

## Systematic Review

## OPEN ACCESS

# Navigating protection and presence: trade-offs around data suppression for small Pacific populations

Helen TURNER,<sup>1,2,3,4,5</sup> Connor FLYNN,<sup>2,4</sup> Lilliana FLYNN,<sup>2</sup> Catherine BROCKWAY,<sup>2,4</sup> Rylan CHONG,<sup>3,4</sup> Apo APOROSA,<sup>6</sup> Johnny T ALDAN,<sup>7</sup> Mata'uitafa FAIAI,<sup>4,8</sup> Chad JANSEN,<sup>1</sup> Alexander STOKES,<sup>4,9,10</sup>

### ABSTRACT

**Introduction:** Datasets, their analytics and their interpretation are key decision support tools for Pacific communities, with the potential to shape public policy, healthcare and social interventions in the Pacific “Blue Continent”. However, in the case of numerically small island populations, privacy concerns have motivated widespread use of data suppression. While suppression safeguards privacy, it also risks erasing the visibility of these populations, leading to “statistical invisibility” that obscures their social, health and economic challenges. This study critically reviews the practice of data suppression, emphasising its rationale in privacy protection, but also highlighting the impacts on resource allocation, advocacy and equitable policymaking for Pacific populations.

**Methods:** We explored the rationale behind data suppression and its legal and regulatory context. Using case studies including the US Census Bureau, Centers for Disease Control and Prevention and the Behavioral Risk Factor Surveillance System, we assess the impact of suppression thresholds and privacy-preserving methods on Pacific communities. We present a novel analysis of data suppression impacts on International Classification of Diseases (ICD) code suppression across different levels of geographical units in the Pacific to illustrate disproportionate impacts. We review alternative privacy-preserving methods, including data smoothing, statistical masking and synthetic data generation, that could mitigate the effects of suppression without compromising individual privacy.

**Finding and conclusions:** We recommend inclusive and transparent data practices needed to prevent data suppression compounding systemic marginalisation of small Pacific populations. By critically evaluating current practices and proposing alternative strategies grounded in Critical Data Theory and Pacific knowledge epistemology, this paper aims to inform policies that balance protection of individual privacy with the accurate representation of small, geographically dispersed populations.

**Key words:** data suppression, population statistics, data privacy, data sovereignty

### INTRODUCTION

Datasets, their analytics and interpretation are key decision support tools that inform public policy, healthcare and social interventions. However, the release of detailed data can compromise individual privacy in small populations. Data suppression is often employed as an approach to withhold or mask data that might reveal personally identifiable information (PII) when population numbers are small. Suppression preserves privacy but obscures challenges faced by small populations such as those in the Pacific. Suppressed data leads to an invisibility in statistics that can skew resource allocation and policy decisions. This paper aims to critically review the practice of statistical data suppression, its rationale and unintended consequences for small populations such as those in the Pacific.

**Corresponding author:** Helen Turner,  
[hturner@chaminade.edu](mailto:hturner@chaminade.edu)

1. Laboratory of Pharmacology and Analytics, School of Natural Sciences and Mathematics, Chaminade University, Honolulu, Hawai'i
2. Data Analytics Research Team, UN CIFAL Honolulu, Chaminade University, Honolulu, Hawai'i
3. Data Science Program, School of Natural Sciences and Mathematics, Chaminade University, Honolulu, Hawai'i
4. NSF INCLUDES ALL-SPICE Alliance program, Chaminade University, Honolulu, Hawai'i
5. Center for Indigenous Innovation and Health Equity, University of Hawai'i at Manoa, Honolulu, Hawai'i
6. Te Huataki Waiora School of Health, University of Waikato, Hamilton, Aotearoa New Zealand
7. Graduate Program in Epidemiology, University of Hawai'i at Manoa, Honolulu, Hawai'i

8. Graduate Program in Chronic Disease Epidemiology, Yale University School of Public Health, New Haven, Connecticut
9. Department of Cell and Molecular Biology, John A. Burns School of Medicine, University of Hawai'i, Honolulu, Hawai'i
10. NIH AIM-AHEAD ARCH Program, John A. Burns School of Medicine, Honolulu, Hawai'i

