**Approximations of medication use history in WHI cohorts**

The quality of observational analyses that use WHI medication data as the exposure variable of interest may depend crucially on constructing the medication use history over time. A strategy that constructs medication use history is described below that allows for analysis of time-to-event outcomes described by Prentice and colleagues[1] (2011).

Medication inventories were collected during the WHI study at baseline and years 1, 3, 6, and 9 in the CT, and at baseline and year 3 in the OS (Figure 1). At each inventory, clinic interviewers entered each medication name and strength directly from the containers into a database that assigned drug codes using Medi-Span software (First DataBank, Inc., San Bruno, CA). The duration of each medication, based on the participant's recollection, was also entered. The collective data can then be used to construct an approximation to each participant's medication use history over time, and subsequently used as a time-dependent exposure variable in outcome association analyses. Medication use data was also collected, by mail or phone, towards the end of the first WHI extension period.

A basic approximation might consist of representing a participant's history by multiple rows of data; each row represents the non-overlapping follow-up after a collection. For example, the medications use history of a CT participant during the WHI study for a particular drug could be approximated by five rows of data: (0, Y1,Z(0)], (Y1,Y3,Z(1)], (Y3,Y6,Z(3)], (Y6,Y9,Z(6)] and (Y9,T,Z(9)]. Here Yi = time from randomization to the ith collection, Z(i) is an indicator variable(s) for medication use, and T is the time to event or censoring. Data in this *counting process* format can readily be analyzed in a *Cox proportional hazards* regression model that allow for time-dependent variables such as Z(i).

An obvious consequence of the basic approximation is that Z(i) is not updated until the next collection, at annual visit i+1. This limitation may not lead to a reliable approximation for medications whose usage pattern changed markedly over time. For example, suppose 20% of the participants being treated for a particular health condition were using drug A at year 1, and the proportion of those treated with drug A increased to 40% by year 3 (Figure 2). While it is very likely that participants not using drug A at year 1, later initiated drug A during the interim, the basic approximation does not address changes in drug use *between* visits. Rather the basic approximation assumes initiation of drug A at year 3.

A better approximation can be obtained by using duration data to refine the time of initiation during follow-up. Specifically, the refined time of initiation is computed as Yi minus the minimum ($\delta$, duration reported at Yi), where $\delta = 1$ for year 3, and $\delta = 2$ for years 6 and 9. Hence the refined initiation times would lie in the intervals (*Y2*, Y3], (*Y4*, Y6], and (*Y7*, Y9], respectively. Limiting the refined initiation time to the aforementioned intervals, extends a participant's recall of medication use to at most one (two) year(s), and consequently does not supersede the preceding collection. A similar refinement can be used at year 3 for the OS where $\delta = 2$, and is illustrated in Example #1. To ensure a similarly reliable approximation of a participant's medication use history during the extension, time to initiation should be refined by no more than two years.

Another challenge when constructing medication use history over time is the potential for collection data to become out-of-date, and thereby no longer accurately representing a participant's medication use. Due to the collection design (Figure 1), a reasonable solution might be to censor the follow-up for

[1] Internal WHI memorandum: Perspectives On Observational Analyses of Medication Use in WHI Cohorts, Ross Prentice and the WHI Medication Analysis Group, 24 September 2011.

a participant 3.5 years after the last collection; 3 years per design and allowing 0.5 years for sample variability. For example, if a CT participant missed her year 6 collection, the history could be approximated by *discontinuous intervals of risk:* (0, Y1,Z(0)], (Y1,Y3,Z(1)], (Y3, Y3 + 3.5, Z(3)], and (Y9,T,Z(9)]. The participant's history during the interval (Y3 + 3.5, Y9] would not be included in the analysis, although the participant would be allowed to re-enter the analysis at year 9. Example #2 illustrates a participant whose out of date medication was censored. Although the collection design differs between the OS and CT, censoring out-of-date medications allows for equally reliable approximations between studies.

