

Online social integration is associated with reduced mortality risk

William R. Hobbs^{a,b,c,1}, Moira Burke^d, Nicholas A. Christakis^{e,f,g,h}, and James H. Fowler^{a,i,1}

^aDivision of Social Sciences, University of California, San Diego, CA 92093; ^bNetwork Science Institute, Northeastern University, Boston, MA 02115; ^cInstitute for Quantitative Social Science, Harvard University, Cambridge, MA 02138; ^dFacebook, Menlo Park, CA 94025; ^eDepartment of Sociology, Yale University, New Haven, CT 06520; ^fDepartment of Medicine, Yale University, New Haven, CT 06520; ^gDepartment of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520; ^hDepartment of Biomedical Engineering, Yale University, New Haven, CT 06520; and ⁱSchool of Medicine, University of California, San Diego, CA 92093

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved September 15, 2016 (received for review April 6, 2016)

Social interactions increasingly take place online. Friendships and other offline social ties have been repeatedly associated with human longevity, but online interactions might have different properties. Here, we reference 12 million social media profiles against California Department of Public Health vital records and use longitudinal statistical models to assess whether social media use is associated with longer life. The results show that receiving requests to connect as friends online is associated with reduced mortality but initiating friendships is not. Additionally, online behaviors that indicate face-to-face social activity (like posting photos) are associated with reduced mortality, but online-only behaviors (like sending messages) have a nonlinear relationship, where moderate use is associated with the lowest mortality. These results suggest that online social integration is linked to lower risk for a wide variety of critical health problems. Although this is an associational study, it may be an important step in understanding how, on a global scale, online social networks might be adapted to improve modern populations' social and physical health.

social networks | social media | health | longevity | social support

People with more friends and more social ties in their community tend to live longer (1–4). Many researchers interpret this association as evidence that greater social support and social network integration lead to better health outcomes (4). For example, social integration is thought to improve health by motivating engagement in healthy behaviors (5, 6), improving immunity (7), and reducing inflammation (8). However, nearly all of this work has been conducted in the context of real-world, face-to-face social interactions. As more and more people use online social media to maintain friendships (as of June 2016, about 1.1 billion people use Facebook daily), an open question is whether or not this new context can be used to measure real world social activity and, distinctly, whether online social interactions are similarly associated with better health and increased human longevity.

Many researchers have shown that online access to friends promotes real world social activities (9). This finding suggests that online social media may increase the amount of overall social integration. To the extent that people use social media to coordinate and engage in healthy face-to-face social behavior, we might therefore expect a positive relationship between online use and health. However, it may also be the case that spending time on social media reduces the amount of time available for offline socializing. If so, then social media use might be more like watching television, which tends to crowd out social activity and, along with other forms of sedentary behavior, has been associated with worse health outcomes (10). Of course, because social media use is not randomly assigned, use might also be a proxy for other unmeasured traits.

To help adjudicate whether online social media use has a positive or negative relationship with health, we test here the associations between online social connection and human mortality. Using several different indicators, we measured deidentified counts of six-months' online social media activity of 12 million people on the website Facebook and evaluated whether these

measures were associated with decreased mortality risk in a two year follow-up.

Results

Before analyzing online social connection and social media behavior, we compare mortality rates for the Facebook population vs. the population-at-large. In these analyses, we control for age and gender differences between the two groups, as well as a coarse proxy for race/ethnicity [based on data from the US Census Bureau (11); results are shown in *SI Appendix*] to account for known health disparities and slightly different levels of Facebook use by race/ethnicity. The age- and gender-matched mortality rate for the “full” population of Facebook users (*Materials and Methods*) was 63% of the rate in the California voter record (our data-matching benchmark). This association may result from difficulties in matching Facebook users to vital records. To more confidently evaluate the relative health of Facebook users compared with the general population, we focus our analysis on the “voter” subpopulation, which includes only those Facebook users also present in the California voter record. That is, we compare voters who are on Facebook to those who are not. The age- and gender-matched mortality rate for Facebook users within the voter record population was ~88% of Facebook nonusers within the voter record population. In other words, the risk of dying in a given year is about 12% less for Facebook users than non-Facebook users.

