Supplementary Material for

Mortality reductions produced by sustained prostate cancer screening have been underestimated

1. The need for time-specificity in the analysis of data from cancer screening studies

Figure 1A and Supplementary Figure A show when it is and is not possible to use ratios of overall (or cumulative) cancer mortality rates to measure the timing and magnitude of the reductions produced by screening. Figure 1B includes a smooth-in-time rate ratio curve that was fitted to the ERSPC data. This appendix contains a more extensive discussion of the need for time-specificity, and provides details on how the smooth curve was fitted.

Since the first cancer screening trials, investigators have tried to avoid the dilution caused by including cancer deaths that occur *after* the window in which the benefits of the regimen become manifest[1 2 3]. The recognition that the dilution caused by including those that occur *before* this window is more recent[4 5 6], and the message to avoid it has gone largely unheeded.

This ongoing time-insensitivity in the analysis of screening trials is all the more surprising today, where reviewers routinely ask whether the data justify the use of a proportional hazards model, i.e., of a single (average) hazard ratio. But even if we did not yet have screening data, biological principles alone suggest that this '*constant-over-time right-from-the-outset reduction*' assumption may govern the time-pattern of the effects of screening for some conditions, such as abdominal aortic aneurysms, but *not for cancers*.

What has not been previously recognized is the considerable influence of the duration of follow-up, particularly in prostate cancer where the time scale is longer than for other cancers. Baker[4] termed the dilution caused by including excessive/superfluous years in the time window

after the effect of the *last* screen become manifest "post-screening noise." In contrast, in trials of sustained prostate cancer screening, such as ERSPC, the attenuation is caused by (i) including the years *before* the impact of the *first* screen become manifest, and (ii) *not having full follow-up information available on the years where the effects of the screening are most apparent.*

These distortions argue for a data-analysis approach that cannot be influenced by, and is insensitive to, the choice of the time horizon of interest, the time window in which the effect of the screening regimen becomes manifest, and the amount of follow-up at the time of data-analysis. The measure should be robust to these and be calculable objectively from the data.

In Figure 1B, we subject the ERSPC data to formal quantitative time-specific analysis. Each time-specific rate ratio is independent of the ratio calculated from any another portion of the follow-up. The curve shows the timing of the delay until the effect of the screening regimen is expressed.

2. Fitting a smooth-in-time mortality rate ratio function.

In screening trials, the yearly observed numbers of cancer-specific deaths from the target cancer in each study arm are small, and so yearly mortality rate ratios fluctuate widely. For example, if the *expected* number in the non-screening arm for a particular year is 25, the *actual* count could vary by more than two-fold: under the Poisson law, it could range from about 15 to 35. Similarly, if the expected number in the screening arm was 16 (a true reduction of 36%), the actual count could vary from maybe 8 to 24, so that the observed rate ratio could vary from 0.3 (70% reduction) to 1.2 (20% increase). With event rates of this order of magnitude, it is difficult even with sample sizes in the tens of thousands to objectively estimate the true timing and extent of

the benefit of the intervention "by eye". Thus, a formal curve-fitting procedure becomes important to smooth out the noise.

In this section, we describe – and show how to fit -- the *simplest* candidate curve for the rate ratios characterizing the results of a cancer screening program comprising several rounds of screening. The assumed form of the rate ratio curve (the fitted version is shown as a dotted line in Figure 1B) is such that it has a value of unity for some unknown number of years, begins to descend after this unknown time point, and descends to an 'asymptote' of unknown value some unknown number of years later, and remains at this value thereafter. If this simplest of all models is postulated, there are only three unknowns to be estimated, when the rate ratio began to be non-null, the value of the RR asymptote and the time at which the asymptote began. More complex curves, such as would be needed to smooth curves that show *transient* reductions, can be fitted in the same way, simply by changing the form of *RR(t)* and adding more parameters. To do so, one would, naturally, require more extensive and more detailed data.

Consider a theoretical rate ratio (RR) curve, of the same shape as the one depicted by a dotted line in Figure 1B. Suppose the RR begins to change (become non-null) at T_c, and that its asymptote has the value RR_a, beginning at time $t = T_{begin}$.

