3. United States Department of Energy, Pacific Northwest National Laboratory, Electricity Infrastructure Operations Center. Situational Awareness. 2013. Available at: http://eioc.pnnl.gov/research/sitawareness.stm. Accessed April 17, 2015.

4. Carley KM. Smart agents and organizations of the future. In: Lievrouw LA, Livingstone S, eds. *Handbook of New Media: Social Shaping and Consequences of ICTs*. Thousand Oaks, CA: Sage Publications; 2002:206–220.

5. Wang F-Y, Carley KM, Zeng D, Mao W. Social computing: from social informatics to social intelligence. *IEEE Trans Intell Syst*. 2007;22(2): 79–83.

6. Hsu CE, Chambers WC, Herbold JR, Calcote JC, Ryczak RS, DeFraites RF. Towards shared situational awareness and actionable knowledge: an enhanced, human-centered paradigm for public health information system design. *J Homeland Security Emerg Manage*. 2010;7(1).

7. Endsley MR, Jones DG. *Designing for Situation Awareness: An Approach to User-Centered Design*. Boca Raton, FL: CRC Press; 2012.

8. US Government Accountability Office. *Public Health Information Technology: Additional Strategic Planning Needed to Guide HHS's Efforts to Establish Electronic Situational Awareness Capabilities. GAO-11-99 Report to Congressional Committees*. Washington, DC: US Government Accountability Office; 2010.

9. Lipowicz A. HHS doesn't have a plan for IT pandemic surveillance, GAO says. *FCW*. December 20, 2010. Available at: http://fcw.com/articles/2010/12/20/hhs-not-fulfilled-pandemic-surveillance-plan.aspx. Accessed December 15, 2014.

10. Yasnoff WA. How to cause information technology disasters. Information Training for Public Health Advisors (lecture online). University of Washington, Seattle. July 2, 1997. Available at: http://faculty.washington.edu/ocarroll/infrmatc/managing/howto/howto.htm. Accessed January 12, 2015.

11. Center for Disease Control, Board of Scientific Counselors. National Center for Public Health Informatics summary report. Atlanta, GA. November 20, 2008. Available at: http://www.cdc.gov/maso/facm/pdfs/BSCNCPHI/20081120_BSCNCPHI_Minutes.pdf. Accessed January 12, 2015.

Addressing the Challenges of Research With Small Populations

Diane M. Korngiebel, DPhil, Maile Taualii, PhD, MPH, Ralph Forquera, MPH, Raymond Harris, PhD, and Dedra Buchwald, MD

Public health policy relies on accurate data, which are often unavailable for small populations, especially indigenous groups. Yet these groups have some of the worst health disparities in the United States, making it an ethical imperative to explore creative solutions to the problem of insufficient data.

We discuss the limits of widely applied methods of data aggregation and propose a mixed-methods approach to data borrowing as a way to augment sample sizes. In this approach, community partners assist in selecting related populations that make suitable "neighbors" to enlarge the data pool.

The result will be data that are substantial, accurate, and relevant to the needs of small populations, especially for health-related policy and decision-making at all levels. (*Am J Public Health*. 2015; 105:1744–1747. doi:10.2105/AJPH.2015.302783)

WHEN PRESIDENT OBAMA signed US Executive Order 13515, he declared that no community should be invisible.[1] Yet for policymakers, the health status of small population groups, especially the indigenous peoples of the United States, remains largely hidden from view. Consistent epidemiological data are needed to inform policy decisions and resource allocation from the community level to the national level. For small population groups, such as American Indians, Alaska Natives, and Native Hawaiians, national reports and public data sets typically fail to provide sufficiently detailed information. Amassing enough accurate data requires innovative solutions, especially because small groups tend to have the largest health disparities. The scarcity of high-quality data means that these groups are often omitted from research agendas—or as the president put it, "Smaller communities in particular can get lost, their needs and concerns buried in a spreadsheet."[1]

