Author's response to reviews

Title: The influence of nativity and neighborhoods on breast cancer stage at diagnosis and survival among California Hispanic women

Authors:

Theresa HM Keegan (theresa.keegan@cpic.org)
Thu Quach (thu.quach@cpic.org)
Sarah Shema (sarah.shema@cpic.org)
Sally L Glaser (sally.glaser@cpic.org)
Scarlett L Gomez (scarlett.gomez@cpic.org)

Version: 4 Date: 29 July 2010

Author's response to reviews: see over
July 29, 2010

Editorial Board, BMC Cancer

Dear Editorial Board:

We appreciate the helpful comments by the reviewers for our paper, “The influence of nativity and neighborhoods on breast cancer stage at diagnosis and survival among California Hispanic women” (212727736362598). Attached please find a response to each of the comments raised by the reviewers. In the revised paper, we have tracked changes made in response to the reviewer’s comments.

In responding to the reviewer’s comments, we realized that the survival analyses were conducted with follow-up data through 2006, so we updated Table 3 with analyses that used follow-up data through 2007. The nativity and neighborhood hazard ratio estimates did not change appreciably with the additional one-year of follow-up data.

We look forward to hearing from you.

Sincerely,

Theresa Keegan, PhD

Responses to reviewers’ comments

Reviewer 1: Carlos Reyes Ortiz

Comment: This an interesting paper on stage at diagnosis and survival related to Hispanic nativity status and controlling for neighborhood SES and enclave (census block level). It is a well designed study with a complete database (except data imputed, justified and explained by authors).

Since the focus of the paper is on Hispanics (foreign vs. US born), including additional analyses on other ethnic groups (Results- page 9) may be not necessary, for stage at diagnosis and survival multivariate analyses. However, if authors still want to include those, they will need to provide some details about those other ethnic groups such as samples and % distribution, and whether they controlled for neighborhood variables or not.

Response: We agree with the Reviewer that the focus of the paper is on Hispanics; therefore, we have removed the additional analyses with other ethnic groups from the results section (pages 9-10).

Reviewer 2: Luisa Borrell
Comment 1: The introduction does not provide a rationale for the examination of neighborhood SE conditions and ethnic enclave. The issue of the Hispanic paradox could be elaborated further and more specifically to the outcomes in which the paradox applies. At times appears that the authors are not clear about the meaning of the paradox- the paradox was not grounded on living conditions at first (given citations 10 and 11) but individual characteristics. Finally, there are several recent papers regarding the validity of the paradox, i.e., Palloni and Arias 2004, Patel et al 2004. Papers specific to neighborhood context are Cagney et al 2007 and Acevedo-Garcia et al 2007.

Response: We have revised the introduction to include a rationale for examining neighborhood factors and expanded our description of the Hispanic paradox to refer to individual characteristics. In the discussion, we elaborate more on the Hispanic paradox and have added several of the Reviewer’s suggested citations (1, 2).

Comment 2: The second sentence of the first paragraph is awkward and long, perhaps the authors could consider rephrased and divided the sentence into two.

Response: The last sentence of the first paragraph was revised for clarity.

Comment 3: It would be important to avoid generalization to all Hispanic women in California when most Hispanics are Mexican Americans. According to the 2000 Census the Mexican population was 77% in California and recent estimates of the 2006-2008 American Community survey indicate an increase to 83.2%. Thus, it would be misleading to speak about Hispanics when the vast majority of this population self-identified as Mexican Americans.

Response: We agree with the reviewer and have stated (page 13) that our findings are generalizable to a larger Hispanic population of primarily Mexican descent (which represents 77% of California’s Hispanic population).

Comment 4: The aims could be more clearly stated.

Response: We have more clearly stated the aims of the study (page 4).

Comment 5: The prevalence estimates of health care access and risk factors measures while seem as a complement to the proposed analysis of the MS, these analyses do not contribute much to the MS. In fact, the MS will be cleaner without these estimates. Finally, the timing of the California Health interview Survey does not fully overlap with the timing of the California Cancer Registry (CCR) data. Thus, the addition of these data does not make a significant contribution to the MS. In fact, it may add noise to the MS’ results.