We disaggregate this comparison by cause of mortality. Mortality due to sexually transmitted diseases, several types of cancer, unintentional injuries, drug overdoses, and suicides did not

Significance

People who have stronger social networks live longer. However, can we say the same about online social networks? Here, we conduct such a study. Using public California vital records, we compare 12 million Facebook users to nonusers. More importantly, we also look within Facebook users to explore how online social interactions—reflecting both online and offline social activity—are associated with longevity. We find that Facebook users who accept more friendships have a lower risk of mortality, but there is no relationship for those who initiate more friendships. Mortality risk is lowest for those with high levels of offline social interaction and moderate levels of online social interaction.

Author contributions: W.R.H., M.B., N.A.C., and J.H.F. designed research; W.R.H. analyzed data; and W.R.H., M.B., and J.H.F. wrote the paper.

Conflict of interest statement: M.B. is a Facebook employee. W.R.H. was a Facebook research intern in 2013.

This article is a PNAS Direct Submission.

Data deposition: We have created a dataset that conforms to the agency's guidelines for working with small cell sizes (20), and we will make this available to researchers who request it from the corresponding author.

Freely available online through the PNAS open access option.

¹To whom correspondence may be addressed. Email: whobbs@ucsd.edu or fowler@ucsd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1605554113/-DCSupplemental.

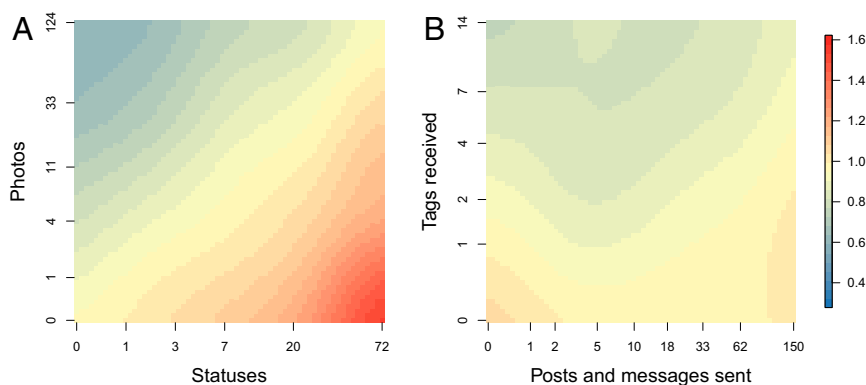


Fig. 2. Mortality risk as a function of two social media activities. (A) All-cause mortality hazard ratio estimates for combinations of text-based activity (as measured by posting statuses) and photo-based activity (as measured by posting photos). (B) All-cause mortality hazard ratio estimates for combinations of text-based directed communications (as measured by posts with tags and messages sent) and photo-based directed communications (as measured by photo tags received). The x and y axes are the quantiles of activities shown in the labels (each tick shows the respective quantile median). Colors show risk of mortality at each combination (red is higher; blue is lower). Results were estimated in a Cox proportional hazard model with interacted indicators for each decile of activity (fewer than 10 categories where the number of users with 0 activity spanned more than one decile). To present similar relative hazard scales (where a hazard of 1.0 corresponds to an average risk), we used the [0,0] interaction quantile as the reference category in the undirected analysis (no photos, no statuses) and the [1, 1] interaction quantile as the reference category in the directed analysis (one sent post or message, one received photo tag). We smoothed the quantile estimates with an approximate Nadaraya-Watson kernel smoother set to a bandwidth of 0.9.

relationships in directed communications on Facebook. Previous research suggests directed communication online is linked to increases in perceived social support, while undirected communication is not (13). Fig. 2 shows all-cause mortality hazard ratio estimates for text-based directed communications (as measured by posts with tags and messages sent) versus photo-based directed communications (as measured by photo tags received). We focus on received tags rather than sent tags for photo-based interactions because they usually indicate that the recipient is in the photo and therefore engaging in a real-world social activity (for comparisons of sent and received tags, see *SI Appendix*). Here, we find, once again, an inverse relationship between photo-based activity and mortality risk. However, there is now a nonlinear relationship with online-directed communications: extremely high and extremely low levels of online-directed communication are both associated with higher mortality. We replicate these results in a proportional hazard model that includes both of these measures and an additional nonlinear term for the online communication measure (*SI Appendix*). This model suggests that mortality risk is lowest for those who are tagged in many photos (highest decile) and who engage in a moderate level of online-directed communication (fourth to sixth deciles).

