HHS Public Access

Author manuscript

J Perinatol. Author manuscript; available in PMC 2023 September 01.

Published in final edited form as:

J Perinatol. 2023 March ; 43(3): 385–386. doi:10.1038/s41372-022-01469-4.

Avoiding Loss of Native Individuals in Birth Certificate Data

Kayla Holloway[1], **Joshua Radack, MS**[2], **Molly Passarella, MS**[2], **Angela M Ellison, MD, MSc**[3,4], **Barbara H Chaiyachati, MD, PhD**[4,5], **Heather H Burris, MD, MPH**[2,4,6], **Diana Montoya-Williams, MD**[2,4,6]

[1]University of Pennsylvania, Philadelphia, PA, USA

[2]Division of Neonatology, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[3]Division of Emergency Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[4]Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

[5]Division of General Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[6]Leonard Davis Institute, University of Pennsylvania, Philadelphia, PA, USA

American Indian or Alaskan Native (AIAN) families are disproportionately affected by preterm birth compared to non-Hispanic White families (11.6 vs. 9.1% in 2020).[1] Persistent inequities across the lifespan result in shorter life expectancies among AIAN individuals than any other racial group in the United States.[2] One challenge to improving health and healthcare of AIAN families is the lack of representation in research.[3] This is partly due to the small proportion of AIAN individuals in the US population (constituting roughly two percent). While the gold standard in analyses of racial disparities in health is to use self-identified race,[4] advanced statistical models often cannot accommodate very small sample sizes.[5] As such, AIAN individuals often are absorbed into an "other" category when analyzing health outcomes, including preterm birth. Such aggregation can mask actionable disparities.

On the Pennsylvania (PA) birth certificate, there are two ways in which parents can indicate their race and ethnicity. Parents can check boxes and they can use a "write-in" option. The aim of this study was to determine whether using write-in data could meaningfully increase the sample size of AIAN populations in perinatal research. We used PA birth certificate data from 2006 to 2014, restricted to singleton births linked to hospital discharge data, to determine how many additional births among AIAN we could capture with examination of each write-in response. We then examined how preterm birth rates and their 95% confidence intervals differed among AIAN people when restricting to individuals who checked the AIAN box or when expanding to include write-in responses. Finally, we performed a power

Corresponding Author: Diana Montoya-Williams, MD, Roberts Building for Pediatric Research, 2716 South Street, Suite 19-361; Philadelphia, PA, 19146; montoyawid@chop.edu.

Author Contributions: KH, DMW and HHB conceptualized of this study. AE and BC contributed to the design of the work. JR and MP acquired and analyzed the data. All authors contributed to the interpretation of the data. KH drafted the manuscript and all other authors critically revised the manuscript. All authors approve of the final version and agree to be accountable.

Disclosures: The authors have no conflicts of interest to disclose.

calculation to demonstrate the importance of increasing sample sizes to study relatively rare events such as preterm birth.

There were 1 089 429 births in the study period, 905 (0.08%) of which were among individuals who checked only the AIAN box. There were an additional 11 895 (1.1%) individuals who checked the AIAN box in addition to another box, who were classified as multiracial and not included in the analysis. Nearly half of these multiracial individuals (n=5 277) identified as Black in addition to AIAN. There were 222 individuals who checked the "other" box and wrote-in a response consistent with AIAN identities. The most common write-in responses in descending order were: Cherokee (n=50), Taino (n=23), Native American (n=21), and Blackfoot (n=8). All other written-in responses had a frequency of one. Including these responses increased the sample size to 1,109 births among AIAN individuals – a 22.5% increase. The preterm birth rate among AIAN people without including write-in responses was 10.1% (95% CI: 8.1%, 12.0%). When write-in responses were included, the preterm birth rate was slightly higher with an expected narrowing of the confidence interval (10.8%, 95% CI: 9.0%, 12.7%). A power curve shows the improvement in power to detect relative changes in preterm birth rates with an alpha of 0.05 when the larger group is used (Figure 1). Using this figure, one can see that if researchers were hoping to test, for instance, whether the preterm birth rate had increased by 20%, the larger group of 1 109 births would have roughly 10% more power than the original group of 905 births to detect this increase.

In conclusion, inclusion of write-in designations of AIAN populations in analyses of birth certificate data has substantial potential to increase representation and avoid loss of information. In addition, future inclusion of individuals who include AIAN among their identities may shed light on intersectional ways in which health is affected by structural and interpersonal racism.

Our findings highlight the importance of better capturing people who identify as AIAN within epidemiologic research. In this study, we relied on manual coding of written responses to improve our accuracy in identifying AIAN birthing people. However, the field of natural language processing has evolved significantly in recent years, with ongoing work to improve its applicability to biomedical research.[6] Funding and incorporating natural language processing, and other machine learning techniques, into epidemiologic studies represents one strategy for more meaningful analyses of health outcomes of AIAN people. Furthermore, efforts to develop more complete national vital statistics datasets should be a public health priority. Improved identification of AIAN individuals in research should continue to be a research goal, given the potential to develop interventions to improve birth outcomes in this marginalized population that faces enormous health disadvantages throughout the lifespan.

Funding:

Dr. Montoya-Williams was supported by the National Institutes of Health K23HD102526.

Data Availability:

Data used in in this study may be acquired from the Pennsylvania Vital Statistics Records office. We also welcome requests for analytic collaborations.