Let it be defined as

$$RR(t) = \begin{cases} 1 & \text{until } t = T_c \\ 1-\{0.1, 0.2, 0.5, 0.8, 0.9\}\times(1- RR_a) & \text{from } t = T_c \text{ to } T_{begin} \text{ [5 equal } t \text{ steps]} \\ RR_a & \text{from t =} T_{begin} \text{ onwards} \end{cases}$$

Suppose the data consist of:

the times, $t_1, t_2, \ldots, t_D$, measured from randomization to screening/not, of each of the D prostate cancer deaths in the two arms combined.

the corresponding indicators, $s_1, s_2, \ldots, s_D$, of whether they occurred to men in the screening arm (1) or control arm (0).

the corresponding denominator-ratios, $dr_1, dr_2, \ldots, dr_D$, where dr_i is the ratio of the numbers of men being followed in the screening and comparison arms at the time of the i-th prostate cancer death.

The values of the three parameters, T_c , RR_a and T_{begin} can be estimated by numerically maximizing the Likelihood constructed by treating $s_1, s_2, \ldots, s_D$ as realizations of D Bernoulli random variables, where the expected value of the i-th such random variable is $dr_i \times RR(t_1)/[1+ dr_i \times RR(t_i)]$.[7] The profile log likelihood can be used to obtain a C% confidence region for the RR_a and T_{begin} parameters by searching for those other pairs of these two parameter values that produce 2×ProfileLogLikelihood values that differ by less than a given amount from the value of the 2×ProfileLogLikelihood evaluated at the MLE (this amount is the C^{th} percentile of the Chi-square distribution with 2 df).

The t's, s's and dr's may not be available at the level of the individual, but the numbers of deaths S and NS in the screening and non-screening arms within each say one or half-year interval of follow-up may be known, along with the value of each "denominator ratio" DR, i.e., the ratio of the person-years lived in the interval by those in the screening and comparison arms. With such data, we can use the same conditioning as above, and regard the value of S for interval centered on t_{mid}, conditional on the total number $S + NS$, of prostate cancer deaths in the interval, as the realization of a binomial random variable with expectation $DR \times RR(t_{mid})/[1+ DR \times RR(t_{mid})]$.

The second derivative of the profile Log-Likelihood can be used, along with cancer-specific and all-cause mortality rates, to calculate in advance what precision/power will be achieved with various numbers of subjects and durations of follow-up.

May 3, 2010

Supplementary Figure A

Timing of effects of screening in disease processes with different natural histories: cumulative cause-specific mortality as reported in the Multicentre Aneurysm Screening Study (MASS) and the European Randomized Study of Screening for Prostate Cancer (ERSPC). The MASS enrolled 68,000 men aged 65-74 and involved a one-time screen with immediate treatment or surveillance of detected abdominal aortic aneurysms. As noted by the authors, "The benefit seen in earlier years of follow-up was maintained in the later years of follow-up, with continued divergence of the cumulative curves of deaths related to abdominal aortic aneurysm in the two groups". The overall mortality rate reduction of 48% (mortality rate ratio 0.52) is an adequate and accurate measure of the impact of screening. The ERSPC enrolled 162,000 men aged 55-69 and involved repeated PSA-based screens 4 years apart. As noted by the authors, "The rates of (prostate cancer) death in the two study groups began to diverge after 7 to 8 years and continued to diverge further over time". The overall mortality rate reduction of 20% (mortality rate ratio 0.80) is an inadequate and inaccurate measure of the impact of screening.

References

[1] Morrison AS. *Screening in Chronic Disease*, first edition New York: Oxford University Press, 1985; 2nd edition New York: Oxford University Press; 1992.

[2] Shapiro S. Evidence on screening for breast cancer from a randomized trial. Cancer. 1977 Jun;39(6 Suppl) 2772-82.

