Sexual minority population density and incidence of lung, colorectal and female breast cancer in California

Ulrike Boehmer, Xiaopeng Miao, Nancy I Maxwell, Al Ozonoff

ABSTRACT

Objective: Risk factors for breast, colorectal, and lung cancer are known to be more common among lesbian, gay, and bisexual (LGB) individuals, suggesting they may be more likely to develop these cancers. Our objective was to determine differences in cancer incidence by sexual orientation, using sexual orientation data aggregated at the county level.

Methods: Data on cancer incidence were obtained from the California Cancer Registry and data on sexual orientation were obtained from the California Health Interview Survey, from which a measure of age-specific LGB population density by county was calculated. Using multivariable Poisson regression models, the association between the age–race-stratified incident rate of breast, lung and colorectal cancer in each county and LGB population density was examined, with race, age group and poverty as covariates.

Results: Among men, bisexual population density was associated with lower incidence of lung cancer and with higher incidence of colorectal cancer. Among women, lesbian population density was associated with lower incidence of lung and colorectal cancer and with higher incidence of breast cancer; bisexual population density was associated with higher incidence of lung and colorectal cancer and with lower incidence of breast cancer.

Conclusions: These study findings clearly document links between county-level LGB population density and cancer incidence, illuminating an important public health disparity.

INTRODUCTION

Lung, colorectal and female breast cancer are three of the most commonly diagnosed cancers in the USA, with incidence rates of 79.5, 51.6 and 121.9 per 100 000, respectively, in 2008. Smoking and alcohol use are well-known risk factors for all three of these cancers; overweight and/or obesity has been linked to risk of female breast cancer and colorectal cancer. In addition, nulliparity is associated with increased risk of female breast cancer.
orientation. To overcome this lack of data on cancer disparities by sexual orientation, three previous studies used county-level ecological analyses to relate breast, lung and colorectal cancer incidence to greater sexual minority population density.\textsuperscript{21–23} Previously, we used SEER registry data and US Census data on individuals living in same-sex partnered households as a proxy measure for sexual minority orientation. These ecological studies concluded that greater female sexual minority density (SMD) in a county is associated with greater incidence of breast and colorectal cancer,\textsuperscript{21, 22} while there was a negative relationship between female sexual minority population density and lung cancer incidence.\textsuperscript{23} Greater density of sexual minority men was associated with higher incidence of lung and colorectal cancer.\textsuperscript{22, 23}

While US Census data on same-sex households are a well-established proxy for sexual orientation, these data also have known limitations. From available Census data, we can enumerate households led by same-sex adults, but this surely represents an undercount since we capture only sexual minority individuals who live with a same-sex partner, thereby excluding sexual minority individuals who are living alone or with non-partners. Further, we do not know how many members of a same-sex partnered household identify as LGB. In particular, it is impossible to assess lesbians separately from bisexual women or to assess gay men separately from bisexual men. This is an important limitation, because studies that analysed lesbian women separately from bisexual women concluded that bisexuals fare worse on some health indicators than both heterosexual and lesbian women.\textsuperscript{24–27}

To address the limitations of US Census data, the present study traded a national scope for an improved measure of SMD. We used statewide population-based data on sexual minority identity (gay, lesbian or bisexual) to estimate its relationship to colorectal cancer, lung cancer and female breast cancer incidence.

MATERIALS AND METHODS

The Institutional Review Board deemed this study exempt from protocol review. Data for this research project were taken from the California Health Interview Survey (CHIS), the largest state health survey conducted in the USA. The CHIS employs a two-stage geographic ally stratified random-digit-dial sample of households, surveying one randomly selected adult from each sampled household. The survey is administered in multiple languages, resulting in a large multiethnic/multiracial sample that accurately represents the California population living in households. The CHIS response rate shows no significant non-response bias by demographic characteristics such as age, sex, income, education or employment status;\textsuperscript{28} however, owing to the absence of a sampling frame, non-response by sexual orientation has not been evaluated. More detailed information about the survey methodology can be obtained from the website: http://www.chis.ucla.edu/.

