sociological science

Supplement to:

Wu, Lawrence L., Steven P. Martin, and Paula England. 2017. "The Decoupling of Sex and Marriage: Cohort Trends in Who Did and Did Not Delay Sex until Marriage for U.S. Women Born 1938–1985." Sociological Science 4: 151-175.

APPENDIX 1

In this appendix, we both: (1) replicate the single-decrement estimates reported by Finer (2007), but (2) show that they do not provide easily interpretable answers to questions such as who did and did not delay sex until marriage. As noted above, two studies (Hofferth, Kahn, and Baldwin 1987; Finer 2007) employed single-decrement procedures to estimate cohort trends in premarital sex using using, as we do, data on age at first intercourse and at first marriage from the National Survey of Family Growth. The difficulty is that estimates obtained from such single-decrement procedures are easily misinterpreted because asking about premarital sex with such data requires multiple-decrement procedures that answer which of two events—sex or marriage—occurs first.

Table A1 illustrates these issues concretely for NSFG women born 1939-48, an exercise that allows us to compare different sets of estimates with one another and with those reported by Finer. Panel A reports single-decrement Kaplan-Meier (1958) estimates for the percent who did and did not delay sex until marriage at exact ages 20, 25, and 30 (240, 300, and 360 months). The single-decrement estimates for not delaying sex until first marriage (column labeled “no”) increase with age, from 44.0% at age 20, to 70.9% at age 25, and to 81.0% at age 30. These estimates are virtually identical to the Kaplan-Meier estimates of 44, 73, and 82% reported by Finer (p. 76). Finer interprets these estimates as if they were simple percentages, thus writing that “Among those turning 15 between 1954 and 1963, 82% had had premarital sex by age 30” (p. 73).

Panel A also reports single-decrement Kaplan-Meier estimates for delaying sex until first marriage (column labeled “yes”). These estimates also increase with age, from 29.6% at exact age 20, to 74.8% at age 25, and to 86.8% at age 30.

What is clear from a cursory glance at these single-decrement estimates is how easily they can mislead the unwary. For example, at age 20 the two single-decrement estimates sum to less than 100%, but at ages 25 and 30, they sum to more than 100%. The standard (Greenwood) 95% confidence intervals, reported in the next columns of Panel A, do not help to unravel this apparent puzzle.

The estimates reported in Panel A, like those in Finer, are simple life table estimates for NSFG women born between 1939 and 1948; hence, these estimates do not contain controls for observed covariates. In Panel B, we provide a second set of descriptive statistics by reporting simple weighted percentages and unweighted frequencies for this same sample of women. Thus for the 3,241 NSFG women born 1938–49,

Table A1: Single-decrement (Kaplan-Meier) and observed percentages for delaying or not delaying sex until marriage by selected ages. NSFG women born 1939–48.

Panel A: Single-decrement (Kaplan-Meier) estimates of the percent of women by exact ages 20, 25, and 30 who reported: (1) delaying sex until marriage and (2) not delaying sex until marriage.

	Delay sex until marriage?			
	percent (weighted)		95% confidence interval	
age	yes	no	yes	no
20	29.6	44.0	(27.7, 31.5)	(42.2, 45.8)
25	74.8	70.9	(72.5, 77.1)	(67.6, 73.1)
30	86.6	81.0	(84.4, 88.7)	(78.4, 83.5)

Panel B: Observed percent and number of women by ages 20, 25, and 30 who reported: (1) delaying sex until marriage, (2) not delaying sex until marriage, or (3) neither.

	Delay sex until marriage?								
	percent (weighted)				*n* (unweighted)				
age	yes	no	neither	total	yes	no	neither	missing	total
20	21.0	39.8	39.3	100.1	500	1405	1023	313	3241
25	40.1	52.7	7.2	100.0	964	1786	178	313	3241
30	42.9	54.6	2.5	100.0	1035	1837	56	313	3241

313 have missing data on either age at first sexual intercourse or age at first marriage. For the remaining 2,928 women comprising the sample analyzed in both panels, we see that by age 20 (exact age 240 months), 21.0% (unweighted *n*=500) reported delaying sex until marriage, 39.8% (*n*=1,405) reported not delaying sex until marriage, and 39.3% (*n*=1,023) reported "neither", i.e., being never-married virgins at age 20. By age 25, the *n*=1,023 reporting "neither" at age 20 had declined by 845 to 178 (7.2%). Of these 845 women, 464 (464=964-500) delayed sex until marriage, while 381 did not delay sex until marriage (381=1786-1405). By age 30, the 178 reporting "neither" at age 25 had declined by 122 to 56, with 71 of these 122 women delaying sex until marriage and another 51 not delaying sex until marriage. The resulting percentages at age 30 are thus that 42.9% reported delaying sex until marriage, 54.6% reported not delaying sex until

marriage, and 2.5% reported being never-married virgins.

