Do People Whose Names Begin with "D" Really Die Young?

Gary Smith

Department of Economics

Pomona College

Correspondence:

Gary Smith

Fletcher Jones Professor

Pomona College

425 N. College Way

Claremont, CA 91711

phone: 909-607-3135; fax: 909-6231-8576

e-mail: gsmith@pomona.edu

Do People Whose Names Begin with "D" Really Die Young?

It has been reported that professional baseball players whose first names begin with the letter D tend to die relatively young (Abel & Kruger, 2010). However, the statistical evidence for this claim is based on selective data and a statistical test that ignores important confounding influences. A valid test applied to more comprehensive data from the same source does not show a statistically significant relationship between initials and longevity. In addition, data for the years 1960 through 2004 for 6.7 million White, non-Hispanic California decedents do not replicate the claim that D's die young.

Do D's Really Die Young?

Several studies (Harari & McDavid, 1973; Levine & Willis, 1994; McDavid & Harari, 1966; Savage & Wells, 1948) have concluded that people with unpopular first names are perceived by themselves and by others as inferior to people with popular names. If health is related to self esteem (McGee & Williams, 2000; Trzesniewsky et al., 2006), then life expectancy may be lower for people with unusual first names. However, Pinzur and Smith (2009) found that there was no relationship between name popularity and life expectancy for Californians who died between 1960 and 2004.

There is presumably even less reason to think that life expectancy might be related to initials, because these are generally not spoken or written. Nonetheless, Christenfeld, Phillips, and Glynn (1999) analyzed California mortality data for the years 1969 through 1995 and concluded that, in comparison to people who names have neutral three-letter initials (such as GNS or EYH), males with positive initials (such as ACE or GOD) lived, on average, 4.5 more years; males with negative initials (such as ASS or BAD) lived 2.8 fewer years; females with positive initials lived 3.4 more years; and that there was no difference for females with negative initials. However, their primary analysis grouped decedents by year of death, which can be misleading if initial frequencies change over time. When California decedents were instead grouped by year of birth, there was no substantial or statistically persuasive evidence of a relationship between initials and longevity, either for the original study period 1969-1995 or for a longer period 1905-2003 (Morrison & Smith, 2005).

Nelson and Simmons (2007) examined M.B.A. students who graduated from a large university during the years 1990-2004 and found that students whose first or last names began

with the letters C or D tended to have lower grade-point averages (GPAs) than did students whose first or last names began with the letters A, B, or E-Z. However, the difference in average GPA for C and D students compared to A, B, and E-Z students was only 0.02 on a 4-point scale.

Abel and Kruger (2010) analyzed mortality data for samples of professional athletes, physicians, and lawyers and concluded that people whose first names begin with the letter D died younger, on average, than did people whose first name begins with the letters E - Z. Their most persuasive statistical evidence was for a relatively large data base of Major League Baseball (MLB) players. They reported that MLB players whose first names began with the letter D died, on average, 1.7 years younger than did players whose first names began with the letters E - Z. Despite the title of their paper, they did not find a statistically significant relationship for physicians and lawyers, or for sports other than baseball.

Abel and Kruger (2010) wrote that professional athletes may be particularly disturbed by their initials because they are "notoriously superstitious" (p. 74). However, the sources they cited to support this assertion refer to superstitious *behaviors* such as wearing certain clothing or following pre-game rituals that have been associated with success in the past, for example, wearing the same (unwashed) socks to every game or touching first base when they enter the field (Mandelbaum, 2004; McClearn, 2004; see also: Buhrmann & Zaugg, 1981; Neil, Anderson, & Sheppard, 1981). Wearing unwashed socks is qualitatively different from feeling badly about the first letter of your name.

The theoretical case for focusing on the letter D is not persuasive. Abel and Kruger (2010) argued that

D is not mentioned as one of the ABC's and is regarded as almost a failure….

> Identifying someone as a D student means that that student has consistently fallen below expectations in academic accomplishment…. We hypothesized that, collectively, individuals whose first names begin with the letter D do not live as long as people whose first names begin collectively with other letters of the alphabet (p. 73).