Last, we examine whether online social activities more strongly predict mortality due to causes that are more likely to be related to social factors. We present cause-specific estimates in order, from least expected to be predicted by social support to most expected. Social support and integration studies tend to find stronger relationships between social support and cardiovascular-caused mortality than cancer-caused mortality (4), and past work on substance abuse and suicide leads us to expect the largest effects for these causes (14–16).

Fig. 3A shows that the number of online friendships is not significantly related to decreased mortality due to cancer but is for cardiovascular disease (91%; 95% CI: 87–96%) and even more so for drug overdose (78%; 95% CI: 70–87%) and suicide (73%; 95% CI: 66–80%). Moreover, when we separately analyze initiated and accepted friendships, the results suggest that accepted friendships are driving the overall relationship, as we previously showed in Fig. 1.

Fig. 3B shows that increased photo uploads is associated with reduced mortality for all major causes except suicide. In contrast, there is a small positive relationship between status updates and

cancer, perhaps reflecting an effort by Facebook users to seek social support or broadcast updates online once they are diagnosed with a chronic illness. Status updates appear to be uncorrelated with mortality for drug overdoses and suicides.

Finally, Fig. 3C shows that sent text-based communications are generally unrelated to mortality risk for these causes, but received communications tend to predict higher risk of mortality due to cancer (108%; 95% CI: 104–112%) and lower risk due to drug overdose (88%; 95% CI: 80–96%) and suicide (82%; 95% CI: 74–90%). Once again, this association suggests that social media is being used by cancer victims to broadcast updates, which elicit received messages, and the contrast between cancer (a positive relationship) and other causes (a negative relationship) may help to explain the nonlinear relationship observed with all-cause mortality in Fig. 2. Meanwhile, received photo tags, our strongest indicator of real-world social activity, are strongly inversely associated with all types of mortality except those due to cancer, and the inverse relationship is strongest with drug overdose (70%; 95% CI: 64–77%) and suicide (69%; 95% CI: 63–76%).

All of the relationships in Fig. 3 remain significant with a Bonferroni correction for 11 comparisons, with the exception of (i) drug overdose and friendships initiated, (ii) drug overdose and sent communications, and (iii) suicide and sent communications.

Discussion

This large-scale study measured the extent to which online social life predicts differences in mortality rates, confirming that it does so in the same way that offline social life does. Our results suggest that people who use online social media experience lower mortality rates than those who do not. Although this finding is very likely to be explained at least in part by differences in socioeconomic status between those who have access to social media and those who do not, we show that among a higher status group (registered voters), this difference in mortality persists. To the extent that online social media platforms like Facebook provide an opportunity to maintain social relationships, they may also indirectly provide people with greater capacity to receive social support and encourage socially-motivated behaviors that may prevent illness.

Among those who do use social media, overall network size is associated with better health. Just like numerous past studies of real world social networks, we find that people with more friends online are less likely to die than their disconnected counterparts. This evidence contradicts assertions that social media have had a