**Rec:** 06.11.24      **Acc:** 16.03.25      **Pub:** 25.08.25

**Citation:** Turner, H, et al. Navigating Protection and Presence: Trade-offs around data suppression for small Pacific populations. *Pacific Health Dialog*. 2025;25(3):121-141. DOI: 10.26635/phd.2025.179

**Copyright:** © 2025 Turner, H et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## 1. Rationale for statistical data suppression

The primary goal of data suppression<sup>1</sup> is to protect individual privacy. The emphasis on privacy in data collection and reporting, particularly for small populations, stems from a deep concern for safeguarding individual identity and protecting against unintended harms that can arise when private information is exposed.<sup>2</sup> Revealing sensitive information, such as health or income data, can lead to discrimination, stigma or economic harm. Even when PII is removed, combining different sources of data can sometimes reveal an individual's identity (e.g., census data broken down by race, gender and location, could inadvertently disclose information about a specific person in a small community).<sup>1,3</sup> Re-identification attacks (the process of matching anonymized data with other information to uncover the identities) are motivated by commercial, political and extortive gains. They exploit vulnerabilities in anonymised datasets by linking seemingly non-identifiable information with external data to reveal individual identities. Re-identification strategies include record linkage<sup>4</sup> (anonymised data matched to identifiable external sources), attribute linkage (known information about individuals is used to infer sensitive attributes) and social network pattern inference. In response, data holders employ advanced privacy-preserving techniques to address these threats (e.g., k-anonymity, l-diversity and differential privacy) by obscuring individual identities while maintaining data utility. This concern is codified legally and ethically (e.g., Health Insurance Portability and Accountability Act [HIPAA] and General Data Protection Regulation [GDPR] laws mandating that data must be suppressed or aggregated to prevent identification when population sizes fall below certain thresholds). Researchers adhere to these practices to avoid legal repercussions and protect vulnerable groups.

## 2. Risks of data suppression for small populations

### *Data invisibility and exclusion*

For Pacific people, data aggregation is a well-understood source of statistical invisibility.<sup>5</sup> When Pacific peoples' data is grouped into broader racial categories, such as "Asian/Pacific person", the distinct health and socio-economic challenges they face become obscured. In health research, for example, Pacific peoples' unique risk profiles may be masked when their data is suppressed or aggregated with other groups. Recent advocacy efforts resulted in US federal rule changes addressing disaggregation; the 2024 OMB Statistical Policy Directive No. 15 mandates more granular data collection for race and ethnicity.<sup>6</sup> However, the separate practice of statistical data suppression is ongoing and unaffected by this rule change. As an illustration, in early 2020 the National Center for Health Statistics (NCHS) applied privacy safeguards suppressing data in counties with COVID-19 death counts <10. This led to early pandemic underreporting in counties with smaller populations or where COVID-19 fatalities were initially low, and Native Hawaiian/Pacific Islander (NHPI) death counts were notably underrepresented. In response, the University of California Los Angeles DataLab used Bayesian imputation techniques to estimate the suppressed mortality rates, incorporating prior distributions and leveraging state- and national-level data to model the likely counts for each county, informing decision support.<sup>7,8</sup>

### *Resource distribution, planning and impact assessments*

Resource allocation decisions (e.g., in healthcare, education and social services) are based on population statistics. Suppressed data can make it appear that certain groups have lower needs than in reality, resulting in inequitable resource distribution. Planning in diverse fields from public health to disaster management, environmental impacts assessment and election re-districting is rooted in understanding how many individuals and their characteristics are present in geographical units ranging from large to highly granular. Suppressed data embedded in decision support datasets is problematic and may have inadvertent effects in small populations. For example, redrawing of county/district lines can sub-divide population groups leading to apparent changes in disease prevalence based on new suppression of certain ICD codes. Data suppression also hampers advocacy. Without access to disaggregated and unsuppressed data, advocating for small

populations becomes challenging, further entrenching systemic exclusion.