However, almost failing is not as bad as actually failing, which is generally denoted by an F. The fact that D is not one of the ABC's seems scant justification for why D should reduce life expectancy by 1.7 years. There are 22 other letters that are not part of the ABCs, many of which might also appear to be "bad" letters. For example, Pepsi-Cola had a blind taste test in which people tasted Pepsi from a glass marked M and Coca-Cola from a glass marked Q ("Coke-Pepsi Slugfest," 1976). More than half of the people preferred Pepsi. Coca-Cola then ran its own test, letting people drink Coke from a glass marked M and Coke from a glass marked Q, They found that most people preferred Coke from the glass labeled M. This prompted their advertising headline: "The Day Coca-Cola Beat Coca-Cola." Evidently, Q is an unattractive letter.

It might also be argued that if uncommon names affect self-esteem, then names that begin with uncommon letters like F, K, Q, X, and U may have detrimental effects. More generally, among 26 letters, we can inevitably expect to turn up some statistical relationships, particularly if we analyze various subsets and permutations of the data, and then think of possible explanations for these coincidental relationships.

The GPA study by Nelson and Simmons (2007) used the five categories A, B, C, D, and E-Z and considered each person's first and last initials, excluding people with conflicting initials. Thus Aaron Jones, James Allen and Aaron Allen were each put into the A category, while Aaron

Davis and David Allen were excluded from their analysis. The only persons put in the E-Z category were people with names like Ethan Fairchild and George Harvey whose first and last initials were both E-Z.

Although they cited Nelson and Simmons (2007) as motivating their study, Abel and Kruger (2010) only looked at first names. For their analysis, Aaron Jones was A, James Allen was E-Z, Aaron Davis was A, and David Allen was D. They also restricted their analysis to persons who were born between 1875 and 1930 and were at least 25 years old when they died. No justification was given for restricting the analysis to first names and to this particular time period,

The present study reexamined the data and tests that Abel and Kruger (2010) used to support their provocative claim. I also attempted to replicate their results with a much larger data base. My first analysis used the same MLB data base used by Abel and Kruger (2010), restricting the analysis, as they did, to the first initials of players born between 1875 and 1930. Where I differed from their approach was in grouping the decedents by birth year. My second analysis removed each of the artificial restrictions and looked at the first and last initials of all deceased MLB players. My third analysis looked at the initials of male and female California decedents who died between 1960 and 2004.

Method

Measures

Following Abel and Kruger (2010), Sean Lahman's Baseball Archive (2010) was used to collect data on the year of birth, year of death, first name, and last name of all deceased MLB players. There were some ambiguities in the identification of initials because many baseball

players are known by names that differ from the names they were given at birth, for example, Bob Gibson (Robert Louis Gibson), Nolan Ryan (Lynn Nolan Ryan, Jr.), and Babe Ruth (George Herman Ruth, Jr.). Here, I followed Abel and Kruger and used the initials of the names the players were known by, in these three example, BG, NR, and BR.

In an attempt to replicate the MLB study, I also looked at the California Department of Health Services (1960-2004) mortality data base that identifies each decedent's name, gender, date of birth, date of death, and race or ethnicity. They also have a mortality data base for 1905 to 1959 that identifies name, gender, date of birth, and date of death—but not race or ethnicity. In practice, these early years have few usable data because the recorded date of birth is usually "unknown." Because mortality varies by race and gender, I followed common practice (Christenfeld, Phillips, & Glynn, 1999) by looking at White, non-Hispanic decedents and separating the decedents by gender.

Procedure

Retrospective studies have many possible pitfalls. Here, one serious problem is that initials frequencies change over time. Suppose, for example, that mortality rates are constant and do not depend on initials. If there happen to be more people with D initials in later birth cohorts than in earlier cohorts, this will reduce the average age at death (AAD) of people with D initials. To use an extreme example, if D initials have only been used in the past 30 years, then all decedents with D initials died before the age of 30. Another confounding factor is that life expectancies have changed over time. If D initials happened to be more common relative to E-Z initials in the 19th century than in the 20th century, this will reduce the average life expectancy of people with D initials relative to people with E-Z initials.

