

- ⁶ Hak E, Verheij TJM, Grobbee DE, Nichol KL, Hoes AW. Confounding by indication in non-experimental evaluation of vaccine effectiveness: the example of prevention of influenza complications. *J Epidemiol Community Health* 2002;**56**:951–55.
- ⁷ Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000;**342**:1887–92.
- ⁸ McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. *Pharmacoevidiol Drug Saf* 2003;**12**:551–58.
- ⁹ Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol* 2003;**158**:280–87.
- ¹⁰ Jefferson T, Rivetti D, Rivetti A, Rudin M, Di Pietrantonj C, Demicheli V. Efficacy and effectiveness of influenza vaccines in elderly people: a systematic review. *Lancet* 2005;**366**:1165–74.

doi:10.1093/ije/dyl068

Advance Access publication 17 April 2006

Statistical fallibility and the longevity of popes: William Farr meets Wilhelm Lexis

From JAMES A HANLEY,^{1*} MARIA PATRIZIA CARRIERI² and DIEGO SERRAINO³

We write to follow up on the editorial¹ on the use of statistical aggregates. We focus on the reaction, in it, to the letter from two of us (MPC and DS) in the same issue suggesting that the 'longer (average) life expectancy of popes relative to artists² reflects the lives characterized by social instability, high-risk behaviours and geographical mobility (and thus infection risk) of the artists.' The Editorial presented evidence that the 'assumption of better behaviour by the popes is perhaps unjustified'. We have now looked behind the summary longevity statistics, and present individualized data showing that the initial, but tentative, conclusion about their longer life expectancy should be reversed. Even if the assumption of better behaviour 'on average' of popes is justified, these behaviours were not—even on average—accompanied by longer life expectancy.

We now show the longevity data in Figure 1, using the Lexis diagram,³ a device sadly neglected by modern epidemiologists. It plots advancing age on the vertical, and calendar time on the horizontal, axis. The 'popes-years' (i.e. *after* they were elected) are shown in black and the artist-years (*all* of them) in grey. In the initial report, the statistics, aggregated over centuries, suggested that popes had a longevity *advantage* of several years. However, if we proceed papacy by papacy, the inter-ocular traumatic test⁴ (IOTT)—another under-used analytic device—applied to Figure 1 reveals that *among those who were alive at the age at which each papacy commenced*, the average remaining life of the popes was *shorter* than that of the corresponding peer artists—at least up until 1750 or so, after which the distributions became more similar.

The principal cause of this reversal is the phenomenon that the first analysis of this dataset sought to remove, namely that 'Popes had to have reached a certain age before being elected to the papacy'. In that analysis, the statistical approach did not

fully address this constraint. Ideally, for each papacy-specific 'longevity competition', the time-clock should start when the pope is elected, and the competition should include the pope, and those artists born the same year as he, who were still alive when he was elected. However, for several papacies, such detailed matching is not possible. Instead, for each of the 1200–1599 papacies, the previous analysis effectively 'started the clock' at age 39—the age at which the *youngest* pope in that era was elected—by excluding artists who died before reaching that age. For the 1600–1900 papacies, it was started at age 38.

Unfortunately, under this broad scheme, as is clear from Figure 1, several artists included in that analysis died before 'their' (and several other) pope(s) were even elected. This inbuilt survival advantage^{5,6} for the popes is an example of what is today called 'immortal time bias'.⁷ William Farr described this fallacy in 1843.⁸ He noted that the average age at death of bishops is greater than that of curates, and thus—concerned for the underprivileged—suggested that curates should be promoted to bishops, and at an early age, 'for the sake of their health.'

Rather than match perfectly on year of birth and age at entry to *each* longevity competition, one could for example proceed half century by half-century, and determine the youngest age (A_{\min}) at which a pope born (or elected) in that half-century was elected, and compare the post- A_{\min} survival of these popes and the corresponding artists. However, these half-century (or even narrower) strata would still contain at least one other pope elected at an age older than A_{\min} , after several artists would already have died, and so the competition would continue to be unfair.

In our new analysis, we circumvented this by creating a separate contest (stratum) for each papacy. We started the clock at the age at which the specific pope was elected. We used as a comparison group those artists, born within 25 years of when the pope was, who had reached that same age. For example, in Figure 2, consider the papacy that began at 1335, when the pope, born in 1280, was 55. Five 'nearby' artists, born in 1260, 1266, 1280, 1284, and 1290, all of whom lived until at least 55, serve as a comparison group. The pope died in 1342, at age 62, after 7 years as pope. His five 'peers' died in

¹ Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada.