As academics who conduct health research in small populations, we use community-based participatory methods within a theoretical framework that encompasses the social determinants of health. Our experience suggests some useful ways to address the problem of scarce data. One approach is to disaggregate data that lump together dissimilar populations, such as Native Hawaiians and Asian Americans, because aggregation can mask health disparities. Another approach is to augment data on extremely small populations by using statistical methods that borrow data from other groups with pertinent similarities to the population of interest. However, given the pitfalls inherent in data borrowing, we recommend qualitative methods that empower small communities to partner with academic researchers in selecting appropriate "neighbors," whose adoption will maintain both the relevance and the distinctiveness of the resulting data pool. In the next sections, we describe a collaborative, multiperspective approach with broad application for small groups throughout the United States.

INDIGENOUS GROUPS AS A CASE STUDY FOR SMALL POPULATIONS

Many data sources are available for monitoring population health trends, but their quality and relevance vary widely. State and national surveys have low response rates from special population groups, especially indigenous peoples.[2] The resulting surveillance data therefore may be insufficient in quantity and lacking in critical details. At least part of this shortcoming can be attributed to culturally discordant survey content and ineffective data collection methods.[3]

Another key shortcoming is ethnic and racial misclassification, a widely reported issue for American Indians and Alaska Natives.[4–7] Racial misclassification of American Indians residing off-reservation is especially common in official documents such as death certificates, with studies reporting misclassification rates that range from 45% to 57% in some areas.[5,8] Such discordance can render population data all but meaningless.

Yet another serious problem ensues when two or more groups of unequal size are collapsed into a single category, such as Asians/Pacific Islanders: data on the smaller groups are overwhelmed by data on the larger groups. In the Asian/Pacific Islander category used in many national reports, Native Hawaiians and other Pacific Islanders represent only 4% of the aggregated sample, whereas Asian Americans represent 96%.[9] The annual age-adjusted death rate for Asians/Pacific Islanders calculated on the basis of this aggregation is 350 per 100 000 persons, which compares very favorably with the rate of 524 per 100 000 for the overall US population. However, the age-adjusted death rate for Native Hawaiians alone is dramatically higher, at 901 per 100 000.[10-13] This instance of misguided aggregation obscures a significant public health disparity for a vulnerable indigenous community.

Nevertheless, more limited approaches to data aggregation can still result in bias. The US Office of Management and Budget disaggregates Pacific Islanders from the Asian/Pacific Islander category to create a smaller subset: Native Hawaiians and Other Pacific Islanders. Tellingly, the bland acronym applied to this category—NHOPI—masks group distinctions.[14] Although the indigenous inhabitants of the Pacific archipelagoes share similar histories, languages, cultures, and lifestyles, their differences are substantial. Chief among them are differential histories of colonization, which leave distinctive traces on population health. Native Hawaiians in particular have a unique relationship with the United States, articulated by US Public Law 103-150 (informally known as the Apology Bill), which recognizes the illegality of US occupation of indigenous lands. As this complex example shows, any approach to aggregation runs the risk of producing biased data and hiding both subtle and overt trends.

Finally, the use of state and national data repositories, such as those generated by the Behavioral Risk Factor Surveillance System and the US Census Bureau, presents numerical challenges. Calculating percentages and disease rates for a given group requires defining a denominator to express the total population.[5,15] However, available population statistics may be outdated or in conflict with other sources, particularly when census or Behavioral Risk Factor Surveillance System data disagree with tribal rolls and clinic records.

Even when indigenous population statistics are well collected, the numbers might be too small to create realistic epidemiological profiles. For example, according to the 2010 US Census, the "American Indian/Alaska Native alone" category represents only about 0.9% of the total US population, whereas the "Native Hawaiians and Other Pacific Islander alone" category represents only 0.2%.[16] Such small numbers render these populations virtually invisible to health planning agencies.