These problems can be circumvented by grouping decedents by birth year. Morrison and Smith (2005) showed that grouping decedents by birth year provides a valid test of the null hypothesis that mortality rates are the same for different groups of decedents. Specifically, if two groups have the same mortality rates, then the expected value of the AAD over any horizon will be the same for both groups. For example, if we look at two groups with the same mortality rates who were born in 1900, the expected value of the AAD increases as we expand the horizon from, say, 25 to 50 to 100 years; however, the expected value of the AAD is the same for both groups whether we look at horizons of 25, 50, or 100 years.

Following Abel and Kruger (2010), I looked at people who were at least 25 years old when they died and compared D initials with E-Z initials because this was the only statistically significant comparison that they found.

<div align="center">Results and Discussion</div>

For each birth year, the average age at death was calculated for decedents with D initials (AAD_t^D) and for decedents with E-Z initials (AAD_t^{E-Z}). Following Morrison and Smith (2005), if there were at least 5 decedents in each category, then the paired difference in the average age at death was calculated for that birth year:

$$D_t = AAD_t^D - AAD_t^{E-Z}$$

For example, for MLB players born in 1900, the average age at death was 63.78 years for those with D first initials and 71.04 years for those with E-Z first initials, a difference of -7.26 years.

The null hypothesis is that mortality rates are not related to initials, so that the expected value of each paired difference is zero: $E[D_t] = 0$. Two-tailed p values were calculated using: (a) a matched-pair t test of the null hypothesis that the population mean of the paired differences is

zero; and (b) a nonparametric Wilcoxon signed-rank test of the null hypothesis that the

population median of the paired differences is zero.

Table 1 shows the results. For MLB players, when the data were restricted to the first

names of players who were born between 1875 and 1930, the p value was less than 0.05 for the

Wilcoxon test (W = 1.97, p = 0.049), but not for the matched-pair test (t(37) = 1.41, p = 0.166).

When either of these artificial restrictions was removed, the p values were consistently above

0.05. For MLB players with 1875-1930 birth years using first or last initials, W = 1.62, p = 0.106

and t(52) = 1.21, p = 0.231. For MLB players with all birth years using only first initials, W =

1.47, p = 0.141 and t(50) = 0.64, p = 0.528. The p values are much larger than 0.05 when both

restrictions are removed. For MLB players with all birth years using first or last initials, W =

0.76, p = 0.450 and t(72) = 0.41, p = 0.684.

For the California data, the p values were less than 0.05 for males, using either the first

initial alone (t(116) = 3.24, p = 0.0016; W = 3.92, p = 0.000088) or the first or last initial (t(117)

= 2.82, p = 0.0056; W =2.56, p = 0.0105), but in each case men with D initials lived slightly

longer on average than did men with E-Z initials. The mean and median differences were small

and the relations were not in the predicted direction. For California female decedents, the mean

and median differences were inconsequential and the p values were above 0.05 for first initial

alone (t(115) = 1.91, p = 0.059; W = 1.67, p = 0.095) and for first or last initial (t(116) = 0.67, p

= 0.502; W = 0.20, p = 0.842).

Evidently, the report that MLB players whose names began with the letter D tended to die

relatively young depends on two artificial restrictions: using first names rather than first and last

names and only considering players who were born between 1875 and 1930. When either of

these artificial restrictions was removed, there was no longer a statistically significant difference in the average age at death of Major League Baseball players whose names begin with the letter D. In addition, data for the years 1960 through 2004 for 6.7 million White, non-Hispanic California decedents did not show any substantial or statistically persuasive evidence that people with D initials die young. These results are consistent with the presumption that initials have little effect on life expectancy and that findings to the contrary may be due to a selective choice of subjects, time periods, and initials.