### **3. Data sets relevant to Pacific people that practice suppression**

#### ***US Census Bureau***

Data collected by the US Census Bureau (USCB),<sup>9-12</sup> which includes the decennial census as well as numerous topical surveys, is one of the most comprehensive datasets available for the United States and its territories, but data suppression is frequently applied when reporting data on small populations. Pacific people are often affected due to their relatively small numbers both in the US-affiliated Pacific Islands (USAPI) and in diasporic contexts in Hawai'i and throughout the continental US. Suppression often occurs at finer geographic levels (e.g., counties or census tracts) or in detailed demographic breakdowns. Until the most recent decennial census, the USCB relied on data suppression, swapping, and partial synthetic data for disclosure avoidance in its tabular summaries.<sup>13</sup> Publications of data from the 2020 Census of Population and Housing in the United States use the new 2020 census disclosure avoidance system (DAS), which uses a differential privacy TopDown Algorithm<sup>14</sup>. The USCB employs a suppression threshold of fewer than three households or individuals in its publicly released datasets. This means that when a subgroup within a particular geographic area has fewer than three households or individuals, the data for that subgroup is either aggregated with other groups or suppressed entirely to prevent potential identification. Additionally, for privacy protection, the USCB applied differential privacy methods in the 2020 decennial census, which introduce statistical noise to small population counts to prevent re-identification. As a result, small populations like those in the Pacific can face further data distortion (Figure 1) and the threshold can lead to their aggregation with other racial or ethnic groups<sup>11</sup> Sasa et al provided further discussion of NHPI underrepresentation in the USCB decennial census.<sup>15</sup>

#### ***American Community Survey***

The Census Bureau American Community Survey (ACS) provides detailed information about US demographic, social and economic conditions including diasporic Pacific populations on the US continent and in Hawai'i. With some exceptions, ACS estimates<sup>16</sup> are suppressed for any geographic area with a population <65,000 in 1-year estimates. ACS suppresses data when the population in a geographic area (e.g., county, ZIP codes, census tract) has <three individuals or households in a category. For data on non-

residence geography (e.g. place of birth), at least 50 cases must be present in the geographic area in order to be released.<sup>16</sup>

#### ***Centers for Disease Control and Prevention (CDC)***

The CDC houses national public health data (e.g., disease prevalence, health behaviors and health outcomes) and it uses geographic suppression to ensure that no PII can be traced to individuals in small and rural populations. CDC data covers some Pacific locales and the Pacific diaspora in the US. Datasets such as the National Health Interview Survey, National Vital Statistics System, and National Health and Nutrition Examination Survey use a suppression threshold of <10 cases or respondents in a specific category. Notably, Healthy People 2010 used a 50-person threshold for suppression.<sup>17</sup>

#### ***Behavioral risk factor surveillance system (BRFSS)***

The BRFSS is a widely used source for data on health behaviors, chronic diseases, and preventive health practices. It is operated by the CDC and is used in some Pacific locales and suppresses data when there are fewer than 50 respondents from a particular subgroup.

### **4. Alternatives to data suppression**

Within the context of valid privacy concerns that trigger data suppression, there are alternative approaches<sup>18-23</sup> that can protect privacy while still providing useful data on small populations:

#### ***Data smoothing, swapping and statistical masking***

Data smoothing techniques (e.g., kernel smoothing, moving averages) apply minor modifications to dataset values to ensure individual data points are not easily re-identifiable but maintain trends and distributions. Noise injection introduces small amounts of random data (or 'noise') to sensitive variables, ensuring the overall statistical properties remain intact but obscuring individual data points.<sup>22</sup> Similarly, differential privacy algorithms add statistical noise to the data, preserving broad patterns but obscuring individual information.<sup>24,25</sup> Bayesian hierarchical models pool data across different levels (e.g., individuals within groups) to produce stable estimates for small populations. They use partial pooling and shrinkage to borrow strength from larger groups, regularizing estimates and protecting privacy by not overfitting to small, identifiable subpopulations.<sup>26</sup>

#### ***Microaggregation***

This technique groups individuals with similar characteristics into small clusters, replacing actual values within each group with average values.<sup>27</sup> For example, if a dataset has income data for individuals, microaggregation might create clusters of people with similar incomes and replace their individual incomes with the group average. Multivariate microaggregation is used when several variables must be protected simultaneously, reducing risk of re-identification across correlated attributes. Microaggregation can be dynamically<sup>28</sup> applied to both numerical and categorical data, making it a flexible tool for protecting privacy in datasets while maintaining a high level of data utility. K-anonymity<sup>29</sup> ensures that any individual in a dataset cannot be distinguished from at least  $k-1$  other individuals<sup>30</sup> by generalizing and suppressing certain data attributes to make individuals indistinguishable within a group of  $k$ .

