

Commentary

# Disease prevalence and survey design effects: A response to Weir and Smith

Sven Wilson<sup>a,\*</sup>, Benjamin L. Howell<sup>b</sup>

<sup>a</sup>*Departments of Political Science and Economics, Brigham Young University, 732 SWKT, Provo, UT 84602, USA*

<sup>b</sup>*Graduate Program in Public Policy, Brigham Young University, 783 SWKT, UT 84602, USA*

Available online 29 June 2007

---

## Abstract

Evidence provided by Weir and Smith, particularly the findings from the National Health and Nutrition Examination Survey (NHANES), leads us to conclude that an increase in arthritis prevalence during the 1990s in the United States is probable, but the trend is likely overstated in the Health and Retirement Study (HRS). We show that a mistake in our earlier method does not change substantively our previous conclusion that survey duration effects are occurring in the HRS, a finding that is also supported by a variety of regression models (including that of Weir and Smith). Furthermore, very little evidence exists for an upward trend among self-reporters in the National Health Interview Survey (NHIS), and less than 25% of the increase in the HRS over the 1990s can be attributed to increases in obesity.

© 2007 Elsevier Ltd. All rights reserved.

*Keywords:* Arthritis prevalence; Survey prevalence; USA; Panel conditioning; Age

---

## Introduction

The motivation for our 2005 paper (Wilson & Howell, 2005) was not the presence of an upward trend in arthritis in the Health and Retirement Study (HRS) but the alarming *magnitude* of that increase (close to one percentage point per year). Weir and Smith argue that the trend in the HRS is real. They find increases among self-reporters in the National Health Interview Surveys (NHIS) and, most convincingly, find a similar trend in the National Health Examination Surveys (NHANES). But given the analysis summarized here, we conclude that, although an upward trend is quite

probable, the magnitude of the trend is most likely overstated in the HRS. Furthermore, the evidence shows that survey duration effects cannot be easily dismissed.

Space considerations allow only a very brief summary of our response here. See Wilson and Howell (2007) for more detail.

### *Comparisons with NHIS results*

Weir and Smith (2007) make quick work of the NHIS. Too quick. Their main point is that self-reporters aged 55–59 show an increase in prevalence from 1992 to 1996 roughly equivalent to the HRS numbers. However, of all the 5-year age intervals between 45 and 69, only the 55–59 group shows this increase among self-reporters; all others are virtually flat. For the combined age interval of 45–64

---

DOI of original article: [10.1016/j.socscimed.2007.04.019](https://doi.org/10.1016/j.socscimed.2007.04.019)

\*Corresponding author.

*E-mail address:* [Sven\\_Wilson@byu.edu](mailto:Sven_Wilson@byu.edu) (S. Wilson).

(the interval used for published NHIS reports), the self-reported rate goes from 30.0% (1992) to 29.1% (1994) to 29.6% (1996).

An even better way to contrast the two data sets is to compare increases in prevalence for the cohort aged 51–61 in 1992 (the sampling frame for the HRS). Prevalence in this cohort increases by 10.9 percentage points between 1992 and 1996 in the HRS but only 6.0 points among self-reporters in the synthetic cohort constructed from the NHIS.

Finally, preliminary analysis of the NHIS (still using self-reporters) in later periods shows that prevalence at aged 51–61 actually *declined* slightly between 1998 and 2000 and between 2002 and 2004 (a change in question wording occurred between 2000 and 2002), though it rose quite rapidly in the HRS over these periods.

#### Comparisons with new HRS cohorts

As noted by Weir and Smith, our original analysis failed to account for the non-representative nature of the new HRS cohorts (taken in 1998 and 2004). However, the consequences of this flaw are not what Weir and Smith claim. In their Table 2, they make an inappropriate comparison between the “original sample” and the war babies cohort in 1998 that shows a 10-point difference (42.8–32.6) for those aged 53–56. However, in 1998 the representative

portion of the original sample includes only those aged 57–67 (born 1931–1941). In both our original and new estimates, our comparisons are made using only age-eligible members of each cohort, which is why we did not calculate prevalence for anyone from the original sample in 1998. Weir and Smith, therefore, concoct a false contrast that we never made.

Furthermore, we can show that the consequences of our error were, contrary to what Weir and Smith claim, relatively benign. The new cases in 1998 were people born in 1942–1947 who do not have a spouse in 1998 who was age eligible (and, hence, already in the sample) in 1992. The 2004 cohort was defined similarly. We can replicate this selection criteria in the 1992 cohort by picking those cases who were born from 1936 to 1941 but do not have a spouse born between 1925 and 1935. We refer to these three similar sub-samples as the “matched samples,” which are almost identical to each other in terms of age, sex, marital status, race, and percent Hispanic (as expected, education and obesity is rising slightly over time, and smoking is falling). We stress that the matched samples are *not* representative of the population; but they *are* representative of each other.