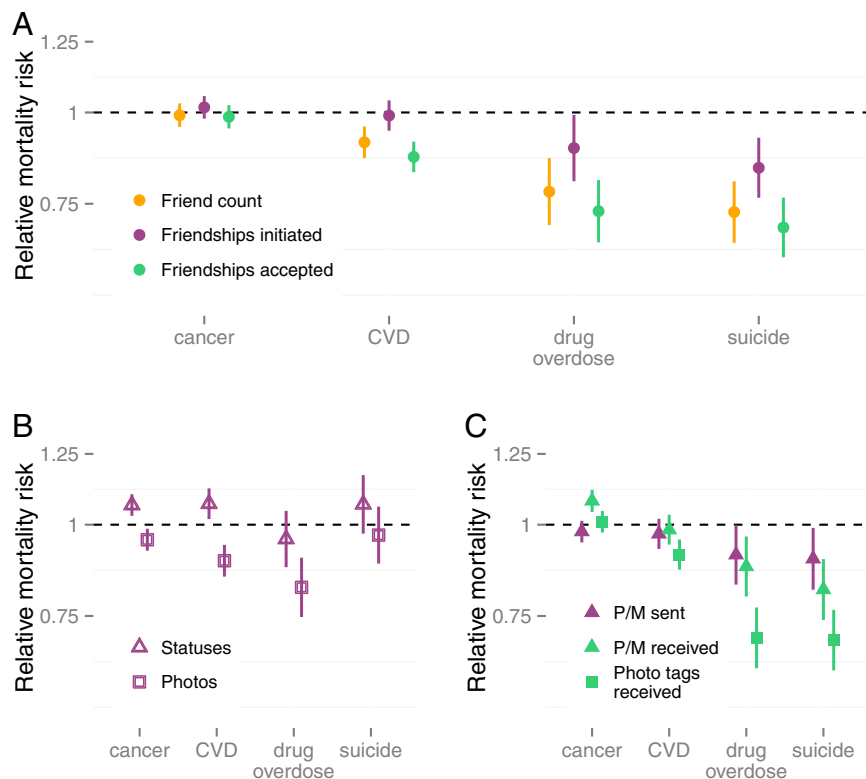


Fig. 3. Cause-specific mortality risk as a function of various Facebook activities. (A) Cause-specific mortality hazard ratio estimates for Facebook friendships (total, initiated, and accepted counts). (B) Cause-specific mortality hazard ratio estimates for undirected activities (social media communications without a specific recipient). (C) Cause-specific mortality hazard ratio estimates for directed activities (social media communications with a specific recipient). Points indicate cause-specific mortality risk estimates and vertical bars show 95% CIs. Actions by the subject are shown in purple, actions by the subject's friend are shown in green, combined actions are shown in orange, circles denote friending activities, triangles denote text-based actions, and squares denote photo-based actions. All variables are logged (after adding one to the activity count), scaled by their SD (so that a unit change is an SD and comparable across activities), and centered at their means. The x axis is the cause of death (an extended table for more and more finely discriminated causes of death is included in *SI Appendix*), and the y axis is the relative mortality risk associated with a SD change in the relevant activity, estimated in a Cox proportional hazard model. All activity categories were estimated separately due to high collinearity among them. "PM" here is an abbreviation for "posts and messages."

net-negative impact on health. Both comparisons between users and nonusers and between low users and high users suggest that social media use is predictive of lower mortality. If social media use were extremely unhealthy, we would expect to find an overall positive relationship between use and mortality, but we do not.

Moreover, those measures of online social activity that are most predictive of reduced mortality are precisely those that are most likely to promote or otherwise indicate offline social interactions. In fact, when we separately analyze activities like posting statuses and posting photos, we find that, underneath the overall relationship between use and reduced mortality, there is an opposite association with behaviors that are likely to be disconnected from the offline world.

Notably, the relationships between offline and online social activity vary by underlying cause of mortality in informative and theoretically consistent manners. Cardiovascular disease is more strongly related to social factors than cancer, and associations between cardiovascular disease and social isolation are stronger in offline interactions than in online ones. Online social interactions and social support most strongly predict better health for underlying causes of death related to mental illness and substance abuse (i.e., drug overdose and suicide), where we expect the largest social support effects.

A major advantage of using online data are that we can distinguish between friendships sought and friendships accepted, and our results show that the relationship between friendship and reduced mortality is driven by others' perception of closeness and desire to connect online. These results suggest that better health is

unlikely to be determined solely by an individual's ability to meet more people or to seek out connection. Instead, it depends on the likelihood that, once having met, social interactions will continue, and/or others will perceive and maintain the friendship. Unfortunately, the finding suggests that interventions that try to increase capacity to seek support may not have the intended effect of improving health.

This study has many limitations. For one, our study may have limited external validity because Facebook is unique among social media sites, and online platforms are constantly changing; by analogy, efforts to predict real world flu using online searches failed when the algorithms were not updated to reflect changing use patterns (17). Another concern is that our measures of mortality risk cover just 2 y for a single state in the United States (California). It is possible that we might find different relationships in studies with longer follow-up or in different states or countries. However, the most notable limitation involves the classic difficulty in distinguishing association from causation that limits all observational studies. Although we show many relationships between social media use and reduced mortality risk, we have not provided evidence of a causal relationship here.

We cannot say that spurring users to post more photos on Facebook will increase user longevity. On the other hand, observational studies are often an important first step for better understanding new phenomena. We hope this study plays a role in spurring interest in online social effects on health, just as Lisa Berkman and Leonard Syme's classic paper in 1979 spurred interest in the relationship between social support and longevity (1).