² INSERM U379/ORS PACA, Marseilles, France.

³ Unit of Epidemiology and Biostatistics, IRCCS CRO Aviano, Italy.

* Corresponding author. Department of Epidemiology, Biostatistics and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal, PQ, Canada H3A 1A2. E-mail: James.Hanley@McGill.ca

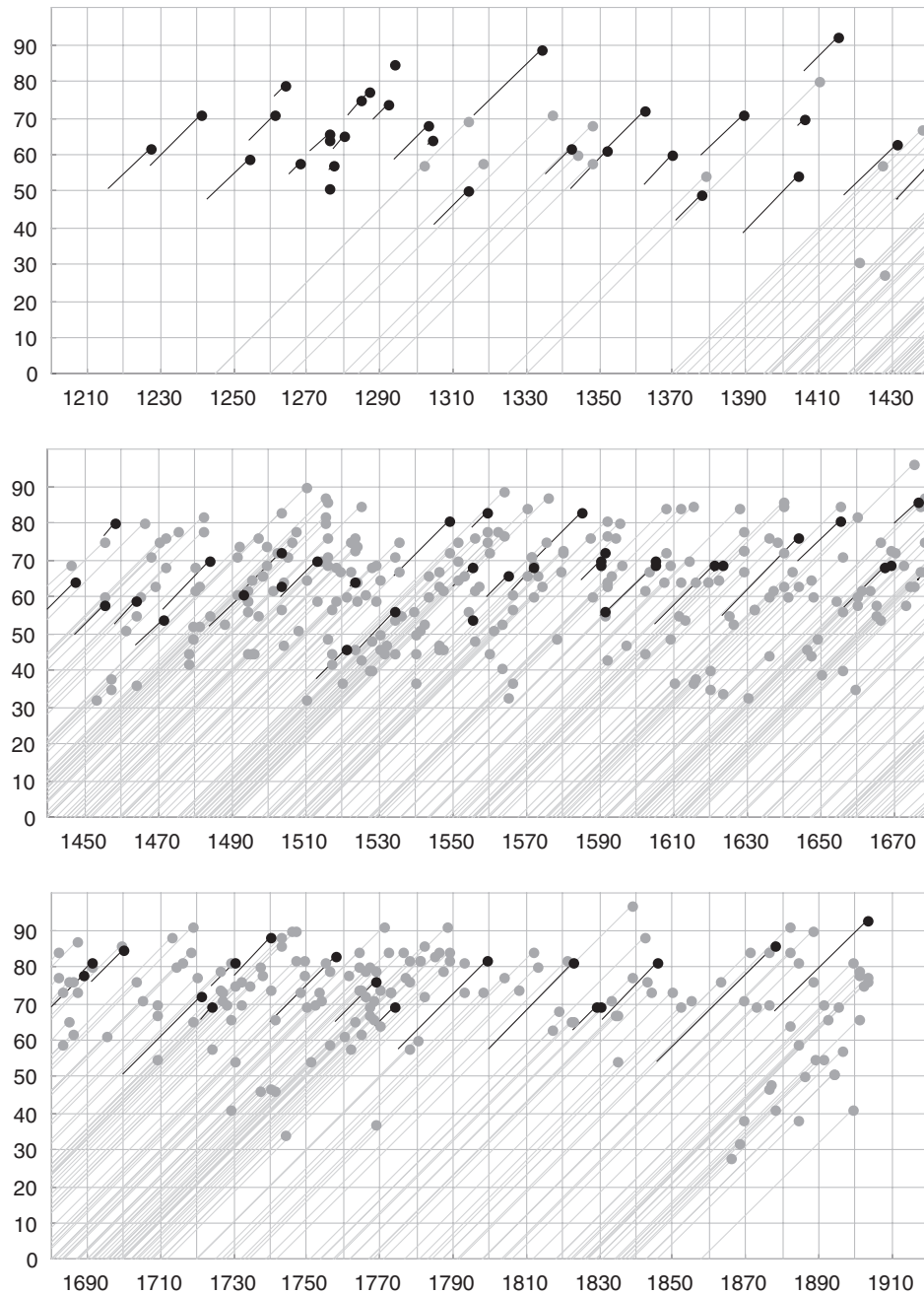


Figure 1 Lexis diagram, with age on vertical, and calendar time on the horizontal, axis. Pope-years (i.e. those post election) shown as black lines and artist-years as grey lines. Age of/year at death indicated by circle

1318, 1337, 1348, 1344 and 1348, respectively, at ages 58, 71, 68, 60, and 58. Thus, their 'post-55' survival times were 3, 16, 13, 5, and 3 years, respectively, so that two lived longer than the pope, by +9 and +6 years, and three lived a shorter amount, i.e. the (artist minus pope) differences were -4, -2 and -4 years. In this approach, some artists serve in several comparisons: for example the artist who lived from 1280 to 1348 competes again in the next papacy, but against a younger pope. One can correct for this 're-use' of some artists, by using robust standard errors, from say a GEE analysis.