### **Synthetic data**

Synthetic data generation creates artificial data mimicking the statistical properties of real datasets but with no actual personal information. Generative adversarial networks (GANs)<sup>31,32</sup> or Bayesian networks can generate synthetic datasets that preserve relationships between variables while ensuring no real-world individuals can be re-identified.

### **Digital twinning**

Digital twinning replicates real-world entities (individuals, communities) in a digital model.<sup>33</sup> A digital twin<sup>34</sup> of a population simulates its demographic, social or health characteristics. Unlike synthetic data, digital twins dynamically model the evolution of the population based on real-world inputs. For instance, a digital twin of a small Pacific population could allow researchers to study health trends or simulate policy interventions while ensuring that no real individual data is used directly. Application of Machine Learning (ML) and predictive modeling to the digital twin then allows sophisticated forecasting, backcasting and analytics without privacy issues.

### **Community-based approaches**

Involving community in data governance can ensure that the collection, sharing and use processes align with cultural values and privacy concerns. Indigenous data sovereignty movements advocate for communities to control the collection, access and use of data concerning their members.<sup>35</sup> Pacific communities, which prioritise collective decision-making, can implement systems where community leaders have oversight of how sensitive data is managed (e.g., through community data governance

boards that work with researchers to ensure that the privacy and values of individuals and community are respected).

### **De-suppression**

It is possible to access suppressed data through an administrative process with agencies such as the CDC. The process is resource-intensive and presents costs, including personnel time, application fees for data access, institutional review board (IRB) fees, ethical and legal advice, and compliance costs with post-release requirements for data security and storage. Formal data access requests and IRB approval are often required, and processing time plus legal review (e.g., for CDC WONDER) may be in the order of months. After gaining access to de-suppressed data, there may also be significant limitations on how the data can be used. De-suppression is therefore resource-intensive and presents prohibitive barriers to many stakeholders.

### **Blockchain**

Blockchain technology offers an alternative to data suppression ensuring data security and decentralisation. Instead of centralised control, blockchain distributes individual data across a decentralised ledger. Individuals (or communities) maintain control of their data through private keys. Each transaction (e.g., update or access) is securely logged and immutable, and smart contracts automate access permissions. Applied to Electronic Health Records (EHR), for example, health data which is currently suppressed by centralised data owners in databases would be part of personalised individual (or community) based programs of data access and control.<sup>36-38</sup> Blockchain's distributed nature mitigates the risk of single-point privacy failures or suppression by any one entity. Such decentralisation and autonomy challenges traditional models of vital statistics and public health data gathering, requiring careful evaluation.

### **Comparing international practices in data suppression**

Data suppression practices are employed globally but some nations have implemented alternative models that balance privacy protection with the need for representative data on marginalised populations such as Indigenous and ethnic minorities. In Australia, the Australian census, Australian Bureau of Statistics and the Australian Institute of Health and Welfare<sup>39</sup> often aggregate Indigenous populations with other ethnic minorities and apply thresholded statistical suppression (e.g., at five persons or households). More recently these groups have

applied geographic and microaggregation in an Indigenous data sovereignty framework to retain visibility of small Indigenous and Pacific populations while maintaining privacy and uses.<sup>40,41</sup> Similarly, Aotearoa New Zealand's data practices for Māori and Pacific populations<sup>42,43</sup> include efforts to disaggregate and balance data suppression with privacy. The New Zealand Integrated Data Infrastructure uses a flexible threshold based on the size of the population and geographic unit. This specialised reporting allows for data disaggregation without overly strict suppression.

