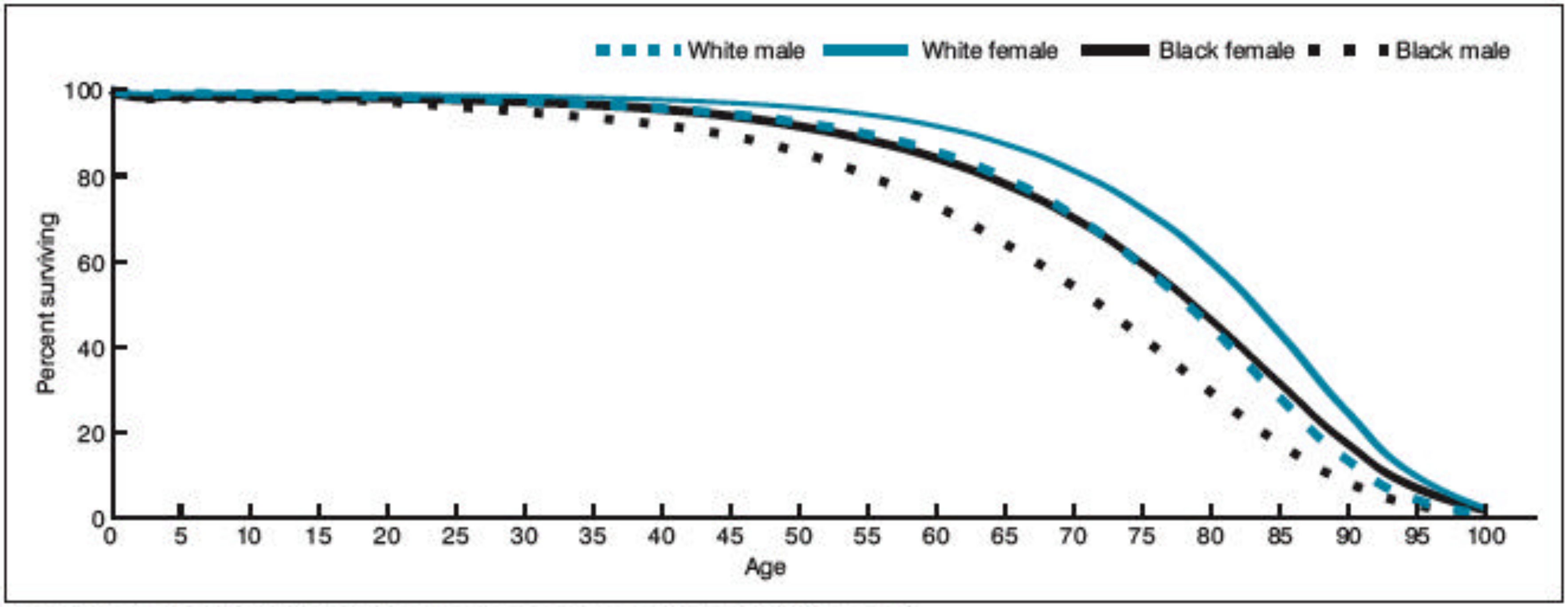


Figure 2. Percent surviving by age, race, and sex: United States, 2000

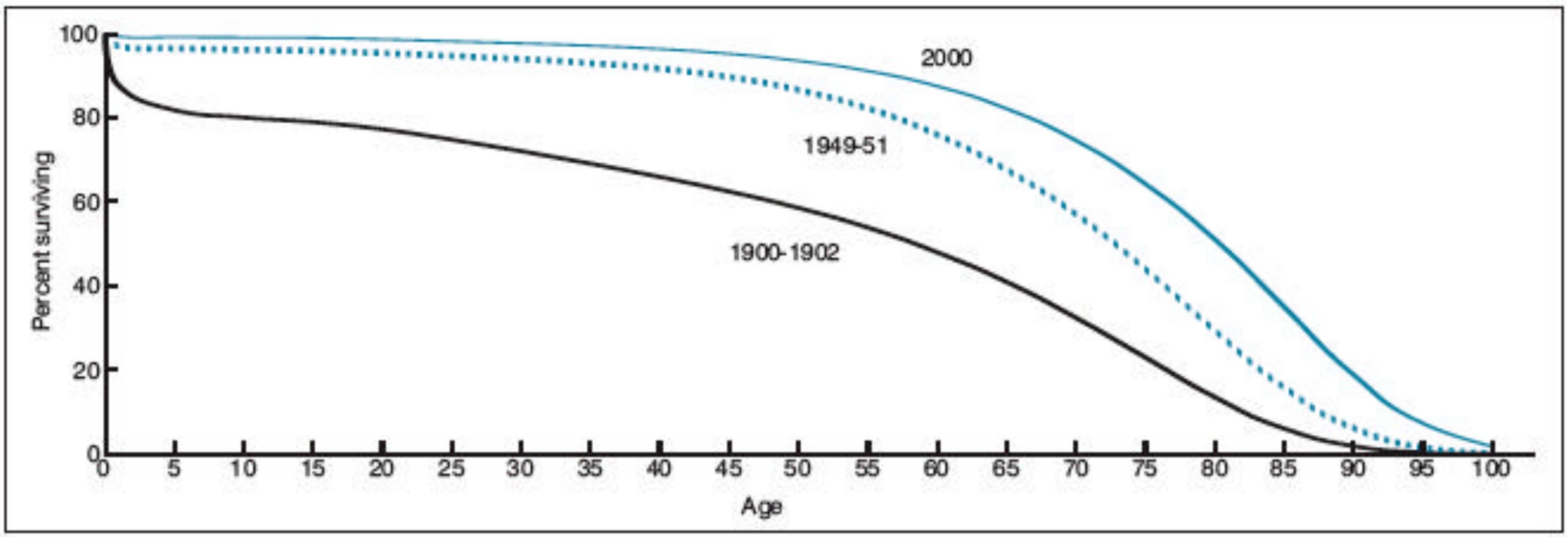


Figure 3. Percent surviving by age: Death-registration States, 1900–1902, and United States, 1949–51 and 2000

Table 6. Life table for white females: United States, 2000

Age	Probability of dying between ages x to x+1	Number surviving to age x
	q_x	l_x
0–1	0.005127	100,000
1–2	0.000414	99,487
2–3	0.000268	99,446
3–4	0.000178	99,419
4–5	0.000154	99,402
5–6	0.000148	99,386
6–7	0.000140	99,372
7–8	0.000134	99,358
8–9	0.000126	99,344
9–10	0.000117	99,332
10–11	0.000109	99,320
11–12	0.000112	99,309
12–13	0.000134	99,298

Table 10. Survivorship by age, race, and sex: Death-registration States, 1900–1902 to 1919–21, and United States, 1929–31 to 2000—Con.

[Alaska and Hawaii included beginning in 1959. For decennial periods prior to 1929–31, data are for groups of registration States as follows: 1900–1902 and 1909–11, 10 States and the District of Columbia; 1919–21, 34 States and the District of Columbia. Beginning 1970 excludes deaths of nonresidents of the United States; see Technical Notes]

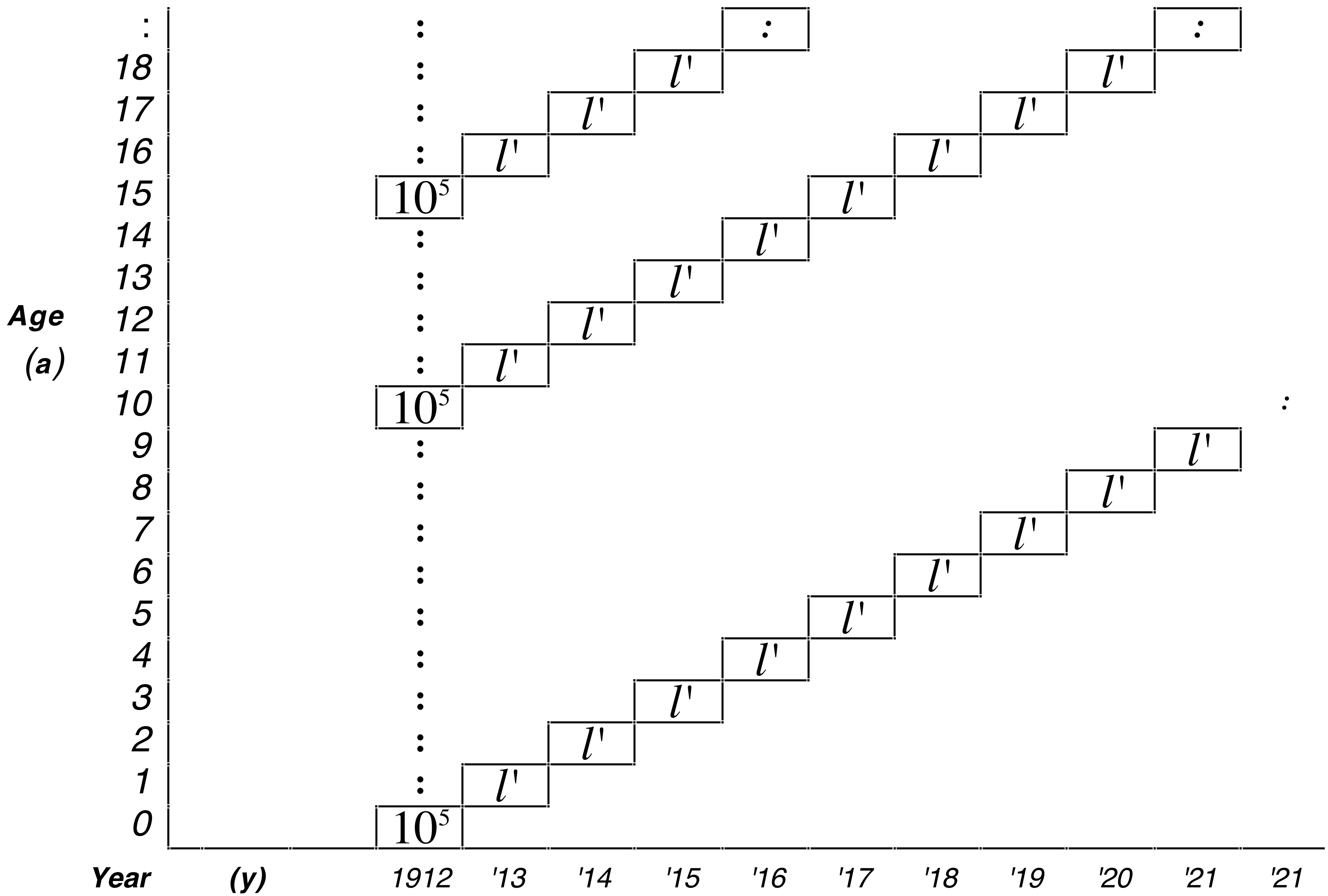
Age, race, and sex	Number of survivors out of 100,000 born alive (l_x)										
	2000	1989–91	1979–81	1969–71	1959–61	1949–51	1939–41	1929–31	1919–21	1909–11	1900–1902
White female											
0	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000
1	99,487	99,333	99,035	98,468	98,036	97,645	96,211	95,037	93,608	89,774	88,939
5	99,386	99,187	98,841	98,203	97,709	97,199	95,309	93,216	90,721	85,349	83,426
10	99,320	99,099	98,725	98,042	97,525	96,960	94,890	92,466	89,564	83,979	81,723
15	99,243	99,007	98,618	97,902	97,375	96,756	94,534	91,894	88,712	83,093	80,680
20	99,046	98,795	98,374	97,618	97,135	96,454	93,984	90,939	87,281	81,750	78,978
25	98,831	98,547	98,093	97,299	96,844	96,072	93,228	89,524	85,163	79,865	76,588
30	98,586	98,283	97,802	96,945	96,499	95,605	92,320	87,972	82,740	77,676	73,887
35	98,268	97,939	97,445	96,474	96,026	94,977	91,211	86,248	80,206	75,200	70,971
40	97,777	97,472	96,913	95,762	95,326	94,080	89,805	84,256	77,624	72,425	67,935
45	97,044	96,768	96,065	94,649	94,228	92,725	87,920	81,780	74,871	69,341	64,677
50	95,970	95,608	94,710	92,924	92,522	90,685	85,267	78,572	71,547	65,629	61,005
55	94,283	93,730	92,594	90,383	89,967	87,699	81,520	74,321	67,323	61,053	56,509
60	91,590	90,789	89,451	86,726	86,339	83,279	76,200	68,462	61,704	54,900	50,752
65	87,385	86,339	84,764	81,579	80,739	76,773	68,701	60,499	54,299	47,086	43,806
70	81,163	79,984	78,139	74,101	72,507	67,545	58,363	49,932	44,638	37,482	35,206
75	72,254	70,834	68,712	63,290	60,461	54,397	44,685	37,024	32,777	26,569	25,362
80	59,792	58,454	55,770	48,182	44,676	38,026	28,882	23,053	20,492	15,929	15,349
85	43,112	42,274	38,774	30,490	26,046	21,348	14,487	10,937	9,909	7,152	7,149
90	24,439	24,270	20,996	14,406	10,219	8,662	5,061	3,719	3,372	2,291	2,322
95	9,638	9,495	7,900	4,526	2,203	2,200	1,109	797	721	434	448
100	2,244	2,239	1,858	872	265	294	139	74	63	44	41

Interpolation ***l*** **for ages 2,3,4, 6,7,8,9, ... in 1910, 1920, ...**

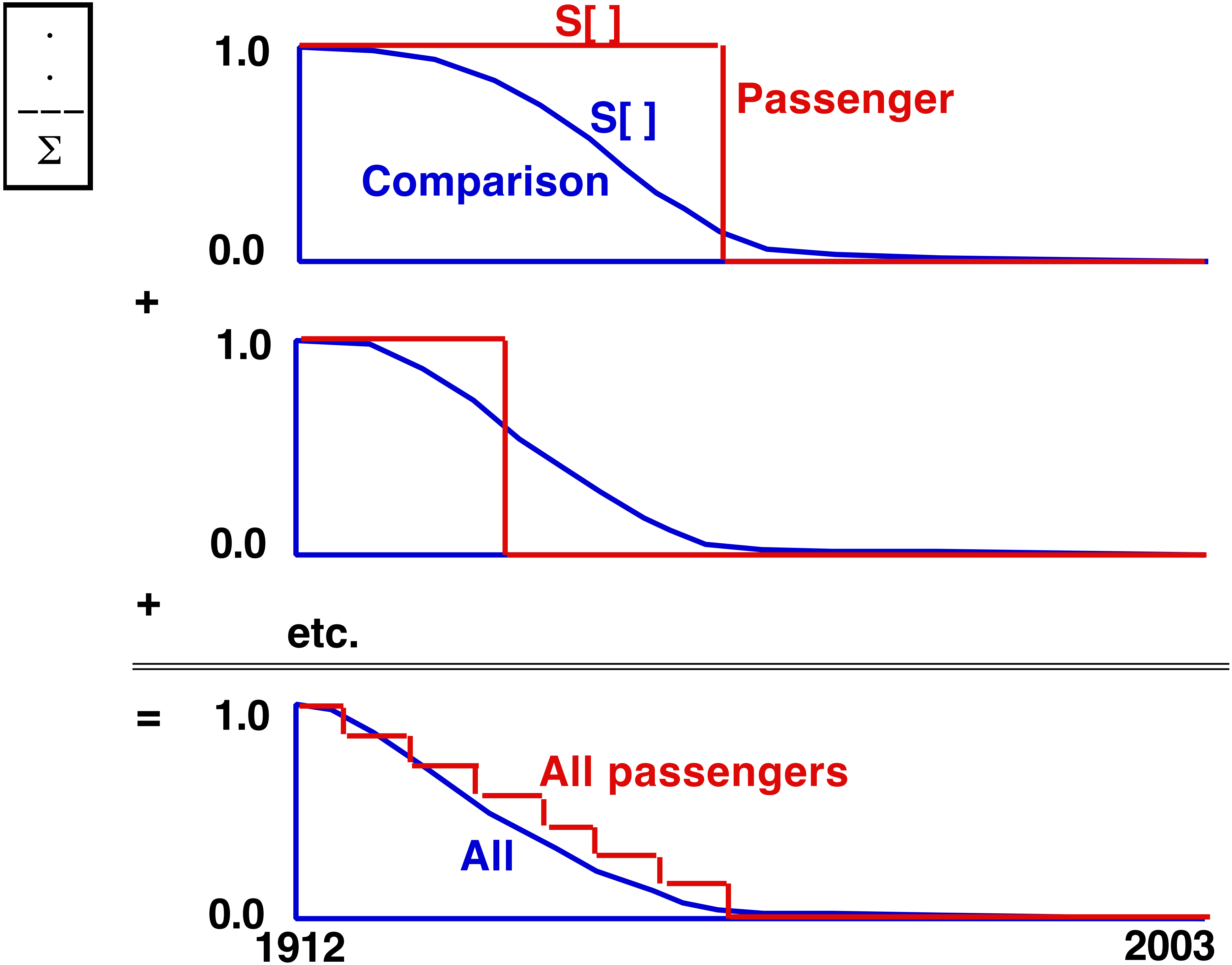
l for entire set of ages for years 1911-1919, 1921-1929, ...