The CHIS collects information biennially, including data on sexual orientation. To ascertain sexual orientation, respondents were asked about their sexual identity, with response choices of heterosexual, LGB along with celibate or other, while recording refusals and do not know responses. We combined 4 years of data, using the adult CHIS surveys from 2001, 2003, 2005 and 2007 to increase the numbers of individuals who report a sexual minority orientation, defined as gay, lesbian or bisexual.

Data on cancer incidence were taken from the California Cancer Registry, which records all cancer cases to monitor the occurrence of cancer among Californians. We chose the cancer data for the years 2001–2008 because these years cover the same time frame as the CHIS sexual minority data. We further restricted our data to men and women aged 18–84, because we focused on adult cancers and also because this is the age range for which the CHIS had sexual orientation data available.

Because there are differences in cancer incidence by age and race, we used the population data from Census 2000 since it has age-specific and race-specific population data available for all the 58 counties in California. From the Census, we also obtained county-level data on poverty, another important confounder of cancer incidence.

Measures

Counts of breast, lung and colorectal cancer incidence were classified into 1 of 11 age categories and four race/ethnicity groups. The 11 age categories were 18–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69 and 70–84, while race and ethnicity consisted of (1) non-Hispanic white, (2) Hispanic, (3) Asian/Pacific Islander and (4) other race/ethnicity. The group of other race/ethnicity combines non-Hispanic blacks (6.79% of cancers), other/unknown race/ethnicity (0.73% of cancers) and non-Hispanic American Indian (0.14% of cancers). We calculated each cancer incidence rate using the total female and male population between ages 18 and 84 in each county using the Census data.

We adjusted for poverty level in our analyses, with poverty level being defined as the percentage of the population living under the Federal poverty level, which has been found to be the most consistent, easily interpretable variable which accurately measures socioeconomic disparities in health outcomes.\textsuperscript{29, 30}

Our main independent variable is derived from the CHIS data on participants’ sexual orientation. We are using these data aggregated at the county level and call this aggregate variable ‘SMD’ to express variation in the density of sexual minority populations in a county. To make these data age specific, we obtained the distribution of sexual minorities, defined as LGBs, across different age groups, and combined this information with the
county-level SMD. Specifically, we obtained the weighted percentage of gay men in a specific age group (denoted as age\textsubscript{weight\_gay}), using the age information on all gay men in the 58 California counties. Then we obtained the weighted percentage of all men in the specific age group (denoted as age\textsubscript{weight\_all}), using the age information on all participants in the 58 California counties. Finally, we obtained the count of all adult gay men (N\textsubscript{gay}) and the count of all adult men (N\textsubscript{all}) in the specific age group, and computed the age-specific gay density as:

\[ \text{Age-specific gay density} = \frac{N_{\text{gay}} \times \text{ageweight}_{\text{gay}}}{N_{\text{all}} \times \text{ageweight}_{\text{all}}} \]

The age-specific lesbian and bisexual population density was computed in a similar way, and we considered bisexual men and women as distinct categories for analyses.

Regression diagnostics indicated that the Los Angeles County, white race, age 70–84, data point was a potentially influential point. We refit all regression models excluding this data point to ensure that our findings are not predominantly dependent on a single observation. Model fit did not improve substantially with the exclusion, and more importantly, changes in estimates of associations between the density measures and cancer incidence were small and well within the SEs. Therefore, we report all regression results with no data exclusions.

