Author's response to reviews

Title: A Perfect Correlate Does Not A Surrogate Make

Authors:
- Dr Stuart G Baker (sb16i@nih.gov)
- Barnett S Kramer (KramerB@OD.NIH.GOV)

Version: 2 Date: 8 Sep 2003

PDF covering letter
Response to Stephen Duffy

We thank Professor Duffy for the helpful comments.

As a result of his comments, we now highlight (with a separate section heading) the fact that we have presented a graphical representation of the Prentice Criterion. We graphically show that the Prentice Criterion is sufficient for correct inference and that "the intervention should not distort the relationship between true and surrogate endpoints."

However, we also clarify (with a new section heading) that one can, in theory, obtain correct inference without the Prentice Criterion when predicting true endpoint based on the surrogate endpoint. We discuss three approaches that do not rely on the Prentice Criterion.

1. Professor Duffy makes a very interesting proposal to use a second screening with the same intervention in both groups (colonoscopy) to obtain a surrogate endpoint that might be more useful for inference than a surrogate endpoint from the first screening. However, there is an implicit assumption that the original intervention does not affect the relationship between surrogate and true outcomes at the time of the second screening. It is not clear to us the circumstances when this implicit assumption would most likely hold.

2. One of the authors (Stuart Baker) is writing on another paper that clarifies the arguments in Begg and Leung. The issues are beyond the scope of this paper.

Response to David DeMets

We thank Professor DeMets for the helpful comments.

1. We are delighted Professor DeMets thinks we have presented a "nice illustrative example"

2. We agree that one problem with using potential surrogate endpoints that have a high correlation with true endpoint is that "a treatment may have other negative effects not captured by the potential surrogate." Another problem, which is illustrated by Figure 1, is that a potential surrogate endpoint that is highly correlated with true endpoint within randomization groups could still give the wrong inference because of large differences between the slopes of the lines.

3. We agree that, due to possible negative effects not captured by the surrogate, inference from a meta-analysis cannot completely "justify" a surrogate endpoint, even in the best situation where the relationship between surrogate and true endpoints in all previous trials is identical.
4. We believe we had made a similar point. In any case, we try to emphasize this point more in our Discussion section.

**Response to Jeremy M. Taylor**

We thank Professor Taylor for the helpful comments.

**Terminology.**
(a) We changed "surrogate" to "potential surrogate" as appropriate. We agree that this will reduce confusion.
(b) We clarify the difference between an auxiliary variable and a potential surrogate that predicts the true outcome. (In short, with auxiliary variables but not potential surrogates, the true outcome is observed in some subjects)

**Plausibility of Figure 1**

We have now added a section to discuss plausibility of Figure 1.

We do not understand why an intervention central to the pathway of the disease implies that the line for E would more likely be lower than the line for C.

Regarding the heterogeneity argument, we think it is most clear with a binary surrogate endpoint and we have carefully formulated sufficient conditions.

Below we give a heterogeneity argument for measured outcomes, although it does not generally apply when the intercept is zero. Suppose that there are two types of surrogate endpoints: high and low risk. For simplicity, suppose that in group $z$, for all subjects at high risk, the value of the surrogate endpoint is $v_{1z}$ and for all subjects at low risk the value of the surrogate endpoint is $v_{0z}$. Let $\pi_z$ denote the probability of having in the high risk surrogate endpoint in randomization group $z$. Then the mean surrogate endpoint is $\bar{s}_z = \pi_z v_{1z} + (1-\pi_z) v_{0z}$. Suppose that true endpoint depends on high or low risk surrogate endpoint so $\bar{t}_z = k_{1z} \pi_z v_{1z} + k_{0z} (1-\pi_z) v_{0z}$. One can substitute $\pi_z = (\bar{s}_z - v_{0z})/(v_{1z} - v_{0z})$ into the latter equation to obtain $\bar{t}_z$ as a linear function of $\bar{s}_z$. With the appropriate choices of $k_{1z}$ and $k_{0z}$, one can obtain any slope (although unlike Figure 1, the intercept is not zero).

**Missing Reference**

We did not know about the graphic in Wang and Taylor, *Biometrics*, 2002, and are glad to reference it.

**Specific Comments**

1. We have clarified that we are discussing perfect correlation within groups.
2. We agree that correlation does not have any meaning for binary endpoints. We have made appropriate changes in the paper.

3. We agree that a different slope may not be sufficient for the reversal of inference with the potential surrogate endpoint. We now say we need a "sufficiently different slope"

4. We have provided a more rigorous mathematical framework for discussing the implications of "bad" and "innocent" adenomas. The intervention needs to change the ratio of probabilities of "bad" to "innocent" adenomas and decrease the total number of adenomas. This would decrease the number of "innocent" adenomas, increase the number of "bad" adenomas, and decrease the total number of adenomas, as mentioned. Furthermore, if the intervention also increased the rate at which "bad" adenomas progress to colorectal cancer, one might obtain the Figure without an increase in the number of "bad" adenomas.

5. We have included the finasteride trial as good example of a mixture "bad" and "less bad" surrogate endpoints.

6. We have dropped the reference to observational studies.