The sheer weight of numbers drives funding, policy decisions, and health care priorities. Public health programs rely on epidemiological data to determine which populations are most in need of intervention. The resulting priorities inform resource allocation at all levels for education, outreach, and clinical activities.[2] Conversely, a lack of epidemiological data is often interpreted as evidence that no problem exists rather than as a reflection of inadequate data collection. As a result, research that produces relevant, accurate, and data-driven descriptions of small populations is indispensable to the mitigation of health disparities.

The limitations of data collection and epidemiological surveillance create a break in the so-called information cycle: inadequate population data impede advocacy on behalf of small groups, restricting their capacity to intervene politically and reducing the availability of resources for research programs and interventions that can improve population health. In this way, the absence of data can both perpetuate and exacerbate poor health for indigenous peoples.

Data to assist in identifying health trends and measuring the effectiveness of health interventions in small populations are vital for federal and state program planning. They are even more crucial to the populations themselves, who need to make informed, empowered choices regarding their health needs at the individual and community level. High-quality, population-specific data are essential for educating communities, both on the existing burden of disease and on locally relevant progress toward improving health.[17]

BORROWING DATA TO INCREASE RELEVANT SAMPLE SIZE

The small sizes of indigenous populations magnify the challenges of health disparities research, especially if the health condition under study is rare, as with certain genetic diseases or cancers.[18] In these instances, available estimates of disease-specific mortality or morbidity often show substantial random variation.[19] Whenever both the number of people affected and the total population size are small, the true rate of disease in a population can be impossible to calculate. Estimates become unstable, and accurate determinations of incidence or prevalence are precluded.[3]

Therefore, to make meaningful inferences about a small group, statisticians often borrow data from similar, neighboring groups. This approach can produce results that are still germane to the small population while enhancing statistical precision by increasing sample size.[3] For data borrowing to work well, however, only suitable neighbors should be incorporated to augment the sample.

Selecting the appropriate groups to combine raises complex questions. Who decides which similarities are relevant and whether two or more populations are similar enough to aggregate: the populations themselves or some outside entity? If the outside entity is an academic institution, how can a small group work collaboratively with researchers to define their own "neighbors"? Who should facilitate the conversation between communities and researchers? How should research questions inform the selection of appropriate neighbors? Which sociodemographic and epidemiological characteristics need to be considered?

Answering these questions involves identifying the most pertinent social determinants of health, which are the political, economic, environmental, social, and cultural dimensions that define the fabric of life for people and communities. By constraining living conditions, these determinants become the "causes of the causes" of health, good or otherwise.[17,20] For example, dispossession from land is a key social determinant of health that applies to most indigenous populations[20] and one that should figure in any process of neighbor selection. Yet it may be overlooked by

outsiders who lack the lived experience that community members share. As a result, the communities themselves must lead efforts to identify appropriate neighbors on the basis of shared determinants of health. Otherwise, data borrowing and aggregation can lead to false or irrelevant conclusions.[18]

To implement an effective data borrowing scheme, researchers need to use approaches that marry qualitative and quantitative methods, because their application in conjunction can yield richer results. The World Health Organization (WHO) explicitly recommends a mix of qualitative and quantitative methodologies to redress health inequities. Every attempt at data borrowing should begin by eliciting community input through standard methods of qualitative data collection, such as key informant interviews and focus groups. Thematic analyses of the resulting data can then assist biostatisticians and epidemiologists in choosing the best quantitative approaches to aggregation. This hybrid approach addresses context and evaluates information in terms of its potential to answer research questions rather than its position in a traditional hierarchy of evidence.[20] Given the potential rewards, we encourage researchers and their community partners to embrace the additional effort needed to combine these approaches.

The key point to remember is that data must be useful, and their use must be appropriate. Ensuring appropriate use depends on leveraging the vital knowledge of community members, including community leaders and local health care professionals. This community knowledge can inform subsequent reporting and advocacy.

As an example, we offer our own experience in a preliminary study with five American Indian tribes in the Pacific Northwest; results are in preparation for publication. Our goal was to identify the factors that tribal members considered most important in aggregating their own health data with those of other American Indian communities. To achieve this goal, we conducted 10 key informant interviews, as well as numerous on-site focus groups that included a total of 39 participants across all five tribes.