**Statistical analysis**

We used descriptive statistics to summarise our outcomes, cancer incidence and our main independent variables, that is, LGB population density, for all 58 California counties. We assessed county-level association between lesbian/gay/bisexual density (LGBD) and cancer incidence rates using multivariable Poisson regression models with the age--race-stratified count of incident cases in each county as the dependent variable, and the LGBD, race, age group and US Census per cent poverty as covariates. We fitted models that considered L/G and B together because the correlation between lesbian/gay density and bisexual density was low—0.36 (p<0.0001) for women and 0.15 (p<0.0001) for men and because our primary parameter estimates remained essentially unchanged when we fitted the two density measures separately. All model selections were based on the goodness of fit of the models assessed by the Akaike Information Criterion (AIC) and residual diagnostics plots. SAS PROC GENMOD was used to fit the models, with the offset term as the logarithm of the US Census age--race-stratified total adult population in the county. The incidence rate ratio (IRR) was presented as the measure of the effect of each predictor, along with its 95% CI and p value. The validity of the assumptions and the goodness of fit of the assumed models were assessed by residual diagnostic plots and goodness-of-fit statistics such as the Deviance statistic. All analyses were carried out in SAS V.9.2 (SAS Institute Inc, Cary, North Carolina, USA).

**RESULTS**

There is considerable variation in the SMD measures by county. Across 58 California counties, the lesbian density measure ranges from 0 to 3.05 (median=0.49; mean=0.66; SD=0.78) and the bisexual density measure from 0 to 4.16 (median=1.24; mean=1.16; SD=0.97). Gay and bisexual male density measures respectively range from 0 to 16.02 (median=0.80; mean=1.10; SD=2.19) and from 0 to 4.07 (median=0.44; mean=0.60; SD=0.76).

**Sexual orientation, demographics and cancer incidence among males**

Controlling for race/ethnicity and poverty, gay density was not associated with county-level lung cancer incidence among men (table 1), although bisexual density was associated with an 8.4% decrease in lung cancer incidence (IRR=0.916, p<0.0001). There were expected significant associations between self-reported race/ethnic category and incidence, and a positive association between poverty and lung cancer incidence.

Controlling for race/ethnicity and poverty, gay density was not significantly associated with colorectal cancer incidence (table 2), but bisexual density among men was associated with a 2.7% increase in incidence of colorectal cancer (IRR=1.027, p=0.02).

**Table 1.** Multivariate Poisson regression analysis for male lung cancer.

| Predictors            | Regression coefficient estimate | IRR    | 95% CI of IRR | p Value |
|-----------------------|---------------------------------|--------|--------------|---------|
| Gay density           | 0.0004                          | 1.0004 | 0.995 to 1.006 | 0.88    |
| Bisexual density      | −0.0875                         | 0.916  | 0.897 to 0.936 | <0.0001 |
| Hispanic vs non-Hispanic white | −0.4243 | 0.654  | 0.637 to 0.672 | <0.0001 |
| Asian/PI vs non-Hispanic white | −0.0674 | 0.935  | 0.910 to 0.960 | <0.0001 |
| Other vs non-Hispanic white | −0.0537 | 1.055  | 1.025 to 1.086 | 0.0003  |
| Poverty               | 0.0134                          | 1.014  | 1.012 to 1.015 | <0.0001 |

*Results also adjusted for age.

IRR, incidence rate ratio; PI, Pacific Islander.
Sexual orientation, demographics and cancer incidence among females

Controlling for race/ethnicity and poverty, each one-unit increase in lesbian density was associated with a 5.1% decrease in incidence of lung cancer (IRR=0.949, p<0.0001; table 3), whereas each one-unit increase in bisexual density was associated with an 11.3% increase (IRR=1.113, p<0.0001). There were significant negative associations between female lung cancer incidence and Hispanic, Asian/Paciﬁc Islander and other races. There was no significant association between female lung cancer incidence and poverty.

Controlling for race/ethnicity and poverty, each one-unit increase in lesbian density was associated with a 2.9% decrease in the incidence of colorectal cancer among women (IRR=0.971, p=0.0095; table 4); bisexual density was not associated with incidence of colorectal cancer among women. Hispanic ethnicity had a signiﬁcant negative association, whereas Asian/Paciﬁc Islander and other race had a signiﬁcant positive association with incidence of colorectal cancer among women. Poverty was not signiﬁcantly associated with incidence of colorectal cancer in women.