Figure 3 plots the (artist minus pope) differences. There are too few artists to serve as comparators for 13th century papacies. From the 14th up until the 18th century, the IOTT confirms that the artists tended to outlive the popes. We heed the editorial warning about the dangers of aggregation (in this case, over *time* rather than people), and thus refrain from giving an overall average; we merely note that the average difference in Figure 3 is positive—statistically so, even when we correct for the 're-use' of some artists. Readers who are not convinced by IOTT's, and who insist on translating the differences into

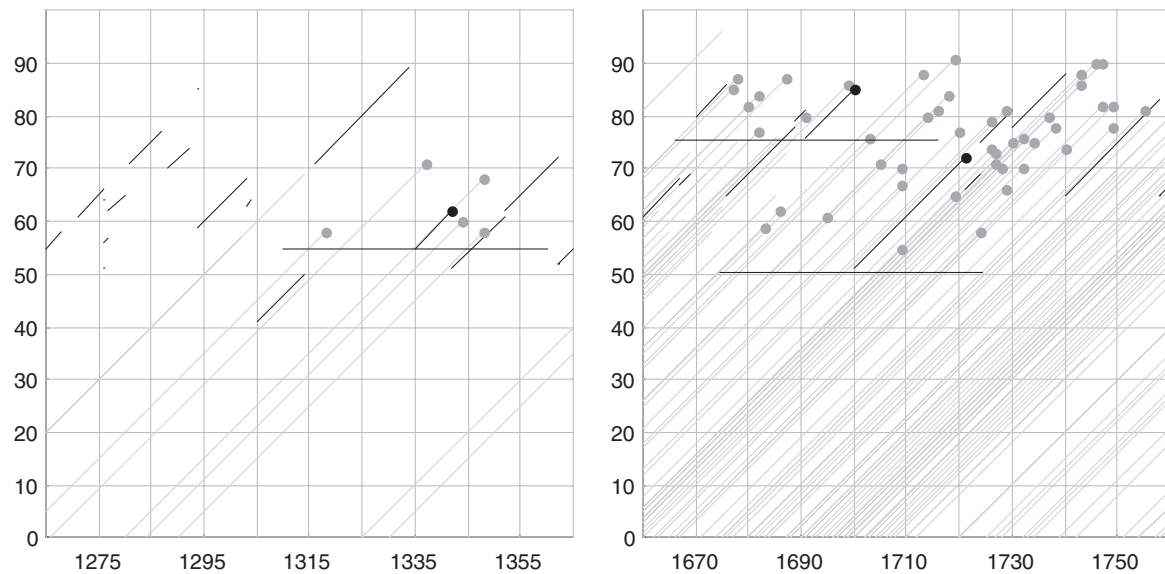


Figure 2 Mini-cohorts based on specific papacies. Left: papacy that began in 1335, when the Pope, born in 1280, was 55, along with (diagonally above and to right of horizontal black line) the five artists, born within 25 years of when the Pope's birth, who reached that same age. Right: papacies that began at 1691 at age 76, and 1700, at age 51

