Author's response to reviews

Title: Epidemiology and Outcomes of Previously Undiagnosed Diabetes in Older Women with Breast Cancer: An Observational Cohort Study Based on SEER-Medicare

Authors:

Robert I Griffiths MS, ScD (bob@outins.com)
Mark D Danese MHS, PhD (mark@outins.com)
Michelle L Gleeson PhD (michelle@outins.com)
Jose M Valderas MD, PhD, MPH (jose.valderas@phc.ox.ac.uk)

Version: 2 Date: 29 November 2012

Author's response to reviews: see over
Dear Dr. Downing:

On behalf of my co-authors, thank you for the opportunity to revise and resubmit our article entitled “Epidemiology and Outcomes of Previously Undiagnosed Diabetes in Older Women with Breast Cancer: An Observational Cohort Study Based on SEER-Medicare” for consideration for publication in *BMC Cancer*. Also, we thank the reviewers for their comments. Below, we detail our responses. In each instance, we have restated (sometimes in abbreviated form) the comment (*italics*) and then provided our response below it.

**Reviewer 1:**

1. *On page 5 in the first paragraph, the statement “there is also evidence many diabetes cases may remain undiagnosed until breast cancer” requires a reference.*

   We have added the reference for this. (Page 5)

2. *In the Inclusion/Exclusion criteria—does the term “had only one primary cancer diagnosed” refer to a single primary during the study period (2001-2005) or ever (so a sequence number of 00)?*

   We have clarified that breast was the first and only type of cancer at the time they were diagnosed. (Page 7) Also, we have clarified that patients were censored from follow-up if there was an occurrence of a second primary cancer. (Page 7)

3. *In the definition of physician outpatient visits, where would OB/GYN visit fall? Is that a primary care physician or medical specialist?*

   We have listed all the types of physicians falling within the category of “primary care physician” including obstetrician/gynecologist. (Page 9)

4. *Since you don’t have any information in your analysis about length of time with diabetes or treatments for diabetes, I think that the discussion about hyperinsulinemia as a cause of advanced stage of cancer at diagnosis should be tempered somewhat. We don’t know that the newly diagnosed cases had uncontrolled glucose for long enough to result in aggressive tumor growth, etc.*

   We have added the following sentences to address this issue. (Page 17)

   “However, since we did not have information on insulin and glucose levels, or on duration of previously undiagnosed diabetes, these findings should be considered as hypothesis-generating requiring laboratory data and information on unobserved confounders for further evaluation. Also, although in the causal diagram previously undiagnosed diabetes precedes advanced stage
cancer diagnosis, since previously undiagnosed diabetes status was ascertained at the same
time as cancer stage, we cannot conclude that previously undiagnosed diabetes caused
advanced stage cancer in our study.”

5. On page 15, at the end of the second paragraph please remove the sentence
“This is consistent with prior research showing that limited health system contact is associated with
advanced cancer stage at diagnosis.”

We have removed this sentence. (Page 16)

6. I would suggest that either you stick to “previously undiagnosed” or change throughout to “newly
diagnosed”.

We have elected to use “previously undiagnosed” throughout the text, including in the abstract and
title. (Multiple Pages)

7. You created an index of preventive services that included a variety of screenings, tests, immunizations.
Did you consider looking at each service separately—in particular mammography and perhaps Pap test?

We did consider this. However, from the outset the scope of this study was ambitious, and involved
conducting 20 multivariate analyses with 4 different outcome measures: previously undiagnosed
diabetes; advanced stage; first treatment; and mortality. Also, unlike other studies in which these
instruments were developed, our primary focus was diabetes and outcome, and not health system
contact and outcomes. Therefore, we did not add this extra level of complexity associated with
examining the effects of individual preventive services.

8. I found the causal diagram piece particularly interesting and Fig 2 clearly shows the complex nature of
the relationships under study. In that regard and given that this is an observational study, did the
authors consider using a propensity score analysis to account for shared risk factors and additionally
unmeasured confounders related to being previously undiagnosed that may also be related to
mortality—BMI, smoking, other access to healthcare variables?