While medication use histories were collected towards the end of the first five year *extension*, for most WHI CT and OS participants, the most recent *study* collection occurred at Y6 and Y3, respectively. Consequently, medication histories collected during the study will become out-of-date and should be censored, until re-entry into the analysis with more recent extension data. To prevent differential follow-up, participants not taking the medication(s) of interest should be allowed to re-enter the risk set two years prior to the date of their extension collection. For participants taking the medication(s) of interest, initiation should be refined by the minimum of 2 years, or the reported duration. If self-reported duration is less than 2 years, the remainder of the 2 year interval should reflect not using the medication(s) of interest. Example #3 illustrates a participant whose reported duration, during the extension, was less than two years. If this participant had not used statins prior to the extension, the time-dependent exposure would have been coded "no statin use" for the interval (12.54, 14.04]. Further refinements to the medication use history might include more complex methods such as imputation as elucidated by Prentice and colleagues (2011).

A comparison of the approximations to participants' medication use histories over time is shown in Figure 2. Medication use was constructed for a cohort of 146K WHI OS and CT participants to examine the association between statin use and cancer death. The proportion of atorvastatin (Lipitor) users, among statin users, based on the collected medication inventory is shown at each follow-up visit including the extension. The usage pattern increases markedly during the study period, but drops precipitously during the extension period; the FDA approved the first generic simvastatin (Zocor) in 2006. It is very likely, that the true proportion of atorvastatin users is grossly underestimated by the basic approximation at year 2; many OS participants may have initiated atorvastatin use before their year 3 collection. Assuming the proportion of the *full* cohort using atorvastatin grew linearly between years 1 and 3 (pink line), the basic approximation is greatly improved, between years 2 and 3, by leveraging duration data to refine the time of initiation. The approximation that censors out-of-date medications improves upon the basic approximation after 3.5 years from enrollment, with further improvements 6.5 and 9.5 years after enrollment. However for this example, censoring out-of-date medications resulted in only a modest *overall* improvement to the basic approximation; integrated squared error (ISE) was reduced by 10%[2]. In situations where the usage pattern increases monotonically over time, censoring out-of-date medications can yield larger improvements; a similar analysis for the anti-diabetic drug, metformin, reduced ISE by 50% relative to the basic approximation.

In summary, the quality and inferences of analyses that associate WHI medication data with disease incidence may be improved with better approximations to medication use history over time. Of course the disease risk models used in association analyses may use these histories in various ways; for example, hazard ratios may be modeled as functions of 'current' medication use status, or of duration of

[2] Assumes the subsample of participants with a medications collection is representative of the full cohort, and the usage pattern of the full cohort changes linearly between collections.

use, among many other possible functions of the preceding medication history. The aforementioned medication history assessment strategies are intended to serve as a guide. Researchers and analysts should incorporate and further refine strategies that meet their research needs.

Prepared by Aaron K. Aragaki on 29 May 2014

**Examples for approximation of medication use histories**

**Example #1:** **Basic Approximation** **Refined* approximation**

| ID | study | year | start | stop | Statin use | ID | start | stop | Statin use |
|---|---|---|---|---|---|---|---|---|---|
| 1 | OS | 0 | 0.00 | 3.01 | fluvastatin | 1 | 0.00 | 1.01 | fluvastatin |
| 1 | OS | 3 | 3.01 | 5.94 | atorvastatin | 1 | 1.01 | 5.94 | atorvastatin |

*Note: Reported duration of statin use was 1 and 2 years at baseline and year 3, respectively. Survival was censored 5.94 years after enrollment.*

**Example #2** **Basic Approximation** **Refined* approximation**

| ID | study | year | start | stop | Statin use | ID | start | stop | Statin use |
|---|---|---|---|---|---|---|---|---|---|
| 2 | CT | 0 | 0.00 | 1.01 | lovastatin | 2 | 0.00 | 1.01 | lovastatin |
| 2 | CT | 1 | 1.01 | 2.98 | lovastatin | 2 | 1.01 | 1.98 | lovastatin |
| 2 | CT | 3 | 2.98 | 8.78 | lovastatin | 2 | 1.98 | 6.48 | lovastatin |
| 2 | CT | 9 | 8.78 | 9.54 | lovastatin | 2 | 6.48 | 6.78 | out -of-date medication *(censored interval)* |
| | | | | | | 2 | 6.78 | 9.54 | lovastatin |