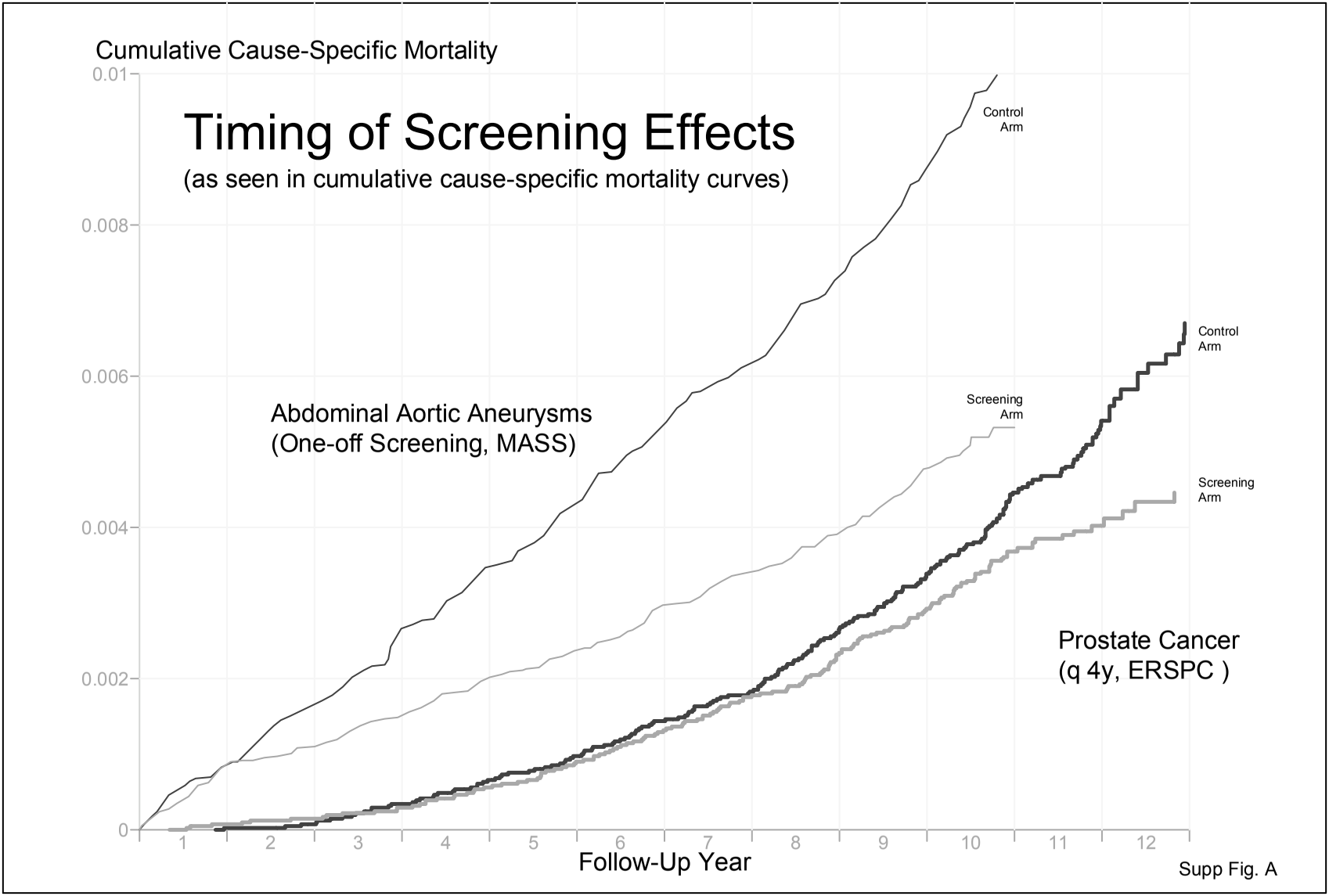
[3] Baker SG, Kramer BS and Prorok PC. Early reporting for cancer screening trials Journal of Medical Screening 2008;15:122–129.

[4] Miettinen OS, Henschke CI, Pasmantier MW, et al. Mammographic screening: no reliable supporting evidence? Available at: http:// image.thelancet.com/extras/1093web.pdf. Accessed July 6, 2005.

[5] Miettinen OS, Henschke CI, Pasmantier MW, et al. Mammographic screening: no reliable supporting evidence? *Lancet* 2002;359:404 – 406.

[6] Hanley JA. Analysis of Mortality Data From Cancer Screening Studies: Looking in the Right Window. Epidemiology 2005; 16: 786-790.

[7] Breslow NE and Day NE. Statistical methods in cancer research. Vol II: the design and analysis of cohort studies. IARC. Lyon, 1987. p 94.


Cumulative Cause-Specific Mortality
Timing of Screening Effects
(as seen in cumulative cause-specific mortality curves)
0.01
0.008
0.006
0.004
0.002
0
Control Arm
Screening Arm
Abdominal Aortic Aneurysms
(One-off Screening, MASS)
Control Arm
Screening Arm
Prostate Cancer
(q 4y, ERSPC)
1
2
3
4
5
6
7
8
9
10
11
12
Follow-Up Year

Supp Fig. A