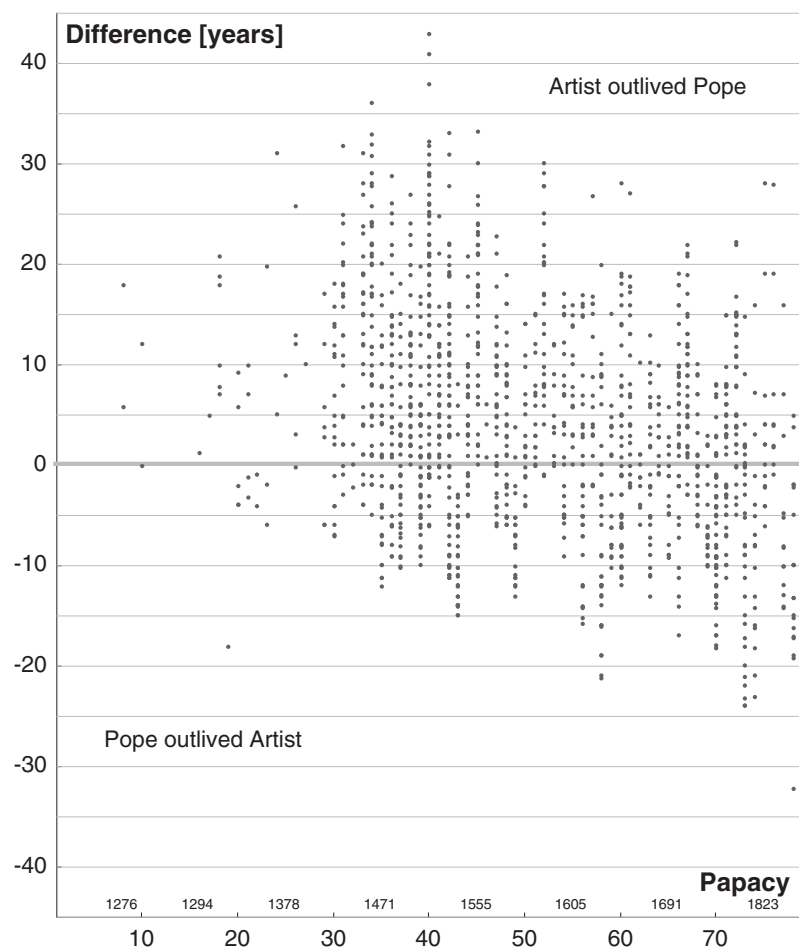


Figure 3 The numbers of years by which artists, who had reached the same age as the Pope was when elected, outlived (positive differences, vertical axis), or were outlived by (negative differences), the Pope

cold hazard ratios (HRs), are asked to estimate the overall and the calendar-time-specific HRs by eye. In the earlier analysis, with popes as the reference category, and observations censored at age 70, the HR for artists was 1.50; here, the average HR is decidedly <1 . The ratio varies considerably: it was ~ 0.3 in 1300 and 1.0 in 1800. Narrower windows give similar results.

Like the Editor, we too believe that Michelangelo would have recognized the dangers of 'assuming that certain aggregate groups, such as popes, were free from the vices of other aggregate groups, such as artists' or—in this instance—of aggregating over time. Michelangelo would also have liked Aaron Levenstein's (<http://politicalgraveyard.com/>) quip about statistics: 'what they reveal is suggestive, but what they conceal is vital.' He might also have admired Lexis' 2-D epi-geometry, and Farr's illustration of statistical fallibility ('bias', nowadays), in relation to (im)mortality and religious careers.

For some situations, even the sharpest and best-designed statistical analysis may fail to uncover the truth. Just as 'confounding by indication' is a near-impossible challenge in non-experimental studies of drug efficacy, a similar phenomenon may have been at play here. It is possible that in some periods cardinals prefer to choose healthier or less healthy popes (depending on political circumstances) to try to influence how long they will be in power. We do not statistically investigate the existence of such 'guided' individualized

choices, preferring instead to let the data in Figures 1 (after Lexis) and 2 and 3 (dedicated to Farr) speak for themselves.

References

- ¹ Davey Smith G. Equal, but different? Ecological, individual and instrumental approaches to understanding determinants of health. *Int J Epidemiol* 2005;**34**:1179–80.
- ² Carrieri MP, Serraino D. Longevity of popes and artists between the 13th and the 19th century. *Int J Epidemiol* 2005;**34**:1435–36.
- ³ Clayton D, Hills M. *Statistical Models in Epidemiology*. Oxford; New York: Oxford University Press, 1993. Chapter 6.
- ⁴ Edwards W, Lindman H, Savage LJ. Bayesian statistical inference for psychological research. *Psychological Review* 1963;**70**:193–242.
- ⁵ Colton T. *Statistics in Medicine*. Boston: Little, Brown and Co., 1974.
- ⁶ Hill AB. *A Short Textbook of Medical Statistics*. London: Hodder & Stoughton, 1977, p. 278.
- ⁷ Walker AM. *Observation and Inference: An Introduction to the Methods of Epidemiology*. Chestnut Hill, MA: Epidemiology Resources Inc., 1991, pp. 59, 60, 161.
- ⁸ Farr W. *Vital Statistics: A Memorial Volume of Selections from the Reports and Writings of William Farr/with an Introduction by Mervyn Susser and Abraham Adelstein*. Metuchen, N.J.: Scarecrow Press, 1975, p. 458.

doi:10.1093/ije/dyl036

Advance Access publication 16 March 2006