	:	:	:	:	:	:	:	:	:	:	:	:	:
	18	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	...
	17	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	...
	16	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	...
	15	***l***	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	***l***	*l*	...
	14	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	...
	13	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	...
Age	12	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	...
	11	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	...
	10	***l***	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	***l***	*l*	...
	9	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	...
	8	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	...
	7	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	...
	6	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	...
	5	***l***	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	***l***	*l*	...
	4	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	...
	3	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	...
	2	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	...
	1	***l***	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	*l*	***l***	*l*	...
	0	$\mathbf{10^5}$	10^5	10^5	10^5	10^5	10^5	10^5	10^5	10^5	10^5	$\mathbf{10^5}$	10^5	10^5
Year		*1910*	*1911*	*1912*	*'13*	*'14*	*'15*	*'16*	*'17*	*'18*	*'19*	*'20*	*'21*	*'21*
		⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑

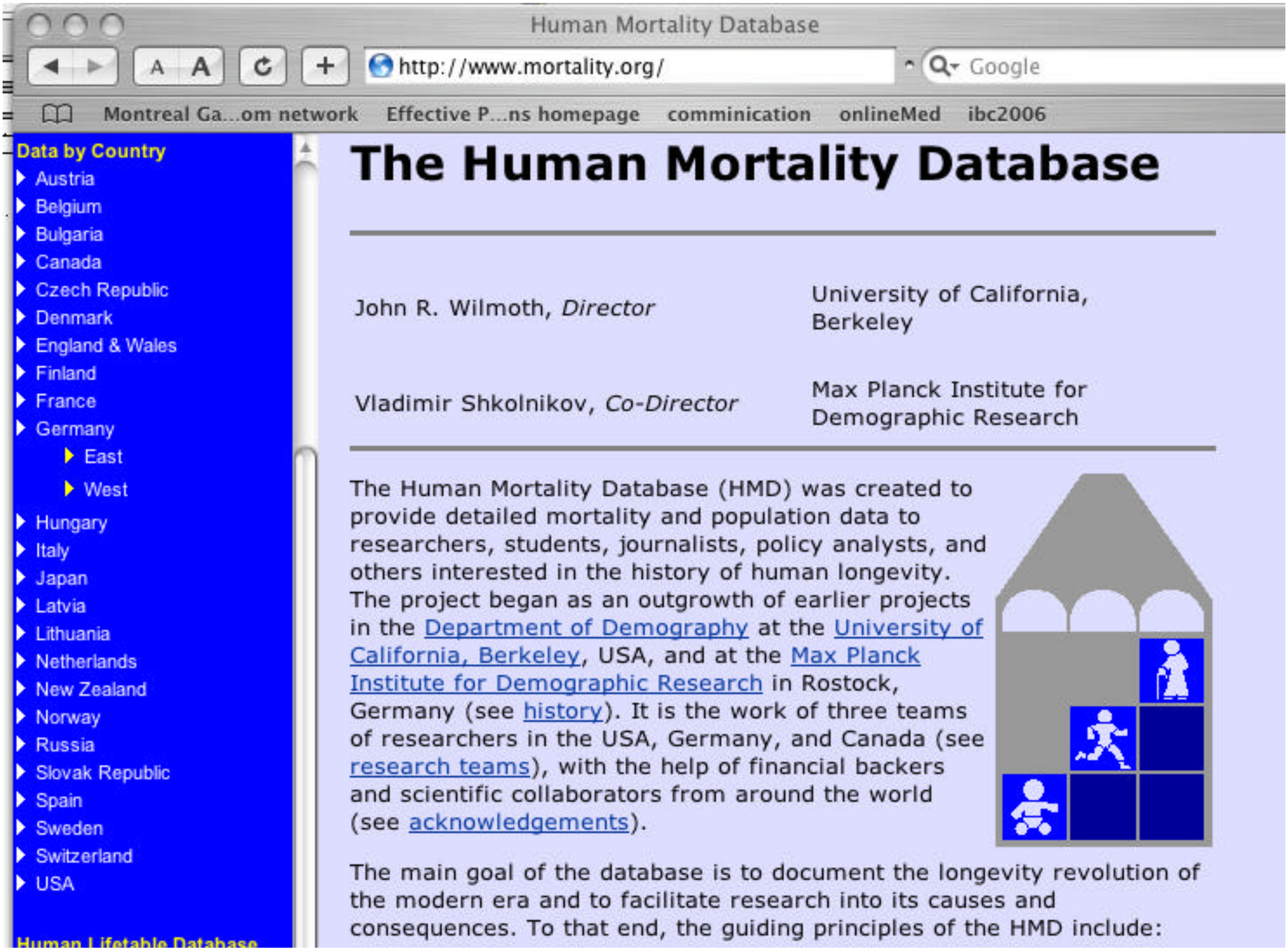
(Synthetic) Cohorts of Persons Alive on April 15, 1912



Pr[> {a+1,y+1} | > {a,y}] = Pr[> a+1 | a] using obs^d mortality in year y.



Σ
S[ ]
1.0
Passenger
S[ ]
Comparison
0.0
+
1.0
0.0
+
etc.
=
1.0
All passengers
All
0.0
1912
2003



Human Mortality Database

http://www.mortality.org/

Google

Montreal Ga...om network | Effective P...ns homepage | comminication | onlineMed | ibc2006

Data by Country

- Austria
- Belgium
- Bulgaria
- Canada
- Czech Republic
- Denmark
- England & Wales
- Finland
- France
- Germany
  - East
  - West
- Hungary
- Italy
- Japan
- Latvia
- Lithuania
- Netherlands
- New Zealand
- Norway
- Russia
- Slovak Republic
- Spain
- Sweden
- Switzerland
- USA

Human Lifetable Database

The Human Mortality Database

John R. Wilmoth, *Director*	University of California, Berkeley
Vladimir Shkolnikov, *Co-Director*	Max Planck Institute for Demographic Research

The Human Mortality Database (HMD) was created to provide detailed mortality and population data to researchers, students, journalists, policy analysts, and others interested in the history of human longevity. The project began as an outgrowth of earlier projects in the Department of Demography at the University of California, Berkeley, USA, and at the Max Planck Institute for Demographic Research in Rostock, Germany (see history). It is the work of three teams of researchers in the USA, Germany, and Canada (see research teams), with the help of financial backers and scientific collaborators from around the world (see acknowledgements).

The main goal of the database is to document the longevity revolution of the modern era and to facilitate research into its causes and consequences. To that end, the guiding principles of the HMD include:

Sweden

WARNING: The quality of the data for 1751-1860 are lower than in later years and should be used with caution. For details, please see the "Data Quality Issues" section of the General Comments file.

Data Files Explanation

General Comments

List of Data Sources

1. Births 1749-2003
2. Deaths 1751-2003 Lexis triangles 1x1 5x1
3. Population size (January 1st) 1751-2004 1-year 5-year
4. Exposure-to-risk

   By year of death (period)

   - 1751-2003 1x1 1x5 1x10 5x1 5x5 5x10

   By year of birth (cohort)

   - 1676-1973 1x1 1x5 1x10 5x1 5x5 5x10

5. Death rates

   By year of death (period)

   - 1751-2003 1x1 1x5 1x10 5x1 5x5 5x10

   By year of birth (cohort)

   - 1676-1973 1x1 1x5 1x10 5x1 5x5 5x10

6. Life tables

   By year of death (period)
   1751-2003

   - Female 1x1 1x5 1x10 5x1 5x5 5x10
   - Male 1x1 1x5 1x10 5x1 5x5 5x10
   - Total 1x1 1x5 1x10 5x1 5x5 5x10

   By year of birth (cohort)
   1751-1912

   - Female 1x1 1x5 1x10 5x1 5x5 5x10
   - Male 1x1 1x5 1x10 5x1 5x5 5x10
   - Total 1x1 1x5 1x10 5x1 5x5 5x10

7. Life expectancy at birth 1751-2003

Sweden, Life tables (cohort 1x1), Females

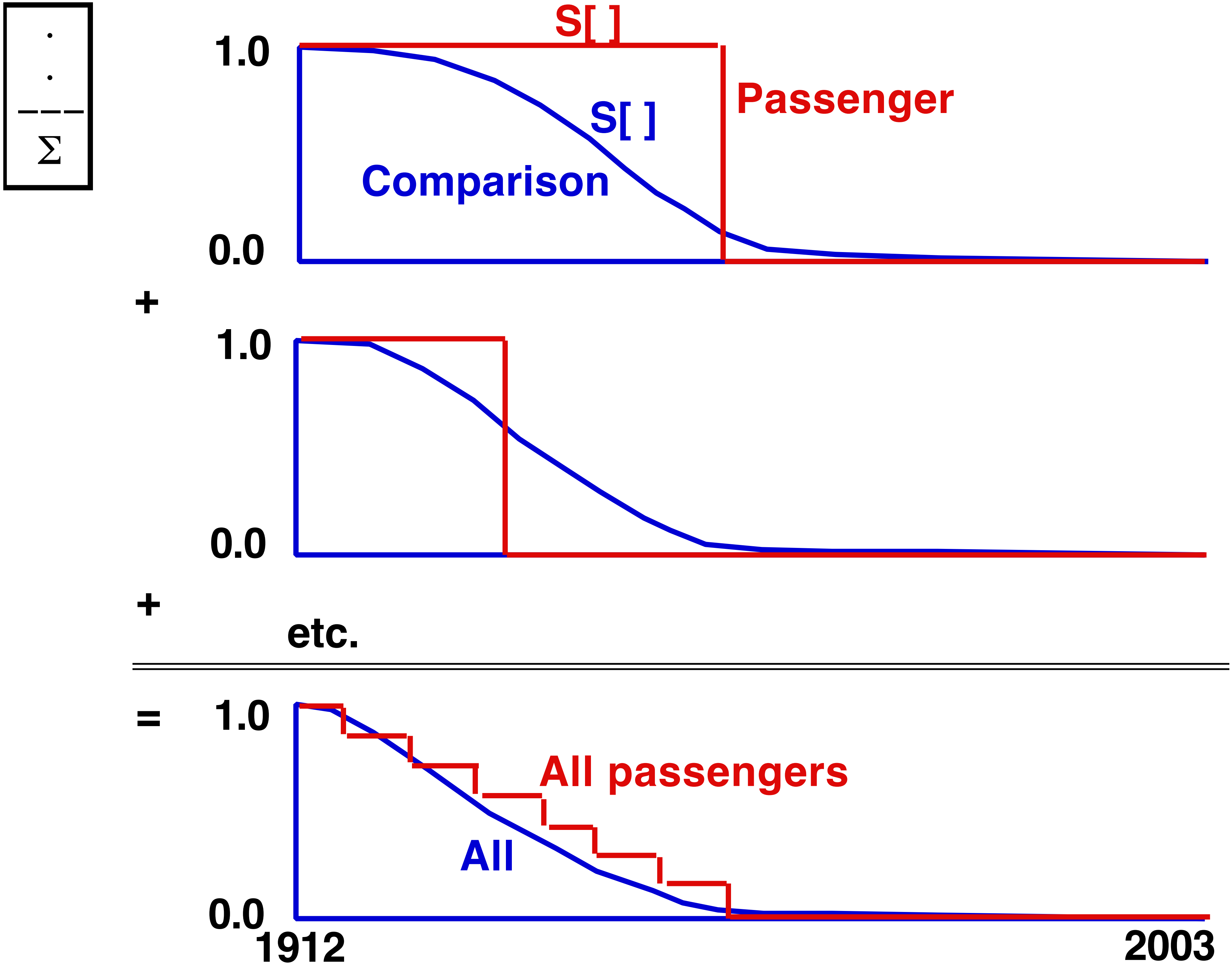
Last modified: 20-Apr-2005, MPv4 (Feb05)

Year	Age		lx	dx	qx	Lx	ex	*l'x* (Re-Scaled)	
1751	0		100000	20834	0.208	86458	35.8		
1751	1		79166	4997	0.063	76416	44.1		
1751	2		74169	2743	0.036	72819	46.0		
....									
1852	0		100000	14957	0.149	90278	46.9		
1852	1		85043	3730	0.043	83014	54.1		
1852	2		81313	2121	0.026	80251	55.6		
....	..		.....	....	.....	.....	....		
1852	60	<	49042(*)	804	0.016	48629	17.3	**100000**	
1852	61		48238(1)	830	0.017	47830	16.6	**98361**	**(1)÷(*)**
1852	62		47408(2)	937	0.019	46937	15.8	**96668**	**(2)÷(*)**
....									
1892	0		100000	9517	0.095	93694	58.0		
1892	1		90483	2514	0.027	89168	63.1		
....	..		.....	....	.....	.....	....		
1892	20	<	79360(*)	410	0.005	79157	52.1	**100000**	
1892	21		78950(1)	341	0.004	78787	51.3	**99483**	**(1)÷(*)**
1892	22		78609(2)	447	0.005	78389	50.5	**99053**	**(2)÷(*)**
1892	23		78162	468	0.006	77932	49.8	**.....**	
1892	24		77694	372	0.005	77509	49.2	**.....**	
1892	25		77322	504	0.006	77091	48.4	**.....**	
1892	26		76818	1185	0.015	76123	47.7	**.....**	
1892	27		75633	419	0.005	75430	47.4	**.....**	
1892	28		75214	410	0.005	75017	46.7	**.....**	
....									

