PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Sexual Minority Population Density and Incidence of Lung, Colorectal, and Female Breast Cancer in California
AUTHORS	Boehmer, Ulrike; Miao, Xiaopeng; Maxwell, Nancy; Ozonoff, Al

VERSION 1 - REVIEW

REVIEWER	Thomas Blank University of Connecticut Storrs, CT, 06269 USA
REVIEW RETURNED	02-Jan-2014

GENERAL COMMENTS	On the one hand, this review can be quite straightforward. The authors' goal is to use data combining the California Cancer Registry about cancer incidence and from the California Health Interview Survey about county-level density of LGB (not T) populations to estimate the relative incidence levels of cancer in those sexual minority populations to the general (presumably heterosexual) population. By and large, that goal is met. That is, the research is carefully done, with appropriate statistical analyses, it is well written, and the data are discussed in relation to the goal. I should also note that the goal is a worthy one, given the general lack of attention to sexual minority populations in research on cancer. The other side of the equation is not weaknesses in the research per se, but the reliance on the so-called "proxy" measure's value, that is, the degree to which the CHIS data at the county level are or are not useful as surrogates for sexual minority status. They might be, of course, but they might also be differentiated on something else (or many "something elses") that may be impacting the incidence of cancer more or at least as much as sexual minority identification or specific behaviors (in terms of sexual activity or in terms of non-sexual activity such as smoking, level of healthy eating). Besides those activity differences, which are not explorable with this data, there also may be non-examined county-level variances, such as density of health care, especially specialized oncology centers, that may, among other things, affect usage of screening, amount of cancer-related health knowledge and literacy, etc. The problem with the argument that these data are useful for the range of purposes indicated is that in previous related studies the authors used a different proxy/surrogate, which was also asserted to be an appropriate measure to be able to extrapolate findings. And, in fact, in the analyses that overlap, the results are different. That begs the question of which (if either) of the approaches is more "accurate" if it could be measured against more individual behavior. Also, the fact that some of the sub-group data seem to go against known variation in behaviors of LGB that are related to specific cancers (that, for example, lung cancer is found to be lower in lesbians and

	not different in gay men to the general population using this method despite the known higher incidence of smoking and/or obesity), makes it difficult to take this approach as indicative of likely individual-level (or even neighborhood level) differences. I would also add that the general pattern of lower or similar incidence in the gay men and lesbian population (with a lot of variation in the bi-men and bi-women groups, in directions that are again different from previous studies) also seems problematic for an access interpretation. I want to emphasize that this kind of article may not be the place for such reflection, and, insofar as the data are obtained and presented well, that may be sufficient to publish the data as a small study that is an interesting step in trying to unravel these important issues. The authors themselves carefully note that, “our study findings…should not be interpreted as evidence of an association between individual sexual orientation and cancer incidence” (p. 13, 51-56). But then the question is, as I have noted, what do they represent? And, in related fashion, what data might in fact reveal associations “between individual sexual orientation and cancer incidence?” There are three other aspects that the authors should address, separate from those larger speculations. One is that there obviously was no way to identify transgender persons in this study. That is unfortunate, and that population (as with bisexuals, who are included here) is even less examined than gay men and lesbians, even though they may have some very specific issues, especially in relation to hormonally-based cancers. Also, the authors should address why the bisexual population seems to be so different, both from gay men and lesbians and between bisexual men and bisexual women. Is a lack of identification with either lesbian or gay sexual minorities but still being a sexual minority somehow related to health behaviors or access? Finally, in the description of strengths and limitations the authors state that, “these county-level differences in cancer incidence provide information for public health planning, previously unavailable” (p. 3, lines 10-12). They do not, however, specify how this would be done. That is, would this lead to targeting by audience (e.g., directly to gays and lesbians—and then how to bisexuals?) or just having more in the way of generic messaging and screening by county for those counties with higher incidence rates?

REVIEWER	Suzanne Haynes DHHS Office on Women's Health Washington, DC 20201
REVIEW RETURNED	06-Jan-2014

GENERAL COMMENTS	LGB density should be defined in the abstract and earlier in the paper. It is difficult to know what the measure is and if it is recorded as a percentage, decimal point, etc or what the range of values are for this unique variable. Please state in common sense terms what this measure is for the readers- who will not be familiar with the term. or concept.The methods on pages 7-8 do not describe it in a way that one can interpret the metric. Perhaps an analogy is the denseness of trees in a forest?