We trace arthritis prevalence for each matched sample in Fig. 1 for ages 55–56, the only age group that is age eligible for the HRS in each wave.

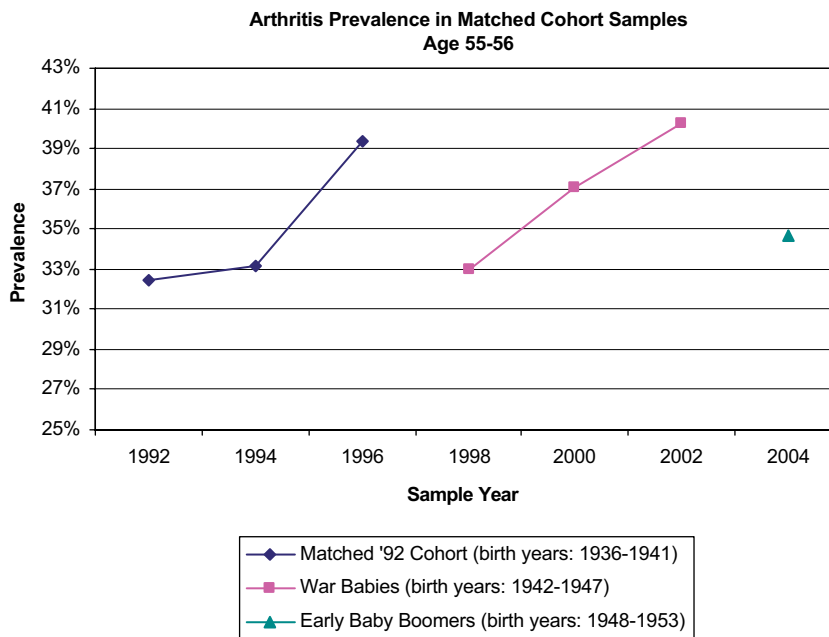


Fig. 1. Arthritis prevalence in matched cohort samples aged 55–56.

As before, we see large increases in prevalence within each cohort and an abrupt decline with the onset of the 1998 and 2004 cohorts (the decline from 1996 to 1998 is statistically significant)—the same pattern we found in our previous work.

#### *Alternative regression results*

Weir and Smith also use a probit model to test for the presence of survey duration effects, which they dismiss because the estimate is not statistically significant. But this estimate is *substantively* significant. A marginal effect of .0044 actually represents an increase of more than 4 percentage points over a 10-year period—close to half the actual increase in the HRS from 1992 to 2002. And when we expand marital status in the Weir and Smith model to capture spousal age (a selection criteria they emphasize), we find an even stronger marginal effect of survey duration of .0074 ( $z = 1.74$ ).

Furthermore, the sample selection used in Weir and Smith—respondents born in 1942–1947 and pooled from 1998 to 2002—is not unreasonable, but it is arbitrary. Alternative legitimate criteria yield estimates of both greater substantive and statistical significance. If we estimate their model only for 1998, for instance, the estimate rises to .0077 ( $z = 2.06$ ). If we run the model again for 1998 for the entire age range (birth years 1931–1947), we get a huge effect: .0127 ( $z = 4.50$ ). Repeating this exercise for 2004, yields an effect of .0085 ( $z = 2.23$ ) using birth years 1948–1953 and .0073 ( $z = 3.04$ ) using birth years 1931–1953. These regression results are essentially a confirmation of the pattern of abrupt discontinuities shown in Fig. 1.

As a final test, we pool data from all survey waves, which is valid as long as we use only representative, age-eligible observations in each year with respect to the cohort they come from. This approach shows an increase in the marginal effect of survey duration to .0092 ( $z = 5.39$ ), while the marginal effect of calendar time (.0034) is much smaller and more reasonable than the highly implausible estimate of .0194 that Weir and Smith report.

The duration effect estimated by Weir and Smith is clearly not robust, but the control variables in the model are remarkably so, suggesting that pooling observations from different birth years and time periods does not significantly alter the effect of the

controls. Certainly we do not pretend to distinguish between age, period, cohort and survey duration effects in the above analyses, but the regression approach employed by Weir and Smith suggests that there is more evidence for duration effects than there is against them.