Sweden, Life tables (cohort 1x1), Females

Last modified: 20-Apr-2005, MPv4 (Feb05)

Year	Age		lx	dx	qx	Lx	ex	l'x (Re-Scaled)
1912	0	<	100000	6248	0.062	95231	68.7	**100000**
1912	1		93752	1400	0.014	93023	72.3	**93752**
1912	2		92352	701	0.007	92004	72.3	**92352**
1912	3		91651	494	0.005	91402	71.9	**91651**
1912	4		91157	416	0.004	90945	71.3	**.....**
1912	5		90741	355	0.003	90569	70.6	**.....**
1912	6		90386	536	0.005	90100	69.9	**.....**
1912	7		89850	330	0.003	89682	69.3	**.....**
1912	8		89520	208	0.002	89418	68.5	**.....**
1912	9		89313	203	0.002	89212	67.7	**.....**
1912	10		89110	135	0.001	89043	66.9	**.....**
1912	11		88975	138	0.001	88904	66.0	**.....**



S[ ]
1.0
Passenger
S[ ]
Comparison
0.0
+
1.0
0.0
+
etc.
=
1.0
All passengers
All
0.0
1912
2003

Hazardous journeys

Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal, QC, Canada H3A 1A2

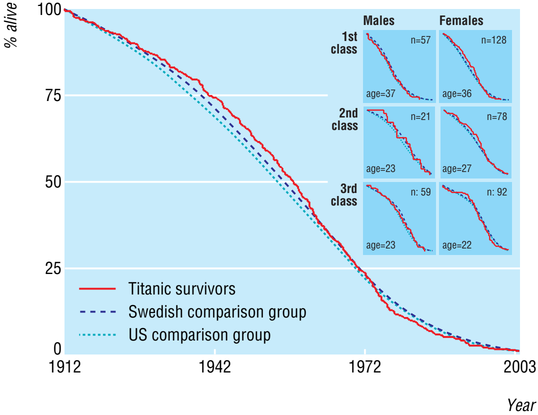
James A Hanley
professor

Carine Bellera
graduate student

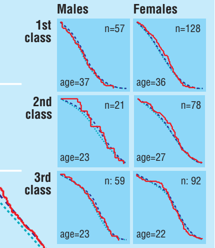
Dana Teltsch
graduate student

Department of Mathematics and Statistics, McGill University

Elizabeth Turner
graduate student



Percentage still alive on each anniversary of sinking of *Titanic* among 435 survivors and Swedish and white American comparison groups matched for age and sex. Inset: analysis by sex and class of travel (n=No of passengers; age=median age in 1912)



Males
Females
1st class
n=57
age=37
n=128
age=36
2nd class
n=21
age=23
n=78
age=27
3rd class
n: 59
age=23
n: 92
age=22

differences were assessed by log rank tests.

The survival of the 435 passengers was slightly, but not significantly, longer than that of the two comparison groups (figure). On average they lived 1.7 years longer than the general population of the United States and 0.5 years longer than that of Sweden. This small advantage was limited to female passengers in first and second class (figure). Five women lived past 100, and the three survivors still alive are now in their 90s. Despite their higher socioeconomic status, male passengers in first class did not outlive similar age males in the general populations.

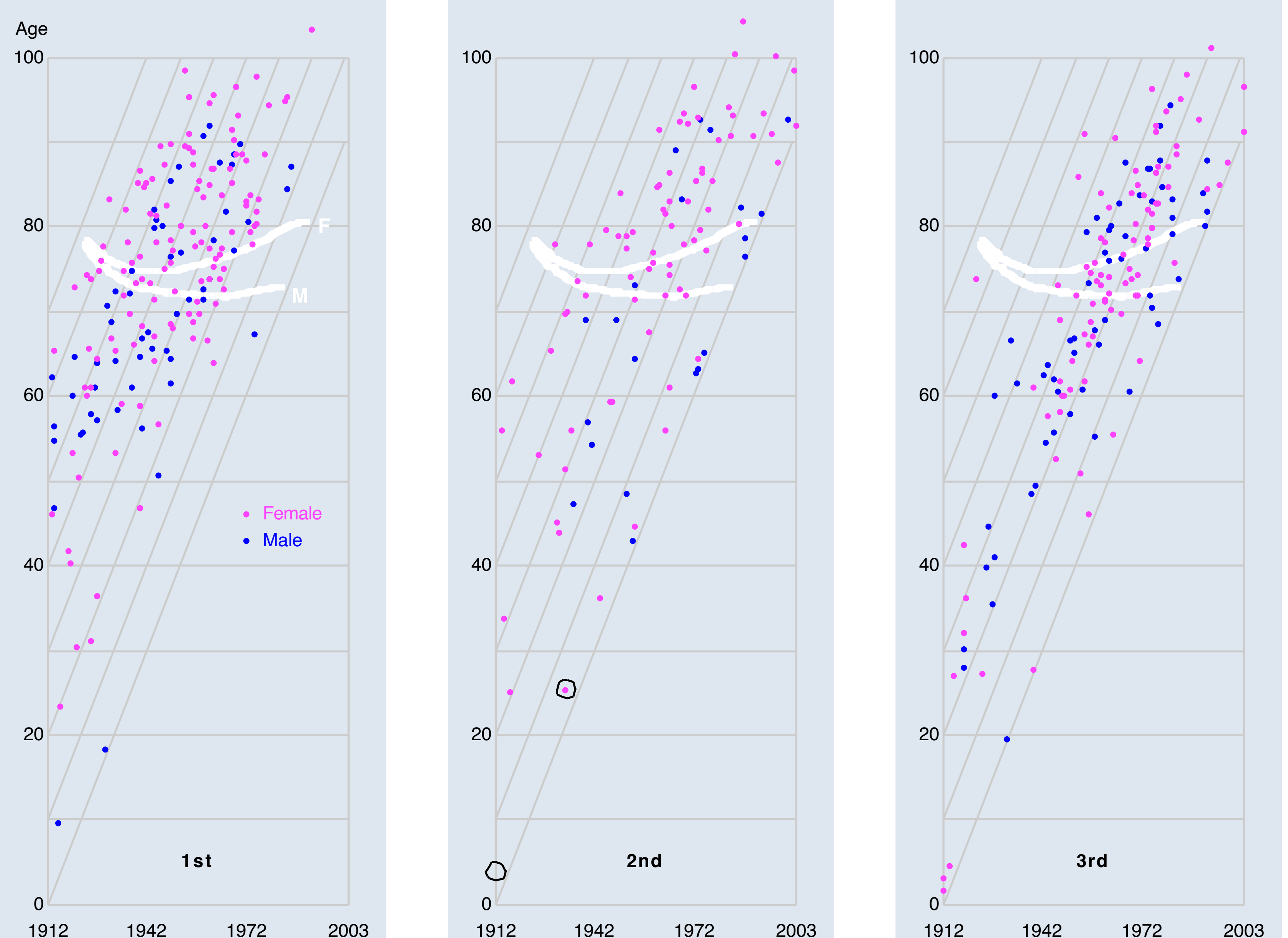
Comment

The longevity of *Titanic* survivors who could be traced was not remarkably different from that of age and sex matched individuals in the general population. The available life table data did not allow us to match on social class. Nevertheless, those who travelled third class had similar survival to our comparison group. We therefore wonder why males (and maybe even females) in first and second class did not fare considerably better than the general population.

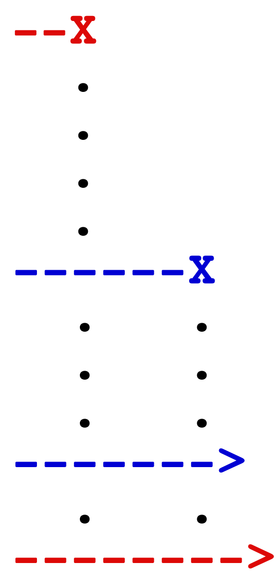
Follow up is complete for 87% of the passengers who survived the sinking; only 65 people, several of them servants to those in first and second class, are still untraced and excluded from our analysis. The quality of the follow up data on those traced seems to be excellent. Most dates of birth, important for age matched comparisons, also seem to be trustworthy.

Although unable to find the perfect comparison group, we avoided errors made in other longevity comparisons.[4 5] For the comparison group, we calculated the remaining lifetimes of people alive in 1912. Since age specific death rates fell substantially during the 20th century, we calculated these remaining lifetimes using the 1912-2000 death rates.

In the closing song of the 1997 film, the heroine tells us that her heart "must go on and on" and tells us twice more that it "will go on and on." The *Titanic* survivors did not have shorter life spans than the general population. Nor did they, despite the determination implied by the lyric, substantially outlive them.


Age
100
80
60
40
20
0
1912
1942
1972
2003
F
M
Female
Male
1st
2nd
3rd

Stratified Log-rank test   in general...

Stratum	n_1	n_0		lifelines & risksets	Observed	Expected | H_0	$V[a \mid H_0]$
				t_0 ⇓ ⇓	X − $\begin{array}{cc\|c} a & b & n_1 \\ c & d & n_0 \\ \hline n_X & n_- & n \end{array}$	X − $\begin{array}{c\|c} a_E & n_1 \\ & n_0 \\ \hline & \end{array}$	$\frac{n_1\, n_0\, n_X\, n_-}{n^2(n-1)}$
1	2	2			$\begin{array}{cc\|c} 1 & 1 & 2 \\ 0 & 2 & 2 \\ \hline 1 & 3 & 4 \end{array}$	$\begin{array}{c\|c} 1/2 & 2 \\ & 2 \\ \hline & \end{array}$	$\frac{2 \times 2 \times 1 \times 3}{4^2(4-1)}$
					$\begin{array}{cc\|c} 0 & 1 & 1 \\ 1 & 1 & 2 \\ \hline 1 & 2 & 3 \end{array}$	$\begin{array}{c\|c} 1/3 & 1 \\ & 2 \\ \hline & \end{array}$	$\frac{1 \times 2 \times 1 \times 2}{3^2(3-1)}$
2	..	..			...	...	...
...	..	..			...	...	...
Σ					$\Sigma\, a$	$\Sigma\, a_E$	$\Sigma\, V[a \mid H_0]$

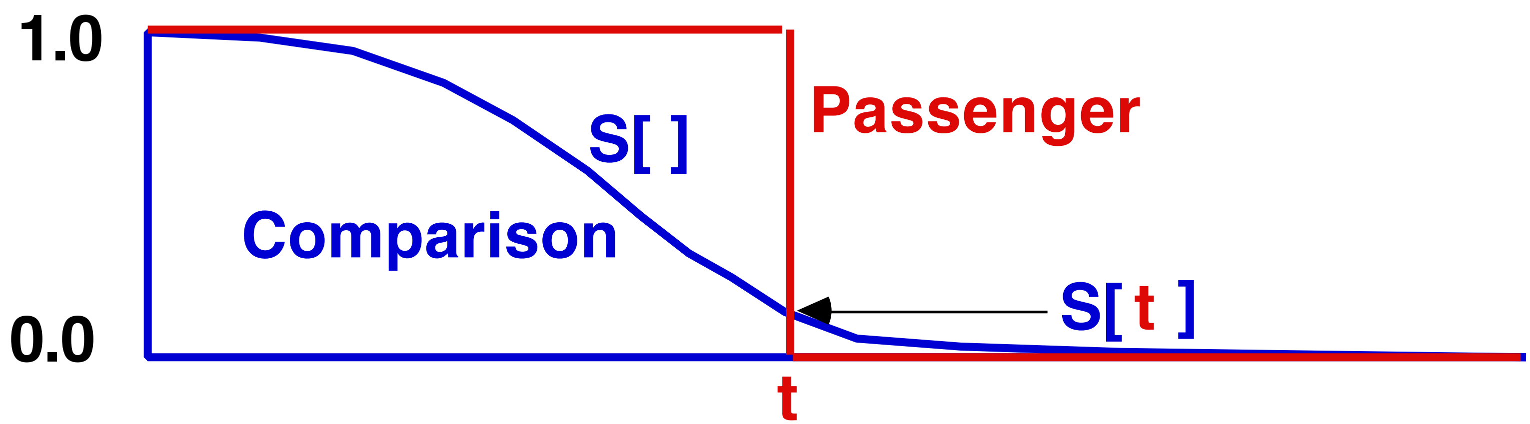
$$\Sigma \textit{ over all strata: } \quad \frac{\{\Sigma\, a \;-\; \Sigma\, a_E\}^2}{\Sigma\, V[a \mid H_0]} \sim \chi_1^2$$

Stratified Log-rank test 1 stratum [passenger&peers] n_1 = 1 and n_0 >> 1 [déjà dead]

n_1	n_0	lifelines & risksets	Observed X - a b | 1 c d | nnnn -------- | ------ 1 nnnn | nnnn+1	Expected | H_0 a_E	V[a|H_0]
1	10^4	X --X ---X ----X -----X ------X -------X* ---------X ...	0 0 0 0 0 0 1	1/10001 .. 1/ 9001 .. 1/ 8001 .. 1/ 7001 .. 1/ 6001 .. 1/ 5001 .. 1/ 4001	
Σ		* S[t] = 0.4	1	0.916	0.916
death at time t : S[t] × 100% of peers still alive			1	-Log[S[t]]	-Log[S[t]]

$$\Sigma \text{ over all 435 passengers: } \frac{\{\Sigma(1 + \text{Log}[S[t]])\}^2}{-\Sigma \text{ Log}[S[t]]} \sim \chi_1^2$$

Alternatively: Combine S[t_1] , S[t_2] ... S[t_{435}] à la Fisher



S[t] = Prob[T > t | Comparison S[ ]) is a 1-sided p-value.

Under Null: -2 log [S[t] } ~ χ_2^2

n (= 435) independent p-values: - Σ 2 log [S[t_i] } ~ χ_{2n}^2

References

Finkelstein DM et al. [2003] Comparing survival of a sample to that of a standard population. *J Nat. Cancer Inst.* 95(1) 1434-1439.

Woolson RF. [1981] Ranks tests and a one-sample logrank test for comparing observed survival data to a standard population. *Biometrics* 37(4) 687-696.

Gail MH and Ware JH. [1979] Comparing observed life-table data with a knows survival curve in the presence of random censorship. *Biometrics* 35(2) 385-391.

Moses LE. [1964] One sam[ple limits for some two-sample tests. *J Amer Stat Association* 59(307) 645-651.

Noname manuscript No.
(will be inserted by the editor)

Longevity comparisons:
fair contests and transparent statistical methods

Manuscript I: Study Design

James A Hanley

Draft: September 08, 2022 / Submitted: date

Abstract Some of the statistical 'longevity contests' that receive media and public attention are more whimsical, while the reported findings of other comparisons have more serious medical and public health implications. Either way, it is important that the statistical methods are transparent and that the accuracy/validity of the findings is maximized. In manuscript I, we review the design of selected longevity contests, the statistical analyses they have employed, and the artifacts they have produced. In manuscript II, using a worked example, we emphasize the clarity and transparency that stem from using matched designs, and analyses that maintain the matching. We highlight the use of Lexis diagrams, the possible time scales that can be used, and the conceptual and practical simplicity of the stratified Cox model in the case of imperfect matching. The use of matched *populations* (rather than individuals) as comparators may give the 'average person in the street' more perspective on the claimed longevity lengthening/shortening of the achievements/activities/occupations/states in question. Population comparators are also used heuristically to show how the Mantel-Haenszel and Log Rank test statistics are connected, and to highlight less well known interpretations of the null expectations involved, along with 'translations' of hazard ratios into longevity differences. In the case of pairwise longevity contests, such as those that pit an Oscar winner against the same-movie same-sex, closest-in-age cast member, the Log-rank statistic and hazard ratio estimator just involve counts.