How might we understand the apparent discrepancies between the estimates in Panels A and B? The textbook answer is that no discrepancy in fact exists because the two panels answer different questions. It is clear that Panel B answers the straightforward and easily understood question: "What are the observed (unconditional) probabilities of being in the origin state ("neither", i.e., a never-married virgin) or one of two destination states (delaying or not delaying sex until marriage) by exact ages 20, 25, and 30. But it then follows that Panel A must answer a different question and as noted previously, the textbook explanation (see, e.g., Cox and Oakes 1984; Preston, Heuveline, and Guillot 2001; Wu 2003) is that the single-decrement estimates in Panel A must be interpreted under a rather convoluted counterfactual in which all other destination states have been eliminated. Thus, the single-decrement estimates in Panel A for not delaying sex until marriage give the probability of premarital sex *among those who will continue to remain never-married*; likewise, single-decrement estimates for delaying sex until marriage give the probability of marriage *among those who will continue to remain virgins*.

How censoring is handled by single-decrement procedures is why the two panels answer different questions. Panel B shows that at age 29, 1,023 NSFG women reported "neither"; hence, at age 20, the single-decrement estimates in Panel A treat these 1,023 women as censored. But Panel B also shows that by age 25, 845 of these 1,023 women had one of the two events of interest, with 464 delaying sex until marriage and 381 not delaying sex until marriage. Thus for the age 25 single-decrement estimates in Panel A, the 464 women delaying sex until marriage are treated as censored for premarital sex because once they enter marriage as virgins, they are no longer at risk of engaging in sex prior to marriage. But censoring in this way also implies that the Panel A single-decrement estimates for delaying sex until marriage will refer only to those who will continue to remain virgins; likewise, the single-decrement estimates for not delaying sex until marriage will refer only to those who will continue to remain never-married. The practical danger is thus to interpret single-decrement estimates as if they were unconditional probabilities, thereby ignoring the condition "*among those who will continue remain.*"

Differences can be large, as shown by the age 30 single-decrement estimate for premarital sex of 81.0% in Panel A vs. the corresponding estimate of 54.6% in Panel B. More generally, the Panel A and B estimates of premarital sex will increasingly diverge as women marry, thus generating larger differences at later ages or in older cohorts when marriage at young ages was more common. Thus for the older birth cohorts examined by both Hofferth, Kahn, and Baldwin (1987) and Finer (2007), this divergence can be substantial at the later ages examined by Finer, but may be less consequential during the teen years examined by Hofferth and colleagues. See also Regnerus and Uecker (2011, Chapter 1, footnote 2, p. 271), who allude to some of these difficulties in interpreting estimates from single-decrement procedures.

APPENDIX 2

In this appendix, we briefly review our procedures for obtaining predicted probabilities from estimated coefficients from our competing-risk regressions (see also England et al. 2014; Wu and Martin 2016).

Let $k = 1, \ldots, K$ index the competing risks of interest and let $T_1, \ldots, T_K$ denote the random variables for the corresponding K event times. The classical competing-risk model in (1) and (2) assumes that the K risks are independent (conditional on the covariates $\mathbf{x}$), but in many respects the more fundamental underlying assumption is that the model posits something akin to an omniscient observer who has knowledge of person i's times at *each* of the $k = 1, \ldots, K$ possible events. The model then acknowledges that what is in fact observed by all non-omniscient observers is the earliest of these event times, with

$$T = \min(T_1, T_2, \ldots, T_K) \tag{A2.1}$$

thus observed and with all other event times latent and thus unobserved.

The basic logic underlying the above is thus identical to what is assumed in formal demographic models for cause-specific mortality and in the multiple-decrement life table. Then consider the classic formal demographic thought experiment that supposes that one cause of death has been eliminated. Given K mutually exclusive and exhaustive causes of death, eliminating one cause (say, cause 1) will imply that

$$T^{(-1)} = \min(T_2, \ldots, T_K)\,, \tag{A2.2}$$

thus making manifest one cause of death that was previously latent. Repeating the above thus makes clear that single-decrement procedures applied to a single T_k must be interpreted under a counterfactual in which all other competing risks have been eliminated.

Then because the kth competing risk is governed by r_k in (1), it can be shown that

$$F_k(t|\mathbf{b}_k, \mathbf{x}) = 1 - S_k(t|\mathbf{b}_k, \mathbf{x}) = 1 - \exp\Big[-\int_0^t q_k(u) \exp(\mathbf{b}_k \mathbf{x})\, du \Big]\,, \tag{A2.3}$$

where F_k denotes the cumulative distribution of event times for the kth competing risk.