We did consider propensity score analyses. However, elucidating and “validating” our pathway diagram
required that we examine the effects of including/excluding specific covariables, e.g. stage and
measures of prior health system contact, on the associations between previously undiagnosed diabetes
and outcomes. Therefore, we did not perform propensity score analyses. We have added the following
as the explanation for not doing propensity score analysis. (Page 19)

“We considered propensity score analysis. However, elucidating our pathway diagram required
that we examine the effects of including/excluding specific covariables, e.g. stage and measures
of prior health system contact, on the associations between previously undiagnosed diabetes
and outcomes, which would not have been possible had we summarized these effects in a single
propensity score.”

2
Also, while we do not claim special expertise in propensity score analysis, and have used it in other studies, it is our understanding that propensity score analysis does not adjust for unmeasured confounders.


9. Also with regard to the causal diagram, though the authors indicate that previously undiagnosed diabetes is a “cause” or precedes advanced stage cancer diagnosis, they could not actually answer that question in their data as diagnosis of diabetes in the “previously undiagnosed” diabetes group was happening at the same time as cancer diagnosis/workup/treatment whereas in the previously diagnosed group diabetes actually did precede cancer diagnosis.

This is an important point, and we have added the following sentence to clarify. (Page 17)

“Also, although in the causal diagram previously undiagnosed diabetes precedes advanced stage cancer diagnosis, since previously undiagnosed diabetes status was ascertained at the same time as cancer stage, we cannot conclude that previously undiagnosed diabetes caused advanced stage cancer in our study.”

10. As for tables and figures—I did not think that Fig 1 or Fig 3 added much. I felt the timeline was well described in the text. And though Table 2 provides a great deal of information already, it was not immediately clear to me if these were multivariate models with prior health system contact not included on the left and prior health system included on the right. I would suggest adding the bivariate associations to either Table 1 or more likely to Table 2 where there is one column for bivariate OR, then multivariate model w/o prior health system contact, then multivariate model w/prior health system contact.

We have deleted Figure 1. However, we have retained Figure 3 (now Figure 2) since it addresses an important issue Reviewer 2 raises. Namely, it shows how introducing measures of prior-health system contact into multivariate survival analysis impacts the associations between previously undiagnosed diabetes and mortality.

We have revised the title for Table 2 so it is clear it contains multivariate analyses of factors associated with the risk of undiagnosed diabetes, and also the headings in the table to clarify that one analysis includes measures of prior health system contact while the other does not. (Page 29)

The bivariate associations between prior health system contact and previously undiagnosed diabetes are included in Table 1 with the other demographic, clinical, and socioeconomic associations. (Page 28)

**Reviewer 2**

1. For the analysis on cancer stage, did the authors look for a significant correlation coefficient between the two chosen variables to evaluate health system contact (health care visits and preventive care score)? Clinically these variables could be highly correlated (i.e. patients who visit doctors more
frequently are also likely to have preventive screening tests done) and if this is proven statistically, one should be removed from the model.

Yes. We have added the following sentences to clarify (Page 12)

“During the exploratory phase of our study, we did sequential analyses in which we introduced first one and then the second measure of prior health system contact into our models. We found that while the effect of the first introduced was attenuated by the second, in almost all instances both remained statistically and clinically significant. Therefore, both were retained in the models that included prior health system contact.”

2. In the analyses section and figure 2, the authors describe a complicated relationship between 4 different sets of variables: demographic, socioeconomic and clinical characteristics, previously undiagnosed DM, cancer stage and mortality. From the diagram, the authors do not make it clear that previously undiagnosed diabetes is the main predictor of interest – this should be clarified. Second of all, either cancer stage or “demographic, socioeconomic and clinical characteristics” are depicted as being colliders. A collider is a variable that is associated with both the “confounder” variable and the outcome creating a point in the causal diagram where two areas meet head on (as is the case for cancer stage in figure 2). The issue of dealing with such variables is complex – this issue should be discussed in both the methods and discussion session. When one identifies a collider variable one should not include it in the adjusted analysis because this introduces bias, instead one can stratify on this variable or else leave it out of the model altogether. Since the authors are showing the models with and without these variables, it is important that more time should be spent discussing this issue. Otherwise, including the multiple models are confusing and I would suggest the authors only report results from what they deem to be the most “correct” model.