#### *Other chronic conditions*

Weir and Smith make much of our brief discussion of panel conditioning. However, our belief was that any panel conditioning in the data would have the most effect on conditions that are either easily (hypertension) or informally (arthritis) diagnosed, not on conditions such as diabetes and heart disease, even though those conditions are typically underdiagnosed. Of course, neither we nor they have a well-developed model of disease reporting behavior that we can rely on to support our conjectures.

#### *The NHANES estimates*

The estimates from waves 3 and 4 of the National Health and Nutrition Examination Survey (NHANES) are the strongest evidence presented by Weir and Smith for an upward trend in arthritis prevalence, although the prevalence estimates differ quite dramatically from the HRS, differences that Weir and Smith believe are due to question wording. However, in wave 3 diagnoses were to come from a “doctor,” and in wave 4, from a “doctor or other health professional.” Weir and Smith claim that the consequences of this claim are likely minimal, though the prevalence of reported professional diagnoses (and possibly self-diagnoses) will surely be higher in wave 4. But how much higher? Enough to account for both an upward trend in prevalence and a 3–5-point increase due to survey duration in the HRS? Neither we nor Weir and Smith have any evidence to answer these questions.

#### *The role of obesity*

Weir and Smith’s estimates in their Table 3 suggest a powerful role for obesity, and they fault us for not recognizing this “clear signal”. But very little of the increase in arthritis can be explained by changes in the BMI distribution. This is true whether we use the unadjusted odds ratios from the HRS (which explain 18% of the increase from

1992 to 2004 for those aged 55–59) or the marginal effects in Table 3 (which explain 24%) as distributional weights. In short, within-category increases dominate compositional changes. For instance, for those with BMI <25, the increase in arthritis prevalence rises from 30.5% (1992) to 39.5% (2002). Furthermore, the odds ratios associated with obesity are falling over the period (which would lower the explanatory effect of increasing obesity even further were we to account for it).

## Conclusions

From our analysis, we conclude that

- (1) In general, arthritis prevalence in the NHIS, even among self-reporters, does *not* increase from 1992 to 1996 among those in late middle age. Only for one narrow age band (where confidence bands are wide) do we see a modest increase.
- (2) Using appropriate matched samples from each HRS sample cohort, we see abrupt declines in arthritis prevalence (among those aged 55–56) when the new cohorts enter the sample, an effect captured in the full HRS sample as well.
- (3) Regression models for duration effects generally show evidence for important duration effects. Even the estimated effect from Weir and Smith is quite substantial.
- (4) The NHANES results show a significant increase in prevalence, though the extent of the increase due to change in question wording is not known.
- (5) Rising prevalence of overweight and obesity can explain only a small part of the trend in arthritis prevalence in the HRS.

The classical hypothesis testing framework of Weir and Smith rejects survey duration effects. But we believe that more attention to a type II error (rejecting a true effect) is warranted. The estimate by Weir and Smith is large in magnitude,

and the posterior probability that the effects are greater than zero is much higher than the probability that the effects are zero or negative. Additionally, our evidence based on alternative regression specifications (where duration turns out to be large and statistically significant) and our analysis of the matched samples further strengthens this conclusion.

Of course this issue would be much less confusing if we had a better understanding of factors affecting the propensity to report health problems in survey data. The most striking feature of Table 1 from Weir and Smith is the tremendous range in estimates across the surveys. Clearly study design issues are profound. And given that there is so much variation in estimation of levels, we ought to be skeptical about observed trends, especially when the trends are in conflict across surveys. More than anything, Table 1 is an intriguing puzzle.

It seems entirely plausible to us that long-term participation in a health survey could (a) increase the comfort level respondents have with respect to reporting conditions and (b) induce respondents to pay more attention to symptoms and conditions between survey waves (especially after they get the HRS participant newsletter). Furthermore, HRS design features such as moving from face-to-face to telephone interviews after the initial interview may play a significant role in explaining the patterns in Fig. 1. Given the evidence, survey duration effects are plausible and troubling.

## References

- Weir, D., & Smith, J. (2007). Do panel surveys really make people sick? A commentary on Wilson and Howell (60: 11, 2005, 2623–2627). *Social Science & Medicine*, 65(6), 1071–1077.
- Wilson, S., & Howell, B. (2005). Do panel surveys make people sick? US arthritis trends in the Health and Retirement Survey. *Social Science & Medicine*, 60(11), 2623–2627.
- Wilson, S. & Howell, B. (2007). Extended Response to Weir and Smith. <<http://fhss.byu.edu/polsci/wilson/papers>>. Weir and Smith.