As we have emphasized throughout, the question of who did and did not delay sex until marriage intrinsically involves asking which of two events—marriage or sex—occurred first. It is possible to obtain simple estimates answering these questions with survey questions that ask directly if sex preceded marriage or if marriage preceded sex, as was true for the survey data analyzed by Klassen et al. In this paper, however, we have posed questions involving covariates such as the two-year birth cohort dummy variables used to operationalize cohort trends but also including background socio-demographic variables common across Cycles 3–7 of the NSFG. For such problems, regression methods, such as the continuous-time competing-risk models in (1), provide a natural analytical framework.

A difficulty with the competing-risk regression in (1) is that estimates from are not easily interpreted. Consider a hypothetical treatment that is estimated to reduce mortality risks from two causes of death, by 50% for a first cause and by 5% for a second cause. It is in fact not obvious what such estimates imply in that answers will depend, at least in part, on levels as modeled by $q_k(t)$, the baseline risks, in (1). That is, if one cause is common but the other rare, then even modest reductions in the risks for a common cause can imply a substantial increase in survival whereas a sizable improvement for a rare condition may increase survival only negligibly. This example thus illustrates the difficulties involved when inspecting estimated regression coefficients in (1) to answer standard *ceteris paribus* questions such as "What is implied by a unit change in a covariate x?' in that a large or small coefficient for a covariate x in a competing-risk regression need not imply large or small consequences for quantities that are typically of key demographic interest. For these reasons, we chose to report the predicted probability by a given age of being in one of the three states depicted in Figure 1. The resulting predicted percentages are easily conveyed to a wide audience and can furthermore be used to answer "what if" counterfactuals as in the results we report in Tables 4 and 5.

We used the following computationally-intensive procedure to obtain predicted probabilities by age τ (see also England et al. 2013; Wu and Martin 2016).

For cases $i = 1, \ldots, n$

A.1: Use estimated model parameters and observed covariates for case i to form the function $F_k(t|\widehat{\mathbf{b}}_k, \mathbf{x}_i)$. Then sample from $F_k(t|\widehat{\mathbf{b}}_k, \mathbf{x}_i)$ to simulate t^*_{ik}, case i's age at the kth event for $k = 1, \ldots, K$ to obtain K simulated event times, $\{t^*_{i1}, t^*_{i2}, \ldots, t^*_{iK}\}$.

A.2: Set $t^*_i = \min(t^*_{i1}, t^*_{i2}, \ldots, t^*_{iK})$; then:

A2.1: if $t^*_i > \tau$, then case i remains in the origin state at age τ;

A2.2: else $t^*_i \leq \tau$, with case i experiencing the event k' satisfying $t^*_i = t^*_{ik'}$.

A.3: Repeat steps **A.1–A.2** until the estimates satisfy a prespecified level of precision.

Note that applying the steps **A.1**–**A.2** for cases $i = 1, \ldots, n$ will yield n counts in the origin state and the K destination states; step **A.3** then repeats these steps until a desired level of precision is reached. As shown in Figure 2, predicted probabilities track observed probabilities closely.

REFERENCES

Cox, D. R., and D. Oakes. 1984. *Analysis of Survival Data*. London: Chapman and Hall.

England, Paula, Lawrence L. Wu., and Emily Fitzgibbons Shafer. 2013. "Cohort Trends in Premarital First Births: What Role for the Retreat From Marriage?" *Demography* 50(6): 2075–104.

Finer, Lawrence B. 2007. "Trends in Premarital Sex in the United States, 1954–2003." *Public Health Reports* 122(1): 73–78.

Hofferth, Sandra L., Joan R. Kahn, and Wendy Baldwin. 1987. "Premarital Sexual Activity Among U.S. Teenage Women Over the Past Three Decades." *Family Planning Perspectives* 19(2): 46–53.

Kaplan, E. L., and Paul Meier. 1958. "Nonparametric Estimation from Incomplete Observations." *Journal of the American Statistical Association* 53(282): 437–81.

Preston, Samuel H., Patrick Heuveline, and Michel Guillot. 2001. *Demography: Measuring and Modeling Population Processes.* Oxford: Blackwell.

Wu, Lawrence L. 2003. "Event History Models for Life Course Analysis." Pp. 477–502 in *Handbook of the Life Course*, edited by J. Mortimer and M. Shanahan. New York: Plenum.

Wu, Lawrence L., and Steven P. Martin. 2016. "Computationally-Intensive Methods for Correlated Cause-Specific Mortality Risks." Paper presented at the 2016 Annual Meetings of the Population Association of America, Washington, DC.