References

1. Osterman M, Hamilton B, Martin JA, Driscoll AK, Valenzuela CP. Births: Final Data for 2020. Natl Vital Stat Rep. 2021;70(17):1–50.
2. Arias E, Xu J, Curtin S, Bastian B, Tejada-Vera B. Mortality Profile of the Non-Hispanic American Indian or Alaska Native Population, 2019. Natl Vital Stat Rep. 2021;70(12):1–27.
3. Pacheco CM, Daley SM, Brown T, Filippi M, Greiner KA, Daley CM. Moving forward: breaking the cycle of mistrust between American Indians and researchers. Am J Public Health. 2013;103(12):2152–2159. [PubMed: 24134368]
4. On Racism: A New Standard For Publishing On Racial Health Inequities. Accessed July 6, 2022. 10.1377/forefront.20200630.939347/
5. McNeish DM, Stapleton LM. The Effect of Small Sample Size on Two-Level Model Estimates: A Review and Illustration. Educ Psychol Rev. 2016;28(2):295–314.
6. Moradi M, Samwald M. Improving the robustness and accuracy of biomedical language models through adversarial training. J Biomed Inform. 2022;132:104114. [PubMed: 35717011]

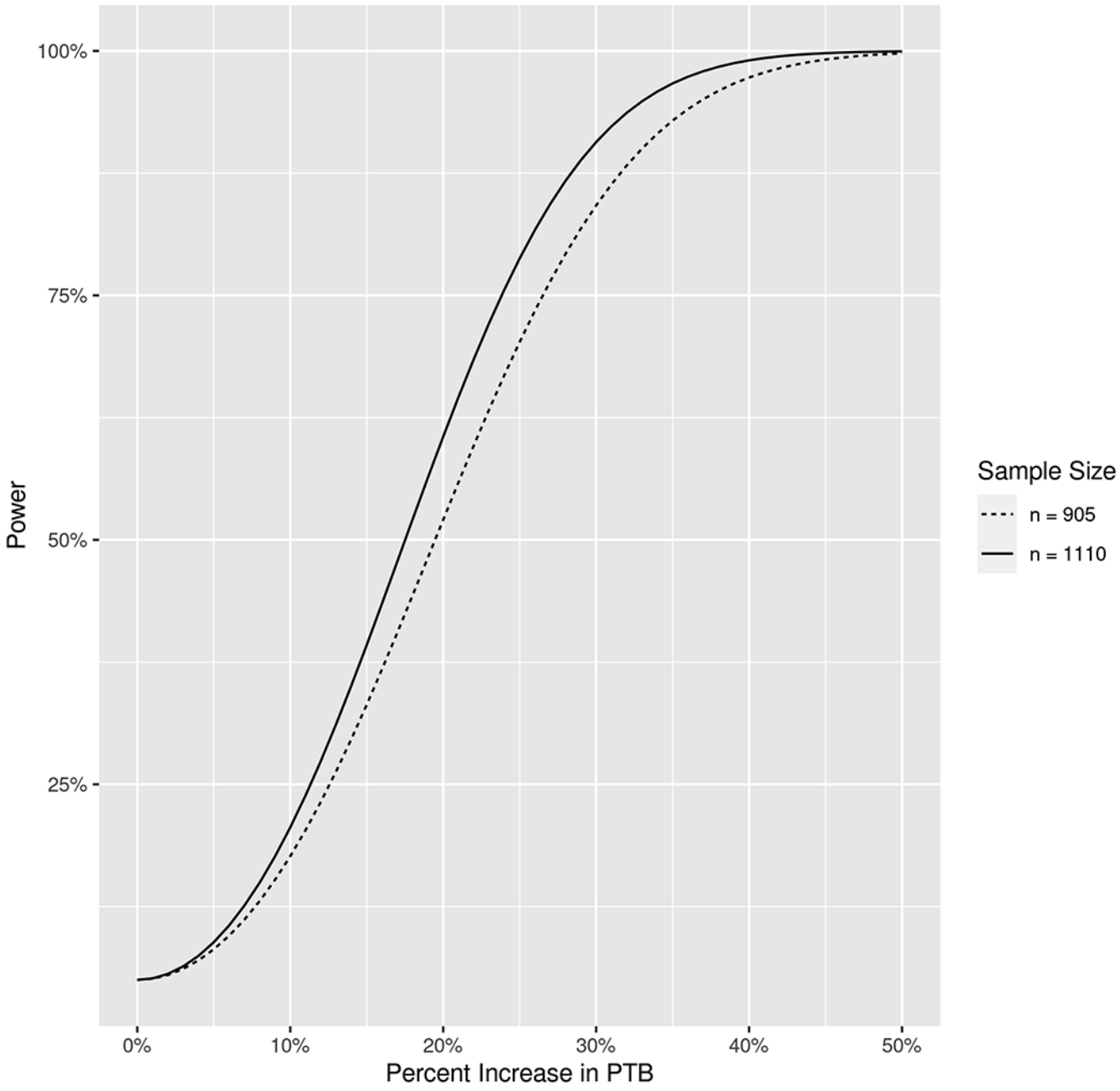

Figure 1.

Power to detect changes in preterm birth rates among American Indian or Alaskan Native individuals using Pennsylvania birth certificate data when including (n=1 109) or excluding (n=905) “write-in” responses to the race question.