Good resources are:

In Figure 2 (now Figure 1), we have added language to indicate previously undiagnosed diabetes is the main predictor of interest. (Page 36)

Regarding the reviewer’s comments on the collider variable, we have added the following to explain that cancer stage is a collider variable, and explaining the rationale for excluding it from the analysis of factors associated with previously undiagnosed diabetes. (Page 11)

“Figure 1 shows that we hypothesized a directed path (A) from a vector of demographic, socioeconomic, and clinical characteristics to previously undiagnosed diabetes. However, since there are also directed paths from both previously undiagnosed diabetes and the vector of demographic, socioeconomic, and clinical characteristics to cancer stage at diagnosis, cancer stage is a collider variable [47-49]. Conditioning on a collider can open a biasing pathway between two variables, in this case between the vector of patient characteristics and previously undiagnosed diabetes, making it appear that there is an association when in fact none exists. Therefore, in the multivariate analyses of factors associated with previously undiagnosed diabetes, we excluded cancer stage from the vector of independent variables in the models.”
We reviewed and have incorporated both references (now 48,49) the reviewer suggested. (Page 26)

Also, we added the following language in the Discussion, again including both citations. (Page 16)

“We did not include cancer stage as a covariate in the multivariate analyses of factors associated with previously undiagnosed diabetes, because our causal pathway diagram indicates it is a collider [47-49] in this instance. Consequently, conditioning on stage could have opened a biasing pathway (the analysis may have identified an association where none exists) between the vector of patient characteristics and previously undiagnosed diabetes.”

Our rationale for excluding stage from the mortality analyses was slightly different, in that since stage is on the directed path from previously undiagnosed diabetes to survival, adjusting for stage in multivariate analyses may be considered as over-adjustment. Nevertheless, we believe a clinical audience will be interested in the impact of stage on mortality (at least to support that the models are “working” correctly). These are the findings presented in Figure 2.

3. The authors need to strengthen their hypothesis that undiagnosed diabetes can have ‘direct’ effects on cancer stage independent of differences in health care behaviour and patterns, given that they were unable to account for metabolic factors in this association. The authors should acknowledge in their limitations the lack of information regarding diabetes severity and glucose control. Is anything known about the “severity” of unknown diabetes – are these patients more likely to have very high glucose and A1c levels or high insulin levels and pre-clinical glucose levels? If this is known, such information would help in the discussion of causative pathways for the association seen between undiagnosed DM and later stage/worse prognosis.

This comment is similar to #4 from Reviewer 1, and we have added language to the discussion acknowledging that we did not have laboratory data for this study.

4. The authors argue that by adjusting for ‘confounders’, they have demonstrated evidence of a direct (causal) relationship between undiagnosed diabetes and cancer stage. This assertion needs to be tempered, as the possibility of residual confounding due to unmeasured and unknown factors cannot be excluded. Moreover, the substantial reduction in strength of association when adjusting for measured confounders indicates a lack of robustness of this association to confounding. Unmeasured confounders include diabetes severity measures, body-mass index (which can affect clinical detection of early-stage cancers), diabetes treatment/medications (e.g. metformin, which has been shown to affect cancer prognosis), health behaviours, etc... This paper should be viewed as ‘hypothesis-generating’ – further studies will have to evaluate the extent to which the association between undiagnosed diabetes and cancer prognosis is related to health care vs. biologic factors.

We have incorporated the reviewer’s language by stating that our findings regarding previously undiagnosed diabetes and stage should be viewed as hypothesis-generating requiring laboratory data and information on unobserved confounders for confirmation. (Page 17)

In the next paragraph, where we raise the general issue of confounding, we list the other unobserved potential confounders mentioned by the reviewer above. (Page 17)
1. The authors clearly show a strong effect of prior health contact on the magnitude of association between their exposure and outcomes. I have a few questions related to this variable.

a. Given that health care contact is a necessary precursor to having ‘diagnosed diabetes’ based on the study definition (at least one diabetes claim in the 2 years prior), please clarify how 65 patients with ‘diagnosed diabetes’ had no health contact in the 2 years prior.