Keywords First keyword · Second keyword · More

James A Hanley
McGill University,
2001 McGill College Avenue Room 1214,
Montreal, Quebec,
H3A 1G1,
Canada
Tel.: +1-514-398-6270
Fax: +1 515-398-4503
E-mail: james.hanley@mcgill.ca

1 Introduction

All the world's a stage,
and all the men and women merely players;
they have their exits and their entrances.
– *Shakespeare, As You Like It*

.

As Groucho Marx once said, "'Getting older is no problem:
you just have to live long enough."
– *Queen Elizabeth II, at her 80th birthday celebration in 2006.*

.

Such an analysis is seriously flawed, because the definition of
one of the two groups to be compared conditions on the future.
As a more comic point, we noted that IJE now quotes P-values with
308-digit precision; we hope that the chi-square approximation
to the distribution of the log-rank statistic is justified!
– *Theis Lange and Niels Keiding, letter to IJE, 2014.*

.

I read an article that said that winning an Oscar could lead
to living five years longer; if that's true, I'd really like to
thank the Academy because my husband is younger than me.
– *Julianne Moore, on winning Oscar for actress in a leading role, 2015.*

.

Statistical 'longevity contests' that report how much a select group of persons outlives its peers receive considerable media and public attention, particularly if the reported longevity difference is substantial. In many instances, these reported differences are misleading, since they result from unfair comparisons or improper statistical analyses, or conceptual misunderstandings.

Some of these comparisons, such as in the Christmas Editions of the British Medical Journal, are merely whimsical, or address very specific circumstances. However, the findings of many longevity contests have broader and more serious medical and public health implications. They involve lifestyle behaviours that ordinary people, who were not the target of the studies, can relate to. Thus, if indeed the reported benefits are correct, these ordinary people might also benefit if they adopted them. The same applies to studies of possible harms: the public at large might also benefit if they too were to avoid these activities/behaviors. Thus, it is important that all longevity contests are valid and that the statistical methods are transparent.

The activities / traits / interventions that have been studied cover a wide range. Some studies have reported shorter lives of those with certain traits, such as left-handedness [27,28], or working in certain occupations, such as playing in the US National Football League [15], being a student [67], a female physician [11,42], a jazz musician [62], a rock star [40], or a pope [12]. Other studies have investigated the effects of early recognition, such as being a class president [55] or a young US president [48], or being inducted into the Baseball

Hall of Fame [2]. Some studies have claimed the life-extending benefits of extreme fame, such as winning an Oscar [53,54], or a Nobel prize [52], or the rigorous training that Olympians [16,79,65], professional cricketers [36], and major league baseball players [30] had to endure. Others have touted the extra longevity that comes from being an orchestra conductor [5,78], a harpist [78], or a politician [6]. Yet others have studied the status that accompanies royalty or peerage [?] or political success, or 'healthy' behaviours such as sunbathing [9] and exercise [7]. Medical examples include heart transplants [44] and taking statins [77]. Some studies have examined whether life events, such as being shipwrecked [29], or the death of one's child [61], affect lifespans.

Since they are too many to examine each study design in detail, Section 2, i.e., (the remainder of Manuscript I) will review the design of selected ones. In Manuscipt II, Section 3 will address statistical methods, while, in the closing act, Section 4 will bring the design and data-analysis principles together by applying them to the 2022 'sequel' to the 2001 study of the longevity of performers who win an Oscar.

I deliberately avoid the term 'bias' and its older-style qualifiers such as 'confounding', 'selection', and 'information', as well as newer terms such as 'backdoor' and 'collider'. Instead, like William Farr, I will try to describe each situation using plain words, so that unfair comparisons are evident even to lay readers. I will also refer to Farr's memorable examples of what is now known as 'immortal' (person)time' – a central topic in the current article – and his tongue-in-cheek attempt at causal inference.

2 Measures of Longevity, Artifacts, and Designs that Avoid Them

I will begin with comparisons of *ages at death*. Even today, this is often the 'go-to' metric, either because the ages of the dead are the only data available, or because investigators do not not know how to deal with the additional data from those still living. In these circumstances, investigators merely compare ages at death of those in the index and reference categories of the factor of interest. Additional considerations, having to do with keeping contests fair, are addressed in section 2.2.

2.1 Cemetery/ Obituary Epidemiology

Examples of the *mean age at death* go back several centuries, to when only the ages of those who have died were known, and the age structure of the populations/samples in which these deaths occurred (sometimes referred to as the 'base') was not. Over his decades at the office of the Registrar General in the 1800s, William Farr repeatedly explained the dangers of using the mean age at death. He repeatedly pointed out that "the mean duration of life, technically known as the expectation of life, differs very widely from the mean age at death." While lamenting that "it is only a pity that the (*mean-age-at-death*) method is not as accurate as it is easy" (Farr, p457), he used natural jargon-free language to describe the limitations and lurking dangers:

> [... More recently] the mean age at death has been relied on to show the healthiness or insalubrity of certain occupations. And this method, as well that of the annual rate of mortality without distinction of age, is applicable in certain definite conditions where only approximations are required. But the mean age at death evidently depends upon many circumstances beside health, and among others, upon the ages of the living which vary in proportions in almost every profession, according as it is a profession that people enter early or later in life, and according the number that enter it annually increase or decrease.

Along with some delightful tongue-in-cheek 1850's style causal inference commentaries, he also gave several compelling and easily understood examples – using the various ranks within the legal, military and religious professions – of why "It requires no great amount of sagacity (or familiarity with directed acyclic graphs) to perceive that the 'mean age at death' or the age at which the greatest number of deaths occurs cannot be depended on in investigations investigating the influence of occupation, rank, and profession upon health and longevity." (p458) And he provided numerous examples of how to calculate life expectancy, of why and by how much it could differ from mean age at death, and of the (seldom-fulfilled) conditions when the two coincide. [As a nice teaching example of the latter, this author recommends the under-appreciated article [10] where a population equilibrium/stability is reached after a number of testing cycles.]

Despite these warnings, mean age at death is still widely used, and indeed the mean age at death in the index category is sometimes compared with life expectancy in the reference category. A few post-Farr longevity examples serve to illustrate the pitfalls.

Musicians Since previous research did not answer the question "why do so many pop musicians die young?" one academic undertook the 'first population study of performing pop musicians (n=12,665) from all popular genres who died between 1950 and June 2014 of whom 90.6% (11,478 musicians) were male. Data were accessed from over 200 sources, including The Dead Rock Stars' Club; Nick Tavelski's (2010) Knocking on Heaven's Door: Rock Obituaries, Pop star mortality; R.I.P. Encyclopaedia Metallicum; Voices from the Dark Side for Dead Metal Musicians; Wikipedia's List of Dead Hip Hop Artists and Hip Hop obituaries."

"Longevity was determined by calculating the average age of death for each musician by sex and decade of death. These averages were then compared with population averages by sex and decade for the US population." Whereas the text description was non-specific as to what "population averages" were being used, the title of the graph ["Life expectancy of pop musicians: average age of death of pop musicians vs. general US population (1950-2014)"], the values plotted, and the sources cited, confirm that they were indeed the life-expectancy numbers. The results (life-shortenings of the order of 20-25 years) were published in a not-for-profit university-supported media outlet [72] whose motto is ‘academic rigour, journalistic flair.’

The source of the unfairness in this comparison does not fall neatly into one of the traditional categories of ‘bias’; nor is not so easily explained using reference to a directed acyclic graph. Nor does it fit into Farr’s list of ‘evidently it depends on’ in inter-profession comparisons . Instead, it is an example where one metric (the mean age at death of just the deceased) is computed for the index category and another metric (the calculated life expectancy in a synthetic or hypothetical life table based on the age-specific death rates in the population-time in question) is computed for the reference category. In addition, as we will expand on later, while those who contribute to the latter are matched on sex and decade, they are from a much broader age range that includes infants and adolescents – born in recent decades, and not yet old enough to be rock musicians (but old enough to die!), as well as older persons – born well before the modern rock era. A contrast of the death-rates in pop musicians vs. those in suitable comparator professionals would be fairer – if indeed one could define a similar entry criterion [74] for the latter.

Politicians Even when data *are* available on *both* the living and the dead, and actuarial methods could be used to deal with the censored lifetimes (see next section), some investigators simply discard the still-living. To address the question, “does political office cause worse or better longevity prospects?”, investigators [6] established the latest vital status of each of the two candidates who received the highest number of votes in US gubernatorial elections from 1945 to 2012. They then “removed” the “772 are still alive as of September 2019,” leaving them with a total of 1092 candidate-year observations from 676 unique elections. From these, they reported that “The results show that politicians winning a close election live 5-10 years longer than candidates who lose.”

Although this study drew severe criticism for its reliance on a regression discontinuity analysis [24], the fact that is was based on the just the dead, and that those who are still alive have information to contribute – and that their inclusion might change the findings – seems to have been overlooked. Fortunately, the authors also provide, without restrictions, their full dataset of 5129 observations.

Academy award winners In a study to investigate whether winning an Academy award (Oscar) for acting was associated with long- term survival [54], the first metric reported on was the “observed life span,” calculated from the age at death of the 1,122 performers (out of a total of 2,111) who had died by July 1, 2020.

> The average age at death for winners was 77.1 years, for nominees was 73.7 years, and for controls was 73.6 years. A t-test comparing winners

> to controls yielded a 3.5 year absolute difference in average life-span (95% confidence interval: 1.2 to 5.8). This simple comparison equaled a 4.8% relative increase in life years (95% confidence interval: 1.6 to 7.9). Similar calculations comparing winners to nominees yielded a 3.4 year absolute difference in average life-span (95% confidence interval: 0.8 to 6.1) equal to a 4.6% relative increase in life years (95% confidence interval: 1.1 to 8.2). [p5]

> Comparisons of winners to controls yielded a 4.8% relative difference average life-span (95% confidence interval: 1.6 to 7.9, p = 0.004), a 5.1 year absolute increase in life expectancy (95% confidence interval: 3.0 to 7.2, $p < 0.001$), and a 41% improvement in mortality hazard (95% confidence interval: 19 to 68, $p < 0.001$). [Abstract p1]

"Life expectancy" was estimated as survival through a multistate model to account for those who had not yet died. The life-expectancy for winners was 81.3 years, for nominees was 76.4 years, and for controls was 76.2 years. The 5.1 year difference was included in the abstract, and the message that "Oscar winners live five years longer than other actors" appeared in a number of media headlines. Some others outlets chose to report the 3.5 year difference in average life-span.

By not using the matching in the analysis, and by including several 'unfair-to-comparator' contests, the design and the analyses raise several issues. These will be addressed in detail in the following sections.

2.2 Artifacts

To make Farr's warnings about the artifacts produced by using the mean age at death more concrete, we provide more recent examples that today's readers might more easily relate to and that include some additional artifacts.

Longevity of left- vs. right-handed persons In the late 1980s, investigators [27] "analysed all baseball players listed in The Baseball Encyclopedia for whom dates of birth and death, as well as throwing and batting hand, are reported." They reported that the "mean age at death for the 1,472 right-handers was 64.64 years (s.d. = 15.5) and 236 left-handers was 63.97 years (s.d. = 15.4). Although this 0.67 year difference is clearly well within the limits of chance variation (its standard error is slightly more than 1 year) the investigators used a (what appears to be a very inappropriate) "nonparametric test of group differences (Wald-Wolfowitz runs test)" that "indicated that the greater longevity for right-handers is significant ($Z = 6.63$, $P<0.001$)."

In order to test this relation between handedness and life span in a general population, they then "obtained death certificates from two counties in southern California.[28] Two thousand questionnaires concerning the handedness of the deceased family member were sent to the listed next of kin, which resulted in 987 usable cases (495 male subjects and 492 female subjects). Subjects were designated as right-handers if they wrote, drew, and threw a ball with the right hand. All other subjects (left-handers and mixed-handers) were assigned to a non-right-handed group." The effects of handedness on life span were "striking in their magnitude. The mean age at death in the right-handed sample was 75 years, as compared with a mean age at death of 66 years in the left-handers. This nine-year reduction in life span for the left-handers is significant ($F_{1,945} = 22.36$, $P<0.0001$)."

The resulting correspondence argued that the prevalence of left-handedness had increased over the period in question, and that since a smaller proportion of the left-handers would have reached the ages where people tend to die, the 9-year difference in the mean ages at death was an artifact. More cogently, data from the Framingham study and from NHANES did not confirm the findings; not did a subsequent large study on UK cricketers. [3]

This intra-profession example fits squarely within Farr's concerns about the differing age structure of the two compared subgroups: "the mean age at death evidently depends upon the ages of the living", and the numbers that enter the compared handed-ness cohorts annually may "increase or decrease" in different ways.

Longevity of female vs. male physicians A similar neglect of the age structure of the living, and of the relatively recent entry of women in larger numbers into the medical profession, is evident in this response to a 2005 BMJ editorial on women doctors and their careers.

> A speculative conversation in BMA House (17 November) led to a feasibility study investigating survival, among that unique population of distinguished professionals who merit an Obituary in the BMJ. This follows a series of studies on the occupational wellbeing of doctors. Between 7 January and 18 November, 297 men and 49 women were described in the printed Journal, with their year of birth and death. [...] However, the main finding from this small study was unexpected. The mean age at death in obituaries of women was 72.7 years and of men was 79.3 years (T-test, unequal variance, p = 0.011, U-test, p = 0.012). In England as a whole, the latest report on the Department of Health website "Tackling health inequalities: Status report on the Programme for Action" suggests women live on average 4.5 years longer than men (80.7 vs 76.2 years).
>
> Let us now praise distinguished women in medicine, who perhaps sacrifice 6.6 years compared to men of similar merit.

Whereas Farr's warnings were limited to the 'age' component of the time dimension, the foregoing examples bring out an important second component, namely 'calendar time.' The Lexis diagram, addressed below, will give these two components equal prominence.

Additional musician contests: To test the commonly held view that jazz players tend to be more liable than other professions to die early deaths from drink, drugs, [,] or overwork, an emeritus professor [62] undertook a statistical study of 86 US-born jazz musicians listed in a university syllabus. "Dates of birth, and of death when it had occurred, were tabulated, and longevity matched with that expected in the United States by year of birth, race, and sex. One musician who had not reached the age of his life expectancy was excluded from the list. Birth years ranged from 1862 to 1938; 16 births occurred before 1900, 23 between 1900 and 1909, 19 between 1910 and 1919, 22 between 1920 and 1929, and five between 1930 and 1939. Comparison with national values showed that 70 (82%) of the musicians exceeded their life expectancy; four-fifths of the Black men, three fourths of the White men, and all the women lived longer than expected.