## **5. Data coverage and suppression practices in the context of COFA and USAPI states**

### ***US suppression practices extend to the Pacific region***

The inclusion of US-affiliated Pacific Islands (USAPI) and states with Compacts of Free Association (COFA)<sup>44</sup> in key datasets such as the CDC and census means that US suppression practices extend over these locales. The US Census Bureau conducts censuses and surveys in US territories (Guam, the Northern Mariana Islands and American Samoa). However, data collection and reporting for these territories differs from the continental US. Guam, the Northern Mariana Islands and American Samoa are also part of CDC NNDSS and BRFSS, and CDC provides limited coverage in Palau, Federated States of Micronesia, and the Marshall Islands through its global health programmes.<sup>45</sup> Since these Freely Associated States are not under US jurisdiction, the CDC works with local health departments and international health organisations (e.g. WHO) to support data efforts, but collection and reporting are less comprehensive and may be inconsistent as a result of limited personnel and technology resources.

### ***USAPI and COFA sovereign data suppression practices***

USAPI and COFA states practice some level of data suppression on their internal data. COFA states are independent nations, and data privacy practices are shaped by their own legal system, international partnerships, and resource constraints. However, there is a strong influence of US standards because as territories the USAPI are under US jurisdiction, and continental laws and regulations may apply. The flow of technical assistance to local actors and tying of grant and aid funding to US-centric compliance guidelines contribute to influence data practice in both USAPI and COFA locales. In addition, partnership with international organisations (e.g., WHO, UN) to collect and report data can impose data

suppression standards leading to de facto suppression of sovereign data. Data may be collected at a hyper-local level but then is upward-aggregated to larger geographical units in response to suppression policies. In these settings the lack of human and technical resources (e.g., data scientists, cyberinfrastructure and information systems) may limit the capacity of states and territories to manage large data sets (e.g., electronic health records) or apply advanced statistical privacy techniques to 'un-mask' suppressed groups.

## **6. Case study: estimating suppressed health data in Pacific countries**

As an experiment, we evaluated population sizes and units of breakdown for Hawai'i's neighbor islands, the USAPI and COFA states. We included commonly applied geographical units (islands, ZIP codes, counties, districts or municipalities, census tracts [US Census]) and BRFSS reporting units (Table 1). We then applied estimates of the numbers of ICD codes (~6,000) that are associated with "rare disorders" (i.e. by US definitions ~1,500<sup>46</sup> individuals affected), and back-calculated the "prevalence threshold" of statistical suppression in these units with a suppression threshold of <10 cases per geographical units. In this analysis, we looked at data suppression likelihood across multiple Pacific regions, accounting for both rare and common disorders. We then evaluated how likely data suppression (at the 10-instance level applied by CDC) would occur for disorders with prevalence rates of 1:1500 "rare", 1:10,000 "very rare" and three example prevalences for diseases/disorder that we would consider "common" (1:100, 1:500, 1:1000). Our findings (Table 2) indicate that larger Pacific population units, such as whole counties or territories, are less prone to data suppression even for rare diseases, as their populations generate enough cases to avoid the privacy-driven suppression threshold of fewer than 10 cases. However, smaller population units including some islands, and counties, census tracts and districts across the region are very vulnerable to suppression, even for relatively common disorders. This analysis demonstrates how smaller regions face greater challenges in reporting health data, as a higher proportion of both rare and common disorder ICD codes may be suppressed due to the limited number of cases, highlighting a critical issue for health data transparency and access in small, vulnerable populations.

## **7. Data privacy and suppression through the lens of pacific epistemology**

This review discusses statistical data suppression practices that primarily evolved

from Eurocentric Judeo-Christian<sup>47</sup> ethical perspectives and are lensed through contemporary American individualism.<sup>48-50</sup> Pacific epistemologies provide relevant alternative lenses through which we can examine regional practices of data privacy and suppression.<sup>51-53</sup> Privacy definitions vary, particularly among Indigenous populations, in which privacy often includes notions of “community, sovereignty, and self-determination”.<sup>54,55</sup> Within these structures, Eurocentric concepts of property, ownership and privacy are unfamiliar, with social and normative systems prioritising interconnectedness and openness. This can easily be misinterpreted as privacy being of little concern to these groups.