	Since the author is using counts of cases as the outcome variable, these counts should be reflected in the tables. Likewise, the values for the LB density variable should be shown in the tables. On page 8, the authors state that a correction of .36 between lesbian density and bisexual density is low. This level of correlation would not be considered low in most fields, so the fitted models combining both together are questionable. It is not clear why White males age 70-84 were excluded form the data analysis. Are the cancer cases much higher in this group? Shouldn't that deviation be discussed further in the paper. What does this mean? On page 11, first paragraph,there is reference to the fact that small IRR's may not represent a substantial difference in real world terms. Normally one calculates an attributable risk estimate to determine real world effects. The statement does not make sense in that regard. The discussion on page 12 would be easier to inerpret is the studies cited were summarized in a table with pluses and minuses for postive or negative results for each cancer and each LGBT group. The paper does not cite an important paper published in 2012 in the Journal of Women's Health (May 528-533)by Cochran and Mays that looked at the " Risk of Breast Cancer Mortality among women cohabiting with same sex partners: Findings from the National Health Interview Survey." They looked at mortaltiy from breast cancer among US women and found that women in same sex couples , compared to women in different sex relationshiops, has greater age adjusted risk for fatal brreast cancer RR=3.2 . -This paper looks like it was written for a statistics or demographics journal. -The authors should be commended for attempting to determne the link between LGBT status and cancer incidence rates. Analysis of Seer data show higher breast cancer rates in "single" persons, a larger proportion of whom will be lesbian than married persons. A call for the collection of and inclusion of sexual identity in the SEER program would be welcomed.

VERSION 1 – AUTHOR RESPONSE

Reviewer Name Thomas Blank
Institution and Country University of Connecticut
Storrs, CT, 06269
USA
Please state any competing interests or state 'None declared': None

On the one hand, this review can be quite straightforward. The authors' goal is to use data combining the California Cancer Registry about cancer incidence and from the California Health Interview Survey about county-level density of LGB (not T) populations to estimate the relative incidence levels of cancer in those sexual minority populations to the general (presumably heterosexual) population. By and large, that goal is met. That is, the the research is carefully done, with appropriate statistical analyses, it is well written, and the data are discussed in relation to the goal. I should also note that the goal is a worthy one, given the general lack of attention to sexual minority populations in research on cancer. The other side of the equation is not weaknesses in the research per se, but the reliance on the so-called "proxy" measure's value, that is, the degree to which the CHIS data at the county level are or are not useful as surrogates for sexual minority status. They might be, of course, but they might also be differentiated on something else (or many something elses) that may be impacting the incidence of cancer more or at least as much as sexual minority identification or specific behaviors (in terms of sexual activity or in terms of non-sexual activity such as smoking, level of healthy eating). Besides those activity differences, which are not explorable with this data, there also may be non-examined county-level variances, such as density of health care, especially specialized oncology centers, that may, among other things, affect usage of screening, amount of cancer-related health knowledge and literacy, etc.
RESPONSE: To address these concerns, we have made changes to the discussion section.

The problem with the argument that these data are useful for the range of purposes indicated is that in previous related studies the authors used a different proxy/surrogate, which was also asserted to be an appropriate measure to be able to extrapolate findings. And, in fact, in the analyses that overlap, the results are different. That begs the question of which (if either) of the approaches is more "accurate" if it could be measured against more individual behavior. Also, the fact that some of the sub-group data seem to go against known variation in behaviors of LGB that are related to specific cancers (that, for example, lung cancer is found to be lower in lesbians and not different in gay men to the general population using this method despite the known higher incidence of smoking and/or obesity), makes it difficult to take this approach as indicative of likely individual-level (or even neighborhood level) differences. I would also add that the general pattern of lower or similar incidence in the gay men and lesbian population (with a lot of variation in the bi-men and bi-women groups, in directions that are again different from previous studies) also seems problematic for an access interpretation.
RESPONSE: The underlying driver for this reviewer's concern is the absence of sexual orientation data in cancer registries and SEER. If such data were available, one could make statements with the accuracy, we all strive for, when calling attention to cancer disparities. In the absence of such data, we have conducted ecological analyses. We acknowledge and discuss differences in the findings of our earlier studies, which used a proxy measure of sexual orientation compared to the present study, which used sexual identity. We think that both the present analyses and our earlier analyses, which used the Census-derived same-sex partner household as a proxy, are valuable. The strength of the earlier analyses is the ability to look at a larger geographic area, that is, all of SEER, which represents the US population. In the present analyses, we are geographically limited to one state, but have the strength of using aggregated data of self-reported sexual minority status, which make it possible to distinguish between lesbian, gay, and bisexual density. We suggest that differences in the main predictor in combination with a different geographic scope explain the differences in findings. We are encouraged by the similarities between these analyses, in that both types of analyses show significant