In particular, the observational relationships reported here suggest that online applications that explicitly promote offline social life may generate positive health outcomes. Indeed, some scholars are currently designing, implementing, and testing such applications (18), and we encourage others to do so as well.

Materials and Methods

To conduct this study, we needed access to Facebook data and also to publicly available vital records. Our study protocol was approved by three different bodies: the institutional review board at the University of California, San Diego (UCSD); the State of California Committee for the Protection of Human Subjects; and the Vital Statistics Advisory Committee at the California Department of Public Health. In addition, the research was reviewed by Facebook's internal review group. The UCSD institutional review board approved a waiver of informed consent for analysis of existing data.

We worked directly with Facebook to analyze aggregate counts of use data. We restricted our analysis to California where we could also access public vital records containing cause of mortality. We identified all California-based Facebook users who (i) signed up before October 2010, so they would have at least 90 d of experience with the site before January 2011, when we started measuring social media use; (ii) listed a first name (or nickname), last name, and date of birth not shared by any other person in California, so that these individuals could be linked to other records uniquely; and (iii) were born between 1945 and 1989, because use at the time was rare among people with earlier birth years, and some people mistakenly reported their birth year as 1990 because it was the default when joining Facebook. We also required first names and last names to appear at least once (independently) in the California voter record, and we omitted users with a birthday of January 1 because it was the default; 12,689,047 profiles fit these criteria (the full population), of whom 4,011,852 were also present in the California voter record (by complete first name, last name, and date of birth: the voter subpopulation). To preserve privacy, after automatically matching to public records, all analyses were performed on deidentified, aggregate data. All data were observational; no one's experience on the site was different from usual.

Once we identified the eligible population, we compared user information (first name or nickname, last name, and date of birth) to California Department of Public Health vital records for 2012 and 2013 to ascertain mortality status and cause of mortality. For the period of study, the Facebook and California voter record populations differed in their age and gender distributions because many people on Facebook tend to be younger than the population of California voters. To ensure age and gender covariate balance in our analyses, we compared all deceased individuals on Facebook to a stratified random sample of nondeceased individuals (SI Appendix, Fig. 5) from the full and voter populations described above. There were 179,345 people in our age- and gender-based probability sample of Facebook users born between 1945 and 1989, of whom 17,990 died between January 2012 and December 2013; 89,597 were also present in the California voter record, of whom 11,995 had died between January 2012 and December 2013.

We categorized underlying causes of mortality in 17 specific categories, as well as 4 broader categories (cancer, cardiovascular disease, drug overdose,

suicide), based on codes of the *International Classification of Diseases, Tenth Revision* (19). We categorized methods of interaction on Facebook using two basic dichotomies: directed (e.g., messages) versus undirected (e.g., status updates) and text-based (e.g., wall posts) versus photo-based (e.g., photo tags). We further distinguished directed actions as outgoing (i.e., sent) or incoming (i.e., received). These activity categories corresponded to the major variance components of a principal component analysis of Facebook activities (shown in SI Appendix).

We used the Cox proportional hazard model to estimate relationships between Facebook activities and mortality. The start time was the user's age on January 1, 2012, and the end time was the user's age when deceased or as of December 31, 2013. We classified the failure event as 1 for deceased in "all-cause" models and 0 otherwise and 1 for mortality from specific causes (e.g., 1 for mortality caused by cancer) and 0 otherwise. All models are stratified by gender and, to account for possible socioeconomic status confounders, include controls for (i) Facebook sign-up date; (ii) access method most commonly used on the site (e.g., the www.facebook.com webpage, the mobile page m.facebook.com, or a native app on a smartphone); and (iii) whether the individual used a smartphone application to access Facebook social tools during the observation period. The additional socioeconomic status control "highest education levels listed by friends on Facebook" (shown in SI Appendix) did not substantively alter the results.

For the proportional hazard models, all Facebook use variables were logged [with $\log(x + 1)$ so that 0 values were well-defined] because online counts of behavior are skewed (most users have low activity and some users have very high activity). Intuitively, this transformation assumes that one additional photo tag (or any other action) is less important for good/bad health and social support once a user already has many of them. For other low, medium, and high use comparisons, we discretized variables by deciles, combining zero-activity deciles when they spanned multiple deciles. This discretization allows us to estimate nonlinear effects of Facebook use, as well as the effects of unusually high levels of online activity. We report details of all models in SI Appendix.