Controlling for race/ethnicity and poverty, each one-unit increase in lesbian density was signiﬁcantly associated with a 2.3% increase, and bisexual density with a 3.2% decrease, in incidence of female breast cancer (lesbian: IRR=1.023, p<0.0001; bisexual: IRR=0.968, p<0.0001; table 5). Hispanic, Asian/Paciﬁc Islander and other race/ethnicity had signiﬁcant negative associations with breast cancer incidence. Poverty was modestly but signiﬁcantly associated with decreased breast cancer incidence.

DISCUSSION

This study’s ﬁndings document disparities in the age-adjusted incidence of three cancers across the counties of California: disparities that are associated with the density of sexual minorities, controlling for race/ethnicity and the prevalence of poverty. Our outcome measure provides effect estimates in the form of the percentage change in cancer incidence per one-unit increase in sexual minority population density.

Among men, bisexual density (but not gay density) was signiﬁcantly associated with cancer incidence. Speciﬁcally, bisexual density was associated with lower incidence of lung cancer and less strongly with higher incidence of colorectal cancer. Among women, lesbian density (but not bisexual density) was signiﬁcantly associated with colorectal cancer. Lesbian and bisexual density was signiﬁcantly associated with female breast and lung cancer incidence. However, these two associations were opposite in direction: lesbian density was associated with lower incidence of lung cancer and higher incidence of breast cancer; bisexual density was associated with higher incidence of lung cancer and lower incidence of breast cancer. For each cancer outcome that we studied, Los Angeles county, white race and age 70–84 emerged as an inﬂuential outlying data point. Because Los Angeles county is also the most populous county in California, we carefully considered issues

| Predictors                                | Regression coefficient estimate | IRR   | 95% CI of IRR       | p Value   |
|-------------------------------------------|---------------------------------|-------|---------------------|-----------|
| Lesbian density                           | −0.0521                         | 0.949 | 0.928 to 0.971      | <0.0001   |
| Bisexual density                          | 0.1067                          | 1.113 | 1.078 to 1.149      | <0.0001   |
| Hispanic vs non-Hispanic white            | −0.7012                         | 0.496 | 0.482 to 0.511      | <0.0001   |
| Asian/PI vs non-Hispanic white            | −0.5264                         | 0.591 | 0.573 to 0.609      | <0.0001   |
| Other vs non-Hispanic white               | −0.1527                         | 0.858 | 0.832 to 0.885      | <0.0001   |
| Poverty                                   | 0.0013                          | 1.001 | 0.999 to 1.003      | 0.18      |

*Results also adjusted for age.
IRR, incidence rate ratio; PI, Paciﬁc Islander.

| Predictors                                | Regression coefficient estimate | IRR   | 95% CI of IRR       | p Value   |
|-------------------------------------------|---------------------------------|-------|---------------------|-----------|
| Lesbian density                           | −0.0521                         | 0.949 | 0.928 to 0.971      | <0.0001   |
| Bisexual density                          | 0.1067                          | 1.113 | 1.078 to 1.149      | <0.0001   |
| Hispanic vs non-Hispanic white            | −0.7012                         | 0.496 | 0.482 to 0.511      | <0.0001   |
| Asian/PI vs non-Hispanic white            | −0.5264                         | 0.591 | 0.573 to 0.609      | <0.0001   |
| Other vs non-Hispanic white               | −0.1527                         | 0.858 | 0.832 to 0.885      | <0.0001   |
| Poverty                                   | 0.0013                          | 1.001 | 0.999 to 1.003      | 0.18      |

*Results also adjusted for age.
IRR, incidence rate ratio; PI, Paciﬁc Islander.

Boehmer U, Miao X, Maxwell NI, et al. BMJ Open 2014;4:e004461. doi:10.1136/bmjopen-2013-004461
around the robustness of our results. We conducted a sensitivity analysis, which suggests our findings are not overly dependent on this single data point, lending strength to the overall reliability of our findings.