Community members identified an extensive list of variables that they considered significant. Subsequent thematic analyses, whose results were presented to the original participants for verification, grouped these variables into seven major categories: geographic proximity, community type (e.g., rural vs urban, coastal vs inland), cultural similarities, presence or absence of local environmental contamination, type and severity of health concerns, similarities in access to health care services, and generational cohort. It is unlikely that researchers without a detailed understanding of community concerns could have arrived at such a categorization, which is far more sensitive than a simple reliance on geographic or temporal proximity. Assistance from community partners is therefore essential to ensure that data aggregation focuses on the characteristics most salient to the community and to the health issues under study.

ENHANCING HEALTH-RELATED DECISION-MAKING

Accurate epidemiological descriptions shed light on the health needs of small populations and thereby assist in health-related policymaking and decision-making at all levels. Small populations will be more involved in their own health if they have relevant local data to inform their health priorities. WHO guidelines support multilevel empowerment to address health needs, in a framework that emphasizes the informed participation of individuals and their communities in their own health-related decisions, often through local ownership of data and health monitoring.[17,20]

The strong likelihood that indigenous health disparities will simply worsen if left unaddressed merits a renewed emphasis on investigating disease and the behaviors that lead to it. In pursuing these goals, we offer a few simple guidelines that draw on the findings of the WHO Commission on the Social Determinants of Health.[20] We endeavor to follow these guidelines in our own research, as illustrated by the preliminary work we reported earlier.

First, community members should be full participants and, if possible, leaders in collecting and analyzing data as well as in designing and implementing evidence-based interventions.[20] In this way, the research process will be enriched by community knowledge and indigenous expertise.[21,22] Second, to the degree possible, data ownership should remain in the hands of the communities themselves. Third, the problem of small numbers should be addressed by a process of data borrowing that is informed at all stages by input from community members. Their participation will help to ensure that data are borrowed from appropriate neighbors who share the most relevant social determinants of health.

The rich data sets that result will enable both top-down (governmental) and bottom-up (community and individual) approaches to improving health outcomes and increasing health equity. This approach is particularly vital for both rural and urban indigenous populations.

Currently, researchers have no alternative to methodologies and data sets that are unsuitable for small populations. The framework that we recommend needs formulation, testing, and application. To achieve this ambitious goal, the National Institutes of Health, the Centers for Disease Control and Prevention, and other entities at the highest levels of US health policy should encourage more creative approaches to data-driven health care. State-of-the-art statistical methods, applied through a community-based approach, have the potential to inform health policies that produce positive health outcomes for indigenous communities and other small populations across the United States. ∎

About the Authors

Diane M. Korngiebel is with the Department of Biomedical Informatics and Medical Education, University of Washington, Seattle. Maile Taualii is with Native Hawaiian and Indigenous Health, Office of Public Health Studies, University of Hawaii, Honolulu. Ralph Forquera is with the Seattle Indian Health Board, Seattle. Raymond Harris is with the Department of Epidemiology, University of Washington, Seattle. Dedra Buchwald is with the Departments of Epidemiology and Medicine, University of Washington, Seattle.

Correspondence should be sent to Diane M. Korngiebel, DPhil, Department of Biomedical Informatics and Medical Education, Box 357240, University of Washington, Seattle, WA 98195 (e-mail: dianemk@uw.edu). Reprints can be ordered at http://www.ajph. org by clicking the "Reprints" link.

This article was accepted May 16, 2015.

References

1. The White House, Office of the Press Secretary. Remarks by the President at AAPI Initiative Executive Order Signing and Diwali Event. October 14, 2009. Available at: http://m.whitehouse.gov/the-press-office/remarks-president-aapi-initiative-executive-order-signing-and-diwali-event. Accessed August 13, 2013.