Although the size and sex distribution of the sample limits the inferences to be drawn, the data suggest that jazz musicians do not die young. Most of the 85 musicians in this study have survived the potential hazards of irregular hours of work and meals, the ready temptation of drugs and alcohol, and the perils of racial prejudice."

A major artifact Rothman's response [57] has a clear description of the *main* artifact: "noted jazz musicians have a head start of several decades of life on the representative citizen, whose longevity was counted from birth. No one is classified as a jazz musician at birth, no matter how auspicious the circumstances. Any death before the age of 20, say, would shorten the average life span considerably, but could not affect the longevity of noted jazz musicians." In today's terminology, jazz players are said to be "*immortal*" during the first 20 or so years, whereas those in the comparison category are not. If, as Lange and Keiding were called on to do, I had to explain the term 'immortal time' to a television audience, this would be one of my go-to examples.

The basis for the widespread belief that *orchestra conductors live longer than average* and the subsequent theory that it may be due to the vigorous physical and mental exercise involved,[1] goes back to a 1978 comparison involving just 35 deceased major symphony leaders compiled form several source books and the author's own experience. [5] Their mean length of life was 73.4 years. The New York Times health reporter told readers that "The life expectancy of American men in general is 68.5 years, [the author] said, and the difference is statistically significant. [...] I am aware that a comparison of the current survival expectancy of American men to that of European born conductors from the last century may be open to question, Nevertheless, since

[1] "Orchestra conductors and harpists appear to live longer than other musicians and they both use their arms more than most other musicians. A review of the mean age of death of 8755 musicians (7371 men and 1404 women) [78] shows that the longest-lived males were conductors (71.1 years), cellists (70.0 years) and violinists (70.0). Among female musicians, those with the longest lives were harpists (80.9 years), pianists (79.9 years) and conductors (79.6 years). The shortest life spans were for rock musicians (45 years), who often die young from drug use." [https://www.drmirkin.com/fitness/arm-exercises-many-conductors-have-long-lives.html]

I have not been able to find a single death in this group at an age younger than 58, I firmly believe that these men were protected by some undetermined factors from the modern scourge of early fatal ischemic vascular disease."

One reader [13] had serious doubts about the validity of the inference that involvement with music lengthens life.

> There is an elementary statistical artifact that I think may seriously vitiate that inference. The mean length of life of Dr. Atlas's sample of 35 conductors was 73.4 years, as compared with 68.5 years, the current life expectancy of American men. The problem is that the latter figure is the life expectancy *at birth*. The fact that a person is an orchestra conductor implies that he or she has reached some minimal age. The average age at which the five eminent deceased conductors mentioned in the article were appointed to their first regular conductorships is 32. The life expectancy conditional upon having attained that age would be 72 years. (This is the figure for white American males as of 1974, as per the 1976 Statistical Abstract of the United States.) Since these five eminent conductors probably received their first appointments at a somewhat earlier than typical age, the correct conditional expectancy is probably greater than 72 years. I rather doubt that Dr. Atlas's figure would prove statistically to be significantly different from that value.

This could be another easily understood example of 'immortal time' for a lay audience, We have previously [33] given several historical examples of artifacts raising from the fact that "certain professions, stations, and ranks are only attained by persons advanced in years," and so we merely update the example in Farr's tongue-in-cheek causal inference: "a strong case may no doubt be made out on behalf of young, but early-dying [assistant-conductors]. It would be almost necessary to promote them earlier – for the sake of their health."

An 'out of synch' artifact Even in the comparison with the more appropriate conditional (remaining) life expectancy that the NY Times reader proposed, there is also another possible artifact – one that, as far as I can tell, has escaped being named or pigeon-holed. In the study of jazz players, the author used a personalized life expectancy derived from the *death rates the year the player was born.* Table 17.1 of the classic textbook by Armitage et al. nicely illustrates the problem with this comparator. It shows the "current" lifetable for 1930-1932, calculated using the observed age-specific mortality rates in England and Wales in 1930-1932. Those who computed this table in the 1930's didn't know how death rates would evolve over the next 50 years. In hindsight, 50 years later, the improvements in public health and medicine in these intervening years were such that some 83% of those born in 1930-1932 were actually alive, whereas the lifetable constructed at the time of their birth calculated that only 75% would be. A more systematic investigation [26] suggests that "current period life expectancy in the industrialized world applies to cohorts born some 40-50 years ago."

Clearly, the comparator should 'move in time/age with the longevity contestant' and not just be limited to the one for the year of *birth* or the year of *death.* The next section will show how this moving contest can be visualized in both time dimensions simulataneously, and how different contestants' races can start at different ages/years.

But before doing so, we emphasize an additional and easily easily overlooked artifact, where even the most modern and appropriate of statistical analyses produced a very wrong answer. This example will also serve as one of our two worked examples, and will also illustrate a further artifact in the second of them – one that can be avoided by limiting the analyses to clean contests.

Longevity of players inducted into the Baseball Hall of Fame Investigators took advantage of the detailed databases of major league players to examine the relation between longevity and awarded achievement.[2] They compared the longevity of the 143 who had been inducted into the Baseball Hall of Fame while still alive with that of 3,430 age-matched players who were alive at the time of the Hall of Famer's induction. Longevity was defined by post-induction survival time. A Cox proportional-hazards survival analysis was used to determine whether differences in survival between the groups were significant, controlling for career length, player position, and body mass index. Median post-induction survival for Hall of Famers was 5 years shorter than for non-inducted players (18 vs. 23 years, respectively).

Some of the explanation for this surprising and "hitherto unrecognized price" of fame can be found in the critique and re-analyses by [59] who pointed out that if no year of death was listed, the authors assumed that the player is still alive. "Unfortunately, the Lahman archive has incomplete data for some players, especially for relatively obscure persons who played in the early years of MLB." When the Lahman archive does not give a death date, the player may be still alive, "but in many cases no death date is listed because the death date is unknown." This *differential quality of the follow-up information* gives the 'less famous' players an artificial longevity advantage.

Unfortunately, in order to remove this artifact, the re-analysis had to limit itself to players with listed death dates. While the "robust test applied to [these] correct data shows that there is no statistically persuasive difference in the life expectancy of players elected to the Hall of Fame and their peers", the restriction to uncensored data creates its own issues. In the next subsection, we will explore other designs.

2.3 Fair contests, visualized simultaneously in two time dimensions

Longevity contests are very profitably visualized in a Lexis diagram, "a (time, age) coordinate system, representing individual lives by line segments of unit slope, joining (time, age) of birth and death. The diagram is an important descriptive tool in epidemiology and demography and it also has several applications in survival analysis and analytical epidemiology as a tool for several classes of statistical models" [39]

Illustration For a concrete illustration, we revisit the question of the longevity of players inducted, while still alive, into the Baseball Hall of Fame. We obtained the list from the `HallofFame` dataset in the `R` `Lahman` package, which contains the voting results for the candidates nominated for the Baseball Hall of Fame in the years 1936 to 2018. To narrow the range in the ages at induction, we limited ourselves to the 120 elected by the Baseball Writers Association of America, Their dates of birth and, when listed, dates of death were obtained from the `People` dataset using the common identifier `playerID`.

The 120 post-induction life-spans are shown as diagonal lines on the grid in Figure 1,[2] along with the marginal frequencies of the ages/years when the players were inducted and when they died or are presumed still alive.

Based on the area under the Kaplan-Meier-type curve calculated using age as the time axis,[3] the 'fitted' mean longevity is 76.4 years. [Incidentally, the mean age at death of the 67 players who have died is 73.8 years].

Comparators appropriate to the question posed Naturally, the design – and, thus, the appropriate comparator – must follow from the question, and from the perspective taken. In the original publication, the question was whether elevation to the Hall of Fame changes the longevity of baseball players. But, just like the ordinary viewers who merely watch the Oscars on television, members of the public – the 'average person on the street' – may well have may well have surmised, when they learned that these players were inducted, as to how much their *own* lives would benefit if *they* could exchange places with the winners. To answer this question, the use of a population-based reference category is appropriate. It also guards against claims, based on possibly misleading within-profession comparisons and complex statistical models, that winners, like the monarchy, live charmed and extra long lives.

This question is more easily and more precisely addressed: one can create player-specific longevity contests with *all* of the persons of the same sex and age who were alive when each contest begins, and use the published contemporaneous mortality rate for each Lexis (calendar-year, age) square that a player traverses [ref] to calculate what proportion of them are still alive at each anniversary of the induction. In the case of the Oscar awards (see below) it also serves as a conservative lower bound, since they are awarded while the performer is still (presumably) a healthy worker, a condition that is not imposed on the population reference. Moreover, the deaths are recorded in a national vital statistics system where obscurity is not a data-quality concern.

[2] From the patterns in the years numbers of deaths in the `People` dataset , we consider that it is close to up to date as as the end of 2021.

[3] Using the `R` call, `survival::survfit(Surv(ageAtInduction,LastAge,Death)` $\sim$ 1), where `death` is an indicator of whether the follow-up was ended (at `Lastage`) by death.

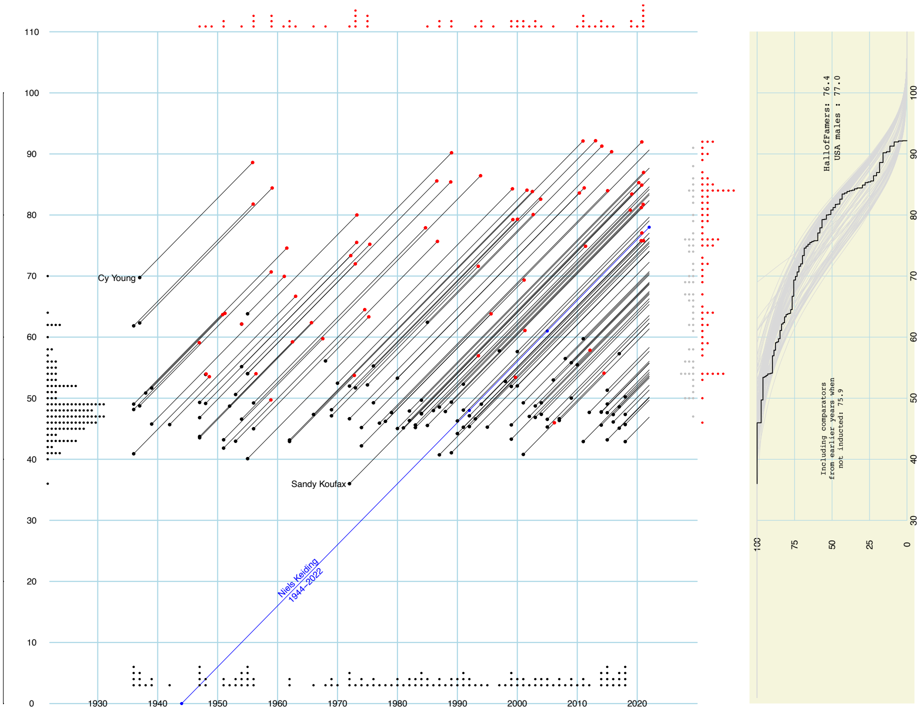


Fig. 1 Lexis diagram showing age at induction (left vertical axis), year of induction (bottom horizontal axis), and post-induction lifespans (diagonal lines) of 120 players inducted into the Baseball Hall of Fame (see text; for orientation, the life-span of a recently-deceased statistician [39] who did much to popularize the Lexis diagram, is shown in blue, along with the year/age when he assumed the presidency of the International Biometric Society, and the International Statistical Institute respectively.
The red and grey dots at the right of the grid are the ages at which players died or are presumed to be still alive, respectively. In the inset (beige) each smooth survival curve in light grey is based on all USA males who were born the same year as one of the inducted players and who were alive when that player was inducted, while the step function is the Kaplan-Meier-type curve calculated using age as the time axis, and the counting method to ensure that only the post-induction survival is considered.

For these reasons, we begin the next section with such a reference group, and then address the added complexities of using within-profession contrasts.

Designs with unfair-to-comparator contests Before doing so, we briefly address the *within-profession* comparison groups used in the study of the longevity of performers who win Oscar awards. Oscar winners *may have been nominated one or more times before they eventually won.* In the design employed, a within-film comparator was selected each time a performer was *nominated*, so some winning performers 'generated' *several* comparators. We can use the baseball data to illustrate what impact this has. The average longevity of 75.9 years shown on its own in Figure **??** was obtained by using *as many population-*

comparators as the number of elections where the eventually-inducted player was a candidate. Since the number of elections the eventually-inducted player had to wait until he was successful varied from 0 to 24 years (median 1, mean 4.2) including these pre-success comparators artificially reduces the average longevity of the comparison populations by just over 1 year.

.

Manuscript II: Data Analysis

3 Statistical Methods

When possible, it is important when communicating with the public that the differences in longevity be expressed in years of age, or in the duration of post-entry life, and not only via SMR's, hazard ratios, test-statistics and p-values. It is also desirable that the age- or calendar-year matching that makes the results more transparent and valid be accounted for in the analysis. We begin with methods that use *population* comparators.

3.1 **Population** comparators

Whereas a 2003 report [22] has a very helpful exposition of the one-sample log-rank test, it, like the two previous articles [69,23] it referred to, and like the example-less 1964 article [46], focuses on null-hypotheses tests.

An initial approach that does produce a 'fitted' longevity difference is shown in the beige inset in Figure 1. The step function is the Kaplan-Meier-type curve calculated using age as the time axis, the age at induction as the entry time, and the age at death or the last date presumed alive as the exit age. The fitted mean of 76.4 years was obtained as the area under this empirical survival curve. (later we will compare the strengths/weaknesses of age-based versus remaining-life-based metrics)

Comparison curves For each of the 120 inducted players, one can compute a comparison survival curve for the USA males who were born the same year as he and who were alive when he was inducted. For example, for a player inducted in year y at age a, the proportion of the males surviving at year $y+t$ at age $a+t$ was $\exp[-\Lambda(t)]$ where the integrated hazard rate $\Lambda(t)$ was computed as $\sum_{i=1}^{i=t} m_{y+i,a+i}\delta t$, where $dt = 1$ year, and $m_{y+i,a+i}$ is the mortality rate for males aged $a+i$ in year $y+i$ as published by the Human Mortality Database [38]. The mean of the areas under the 120 such curves was 77.0 years.

This comparison points up a subtlety associated with the Kaplan-Meier estimator in general, and with this version in particular. Whereas it is often stated that a person who is censored only contributes information up until (s)he is censored, and is 'removed' from the calculations thereafter, in fact that person 'inherits' the life experience of those who are followed beyond that censoring time point. In effect, their additional longevity is imputed from the experience of those 'to the right' of the censored lifespan [20,25,**?**] Thus, how much longer the players who are still alive (and shown as grey dots at the right of Figure 1) will live is *imputed* from those who have already died beyond these ages. For our population based comparator, we (arbitrarily) subjected who are alive as of the last year for which mortality rates are available (2019) to the rates prevailing in 2019. This gives the comparator group a slight longevity advantage.