The differences between results for lesbian/gay density and bisexual density are consistent with an increasing understanding among sexual orientation researchers that—after years of combining lesbian/gay and bisexual individuals into one group due to small sample sizes—differences between these groups come to the forefront once they are separated.8 31–34 These findings reflect methodological improvements on our previous studies of SMD and cancer incidence21–25; those analyses used US Census data on whether a respondent lived in a same-sex partnered household, whereas data for the current analysis rest on self-reported sexual identity from the CHIS, as described above, and distinguish between gay/lesbian and bisexual respondents. A comparison of findings from our other analyses to those of this study shows a mix of consistencies and inconsistencies (figure 1).

Using SEER data, colorectal cancer incidence was significantly and positively associated with both sexual minority men density and sexual minority women density26; that finding aligns with this study’s finding for male and female bisexual density and colorectal cancer, but is inconsistent with the present findings for gay and lesbian density. In the present study, we identified among men no association between gay density and lung cancer, and a negative association between bisexual density and lung cancer; this conflicts with our finding of a positive association for sexual minority men density and lung cancer using a different dataset.23 For female lung cancer, our current findings for lesbian women are consistent with our previous finding of a significant negative association between lung cancer and sexual minority women density,23 although the present study also found a positive association with bisexual density specifically. Finally, findings for breast cancer show some consistency: in the present study, lesbian density has a significant positive association with breast cancer incidence, confirming our earlier findings,21 although the present study also identified a negative association with bisexual density.

A number of factors are likely to contribute to differences between results of earlier Census-based analyses, which did not distinguish between gays/lesbians and bisexuals, and those of the present study, which relied on self-reported sexual identity. The Census data on same-sex partnered households are probably better indicators of gay and lesbian than bisexual individuals, given the data showing that the majority of partnered gay and lesbian individuals have a same-sex partner, whereas partnered bisexual individuals are more likely to be in heterosexual relationships.34 The present study’s ability to distinguish lesbian/gay density from bisexual density is an improvement in the ascertainment of sexual minority orientation. This most likely accounts for identifying effects that are opposite in direction, as for example, lesbian density’s association with a 5.1% decrease and bisexual density’s 11.3% increase in lung cancer incidence. Another factor may be the relative size of the

---

**Table 4** Multivariate Poisson regression analysis for female colorectal cancer*

| Predictors                       | Regression coefficient estimate | IRR   | 95% CI of IRR     | p Value |
|---------------------------------|---------------------------------|-------|------------------|---------|
| Lesbian density                 | −0.0290                         | 0.971 | 0.950 to 0.993   | 0.0095  |
| Bisexual density                | 0.0268                          | 1.027 | 0.994 to 1.062   | 0.1105  |
| Hispanic vs non-Hispanic white  | −0.0670                         | 0.932 | 0.909 to 0.957   | <0.0001 |
| Asian/PI vs non-Hispanic white  | 0.1133                          | 1.120 | 1.090 to 1.151   | <0.0001 |
| Other vs non-Hispanic white     | 0.1616                          | 1.175 | 1.139 to 1.213   | <0.0001 |
| Poverty                         | 0.0020                          | 1.002 | 1.000 to 1.004   | 0.056   |

*Results also adjusted for age.

IRR, incidence rate ratio; PI, Pacific Islander.

---

**Table 5** Multivariate Poisson regression analysis for female breast cancer*

| Predictors                       | Regression coefficient estimate | IRR   | 95% CI of IRR     | p Value |
|---------------------------------|---------------------------------|-------|------------------|---------|
| Lesbian density                 | 0.0225                          | 1.023 | 1.014 to 1.031   | <0.0001 |
| Bisexual density                | −0.0328                         | 0.968 | 0.955 to 0.981   | <0.0001 |
| Hispanic vs non-Hispanic white  | −0.2691                         | 0.764 | 0.755 to 0.774   | <0.0001 |
| Asian/PI vs non-Hispanic white  | −0.1790                         | 0.836 | 0.824 to 0.848   | <0.0001 |
| Other vs non-Hispanic white     | −0.2533                         | 0.776 | 0.763 to 0.790   | <0.0001 |
| Poverty                         | −0.0106                         | 0.990 | 0.989 to 0.991   | <0.0001 |

*Results also adjusted for age.