2. Knutson K, Zhang W, Tabnak F. Applying the small-area estimation method to estimate a population eligible for breast cancer detection services. *Prev Chronic Dis.* 2008;5(1):A10.

3. Andresen EM, Diehr PH, Luke DA. Public health surveillance of low-frequency populations. *Annu Rev Public Health.* 2004;25:25–52.

4. Sugarman JR, Lawson L. The effect of racial misclassification on estimates of end-stage renal disease among American Indians and Alaska Natives in the Pacific Northwest, 1988 through 1990. *Am J Kidney Dis.* 1993;21:383–386.

5. Thoroughman DA, Frederickson D, Cameron HD, Shelby LK, Cheek JE. Racial misclassification of American Indians in Oklahoma State surveillance data for sexually transmitted diseases. *Am J Epidemiol.* 2002;155:1137–1141.

6. Rhoades DA. Racial misclassification and disparities in cardiovascular disease among American Indians and Alaska Natives. *Circulation.* 2005;111:1250–1256.

7. Bertolli J, Lee LM, Sullivan PS, AI/AN Race/Ethnicity Data Validation Workgroup. Racial misidentification of American Indians/Alaska Natives in the HIV/AIDS Reporting Systems of five states and one urban health jurisdiction, U.S., 1984-2002. *Public Health Rep.* 2007;122:382–392.

8. Arias E, Schauman WS, Eschbach K, Sorlie PD, Backlund E. The validity of race and Hispanic origin reporting on death certificates in the United States. *Vital Health Stat 2.* 2008;(148):1–23.

9. US Census Bureau. 2010 Census Redistricting Data, Pub L No. 94–171. Summary File. US Department of Commerce (2011).

10. Hoyert DL, Kung HC. Asian or Pacific Islander mortality, selected states, 1992. *Mon Vital Stat Rep.* 1997;46(1 suppl): 1–63.

11. Ghosh C. Healthy People 2010 and Asian Americans/Pacific Islanders: defining a baseline of information. *Am J Public Health.* 2003;93(12):2093–2098.

12. Johnson DB, Oyama N, LeMarchand L, Wilkens L. Native Hawaiians mortality, morbidity, and lifestyle: comparing data from 1982, 1990, and 2000. *Pac Health Dialog.* 2004;11:120–130.

13. National Center for Health Statistics. *Health, United States, 2005: With Chartbook on Trends in the Health of Americans.* Hyattsville, MD: Centers for Disease Control and Prevention; 2005.

14. Office of Management and Budget. Federal Register notice of October 30, 1997, 62 *Federal Register* 58782-587901997 (1997).

15. Srinivasan S, Guillermo T. Toward improved health: disaggregating Asian American and Native Hawaiian/Pacific Islander data. *Am J Public Health.* 2000;90(11):1731–1734.

16. US Census 2010. 2011. Available at: http://www.census.gov/2010census. Accessed June 3, 2011.

17. Forde I, Raine R. Placing the individual within a social determinants approach to health inequity. *Lancet.* 2008;372:1694–1696.

18. Sue S, Dhindsa MK. Ethnic and racial health disparities research: issues and problems. *Health Educ Behav.* 2006;33: 459–469.

19. Riggan WB, Manton KG, Creason JP, Woodbury MA, Stallard E. Assessment of spatial variation of risks in small populations. *Environ Health Perspect.* 1991;96: 223–238.

20. Commission on the Social Determinants of Health. *Closing the Gap in a Generation: Health Equity Through Action on the Social Determinants of Health.* Geneva, Switzerland: World Health Organization; 2008.

21. Israel BA, Schulz AJ, Parker EA, Becker AB. Review of community-based research: assessing partnership approaches to improve public health. *Annu Rev Public Health.* 1998;19: 173–202.

22. Marmot MG, Bell R. Action on health disparities in the United States: commission on social determinants of health. *JAMA.* 2009;301:1169–1171.

September 2015, Vol 105, No. 9 | American Journal of Public Health · *Korngiebel et al.* | Peer Reviewed | Commentaries | 1747