An alternative, and more genuinely-matched, approach An alternative approach is to extend the existing methods for 'sample vs. population' contrasts [23,69,22] so that the player-specific contests are matched on both age and calendar time,[4] not just at the *design* stage, but also explicitly in the *analysis*. However, all three of these articles were limited to null-hypothesis tests and p-values. And while they use the (null) proportional hazards model to derive the equivalent of the log rank test, they did not explicitly provide an estimator of the hazard ratio. Given that the–recently reviewed [50]–conditions for translating a hazard ratio into a longevity difference are likely to apply to many longevity contests, we now provide such an estimator. It was already referred to in passing on page 193 of [51], but does not seem to have been applied to any 'individual vs. population' contests. Along the way we fill in the now-largely-lost connection between the Mantel-Haenszel test [45] and the Log Rank test [51], and give a heuristic meaning to the player-specific expected values in the latter.

Mantel-Haenszel test ↔ LogRank test The link is best appreciated by first examining the two selected player-vs.-population longevity contests in Figure 2, in which the lifespans of the populations, ranked from shortest to longest, form survival curves. The initial population size is arbitrarily set to 100,000. For every 10,000th riskset, the *obs*erved number of player-deaths, along with the *expected* number of player deaths under the null hypothesis, is shown. The summary statistics, following the Mantel-Haenszel approach [45], are shown at the bottom. One can check that for each of the two selected players, the sum of the expected values equals the negative of the log of the proportion surviving at the time the player lifespan is censored or ends. Thus, if we denote the surviving fraction of the reference population-fraction by S_0, then a player who dies at age a contributes a score of $1+\log[S_0(a)]$, while a player still alive at age a contributes a score of $\log[S_0(a)]$.

Score statistics The score statistics for all 120 contests are shown in the right-most panel, with scores to the *left* (right) of the null indicating players who *outlived* (were outlived by) a large proportion of their population peers. The mean (sd) score across all 120 scores is -0.03(0.63), only half a standard error from the null.

Expected values and relay-races: a different heuristic For those who like heuristics, there is a different and less well known interpretation of the player-specific expected values, i.e., the null expectations under the assumption that their post entry survival follows the same distribution as the age-matched general population (S_0). [35] This interpretation is scattered throughout the renewal process and repeated events literature [1]. Consider the comparator for a player aged a_0 inducted in year y_0. Under the null, a human 'time-chain' or 'relay' is

[4] In the worked examples in these articles, the calendar period was short enough that this dimension was ignored, and just age-matching was used.

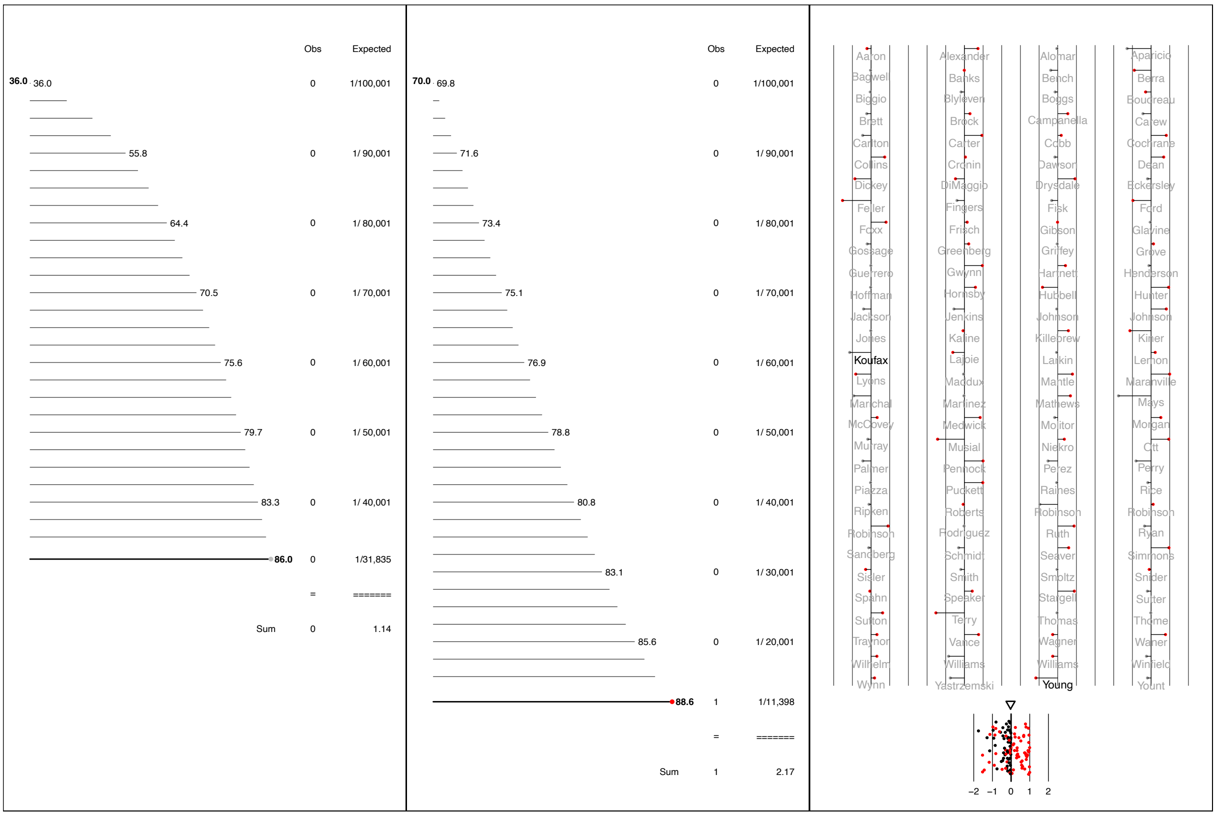


Fig. 2 Links between calculations for Mantel-Haenszel and Log-Rank tests in two selected player-vs.-**population** longevity contests (leftmost and middle panels) and score statistics for all 120 contests (rightmost panel). See text for full details. For each of the two selected players (shown in bold), the sum of the expected values equals the negative of the log of the proportion surviving at the time the player lifespan is censored or ends. Thus, if we denote the surviving fraction of the reference population-fraction by S_0, then a player who dies at age a (red dot) contributes a score of $1 + \log[S(a)]$, while a player still alive at age a (grey dot) contributes a score of $\log[S(a)]$. Under the proportional hazards model the ML estimate of the hazard ratio is the number of deceased players, i.e., 67, divided by the negative of the sum of all of the $\log[S_0(a)]$'s, i.e., 74.5. This same ratio of 0.95 is obtained as the ratio (P/Q) of the double-Mantel-Haenszel sums, i.e. the sums of the numbers shown at the foot of the 'Observed' and 'Expected' columns. For alternative interpretations of the expected values and of the hazard ratios, see text.

begun by randomly selecting a 'starter' from that comparator population, and it continues up until age a. If and when the starter dies (at age a_1 say), he is replaced by a still-living person aged a_1, and so on. [In the words of Edmonds [19], it is 'one person *constantly living*'; in the words of Miettinen, the chain forms a 'dynamic population of constant size 1']. In some instances, the randomly selected starter will be still in the endurance race at age a; in others, there will have been 1 replacement, or 2, or more. The expected value (E) is the expected number of replacements needed to keep the time-chain going until age a, and it can be calculated as the integral of the force of mortality of hazard function over the interval a_0 to a, or by the sum of a large number of small contributions, as shown in Figure 2. And since the sum of many Poisson random variables with small (but increasing over age) expectations is itself a Poisson random variable, the number of replacements can be take as a Poisson random variable with expected value E. Thus, the probability $[S_0(a)]$ that the *starting* person is still alive at (survives to) age a is also the Poisson probability that 0 replacements are required, i.e., $S_0(a) = \exp(-E)$.

The interpretation is even simpler when the cohort of interest is 'extinct', e.g., the passengers who survived the sinking of Titanic (435 of whom have been traced). Across the 435 endurance contests pitting passengers (average age yy) against their age and sex matched population comparators, it look an average of x.x replacements to keep the contests going until the last passenger died.

ML Hazard Ratio Estimator If we denote the (assumed constant over age and calendar year) hazard ratio (HR) by θ, and players by the subscript p, then $S_p(a) = [S_0(a)]^\theta$. From this, we can derive the ML estimator of θ. The log-likelihood contribution [22] from a player with vital status d (0 if alive, 1 if dead) at age a is $d\log[\theta] + \theta\log[S_0(a)]$. Thus, summing over all players, the estimator satisfies the estimating equation

$$\sum\{d/\theta + \log[S_0(a)]\} = 0.$$

This yields the ML estimator $\hat{\theta} = \{\sum d\}/\{\sum -\log[S_0(a)]\}$, i.e., the number of deceased players divided by the negative of the sum of all of the $\log[S_0(a)]$'s. Thus, in this example, the fitted HR is 67/74.5, or 0.95, and its standard error (based on the Fisher information regarding θ, rather than $\log\theta$) is $0.95/\sqrt{67} = 0.12$.

Mantel-Haenszel Hazard Ratio This same value of 0.95 is obtained as the ratio (P/Q) of the double-Mantel-Haenszel sums, $P = \sum P_i$ and $Q = \sum Q_i$, where, for player i, who survives until only a proportion S_i of the reference population are still living, $P_i = \sum_j d_{ij} \times j/(j+1)$, $Q_i = \sum_j 1 \times 1/(j+1)$, and j runs from 10,000 to 10,000S_i. The d_{ij} are the values in the 'Observed' column in Figure 2, so the sum, P_i, of the $d_{ij} \times j/(j+1)$ products, shown at the foot of the 'Observed' column is effectively d=1 or d=0. We have already seen that the sum, Q_i, of the $1 \times 1/(j+1)$ products is $-\log[S(a)]$.

'Translating' the hazard ratio into a difference in years If, as is described in [50], the age-specific mortality rates in the reference populations follow the log-liner in age pattern (discovered by Gompertz) with (say) a slope of 0.1/year of age, then under the PH model, a hazard ratio of 0.95 'translates' to a player longevity advantage of $\log(1/0.95)/10$ or approximately 0.5 years.

A less precise and less exacting 'translation' As is also described in [50], if we are *reluctant* to assume that Gompertz' Law is operating, but *are* willing to assume the PH model is, then a simpler but less assuming interpretation is possible. The hazard ratio of 0.95 can be used to calculate that the probability that a player will *outlive* a randomly selected population peer is 100/(100+95) = 51%, or that the probability that a player will *be outlived by* a randomly selected population peer is 95/(100+95) = 49%.

As one author [63] nicely put it (when dealing with a desired clinical outcome), the greater generality (its nonreliance on equal-slope Gompertz distributions) of the PH-only assumption comes at a cost: "When the hazard ratio is thought of as the odds that a patient will heal faster with treatment, a unitless term not directly reflective of the fundamental time units of the study, it also becomes more evident that the hazard ratio cannot convey information about how much faster this event may occur. The difference between hazard-based and time-based measures is analogous to the odds of winning a race and the margin of victory."

3.2 Local, **within-profession** comparators

To avoid the artificial longevity advantage of same-aged but more obscure baseball players [2], we created 120 within-profession longevity contests where, for each player voted into the Baseball of Fame by the BBWAA, the comparator(s) is(are) the players(s) *nearest in age and votes obtained* among the *unsuccessful candidates* that same year. By default we used a tolerance of 5 years of age, and at least 1/2 the required number of votes. For any inducted player for whom this did not produce any match, we successively relaxed the matching criteria under at least one was found. This procedure yielded a mean of 256/120 = 2.1 matches per inducted player. The top portion of Figure 3 shows 10 such contests. We begin by using the calendar year time scale, and ignore for now any differences in age.

3.2.1 Calendar-year as the time scale

The 3 displayed contests in which all players are still alive are uninformative. Of the 7 informative contests displayed, five involved 1 risk set each, and 2 contests involved 2 risk sets each, so in total there are 9 informative risk sets. These 9 within-contest risk sets can serve as a guide to the various data analysis options, and illustrate the calculations involved

The stratified Mantel-Haenszel or PH-based score test A null hypothesis test can be based on the contest-stratified Mantel Haenszel test or the score statistics from on a stratified proportional hazards model. Hall of Fame member Collins provides a good example of the calculations involved. He outlived the 1st of his 2 comparators, so the first risk set in which he was involved contributes $O_1 = 0$ and $E_1 = 1/3$. He was the loser in the 2nd risk set, so $O_2 = 1$

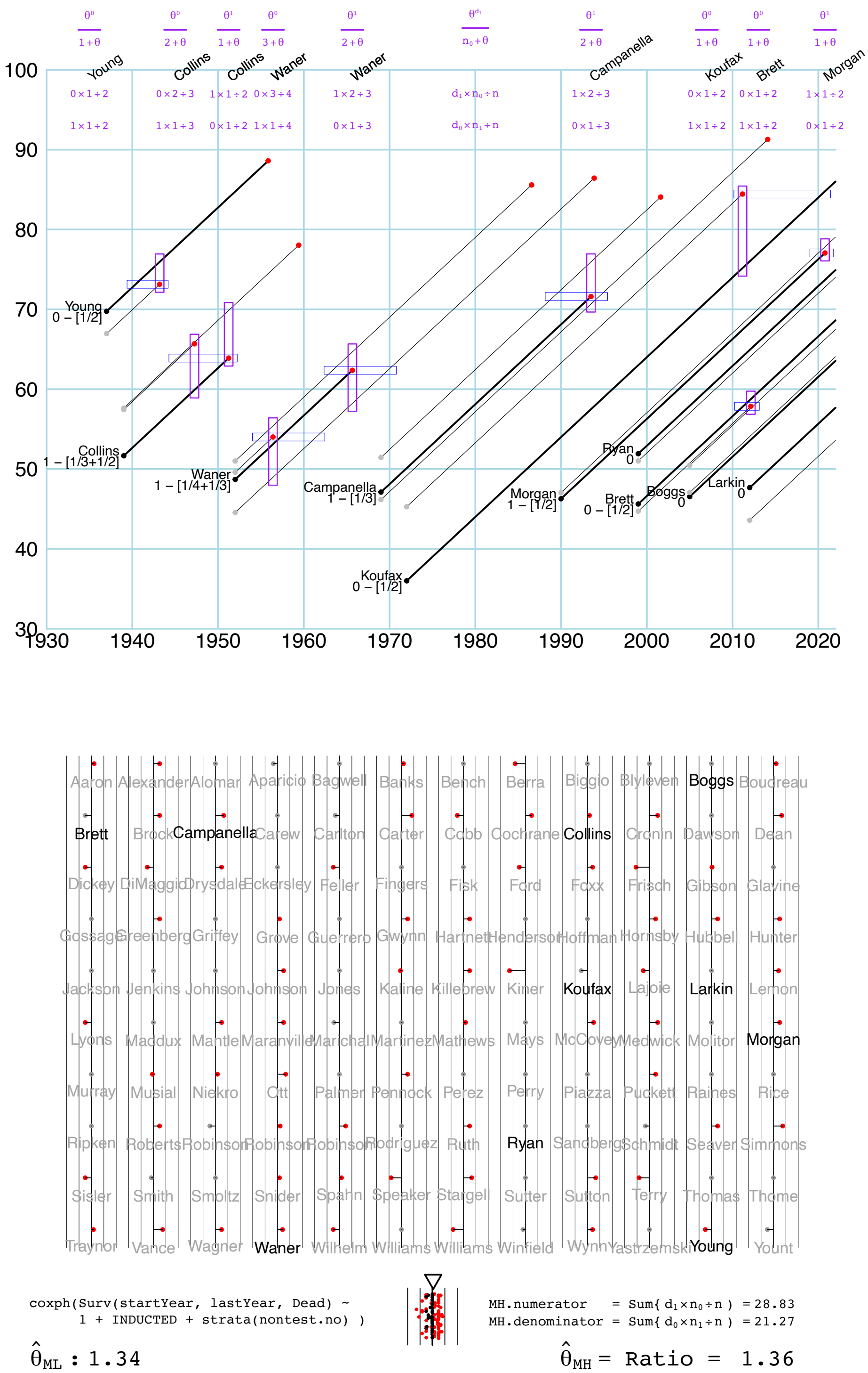


Fig. 3 Within-profession longevity contests using as comparators the players nearest in age and votes obtained among the unsuccessful candidates that same year.
Top: 10 selected contests (7 of them informative); the 9 informative risk sets (purple) associated with these 7 contests, the 9 contributions to the likelihood $L(\theta)$ and to the numerator and denominator of the Mantel-Haenszel summary hazard ratio; and the calculated score statistic, $O - E$ for each player. For example, Collins outlived the 1st of his 2 comparators, so the first risk set in which he was involved contributes $O_1 = 0$ and $E_1 = 1/3$. He was the loser in the 2nd risk set, so $O_2 = 1$ and $E_2 = 1/2$. Thus, overall, $O = 0 + 1 = 1$ and $E_1 = 1/3 + 1/2$, and so his score statistic is $O - E = 1/6$. The 3 contests in which all players are still alive have no associated risksets and are uninformative.
Bottom: score statistics for all 120 contests, as well as estimates of the overall hazard ratio: under the stratified proportional hazards model the ML estimate of the hazard ratio is 1.34; a close approximation is provided by the ratio of the sums of the 113 Mantel-Haenszel numerators $(d_1 n_0/n)$ and the 113 denominators $(d_0 n_1/n)$, namely $28.83/21.27 = 1.36$. Some 43 contests were uninformative, 51 contributed 1 risk set, 18 contributed 2; 6 contributed 3, and 2 contributed 4.