We have explained that there are two outpatient files in Medicare – physician/supplier and “outpatient.” Since only one contains information on specialty, the absence of a primary care or specialist visit in the physician/supplier file should not be interpreted as absence of any outpatient health system contact. (Page 9)

b. The majority of patients with ‘undiagnosed diabetes’ had some health contact in the 2 years prior to cancer diagnosis – clearly these visits were for something other than diabetes (as there was no claim for diabetes) – is there information on what these visits were for?

Yes. Each claim for services contains ICD-9-CM diagnosis and procedure codes. As discussed above in the response to Reviewer 1, Comment 7, rather than report individual diagnoses and procedures, we summarized other clinical conditions prior to cancer using the NCI Comorbidity Index, and we created the two indices of prior health system contact. The majority of studies based on SEER-Medicare use the NCI Comorbidity Index, based on Charlson, to summarize diagnoses contained in visits prior to cancer diagnosis.

c. Did the authors assess for an interaction between diabetes status and health care contact, to determine its role as an effect modifier in the association?

Yes. In the analysis of mortality, we conducted sensitivity analyses in which we examined the impact of adding the measure of health care contact to models that included diabetes status. These findings are discussed in the Results under the Mortality sub-heading and also presented in Figure 2 of the paper. We show that as expected based on the causal pathway diagram the associations between diabetes status and mortality were attenuated when the measures of health care contact were added to the model. However, as shown measures of health care contact were included in the “base-case” models.

2. The authors should provide a rationale for limiting the cohort to age 65 years or older (due to availability of Medicare data?). Were there other inclusion/exclusion criteria?

We have explained that 65 is the minimum age required to qualify for Medicare benefits. We also explain that since all patients also were required to have 24 months of claims prior to cancer diagnosis, the minimum age in the cohort at cancer diagnosis was 67 years. (Page 7)

Also, we added to this paragraph that patients qualifying for Medicare solely on the basis of disability (< 65 years old) were excluded from the study. (Page 7)
3. It is unclear whether all 93% of eligible Medicare patients > age 65 use Medicare solely for their medical needs, i.e. would all doctor visits and diagnostic claims be captured using Medicare data? For readers who are not familiar with the US databases it would be helpful to clarify this point.

We have added a sentence to the Inclusion and Exclusion Criteria explaining that we restricted the cohort to those with Part A and B coverage because the vast majority of inpatient and outpatient services for these patients are captured within the SEER-Medicare database. (Page 7)

4. The authors state that ‘undiagnosed diabetes’ was based on a ‘new’ claim for diabetes (based on a validated algorithm) within the 6-month peri-diagnostic period. There is no mention of whether this definition has been validated (e.g. based on chart review or survey?). Given that the sensitivity of the diabetes case definition is only 74.4%, it is possible that up to 16% of previously undiagnosed cases were actually pre-existing (or prevalent) diabetes cases that did not have a claim for diabetes in the 24-month look-back period prior to diabetes. As shown by the authors, these patients were less likely to have seen a physician in that period. One suggestion would be to conduct a sensitivity analysis with a longer look-back period to exclude those with more remote prior diabetes claims. The authors should mention the lack of validation and limited specificity of this definition.

We have clarified in the Methods text pertaining to the description of the algorithm, and in the Discussion, that it is a validated algorithm. (Page 8 and 19)

We have not performed a sensitivity analysis extending the look-back period since the algorithm is only validated for 24 months. Also, extending the look-back period, say from 24 to 36 months, would require dropping younger patients with less than 36 months of Medicare eligibility prior to cancer diagnosis.

We have added the following language addressing this issue. (Page 19)

“Further, it is possible that some of the diabetes cases we identified as previously undiagnosed would have been reclassified as previously diagnosed had we extended the look-back period of the algorithm from 24 to 36 months. However, this would have resulted in excluding all patients age 67, who would not have had at least 36 months of Medicare eligibility prior to the diagnosis of cancer.”