Response: While the California Health Interview Survey (2001, 2003) data do not overlap with our entire study period (1988-2005), these data allow us to consider a variety of characteristics in California Hispanic women, by nativity, which can help inform the associations seen in our study. Because the population of California Hispanic women is primarily of Mexican descent, as noted in Comment 2 above, it is important to describe the characteristics of this underlying population in detail. Therefore, we believe including this data strengthens the manuscript.
Comment 6: Study population: This section needs some details on how the authors arrived to a sample of 38,555. Information such the total sample, how many were female, exclusion (this info in the MS on the exclusion needs to be consistent and clearer) and age range included in the analyses.

Response: Because data for this study are from the population-based cancer registry, 38,555 is all Hispanic female residents of California newly diagnosed with invasive breast cancer and reported to the California Cancer Registry (CCR) during the period January 1, 1988 through December 31, 2005, as stated on page 4. From this group of 38,555, we excluded the 860 women who were diagnosed at autopsy or by death certificate only, who had zero or invalid survival time (n=152), or whose address at diagnosis could not be precisely geocoded to determine neighborhood Hispanic enclave (n=708). There were no further exclusion criteria and the characteristics of the study population are presented in Table 1.

Comment 7: Patient and tumor information: The imputation for nativity status is troublesome because this is the main predictor of the analysis. There is a high percent of data missing (28.7%) and method used may have introduced as much bias as excluding those cases. The reviewer is curious about the direction of the bias introduced in citations 20 and 21, i.e., Did the bias under or overestimate the results?

Response: Our previous research indicates that unrecorded birthplace in the cancer registry is more likely among US-born than foreign-born, and is further associated with being alive, less extensive disease, younger age, higher education, English language preference, and admission at certain hospitals (3). We have previously shown that because of the joint association of missing birthplace with birthplace itself and vital status, excluding cases with missing birthplace in this analysis would bias the results away from the null (4, 5). Therefore, these data are not missing at random and our imputation method, which was validated against self-reported nativity data from a series of previously interviewed Hispanic cancer cases (noted on pages 5-6 of the manuscript), would likely be associated with less bias than excluding patients with missing nativity data.

Comment 8: The authors mentioned in page 4 that 708 records were excluded because their data could not be geocoded. However, in the first sentence the authors stated that 96% of the records could only be precisely geocoded to a census block group. This does not add up.

Response: Because we excluded 708 women whose address at diagnosis could not be precisely geocoded to determine neighborhood Hispanic enclave, we revised the description of the previously developed neighborhood socioeconomic status variable to be consistent (page 6).

Comment 9: The authors could provide few details on how the neighborhood SE was developed, i.e., were all the variables added? Were z-score created?

Response: The methods utilized and the variables included in the previously developed neighborhood socioeconomic status variable have been added to the methods (page 6). In addition, we have clarified that all the census variables listed were included in the neighborhood Hispanic enclave index (pages 6-7).
Comment 10: The reference cited for the Hispanic enclave index does not seem to match, please advise.

Response: We previously cited the work by Hu on dietary pattern analysis because the author utilized similar principal components methodology to our study. However, now that we include more information on the development of the neighborhood socioeconomic status variable (Review 2, comment 9 above), we have cited this study instead (page 6).

Comment 11: Some information about the covariates included and justification would be helpful.

Response: We included variables collected by the California Cancer Registry and now provide a reference that provides more details on these variable (6) (page 5). As noted on page 8, we included all variables significant at p<0.10 in univariate models or with a priori hypotheses for inclusion (e.g., age and stage at diagnosis, first course of treatment) in the multivariable survival models.

Comment 12: It is unclear how the Hispanic enclave variable was specified for analysis and why the neighborhood SE was dichotomized.

Response: We now note (page 7) that each breast cancer patient was assigned a Hispanic enclave quintile based on the distribution of this variable across census block groups in California. As noted on page 7, due to high correlations between ethnic enclave and SES, we created a four-level combined variable (low SES and low enclave, low SES and high enclave, high SES and low enclave, and high SES and high enclave) based on the sample size distribution for each variable. These two variables were dichotomized to preserve sample size.

Comment 13: The authors should present the incidence on stage of diagnosis to provide readers with info on why OR was the best measure of association to be used here. In fact, by looking at the incidence of mortality in Table 1, relative risk should have been used as the estimates are >10%.