To compare the Facebook population to the California voter record population, we weighted the population in the California voter record to exactly resemble the distribution of age and gender on Facebook.

P values for all estimates are two-tailed. We computed CIs using robust SEs. Because we are more confident that the voter subsample represents real people, we show the results for this subsample in the main text, but we also show results for the full sample in SI Appendix.

The California Department of Public Health prohibits the release of individual-level data. However, the agency does allow the release of aggregated data that protects data confidentiality. We have created a dataset that conforms to the agency's guidelines for working with small cell sizes (20), and we will make this available to researchers who request it from the corresponding author.

ACKNOWLEDGMENTS. We thank Cameron Marlow, Lada Adamic, Danny Ferrante, Arturo Bejas, Pete Fleming, Will Nevius, and Molly Jackman for their support on this project.

- Berkman LF, Syme SL (1979) Social networks, host resistance, and mortality: A nine-year follow-up study of Alameda County residents. *Am J Epidemiol* 109(2): 186–204.
- House JS, Landis KR, Umberson D (1988) Social relationships and health. *Science* 241(4865):540–545.
- Smith KP, Christakis NA (2008) Social networks and health. *Annu Rev Sociol* 34(1): 405–429.
- Holt-Lunstad J, Smith TB, Layton JB (2010) Social relationships and mortality risk: A meta-analytic review. *PLoS Med* 7(7):e1000316.
- Ford ES, Ahluwalia IB, Galuska DA (2000) Social relationships and cardiovascular disease risk factors: Findings from the third national health and nutrition examination survey. *Prev Med* 30(2):83–92.
- Umberson D, Crosnoe R, Reczek C (2010) Social relationships and health behavior across life course. *Annu Rev Sociol* 36(1):139–157.
- Cohen S, Doyle WJ, Skoner DP, Rabin BS, Gwaltney JM, Jr (1997) Social ties and susceptibility to the common cold. *JAMA* 277(24):1940–1944.
- Loucks EB, Berkman LF, Gruenewald TL, Seeman TE (2006) Relation of social integration to inflammatory marker concentrations in men and women 70 to 79 years. *Am J Cardiol* 97(7):1010–1016.
- Rainie H, Wellman B (2012) *Networked: The New Social Operating System* (MIT Press, Cambridge, MA).
- Thorp AA, Owen N, Neuhaus M, Dunstan DW (2011) Sedentary behaviors and subsequent health outcomes in adults: a systematic review of longitudinal studies, 1996–2011. *Am J Prev Med* 41(2):207–215.
- US Census Bureau (2000) Genealogy data: Frequently occurring surnames from census 2000. Available at www.census.gov/topics/population/genealogy/data/2000_surnames.html. Accessed March 24, 2014.
- Jones JJ, et al. (2013) Inferring tie strength from online directed behavior. *PLoS One* 8(1):e52168.
- Burke M, Kraut R (2016) The relationship between Facebook use and well-being depends on communication type and tie strength. *J Comput-Mediat Comm* 21: 265–281.
- Conwell Y, Duberstein PR, Caine ED (2002) Risk factors for suicide in later life. *Biol Psychiatry* 52(3):193–204.
- Rosow I, Lauritzen G (1999) Balancing on the edge of death: Suicide attempts and life-threatening overdoses among drug addicts. *Addiction* 94(2):209–219.
- Hawton K, Houston K, Haw C, Townsend E, Harriss L (2003) Comorbidity of axis I and axis II disorders in patients who attempted suicide. *Am J Psychiatry* 160(8):1494–1500.
- Lazer D, Kennedy R, King G, Vespignani A (2014) Big data. The parable of Google Flu: Traps in big data analysis. *Science* 343(6176):1203–1205.
- Patrick K, et al. (2014) Design and implementation of a randomized controlled social and mobile weight loss trial for young adults (project SMART). *Contemp Clin Trials* 37(1):10–18.
- World Health Organization (1990) *International Statistical Classification of Diseases and Health Related Problems* (World Health Organization, Geneva), 10th Ed.
- California Department of Public Health Office of AIDS (2014) *CDPH/OA Internal Guidelines for Working with Small Cell Sizes* (California Department of Public Health Office of AIDS, Sacramento, CA). Available at <https://www.cdph.ca.gov/programs/aids/Documents/SmallCellGuidance.pdf>. Accessed June 13, 2016.