and $E_2 = 1/2$. Thus, overall, $O = 0 + 1 = 1$ and $E_1 = 1/3 + 1/2$, and so his score statistic is $O - E = 1/6$. Over the 77 informative contests (see bottom of Figure 3), the 0's and the E's sum to 51 and 43.4 respectively, i.e., on average the Hall of Fame members die earlier than their comparators. Using Mantel's *hypergeometic-based* variances, "computed conditionally on the separate contingency-table marginal totals," [45] [5] the 1 df. X^2 statistic is $(51 - 43.4)^2/43.4 = 1.31$. Using the *binomial*-based variances, (as `coxph` does) the 1 df. X^2 statistic is $(51 - 43.4)^2/25.25 = 2.26$. As Mantel emphasized right from the outset, which version one uses ($E_j[1 - E_n]/n_j$ or $E_j[1 - E_j]/[n_j - 1]$) becomes important when the average size of a riskset is just 2 or 3, and especially when using matched pairs, where "the variances would have been understated by a factor of 2, had $n - 1$ been replaced by n in the variance formulas" [43]

Fitting the hazard ratio using dedicated software If we assume a common hazard ratio θ over the follow-up in each contest, it can be estimated as 1.34 using the contest-stratified model fitted by the `R` code shown at the bottom left of Figure 3. The z statistic based on the Wald standard error is 1.5. The output reports that it is based on a total of 178 events. However, as is illustrated by the discrepancy between the number of deaths (15) and the number of informative risk sets (9) in the selected contests shown in the top of Figure 3, the SE is based on the latter.

Fitting the hazard ratio using standard GLM software Again Hall of Fame member Collins illustrates the partial likelihood contribution(s) per contest. The 1st is $1/(2+\theta)$ and the 2nd is $\theta/(1+\theta)$. These Bernoulli forms suggest that one can fit this specialized stratified Cox – or conditional logistic – model using standard GLM software. [34] As in the top row of Figure 3, let d_1 indicate, using 1 or 0, whether the risk set with $n_1 = 1$ inducted member, and n_0 comparators, was formed by the death of the Hall of Fame member. Then the $\{d_1\}$ are Bernoulli random variables with expectations $\{\pi = \theta/(n_0 + \theta)\}$, so that $\log[\pi/(1 - \pi)] = \log[\theta/n_0] = \log[\theta] - \log[n_0]$. Thus, $(\log)\theta$ can be fit using standard logistic regression software, for example, in `R`, via the statement `glm(d`$_1 \sim 1 +$ `offset(-log(n`$_0$`)), family=binomial)`.

An 'almost ML' hazard ratio, and a deluxe iteratively reweighted Mantel-Haenszel version As is shown in the bottom right of Figure 3, the hazard ratio estimator $\{\sum d_1 n_0/n\}$ / $\{\sum d_0 n_1/n\}$ provides a close-to-ML hazard ratio estimate. However, as was elegantly shown by Clayton [17,18], it is possible to use an iterative version of this to arrive exactly at the ML estimate. The key is the fact that at the ML value, the following equilibrium obtains

$$\theta_{ML} = \frac{\sum d_1 n_0/(n_0 + \theta_{ML})}{\sum d_0 n_1/(n_0 + \theta_{ML})}.$$

[5] For a not-entirely-modest self-appreciation of this 1966 paper, see here.

Thus, the traditional $\hat{\theta}_{MH}$ can be seen as the version in which each cross product is down-weighted [49] by a factor of $n_0 + 1$ rather than the general factor $n_0 + \theta$ in the ML version. Starting from this first iteration, ratios of iteratively re-weighted cross products quickly converge to the ML ratio.[34]

We switch now to the *age* time scale.

3.2.2 Age as the time scale

For those who regard age as the more natural time scale, the top of Figure 3 also shows in blue the (now 8, rather than 9) risk sets associated with the 10 selected contests (with calendar time as the time scale, Collins was a member of 2 risk sets; in the age scale he is a member of just 1).

Of the 120 contests, 73, involving 101 risk sets, were informative. The reason for slightly smaller numbers when using the age time scale is best illustrated by the contest involving Yogi Berra. When he was elected, at age 46, his 2 comparators were aged 47 and 49 The first of the comparators died very soon afterwards. If we were proceeding forward in *calendar* time, all three would have been members of that first riskset. But, when we use *age*-matching, the second comparator is not a member of it. He is a member of the second (and final) age-matched riskset that is formed when Berra dies at age 90.

Across all the contests. the 0's and the E's sum to 49 and 39.18 respectively, again indicating that on average the Hall of Fame members die earlier than their comparators. Using the Mantel version of the variances, the 1 df. X^2 statistic is $(49 - 39.18)^2/22.61 = 4.26$.

With `startAge` and `lastAge` replacing `startYear` and `lastYear` in the call to `coxph`, the fitted hazard ratio θ over the follow-up ages in each contest, was 1.53. The z statistic based on the Wald standard error is 2.05.

The 'almost ML' hazard ratio based on null weights in the Mantel-Haenszel summary ratio, is $28.08/18.27 = 1.54$, and the re-weighted version matches the ML version to 4 decimal places after just 2 iterations.

3.2.3 Further leveling of the playing field

If we merely synchronize (match) on calendar year, as we did first, the contrasts are not necessarily matched on age. Just as [2] did, it is common in such contexts to abandon the matching and deal with both the matched and the imbalanced variables through multivariable regression models. Even if the models do not lose any efficiency, or introduce bias, they makes the process less transparent: when explaining our methods to lay people, it is less easy to explain modelling than matching. Indeed, just like Fisher argued that in design, we should 'match first and randomize second,' for data-analysis we might adopt a similar 'match first, then model if you must' approach.

Fortunately, with the aid of the stratified Cox model, in longevity comparisons we can have the best of *several* worlds: whichever time scale we select automatically matches on this scale, and means we do not have to model it.

Second, the stratified version allows one or more important variables that were matched on at the design stage to also be matched on, rather than modeled, in the analysis.

The stratification is also easily explained to non-statisticians: If one had to defend one's statistical analysis in a legal case, it would be less open to criticism if, instead of 'mathematically adjusting' for age-sex differences [?], one could point out that one did not globally compared the mathematically 'transmuted' [31] lifespans of (possibly younger) winning actresses with (possibly older) actors in the comparator category. In a Cox model that stratifies on age and sex, one directly compares women with similar-aged women, and men with similar-aged men: the ages and sexes are always *segregated*, and the within strata results are then aggregated by summing their log-likelihood contributions. With the reduced computations involved, the smaller risk sets emphasize how sharp the competitions are, and that they are not artificial mathematically-created contests that depend on additional and unneeded model assumptions.

Third, if despite the stratification, some important covariates are imbalanced, or generate substantial noise, they can be included in the stratified Cox model, or its equivalent, the conditional logistic regression model.

If we use the calendar time scale and correct for age differences As an example, we recall the analysis, in section 3.2.1, which matched on calendar time, but ignored the players' ages. To describe the imbalance, most publications would show the marginal distributions of the ages in the 120 Hall of Famers and the 256 competitors, here 48.8 years and 49.4 years respectively, with the Hall of Famers being 0.60 years younger on average. However, as should now be evident, the extent of the age differences in the 113 informative risk sets is the relevant information. In a typical risk set, the Hall of Famer was 70.48 years, and the competitor(s) 71.20, i.e.., on average, the Hall of Famer was 0.72 years younger than the competitor(s).

Since age-differences within a contest are preserved as we proceed in calendar time, the age imbalances are easily handled by including each player's age-at-the beginning of the contest as a modeled regressor in the stratified Cox model

```
coxph( Surv(startYear, lastYear, Dead) ~
         INDUCTED + Age.at.entry + strata(contest.no) ) .
```

The resulting age-adjusted hazard ratio is 1.38. The direction of the correction is as expected, since the Hall of Famers had an age advantage. its magnitude is also broadly in line with what might calculate from the fitted age coefficient of 0.07, similar that found by Gompertz [?] two centuries years ago. This, together with the age advantage of 0.72 years, would have suggested an approximate upwards correction of approximately $0.07 \times 0.72 = 0.05$, or $100 \times (\exp(0.05) - 1) \approx 5\%$.

If we use the age time scale, and correct for year of birth differences In the analysis in section 3.2.2, in which risksets were formed at the ages at which players died, the hazard ratio of 1.53. However, within each 'horizontal' risk set there was a spread in the years that the players were born. We can take these temporal variations into account by including year of birth in the stratified regression model

```
coxph( Surv(startAge, lastAge, Dead) ~
        INDUCTED + Year.of.Birth + strata(contest.no) ) .
```

The tiny correction did not alter second decimal of the hazard ratio.

4 The Oscars longevity contest: Act II

As we have already noted, the matching in this study was not as straightforward as in Baseball Hall of Fame study. Moreover, even in the database that, in 2005, the authors shared with its readers, it was not easy to reconstruct the matched sets. To fulfill the PLOS requirements for supporting material for the 2022 study, the authors have used the Harvard Dataverse to share their extensive code. But they have told those seeking access to the dataset itself that "you will need to send approval from Institutional Review Board operating under an Office of Human Research Protections; you will also need to be patient with possible additional institutional terms and conditions for materials distribution agreements and processing times for outside requests."

Rather than undertake this extended process, we took advantage of the fact that the selection of comparators was well described in the 2001 and 2022 articles, and that the information can be found in the public domain.[6] Thus, we have used the Internet Movie Data Base (IMDb) to assemble the basic data for what can be argued is the cleanest longevity contrast, namely that between the winner of the Oscar and the same-sex closest-in-age performer in the movie the winner won for. Moreover, in the spirit of PLOS, we are making this dataset available online, without restrictions, to those who wish to check our work or analyze it differently.

We identified 302 unique performers who won Oscars between 1928 and 2018, and were alive when the award was announced. Time 0 was taken to be when the performer won the (first) Oscar.

Figure 4 shows the Kaplan-Meier-type curve calculated using age as the time axis, the age at winning the award as the entry time, and the age at death or the last age presumed alive as the exit age. A fitted mean longevity of 81.5 years was obtained by computing the area under this empirical survival curve.

4.1 **General-population** comparators

Survival curves, test statistics, longevity/force of mortality differences For each of these winners, we computed a comparison survival curve for the USA (fe)males who were born the same year as the winning performer and who were

[6] Soon after the 2001 article appeared, we were denied access to the dataset the article was based on, since was stated to be subject to Ontario's patient privacy laws. Yet, individual individuals' names and their personal information had been disclosed in the article.

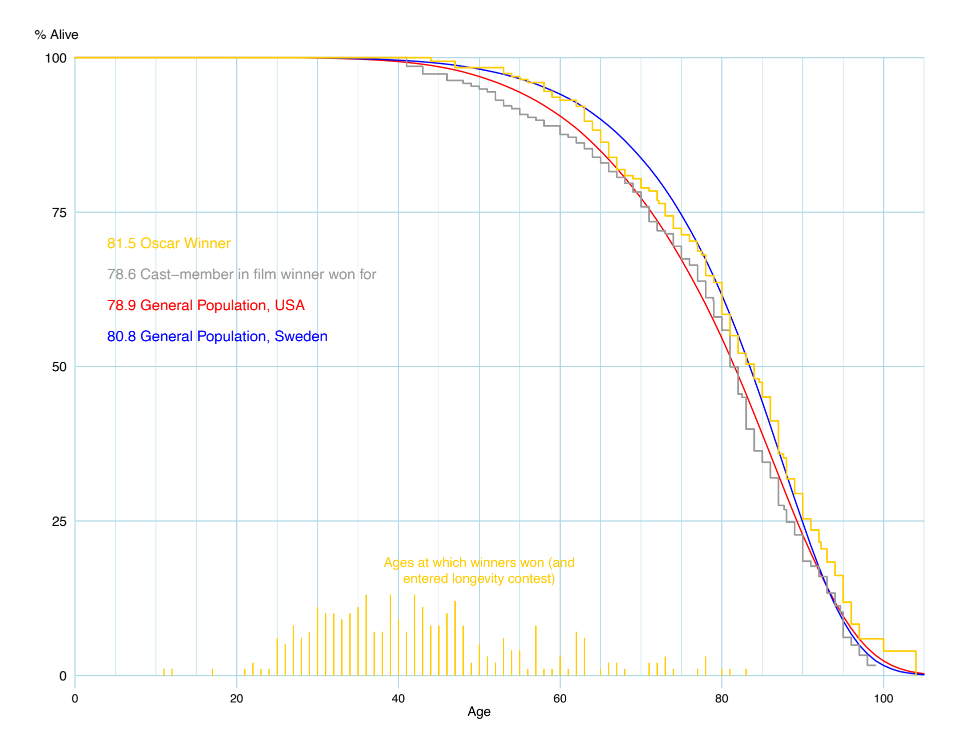


Fig. 4 Longevity contests: Fitted survival curves for Oscar winners vs. sex-age-year-matched national (USA/Sweden) populations and vs. a nearest-match member from the film the winner (first) won for, along with life expectancies (years) calculated as the areas under these curves.

alive when (s)he was inducted. The mean of the areas under the 302 such curves was 78.9 years. Had we used the slightly more upscale population of Sweden as a comparator, the mean would be 80.8 years. Either way, people who win at the Oscars ceremonies *do not live that much longer* that the '*ordinary*' person who watches or listens to the ceremonies remotely. Winning an Oscar does not ensure the further 'jubilees' that some privileged celebrities look forward to when marking their 80th birthdays.