Response: This manuscript does not look at incidence of breast cancer. Logistic regression was chosen to determine the association between nativity and neighborhood factors and our dichotomous outcome, regional/distant stage at diagnosis. Because the outcome of regional/distant stage at diagnosis, as presented in Table 1, is not rare (39.1% of US-born Hispanics and 41.7% in foreign-born Hispanics were diagnosed with regional/distant disease), we were careful not to interpret the odds ratio as a relative risk in our manuscript.

Comment 14: First sentence of page 7 (Models...): If nativity status was one of the main predictors, why were models adjusted for nativity? Perhaps some info is missing here.
Response: For clarity, we now state that the models “included,” rather than were adjusted for, nativity (page 7).

Comment 15: It is unclear how the authors could evaluate overall survival when cases eligible for the study were 38,555 Hispanic women newly diagnosed with invasive breast cancer. Please explain this issue, what overall exactly means here?

Response: Overall survival refers to our analyses that considered death from all causes, rather than deaths from breast cancer, in women diagnosed with invasive breast cancer.

Comment 16: The analyses performed here were multivariable and not multivariate.

Response: We have replaced the term multivariate with multivariable in the manuscript.

Comment 17: Second sentence, second full paragraph, the effect modification should be evaluated first through interaction terms in multivariable analyses. Why was the proportional hazards assumption assessed visually when a cross-product term between each covariate and the log of the time variable could be used for the test?

Response: As noted on page 8, we evaluated effect modification through interaction terms in the multivariable analyses. The visual assessment was an additional test to ensure that we did not violate the proportional hazards assumption of the Cox model (i.e., survival curves for two strata must have hazard functions that are proportional over time - constant relative hazard).

Comment 18: What was the median follow up time for the study?

Response: The median follow-up time of 69.9 months has been added to the first sentence of the results.

Comment 19: It would be helpful to present the number of block-groups in which the participants were distributed and the range of participants across the block groups. The 45.7% block groups with only one case and 70.5% with two cases add to more than 100%.

Response: We have revised this sentence for clarity; 70.5% of block groups had 2 or fewer cases, with 45.7% of block groups having only one case (page 8).

Comment 20: I will suggest not including the prevalence info from CHIS in this MS.

Response: Please refer to our response to this Reviewer’s Comment 5 above.

Comment 21: Table 1 should present row percents instead of column percents.

Response: In this study, we chose to present column percentages to highlight the distributions of demographic, tumor and neighborhood characteristics among foreign-born and US born Hispanics separately. However, row percents can be calculated with the data provided in Table 1.
Reviewer 3: James S Goodwin

Comment: My major concern is with selection biases. the authors should read an article by Sharon giordano in Cancer published in 2008 (or perhaps 2007 or 2009 – it shouldn't be too hard to find) on selection biases in such studies. It is called something like "use of observational data to assess treatment outcomes: a cautionary note". In it, the authors show how these biases can persist in the face of propensity scores, etc. They strongly recommend that such studies divide deaths by cause of death (cancer vs non-cancer). If the treatment has a strong effect on deaths from non-cancer causes, even after all the controls such as propensity score matching, then that casts real doubt on the validity of the findings. For example, aggressive treatment for cancer shouldn't decrease deaths from diabetes, heart disease, etc. I realize that in this study of metastatic breast CA that the overwhelming percentage of deaths may be from cancer, so maybe this strategy will not work because of small numbers. But the authors should at least try it, and also report on what percentage of the deaths were from cancer, the cause of death data in SEER-Medicare are from the national death index (death certificate data), and have been validated in several studies of SEER data. Otherwise I have no suggestions for change. Nicely done.

Response: Rather than comparing outcomes of different therapies, which was the primary focus in the Giordano et al. (2008) paper (7), the goal of our study was to consider the effects of nativity and neighborhood factors on survival after breast cancer while adjusting for first course of therapy. However, to address the reviewer’s suggestion to look at other causes of death, we conducted additional analyses considering the nativity and neighborhood associations with all cancer deaths and deaths from non-cancer causes. In these analyses (survival based on deaths due to all cancers and to non-cancer causes of death), the hazard ratio estimates for nativity and neighborhood associations (data not shown) were similar to the estimates presented in Table 3. Nativity-specific cause of death distributions were already included in Table 1. As noted in the methods (page 5) and discussion (page 14), we were limited to first course of treatment data in the cancer registry. In response to the reviewer’s concern, we have expanded upon our treatment limitation and cited the Giordano et al., paper (7) (page 14).

References