associations between county-level sexual minority population density and the three cancers.

I want to emphasize that this kind of article may not be the place for such reflection, and, insofar as the data are obtained and presented well, that may be sufficient to publish the data as a small study that is an interesting step in trying to unravel these important issues. The authors themselves carefully note that, "our study findings…should not be interpreted as evidence of an association between individual sexual orientation and cancer incidence" (p. 13, 51-56). But then the question is, as I have noted, what do they represent? And, in related fashion, what data might in fact reveal associations "between individual sexual orientation and cancer incidence?"
RESPONSE: The reviewer raises the question what the findings represent. These findings are the result of an ecological study that examined the association between county-level sexual orientation data and cancer incidence. Therefore, these findings provide information about sexual minority orientation at the county-level only and do not describe the association between individual sexual orientation data and cancer incidence. Estimation of the individual-level association between sexual orientation and cancer outcomes would require outcomes and reported orientation at the individual level from a representative sample, data which are not presently available for us to study on a large scale. We see these analyses as formative work and a basis to build on. Ecological models have been used for other populations, our contribution is to apply ecological modeling to sexual orientation data.

There are three other aspects that the authors should address, separate from those larger speculations. One is that there obviously was no way to identify transgender persons in this study. That is unfortunate, and that population (as with bisexuals, who are included here) is even less examined than gay men and lesbians, even though they may have some very specific issues, especially in relation to hormonally-based cancers. Also, the authors should address why the bisexual population seems to be so different, both from gay men and lesbians and between bisexual men and bisexual women. Is a lack of identification with either lesbian or gay sexual minorities but still being a sexual minority somehow related to health behaviors or access? Finally, in the description of strengths and limitations the authors state that, "these county-level differences in cancer incidence provide information for public health planning, previously unavailable" (p. 3, lines 10-12). They do not, however, specify how this would be done. That is, would this lead to targeting by audience (e.g., directly to gays and lesbians—and then how to bisexuals?) or just having more in the way of generic messaging and screening by county for those counties with higher incidence rates?
RESPONSE: We provide information about research that shows bisexuals are different from lesbian or gay individuals. While the reasons for such differences are complex and go beyond the scope of this study, we refer in this study to one of the most important aspects: bisexual individuals less likely cohabite with a same-sex partner. This fact likely contributed to the differences in findings, when comparing our earlier analyses to the present analyses. With respect to policies, the present study alone does not provide sufficient details to recommend specific policies. We now include more details about other ecological factors that shall be considered in future research to work towards the identification of factors that can be intervened on, such as the county-level equality of health care for LGB individuals, availability of health care, access issues, as well as risk factors for cancer.

Reviewer Name Suzanne Haynes
Institution and Country DHHS Office on Women's Health
Washington, DC 20201
Please state any competing interests or state 'None declared': NONE

LGB density should be defined in the abstract and earlier in the paper. It is difficult to know what the measure is and if it is recorded as a percentage, decimal point, etc or what the range of values are for this unique variable. Please state in common sense terms what this measure is for the readers- who will not be familiar with the term. or concept.The methods on pages 7-8 do not describe it in a way that one can interpret the metric.