*Note: Reported duration of statin use was 4, 5, 7 and 5 years at baseline, years 1, 3 and 9, respectively. Survival was censored 9.54 years after enrollment.*

**Example #3** **Basic Approximation** **Refined* approximation**

| ID | study | year | start | stop | Statin use | ID | start | stop | Statin use |
|---|---|---|---|---|---|---|---|---|---|
| 3 | CT | 0 | 0.00 | 0.99 | no statin use | 3 | 0.00 | 0.99 | no statin use |
| 3 | CT | 1 | 0.99 | 3.10 | no statin use | 3 | 0.99 | 3.10 | no statin use |
| 3 | CT | 3 | 3.10 | 6.00 | no statin use | 3 | 3.10 | 5.00 | no statin use |
| 3 | CT | 6 | 6.00 | 8.97 | atorvastatin | 3 | 5.00 | 6.97 | atorvastatin |
| 3 | CT | 9 | 8.97 | 14.54 | atorvastatin | 3 | 6.97 | 12.47 | atorvastatin |
| 3 | CT | X | 14.54 | 17.89 | simvastatin | 3 | 12.47 | 12.54 | out -of-date *(censored interval)* |
| | | | | | | 3 | 12.54 | 14.04 | unknown statin type |
| | | | | | | 3 | 14.04 | 17.89 | simvastatin |

Note: Reported duration of statin use was 1, 5 and 0.5 years at years 6, 9 and during the extension, respectively. *Survival was censored 17.89 years after enrollment.*

* Approximation censors out-of-date medication and incorporates duration data to refine time to initiation of medication use.

**Figure Legends.**

**Figure 1: Medication collection design and refinements.**
The top and bottom panels show the medication collections for the OS and CT, respectively. The vertical bars indicate when medications were collected during the study (blue) and the extension period (black). The heights of the vertical bars indicate what fraction of women, relative to baseline, had a collection. For example, a much smaller % of women in the CT had a collection at year 9. The shaded regions indicate variability when each annual visit occurred (+/- 1 SD). Left-pointing green arrows indicate the extent to which duration information, for a particular collection year, could be used to refine time to medication initiation. For example, refinement would be the minimum ($\delta$, reported duration), where $\delta$ = one for year 3 and two years for years 6 and 9. The right-pointing arrows indicate when a recent medication collection is considered reliable (green) and when it is considered out-of-date (red dotted line w/ x); for simplicity, similar arrows that indicate out-of-date medication use during the extension are not shown.

**Figure 2: A comparison of approximations to participants' medication use history over time.**
Red circles indicate the proportion of atorvastatin users (among statin users) based on the collected medication inventory. Blue(dotted), green and black lines are estimates of the proportion of atorvastatin users that were computed using the basic approximation, approximation that censors out-of-date medications, and the approximation that also refines initiation, respectively. The pink line interpolates the observed data, collected on a subsample of the cohort, and is assumed to be the true proportion of atorvastatin use of the full cohort. The integrated squared error is the squared difference between an approximation (blue, green or black lines) and the linear interpolated estimate between each collection (truth; pink line), integrated between 0 and 16 years.

Figure 1.