It was surprising to find statistically significant differences in the life expectancies of men with D initials in the large California data base, although they tended to live longer, which contradicted the theory that D's die sooner. Nonetheless, the observed differences were quite small, averaging approximately one month, and we cannot infer causality from these statistical correlations. If there are initials-related differences in life expectancies, beyond mere coincidence, these may be due to people with different socioeconomic backgrounds and/or parenting styles related to diet and hygiene having different initials frequencies.

An investigation of this possibility would require socioeconomic and parental data that are not contained in the MLB or California data bases used in this study. On the other hand, if initials are a proxy for socioeconomic status and other fundamental factors that influence life expectancy, initials are evidently such a noisy proxy that the statistical relationship between initials and life expectancy is too small to be of practical interest. It is probably more productive to investigate the fundamental factors rather than a noisy proxy.

References

Abel, E. L., & Kruger, M. L. (2010). Athletes, doctors, and lawyers with first names beginning with "D" die sooner. *Death Studies*, *34*, 71-81.

Buhrmann, H. G., & Zaugg, M. K. (1981). Superstitions among basketball players, *Journal of Sports Behavior*, *4*, 163-173.

California Department of Health Services. (1960-2004). *Death statistical master files*. Sacramento, CA: Author.

Christenfeld, N., Phillips, D., & Glynn, L. (1999). What's in a name: mortality and the power of symbols. *Journal of Psychosomatic Research*, *47*, 241-254.

Coke-Pepsi Slugfest, July 26, 1976. *Time*, 64-65.

Erwin, P. (1993). First names and perceptions of physical attractiveness. *The Journal of Psychology*, *127*, 625-631.

Harari, H., & McDavid, J. (1973). Name stereotypes and teachers' expectations. *Journal of Educational Psychology*, *65*, 222-225.

Levine, M., & Willis, F. (1994). Public reactions to unusual names. *The Journal of Social Psychology*, *134*, 561-568.

Mandelbaum, M. (2004). *The meaning of sports*. New York: PublicAffairs.

McClearn, D. G. (2004). Interest in sports and belief in sports superstitions. *Psychological Reports*, *94*, 1043-1047.

McDavid, J., & Harari, H. (1966). Stereotyping of names and popularity in grade-school children. *Child Development*, *37*, 453-459.

McGee, R., & Williams, S. (2000). Does low self-esteem predict health compromising

behaviours among adolescents? *Journal of Adolescence*, *23*, 569-582.

Morrison, S. & Smith, G. (2005). Monogrammic determinism? *Psychosomatic Medicine*, *67*, 820-824.

Neil, G., Anderson, B., & Sheppard, W. (1981). Superstitions among male and female athletes of various levels of involvement, *Journal of Sports Behavior*, *4*, 137-148.

Nelson, L. D., & Simmons, J. P. (2007). Moniker maladies: When names sabotage success. *Psychological Science*, *18*, 1106-1112.

Pinzur, L., & Smith, G. (2009). First names and longevity, *Perceptual and Motor Skills*, *108*, 149-160.

Savage B. M., & Wells, F. L. (1948). A note of singularity in given names, *Journal of Social Psychology*, *27*, 271-272.

Lahman, S. (2010). The baseball archive database. Retrieved January 26, 2010, from http://www.baseball1.com/

Trzesniewsky, K., Donnellan, M., Moffitt, T., Robins, R., Poulton, R., & Caspi, A. (2006). Low self-esteem during adolescence predicts poor health, criminal behavior, and limited economic prospects during adulthood. *Developmental Psychology*, *42*, 381-390.

Table 1

Paired Difference in Average Age at Death (AAD) for Decedents with D Initial Minus AAD for

Decedents with E - Z Initials

	First Initial		First or Last Initial	
	Mean Difference	Median Difference	Mean Difference	Median Difference
Major League Baseball players				
1875-1930 birth years	-1.31	-2.13*	-0.81	-1.67
All players	-0.57	-1.82	-0.26	-1.34
California 1960-2004 Decedents				
Males	0.10**	0.10***	0.07**	0.07*
Females	0.07	0.05	0.02	-0.02

* p < .05. ** p < .01. *** p < .001