IRR, incidence rate ratio; PI, Pacific Islander.
subgroups within the LGB adult population: among women, more identify as bisexual than lesbian, whereas among men, more identify as gay. Differences in the findings of this study compared to the earlier studies are also attributable to the difference in geographical scope. The present study is limited to cancer incidence in California, while the previous SEER-based study was representative of the USA. Moreover, the earlier studies used a roughly fourfold larger sample (215 counties compared to the present study’s 58 counties); thus, in the current study, we had lower statistical power to detect associations. Both ecological studies failed to achieve a complete ascertainment of sexual minority status, albeit for different reasons; the Census-based studies relied on an enumeration of same-sex partnered individuals, while the present study relied on population estimates of sexual identity. Methodological differences aside, there may be real-world differences between California and the country as a whole in connections between SMD and cancer incidence.

This study has the inherent limitations of the ecological study design. In particular, because data are at the county level, it is not possible to link sexual minority status to cancer risk at the level of the individual. Thus, our study findings clearly describe links between county-level density measures and cancer incidence and should not be interpreted as evidence of an association between individual sexual orientation and cancer incidence. A finer scale—for example, at the level of the census tract—may provide more insight into patterns of SMD and cancer incidence, but it is not yet clear what is the most appropriate geographic scale for such studies.

Despite these limitations, and some inconsistencies between the work described here and in our previous analyses, a particular strength is that these ecological analyses identified the existence of sexual orientation disparities in cancer incidence at the county level, while the consideration of gender minority density was outside the scope of this study. Future studies are needed to identify ecological causes for the disparity in cancer incidence, which are most likely complex, possibly examining county-level factors related to sexual minorities’ access to the healthcare system and quality of care delivery. So far, a contextual understanding of sexual minorities’ cancer incidence is lacking as well as knowledge about the county or neighbourhood effects on sexual minority populations’ health and health behaviours more broadly. We suggest that the consistency with which our ecological analyses identified disparities in cancer incidence is an opportunity for public health policy interventions, which are larger in scale, considering county-level programmes, rather than interventions that focus on individual behaviour change. We hope additional research can be performed to identify county-level factors, such as density of healthcare, the equality of healthcare for sexual minorities, along with other known cancer prevention behaviours, such as smoking, that may affect cancer incidence.

While increasing attention is being paid to the collection of sexual orientation data in the context of state or federal health surveys, from which one can derive differences in the prevalence of health risk factors, at present, there is no systematic surveillance of sexual or gender minorities with respect to cancer. Owing to this omission, our goal has been to examine the question about cancer disparities by sexual orientation using ecological analyses. Similarly, motivated by a lack of individual-level data on sexual orientation and cancer, a recent study analysed data on women in same sex and opposite sex relationships, that is, a proxy for sexual orientation, concluding that women in same-sex relationships have greater risk for breast cancer-related mortality. So far, evidence is accumulating that sexual minorities carry a disproportionate burden of cancer; therefore, calls are growing louder that cancer registries and SEER ought to collect sexual orientation data to adequately fulfil their mission of monitoring population health, which has to include sexual minority populations as well.

**Contributors** UB originated the study and interpreted the findings. XM conducted the analyses and helped to interpret the findings. NIM led the writing. AO used his statistical expertise to direct the analysis and helped to interpret the findings. All authors contributed in significant ways to the final version of the article by discussing earlier drafts, reviewing and revising the manuscript.

**Funding** This work was supported by the National Cancer Institute at the National Institutes of Health (grant number 1 R03 CA153063-01).

**Competing interests** None.

**Ethics approval** Boston University Institutional Review Board (IRB).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data are available.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

**REFERENCES**

1. U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2008 incidence and mortality web-based report. Atlanta, GA: Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute, 2012.