For those who consider the USA as the more appropriate comparator, the log-rank or score statistic is $z = -51.8/204.8^{1/2} = -3.6$. The ML estimate of the hazard ratio is 153/204.78, or 0.75, which would 'translate' to a 2.9 year longer lifespan if one were to a use a Gompertz slope of 0.1/year of age. Its standard error is $0.75/\sqrt{153} = 0.06$, which places it at about 4 SEs from the comparator. Measured against the longer lived Swedish population, the performers' statistics are still 'statistically' significant, but less impressive: $z = -36.2/189.2^{1/2} = -2.6$, and HR point estimate 0.81 with SE 0.07.

4.2 Local, **within-movie** comparators

For each of the 302 within-profession contests, we identified a same-sex closest-in-age performer from the same movie the winner won the (first) Oscar for, and was alive when the winner won. For the reasons explained above, and unlike the original authors, we did not include as a control a performer from a movie that the winner had previously been nominated for but not won for.

The area under this within-movie comparator survival curve was 78.6 years. In this simple and cleaner matched pairs design, only those (183/302) pairs where at least one of the pair has died are informative for the score statistic and the hazard ratio estimator, both of which have very simple forms. In 95 pairs the winner outlived the comparator, and in 88 it was the converse. The score statistic is $X^2 = (95 - 88)^2/183 = 0.27$, and the hazard ratio is $88/95 = 0.93$. The standard error of its log is $(1/88 + 1/95)^{1/2} = 0.15$ and so, using the 95% multiplicative margin of error of 1.34, the 95% limits for the hazard ratio are $0.93 \div 1.34 = 0.69$ to $0.89 \times 1.34 = 1.24$.

The statistical precision could be improved by having more than 1 cast member in each contest. We leave such analyses to interested readers.

Adjustment for age-differences Since it was not possible to identify a same-age same-sex comparator in each contest, the winners started out with a slight age disadvantage (median 41 vs 40; mean 42.6 vs. 42.0). Adjustment for this imbalance moved the hazard ratio to 0.85 (95% CI 0.62 to 1.16). Interestingly, the coefficient for age was 0.1, the same slope/year of age slope we used above. This exchange rate would 'Gompertz-translate' [50] the hazard ratio of 0.85 to a longevity advantage of 1.6 years.

5 Epilogue

Deficiencies in study design, unequal data-quality, and inappropriate statistical analyses continue to produce misleading claims of longevity benefits and harms. This article has emphasized the importance of clean contests and transparent analyses in longevity comparisons, and the tight integration of the design and the analysis. It advocates maintaining the matching in the analysis, or, if not all of the matching can be retained, using a combined 'match then model' approach. It has also brought together a number of connections between what appear to be seemingly quite separate statistical techniques in the 'survival analysis' and 'classical epidemiology' cultures. Lastly, it pays tribute to a 'lifetime' statistician who did much to promote the benefits of locating, visualizing and analyzing longevity data within a Lexis coordinate system.

By examining the structure of the statistics involved in population-based comparisons, we gain valuable statistical insights into the connections between the integrated hazard function, the risk function, the score test, the Cox model, the Mantel-Haenszel procedures, the log-rank test, and the Gompertz model.

From population-based comparisons, we learn that the highest-honored baseball players and actors do not live any longer than the general public who merely watch them perform. Their diminished/extra vitality turns out to be a statistical illusion.

Some 121 years ago, Francis Galton, when helping Karl Pearson launch Biometrika (1901), wrote these introductory words

> This journal, it is hoped, will justify its existence by supplying these requirements either directly or indirectly. I hope moreover that some means may be found, through its efforts, of forming a manuscript library of original data. Experience has shown the advantage of occasionally rediscussing statistical conclusions, by starting from the same documents as their author. I have begun to think that no one ought to publish biometric results, without lodging a well arranged and well bound manuscript copy of his data in some place where it should be accessible, under reasonable restrictions, to those who desire to verify his work.

In line with Galton's wishes, we are making available the data we used in sections 3 and 4 – all of them from public sources. In addition to being used to verify our work, they can also be used to verify the reported "life expectancy" of 76.4 and 76.2 years of the two within-profession control groups in the 'Oscars-longevity:the-sequel'. Do less successful cast members truly have shorter lifespans than everyone else or are they being defamed by 'lifetime' statisticians'?

References

1. Andersen PK, Keiding N. (ed.) (2006). Survival and Event History Analysis. Chichester: Wiley.
2. Abel EL, Kruger ML (2005) The longevity of baseball hall of famers compared to other players. Death Studies, 29:959-963.
3. Aggleton JP, Bland JM, Kentridge RW, Neave NJ. (1994) Handedness and longevity: an archival study of cricketers. BMJ 309:1681-1684.
4. Albertsen PC, Hanley JA, Murphy-Setzko M (1999) Statistical considerations when assessing outcomes following treatment for prostate cancer. J Urology 162:439-444.
5. Atlas, DH (1978). Longevity of orchestra conductors. Forum on Medicine, 1(9):50-51.
6. Barfort S, Klemmensen R, Larsen EG (2020) Longevity returns to political office. Political Science Research and Methods, page 1-7
7. Bauman AE, Steven N Blair SN (2012) Everyone could enjoy the "survival advantage" of elite athletes [Editorial] BMJ 345:e8338
8. Bellis MA, Hennell T, Lushey C, Hughes K, Tocque K, Ashton JR (2007) Elvis to Eminem: quantifying the price of fame through early mortality of European and North American rock and pop stars. J Epidemiol Community Health 61:896-901.
9. Brøndum-Jacobsen P, Nordestgaard BG, Nielsen SF, Benn M (2013) Skin cancer as a marker of sun exposure associates with myocardial infarction, hip fracture and death from any cause. International Journal of Epidemiology 42:1486-1496.
10. Brown GW, Flood MM (1947) Tumbler Mortality. Journal of the American Statistical Association, 42:562-574.
11. Caan W (2006) Let us now praise famous men, and our fathers that begat us. Rapid responses to BMJ 02 December
12. Carrieri MP, Serraino D (2005) Longevity of popes and artists between the 13th and the 19th century. Int J Epidemiol 34:1435-36.
13. Carroll JD. (1979, January 23). Music and age. The New York Times, p. C2.
14. Cerhan JR, Folsom AR, Potter JD, Prineas RJ (1994) Handedness and mortality risk in older women. Am J Epidemiol 140:368-374.
15. Chakradhar S (2019). Former NFL players die at a faster rate than other professional athletes, study finds. STAT. https://www.statnews.com/2019/05/24/former-nfl-players-die-at-faster-rate-study/
16. Clarke PM, Walter SJ, Hayen A, Mallon WJ (2012) Survival of the fittest: retrospective cohort study of the longevity of Olympic medallists in the modern era BMJ 345.
17. Clayton DG. The analysis of prospective studies of disease aetiology. Commun Stat Theory Methods. 1982;11:2129–55.
18. Clayton D, Hills M. Statistical models in epidemiology. New York: Oxford University Press; 1993. p. 178.
19. Edmonds TR (1832). Life tables, founded upon the discovery of a numerical law regulating the existence of every human being, illustrated by a new theory of the causes producing health and longevity. London, James Duncan.
20. Efron, B (1967). The two-sample problem with censored data. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of Berkeley Press, 4, 831-853.
21. Farr W (1843) Fifth Report of the Registrar General 1843
22. Finkelstein DM, Muzikansky A, Schoenfeld DA (2003) Comparing survival of a sample to that of a standard population. J Nat. Cancer Inst 95:1434-1439.
23. Gail MH, Ware JH (1979) Comparing observed life-table data with a knows survival curve in the presence of random censorship. Biometrics 35:385-391.
24. Gelman A (2020) No, I don't believe that claim based on regression discontinuity analysis that . . . https://statmodeling.stat.columbia.edu/2020/07/02/
25. Gerds TA (2016) The Kaplan–Meier theatre. Teaching Statistics 38:45-54.
26. Goldstein JR, Kenneth W. Wachter KW (2006) Relationships between period and cohort life expectancy: Gaps and lags. Population Studies. 60:257-269.
27. Halpern D, Coren S (1988) Do right-handers live longer? Nature 333:213-213.
28. Halpern D, Coren S (1991) Handedness and lifespan. New England Journal of Medicine, p998

29. Hanley JA, Turner E, Bellera C, Teltsch D (2003) How long did their hearts go on? A Titanic study. British Medical Journal, 327: 1457-1458.
30. Harvard Gazette (2019) Baseball players live longer than football players, average men. https://www.hsph.harvard.edu/news/hsph-in-the-news/ baseball-players-live-longer-than-football-players-average-men/ .
31. Hanley JA (2004) Transmuting' women into men: Galton's family data on human stature. The American Statistician 58:237-243..
32. Hanley JA, Carrieri MP, Serraino D (2006) Statistical fallibility and the longevity of popes: William Farr meets Wilhelm Lexis. Int J Epidemiol 35:802-805.
33. Hanley JA, Foster BJ (2014) Avoiding blunders involving 'immortal time' [Education corner] Int J Epidemiol xx:
34. Hanley JA (2018) Individually-matched etiologic studies: classical estimators made new again European Journal of Epidemiology 33:897–907 https://doi.org/10.1007/
35. Hanley JA (2018) Going from the incidence rate (or hazard rate) function to the risk function. Class Notes.
path: http://www.biostat.mcgill.ca/hanley/bios601/SurvivalAnalysis/
filename: incidenceFunctionToRisk2018.pdf
36. Hill AB (1927). Cricket and its relation to the duration of life Lancet 949-950.
37. Hill B (1937) Principles of medical statistics. XIV: Further fallacies and difficulties. Lancet 229:825-827.
38. Human Mortality Database (2022). 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 May 14, 2022).
39. Keiding N (2005) Lexis Diagram. Encyclopedia of Biostatistics, Wiley.
40. Kenny DT (2014) Stairway to hell: life and death in the pop music industry The Conversation. https://theconversation.com/stairway-to-hell-life-and-death-in-the-pop-music-industry-32735
41. Lange T, Keiding N (2014) Skin cancer as a marker of sun exposure: a case of serious immortality bias [Letter to the Editor] Int J Epidemiol 43:
42. Lewis CE (2000) Elegy for fifty-six black female physicians (Editorial). American Journal of Preventive Medicine 19:206.
43. Mantel N, Haenszel W (1959) Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease. JNCI 22:719-748.
44. Mantel N, Byar DP (1974) Evaluation of response-time data involving transient states: an illustration using heart-transplant data. JASA 69:
45. Mantel N (1966) Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer chemotherapy reports 50:163-170.
46. Moses LE (1964) One sample limits for some two-sample tests. J Amer Stat Association 59:645-651.
47. Nguyen VT, Zafonte RD, Kponee-Shovein KZ (2019) All-cause and cause-specific mortality among Major League Baseball players. JAMA Intern Med. 179:1298-1301.
48. Olshansky SJ (2011) Aging of US Presidents. JAMA 306:2328-2329.
49. Hanley JA. (2011) [minute 17 of A Conversation With Olli Miettinen https://www.youtube.com/watch?v=FYKESrqGnlU (Epidemiology ('Voices' series) 23(3):495-499, May 2012)
50. Pang M, Hanley JA (2021) "Translating" All-Cause Mortality Rate Ratios or Hazard Ratios to Age-, Longevity-, and Probability-Based Measures. American Journal of Epidemiology 190:2664-2670.
51. Peto R, Peto J (1972) Asymptotically efficient rank invariant test procedures. Journal of the Royal Statistical Society Series A 135:185-207.
52. Rablen MD, Oswald AJ (2008) The Nobel Prize as an Experiment into the Effect of Status upon Longevity. Journal of Health Economics 27:1462-1471.
53. Redelmeier DA, Singh SM (2001). Survival in Academy Award-winning actors and actresses. Ann Intern Med 134:955-962.
54. Redelmeier DA, Singh SM (2022). Long-term mortality of academy award winning actors and actresses. PLOS ONE Published: April 13, 2022 https://doi.org/10.1371/journal.pone.0266563
55. Redelmeier D (2004) Death rates of medical school class presidents. Soc Sci Med.

56. Rothman KJ (1991) Left-Handedness and Life Expectancy. N Engl J Med 325:1041-1041.
57. Rothman KJ (1992) Longevity of Jazz Musicians: Flawed Analysis. Amer J Public Health. 82:761.
58. Wolf PA, d'Agostino RB, Cobb J (1991) Left-Handedness and Life Expectancy. N Engl J Med 325:1042-1042.
59. Smith G (2011) The Baseball Hall of Fame is not the kiss of death. Death Studies 35: 949;955.
60. Sommers PM (2003). The Writing on 'The Wall.' Chance, 16:35-38.
61. Song J, Mailick MR, Greenberg JS, Floyd FJ (2019) Mortality in parents after the death of a child. Soc Sci Med. 239:1-18
62. Spencer FJ (1991) Premature death in jazz musicians: fact or fiction? Am J Public Health 81:804-805.
63. Spruance SL, Reid JE, Grace M (2004) Hazard ratio in clinical trials. Antimicrob Agents Chemother 48:2787-2792.
64. Sylvestre MP, Huszti E, Hanley JA. (2006). Do Oscar winners live longer than less successful peers? A reanalysis of the evidence. Annals Int. Med.
65. van der Wall EE (2016) A longer life at the top of Mount Olympus. Netherlands Heart Journal 24:561-562.
66. van Walraven C, Davis D, Forster AJ, Wells GA (2004) Time-dependent bias common in survival analyses in leading clinical journals J Clin Epidemiol 57:672-682.
67. Wainer H (1999) The Most Dangerous Profession: a note on non-sampling error. Psychological methods 3:250-256.
68. Wikipedia (2022) The Conversation. https://en.wikipedia.org/wiki/The_Conversation_(website)
69. Woolson RF (1981) Ranks tests and a one-sample logrank test for comparing observed survival data to a standard population. Biometrics 37:687-696.
70. Wolkewitz et al. Am. Statistician 2010
71. Wolkewitz M, Allignol A 1, Graves N, Adrian G Barnett AG (2011) Is 27 really a dangerous age for famous musicians? Retrospective cohort study. BMJ 343:d7799 doi: 10.1136/bmj.d7799
72. Wikipedia (2022) The Conversation. https://en.wikipedia.org/wiki/The_Conversation_(website)
73. Wolkewitz et al. Am. Statistician 2010
74. Wolkewitz M, Allignol A 1, Graves N, Adrian G Barnett AG (2011) Is 27 really a dangerous age for famous musicians? Retrospective cohort study. BMJ 343:d7799 doi: 10.1136/bmj.d7799
75. Han et al. Applied Statistics 2011.
76. Vietnam Wall
77. Yee (2004) Diabetic Medicine 21:962-67.
78. G. M. Zharinova GM, Anisimov VN (2014) Music and Longevity. Advances in Gerontology, 4:283-289.
79. Zwiers R, Zantvoord FWA, Engelaer FM, van Bodegam D, van der Ouderaa FJG, Westendorp RGJ (2012) Mortality in former Olympic athletes: retrospective cohort analysis. BMJ2012:345:e7456