Perhaps an analogy is the denseness of trees in a forest?
RESPONSE: We appreciate this suggestion and recognize that our measure is not a common "household name." As requested, we have made revisions to the abstract. We now explain that we aggregated data at the county-level and refer to our measure as "LGB population density." The term "population density" is a term frequently used in common everyday language, from which we infer that readers will be able to relate to the term "LGB population density" similar to more commonly used terms such as the racial minority density of a neighborhood, city, or county. Further, as requested, we now provide summary statistics for the LGB population density measures. The reviewer's request about the metric is difficult to answer, in that there is no standardized metric for this measure. As we discuss in the work where this metric was first defined, a strict percentage of LGB population is difficult to calculate from available data. Our derived density measure is a fraction, and behaves similarly to a percentage, but it does not range between possible values of 0% to 100%.

Since the author is using counts of cases as the outcome variable, these counts should be reflected in the tables. Likewise, the values for the LB density variable should be shown in the tables.
RESPONSE: It seems we have been unclear in our manuscript. To enhance the clarity of our manuscript, we have now removed the description of the number of cancer cases in California, as it apparently misled this reviewer. We hope this enhances the clarity, in that when we describe this measure, we clearly stated, "We calculated each cancer incidence rate using the total female and male population between age 18 and 84 in each county using the Census data." As requested, we now summarize the descriptive statistics for the density measures. We considered the reviewer's request to provide a table. However, given the complexity and age-gender-specificity of the density measure, as explained in the section on measures, we found no way of showing the density measure in more detail. Indeed we have 58 counties, and within each county 11 age categories, and the 4 race/ethnicity groups.

On page 8, the authors state that a correction of .36 between lesbian density and bisexual density is low. This level of correlation would not be considered low in most fields, so the fitted models combining both together are questionable.
RESPONSE: Using traditional classification for strength of association, a correlation coefficient of 0.36 might be called weak-to-moderate. However in the context of collinear predictors of a regression model, most of the statistical evidence (e.g. Booth 1994 and subsequent citations) suggests a common-sense threshold of roughly 0.7 for concern and attention to the collinearity of a particular pair of predictors. For this reason we do not believe collinearity to degrade the performance of the regression model. Despite our confidence, we appreciate the reviewer suggestion. We fit models after removing one of the two potentially collinear variables, and our primary parameter estimates remained essentially unchanged.

It is not clear why White males age 70-84 were excluded form the data analysis. Are the cancer cases much higher in this group? Shouldn't that deviation be discussed further in the paper. What does this mean?
RESPONSE: The white race, age 70-84, group of LA county, appeared as an influential point on the regression diagnostic plots for the cancer incidence data under study. We then removed this point from the analyses. However, we then retained the data point for all cancers, after we were able to reject any concern that this data point had an undue impact on the results.

On page 11, first paragraph,there is reference to the fact that small IRR's may not represent a substantial difference in real world terms. Normally one calculates an attributable risk estimate to determine real world effects. The statement does not make sense in that regard.
RESPONSE: Per the reviewer's request, we have removed this statement.

The discussion on page 12 would be easier to inerpret is the studies cited were summarized in a table with pluses and minuses for postive or negative results for each cancer and each LGBT group.
RESPONSE: Thank you for the suggestion, we have added a figure to summarize the various results.

The paper does not cite an important paper published in 2012 in the Journal of Women's Health (May 528-533)by Cochran and Mays that looked at the " Risk of Breast Cancer Mortality among women cohabiting with same sex partners: Findings from the National Health Interview Survey." They looked at mortaltiy from breast cancer among US women and found that women in same sex couples , compared to women in different sex relationshiops, has greater age adjusted risk for fatal brreast cancer RR=3.2 .
RESPONSE: We have added this reference and refer to these findings in the discussion.

This paper looks like it was written for a statistics or demographics journal. The authors should be commended for attempting to determne the link between LGBT status and cancer incidence rates. Analysis of Seer data show higher breast cancer rates in "single" persons, a larger proportion of whom will be lesbian than married persons. A call for the collection of and inclusion of sexual identity in the SEER program would be welcomed.
RESPONSE: We agree with the reviewer and have made the requested changes.

VERSION 2 – REVIEW

REVIEWER	Suzanne Haynes Department of Health and Human Services' Office on Women's Health, Washington, D.C., United States
REVIEW RETURNED	02-Mar-2014

- The reviewer completed the checklist but made no further comments