### ***Community, reciprocity and knowledge***

Pacific epistemology<sup>56-58</sup> is a worldview grounded in community, interconnectedness and reciprocity. Knowledge is not viewed as an individual possession but as a collective resource, held by community and shared in the context of relationships and responsibilities. This view contrasts with Western epistemologies, where knowledge is a commodity that can be owned, privatised, or controlled by institutions or individuals (leading to extractive ideas of data as the “new oil”).<sup>59</sup> In Pacific societies, sharing of knowledge is intertwined with respect for family, lineage, and land (*vanua*, *whenua*, or *aina* in different Pacific cultures), and the concept of privacy is deeply embedded in the protection of communal relationships and obligations.<sup>60</sup> In this sense, privacy is less about individual autonomy and more about safeguarding the dignity and well-being of the community.

### ***Data privacy in the context of collective identity***

Western approaches to data privacy protect the rights of the individual, ensuring that PII is not disclosed or misused. For Pacific people, privacy may operate at the level of the community rather than the individual. Concerning data, a comment from the Pacific Data Sovereignty Panel<sup>61</sup> at the 2019 Moana Research Seminar Series in Aotearoa New Zealand, reflects that “Data is our identity”. Pacific epistemologies emphasise collective identity and interdependence, which means that the protection of data privacy involves safeguarding the integrity of the community as a whole. When data suppression is applied to Pacific populations, it is not just individual privacy at stake but the visibility and recognition of the community itself. Data suppression, especially when it aggregates Pacific people’s data into broader racial categories, can be seen as a form of epistemic violence that erases the unique identity,

experiences, and needs of Pacific communities. Pacific epistemology recognises that privacy is not simply about withholding information but about managing the flow of knowledge in ways that protect the community and enhance its collective well-being. In this context, data suppression should not be viewed as an all-or-nothing proposition. Instead, it must be tailored to ensure that sensitive information is protected without compromising the visibility and representation of Pacific communities.

### **8. Mitigation of data suppression as a key component of an emerging ‘critical data theory’ for the Pacific and beyond**

In discussions of data suppression and its impacts on small populations, it is crucial to understand how power structures shape the creation, control and dissemination of data.<sup>62-65</sup> Critical data theory (CDT)<sup>66</sup> and critical data studies<sup>67,68</sup> share a common foundation in critical social theory, emphasising the need to question the neutrality of data and its role in reinforcing or challenging existing power structures. CDT is an emerging field that critiques the role of data in perpetuating inequalities and explores how data practices can either reinforce or challenge hegemonic power structures. Drawing on insights from critical race theory, feminist theory and postcolonial studies, CDT examines how data is not neutral but is instead embedded with the interests, biases and power dynamics of those who control it. CDT provides a framework to analyse how data suppression can reinforce or challenge existing hegemonies. For small and marginalised populations, data suppression can be seen as part of a larger structure of control and power that limits representation, autonomy and influence in decision-making processes. Data is both a powerful tool and a tool of power, and groups that collect and control data can benefit from suppressing or shaping it in their interests. CDT posits that the ownership and control of data are central to understanding power relations. In many cases, marginalised groups are not the primary owners of their own data. Instead, powerful institutions collect, store, and use data about these groups without their input or control. This creates an inherent power imbalance, where controlling when and how data is suppressed, power structures maintain control over whose stories are told and whose are ignored.

Data suppression and aggregation can be seen as forms of epistemic injustice<sup>69-71</sup> where certain groups are systematically excluded from contributing to and benefitting from knowledge production. Pacific people and other marginalised communities may be excluded from

representation in datasets, and from influencing how data is collected, interpreted and used. This further entrenches marginalisation, as they are denied the opportunity to shape the narratives that emerge from the data. Transfer of power from hegemonic institutions to communities is central to “critical” movements and in this case a major goal is to shift the power balance in data governance. In practice, this means Pacific, COFA states, and other small populations realising autonomy to manage data, set privacy thresholds, and decide how and when suppression is applied. Resourcing data collection efforts that are by communities, for those communities is a key component of data autonomy, as is the development of autonomous data science capacity to engage in advanced practices such as smoothing, microaggregation, synthetics and digital twinning. Building awareness of data suppression, its practices, consequences, and mitigation strategies requires a holistic wayfinding effort by Pacific communities and other small populations. Figure 2 offers a framework for communities and individuals to navigate this field, including assessment questions, knowledge assets and associated needs for new tool development, new open-source data products and further resource allocation.