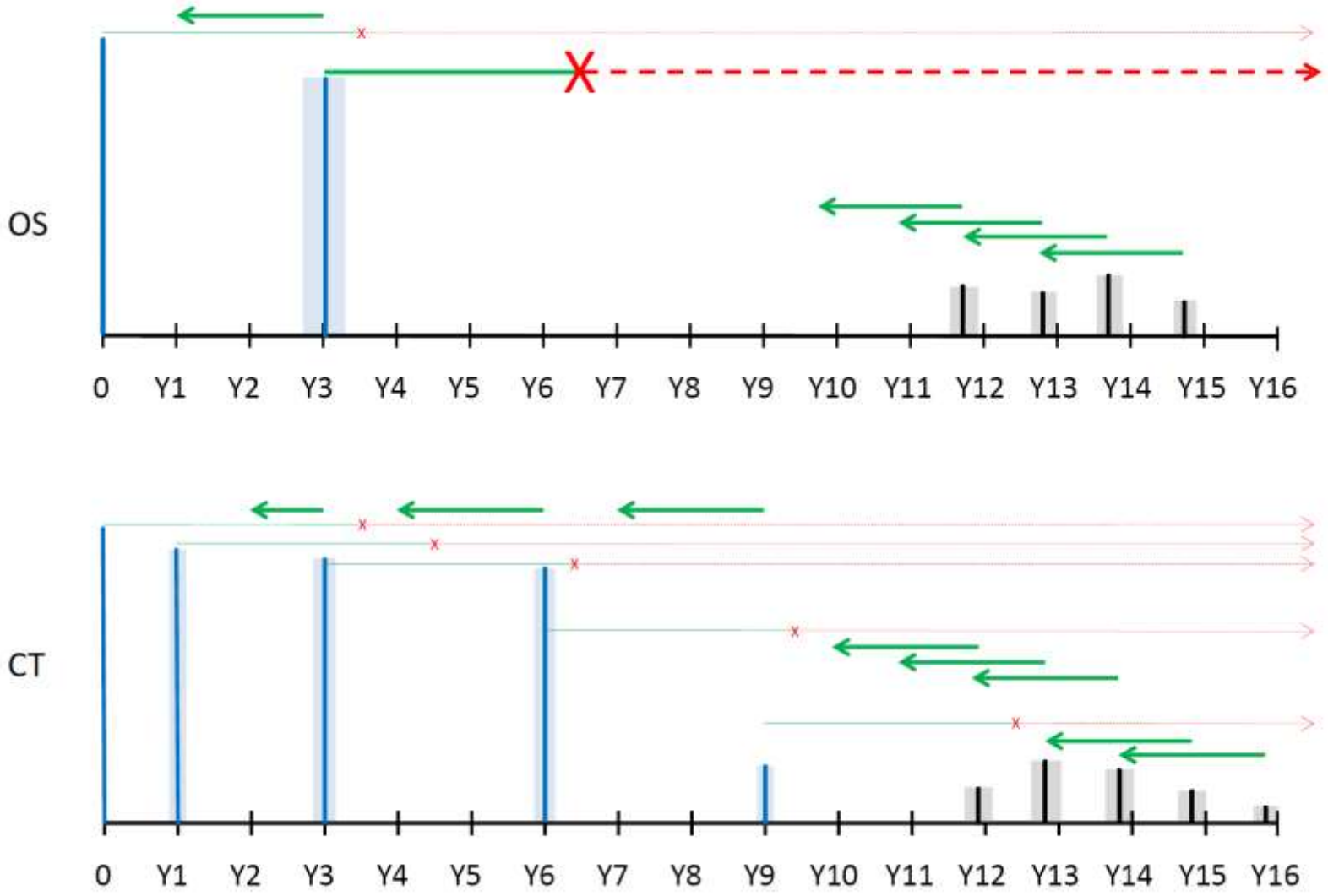
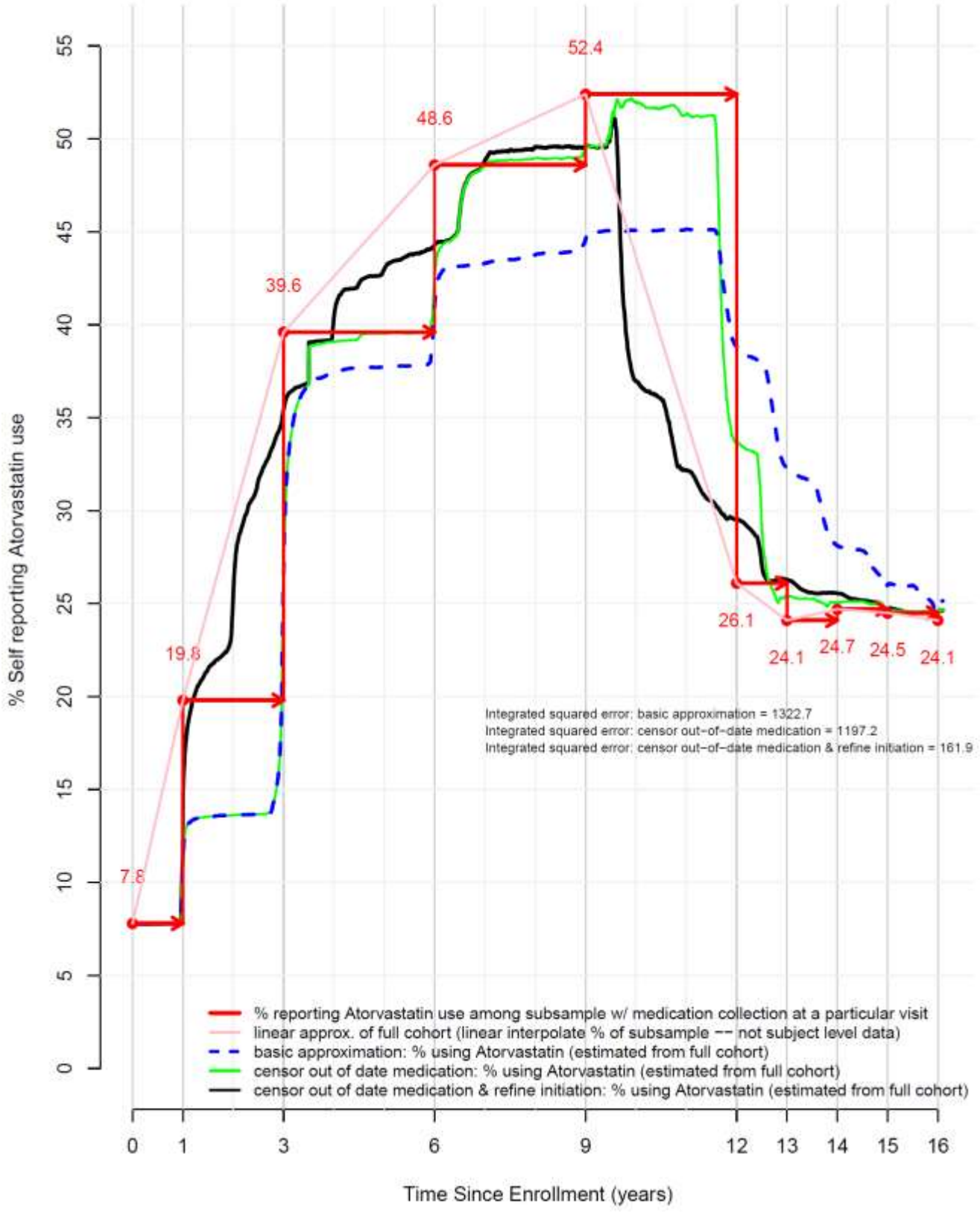

OS
0
Y1
Y2
Y3
Y4
Y5
Y6
Y7
Y8
Y9
Y10
Y11
Y12
Y13
Y14
Y15
Y16
CT
0
Y1
Y2
Y3
Y4
Y5
Y6
Y7
Y8
Y9
Y10
Y11
Y12
Y13
Y14
Y15
Y16

**Figure 2.**


% Self reporting Atorvastatin use
0
5
10
15
20
25
30
35
40
45
50
55
7.8
19.8
39.6
48.6
52.4
26.1
24.1
24.7
24.5
24.1
Integrated squared error: basic approximation = 1322.7
Integrated squared error: censor out−of−date medication = 1197.2
Integrated squared error: censor out−of−date medication & refine initiation = 161.9
% reporting Atorvastatin use among subsample w/ medication collection at a particular visit
linear approx. of full cohort (linear interpolate % of subsample −− not subject level data)
basic approximation: % using Atorvastatin (estimated from full cohort)
censor out of date medication: % using Atorvastatin (estimated from full cohort)
censor out of date medication & refine initiation: % using Atorvastatin (estimated from full cohort)
0
1
3
6
9
12
13
14
15
16
Time Since Enrollment (years)