5. The authors should explain why they did not include a non-diabetes comparison group.

We have explained that since it was not our intent in this study to compare the outcomes of those with versus without diabetes, we did not include a control group of non-diabetes patients. (Page 20)

6. The authors adjusted for a measure of prior preventive services, which includes mammography. However, given that mammography is a strong predictor of early diagnosis and stage, it would be useful to also compare rates of mammograms between the 2 groups and how that factor specifically influences the association between the diabetes groups and outcomes, if possible.

As discussed above, the measure of preventive services does include mammography, and since our paper already includes a complex array of independent and dependent variables with no fewer than 20
multivariate analyses, we elected not to include a detailed analysis of the components of either the preventive services or the physician contact instruments.

7. The authors examine mortality outcomes, but fail to consider post-cancer variables during the follow-up in their model, such as cancer treatment, surveillance, diabetes-related complications. These are important predictors of survival following cancer, and should be mentioned in the limitations if they were not available.

We have mentioned these in the context of “unmeasured confounders” in the mortality analysis. (Page 19)

8. Where are the results from the analysis relating to effect of undiagnosed diabetes on cancer stage? The text refers readers to table 1, but table 1 is the table of baseline characteristics.

The reference to Table 1 in this section pertains to the bivariate analyses presented there. We did not include a Table for the multivariate analysis of stage. Rather, we report the principal findings in the text. We have clarified this in the Results section. (Page 14)

9. The authors only briefly mention the outcome of time to chemotherapy or radiation, and provide hazard ratios. However there is no information on proportions who received chemotherapy/radiation between groups.

We have added this information to the appropriate section. (Page 15)

“Overall, 479/2,418 (19.8%) received chemotherapy: 18.6% of those with previously diagnosed diabetes and 23.3% of those with previously undiagnosed diabetes. In addition, 662/2,418 (27.4%) received radiation: 28.1% of those with previously diagnosed diabetes and 25.2% of those with previously undiagnosed diabetes.”

10. The follow-up period appears short for mortality (2-6 years). Please provide information on number and rates of death by group, and follow-up time (median, etc.).

We have added this information to the appropriate section. (Page 15)

“Overall, 980/2,418 (40.5%) died during the observation period: 40.2% of those with previously diagnosed diabetes and 41.5% of those with previously undiagnosed diabetes. The estimated median survival based on Kaplan-Meier analysis was 68.6 months in those with previously diagnosed diabetes and 62.3 months in those with previously undiagnosed diabetes.”

11. The authors should elaborate on the implications of the finding that undiagnosed diabetes is only statistically associated with all-cause mortality, and not cancer-specific mortality after adjustment for stage and health care patterns. This suggests that patients with undiagnosed diabetes are overall sicker
and dying of causes other than their cancer. How are these findings specific to breast cancer? This should be addressed more specifically in the discussion.

We have incorporated these points into the section of the Discussion on the mortality. We emphasize that it is not clear whether our findings are specific to breast cancer because the impact of previously undiagnosed diabetes on cancer treatment and cancer mortality may differ for other types of cancers. (Page 20)

Discretionary Revisions

1. The authors conducted a thorough literature review that is well described in the Background. However, there is a lot of detail regarding previous studies that might be better placed in the Discussion section, when comparing this paper with prior research.

Much of the literature cited in the Introduction is referenced again in the Discussion section in the context of our findings: for example, a comparison of our findings on the association between prior health system contact and stage, and those of prior investigators.

2. Figure 3 – it would be helpful to have HR and 95% CI within the graph.

As described in the legend, the triangles in this figure represent the adjusted hazard ratios and the hash marks and lines the 95% confidence intervals.

Again, we thank the reviewers for their comments and the editors for the opportunity to revise and resubmit our manuscript. We have endeavored to address all of the comments while retaining the main focus of the manuscript and managing its length. However, if there are additional comments, please do contact us.

We hope that our article is now acceptable for publication in BMC Cancer.

Kind Regards,

Robert I. Griffiths, MS, ScD
Vice President
Outcomes Insights Inc.
Faculty - Part Time, Division of General Internal Medicine
Johns Hopkins University School of Medicine, Baltimore, MD

Department of Primary Care Health Sciences,
University of Oxford, Oxford, UK

Office Address
340 North Westlake Blvd, Suite 200
Westlake Village, CA, USA 91362

Voice: 805 367 7488
email: bob@outins.com