## CONCLUSION

Data suppression, though intended to protect privacy, often has negative consequences for small populations including regional and diasporic Pacific people. The practice leads to data invisibility, inequitable resource distribution, and reduced capacity for advocacy. While privacy concerns are serious and compelling, methods such as data smoothing, microaggregation and community-based governance offer promising balance that can ensure both privacy and representation. Moving forward, researchers, policymakers, data scientists and statisticians must work together to find culturally responsive ways to balance privacy with the need for accurate, disaggregated data on small populations. Recommendations include:

- Reevaluating data suppression policies to ensure that they do not disproportionately harm small or marginalised populations.
- Generating transparency for communities to know the extent and impacts of suppression.
- Facilitating data de-suppression pathways for low resource communities.
- Acknowledging and addressing hidden legal and technical costs for communities to request de-suppression or use algorithmic approaches to.<sup>72,73</sup>

Pacific culture and practice can positively inform communities outside the region, with the Blue Continent increasingly positioned as a source of globally applicable models in sustainability and social justice rather than a sink for aid and neocolonialism. Developing Pacific models for data inclusion will be transferrable and benefit other ‘small’ populations (such as sex and gender minorities, ethnic minorities) to mitigate their masking in critical discussions about health, education and social policy.

## Figure legends

### Figure 1: Statistical data suppression overview

**A. Schematic of the practice of statistical data suppression.**<sup>1,18,74</sup> A threshold for visibility is applied (left panel) to prevent disclosure of private information about individuals or small groups when displaying statistics for a given geographical unit. Strong rationales (center panel) based on individual privacy and the need to counter re-identification attack are juxtaposed with inadvertent consequences such as population masking and distortion of information such as disease prevalences or population composition. Illustrative examples include the following.

Suppression in cancer registries: rare pediatric cancers (e.g., retinoblastoma or Ewing sarcoma) are particularly vulnerable to suppression in cancer registries. Data is suppressed at state and county levels to protect the privacy of affected families. This suppression makes it difficult to assess the true incidence of these cancers, particularly in regions where genetic or environmental factors may contribute to higher-than-expected rates.

Early epidemic reporting: early COVID-19 mortality statistics in small populations were suppressed according to standard practices, potentially masking impacts and affecting decision support.

Kauai birth defects controversy: there is ongoing concern in Kauai regarding the potential link between pesticide exposure (~17x national averages) and excess birth defect rates. Agency statistics (e.g., Hawaii’s Birth Defects Program, CDC WONDER), are suppressed at the 10 case/year/geographical unit and are at odds with physician-advocates reports of concerning birth defect rates. Given Kauai’s population (~40,000) and the low prevalence of the specific birth defects, this statistical system is likely insufficient to identify excess cases that are numerically few but of profound human impact.

**Cluster masking:** data suppression, especially in small populations, can obscure potentially concerning disease clusters by preventing the identification of localised health anomalies. When data is withheld or aggregated to meet thresholds for privacy protection, such as the 10 case/year/geographical unit rule, small but significant increases in disease rates can go unnoticed. In communities with low population numbers, even a few additional cases of rare conditions may indicate an emerging public health issue.

**B. Example: Hawai'i impacts of differential privacy practices.** This figure presents findings from the Hawai'i Data Collaborative's analysis of U.S. Census 2010 Demonstration Data, which implemented differential privacy techniques to safeguard individual privacy. The analysis focuses on the effects of differential privacy on smaller subpopulations in Hawai'i, leading to the apparent loss or alteration of population counts, particularly in minority racial and ethnic groups.

### **Figure 2. Wayfinding through Pacific Data Suppression: A Framework**

This figure presents a decision-making framework to guide stakeholders through issues of data suppression, particularly in Pacific Island contexts and other small populations. The framework provides a pathway for understanding and addressing the impact of data suppression on geographically and demographically small populations. The steps in the framework include:

*Questions:* Starting with the inquiry, the framework prompts users to consider their specific data context, focusing on population size and geographic units (e.g., Census Tract, Zip Code, County, Island).

*Knowledge Assets:* Users are directed to relevant datasets (e.g., CDC, BRFSS, Census, AHS) that correspond to the geographic unit of interest. A key decision point is identifying the suppression threshold in terms of instance size (e.g., 10, 50, etc.) and determining what privacy protections are in effect, such as microaggregation, differential privacy, smoothing, digital twins, and Bayesian methods.

*Development of new impact assessment workflows:* This step is needed for stakeholders to evaluate impacts of suppression. For example, an ICD-code suppression Rate (SR) by Geographical Unit (GU) calculator would be one impact assessment tool. It would be used to help quantify the effect of data suppression by calculating the ratio of ICD codes that fall below the suppression threshold for a population, with associated visual tools like gridded SR products

and indices of suppression by location and demographic.

*De-suppression toolkits:* These suggested tools would mitigate the complexity of the processes that exist for de-suppression including identification of agency-specific processes, necessary documentation, and worked examples.

*Resource needs:* The right side of the framework highlights critical resource needs for communities to successfully navigate suppression, including curated data resources, data science technical assistance, legal and regulatory expertise, and advocacy and feedback processes to inform decision-making.

### **Funding Acknowledgements.**

This work was supported by the following grant awards: NSF HRD-2217242 (PI HT, AJS, RC, CF, CB); NSF SCIP E OAC 2417946 HT, AJS); NIH AIM-AHEAD 10T2OD032581-01 (AJS, HT). New Zealand HRC 23/153 (AA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the NIH. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the NSF.

### **Competing Interests of Interest.**

The authors report no competing interests.

### **Author Contributions (CRediT Taxonomy).**

Conceptualization: HT, AJS. Funding Acquisition: HT, AJS. Investigation, HT, AJS, CF, LF, JTA. Writing, Review and Editing: HT, AJS, CJ, CB, AA, RC, MF.

### **REFERENCES**

1. Emam, K. E., Jonker, E., Arbuckle, L. & Malin, B. A Systematic Review of Re-Identification Attacks on Health Data. *PLOS ONE* 6, e28071 (2011).
2. McGraw, D., Dempsey, J. X., Harris, L. & Goldman, J. Privacy as an enabler, not an impediment: building trust into health information exchange. *Health Aff. Proj. Hope* 28, 416–427 (2009).
3. Loukides, G., Denny, J. C. & Malin, B. The disclosure of diagnosis codes can breach research participants' privacy. *J. Am. Med. Assoc. Inform. Assoc. JAMIA* 17, 322–327 (2010).
4. Brownstein, J. S., Cassa, C. A. & Mandl, K. D. No place to hide--reverse identification of patients from published maps. *N. Engl. J. Med.* 355, 1741–1742 (2006).

5. Bhakta, S. Data disaggregation: the case of Asian and Pacific Islander data and the role of health sciences librarians. *J. Med. Libr. Assoc. JMLA* 110, 133 (2022).
6. U.S. Office of Management and Budget Statistical Policy directive 15 on Race and Ethnicity Data Standards. The 2024 Statistical Policy Directive No. 15. <https://spd15revision.gov/content/spd15revision/en/2024-spd15.html> (2024).
7. Kao, S.-Y. Z., Tutwiler, M. S., Ekwueme, D. U. & Truman, B. I. Better data for decision-making through Bayesian imputation of suppressed provisional COVID-19 death counts. *PLOS ONE* 18, e0288961 (2023).
8. UCLA Center for Health Policy Research. Native Hawaiian and Pacific Islander Data Policy Platform: No Health Equity Without Data Equity. <https://healthpolicy.ucla.edu/sites/default/files/2023-03/nhpi-data-policy-platform-report-mar2023.pdf> (2023).
9. US Census Bureau. Understanding Differential Privacy. <https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/differential-privacy.html> (2020).
10. US Census Bureau. Detailed Look at Native Hawaiian and Other Pacific Islander Groups. <https://www.census.gov/library/stories/2023/09/2020-census-dhc-a-nhpi-population.html> (2020).
11. Hong, J. Census 2020 + / - 2: Census, Differential Privacy, and the Future of Data. Hawai'i Data Collaborative <https://www.hawaiidata.org/news/2020/9/24/census2020-census-differential-privacy-future-of-data> (2020).
12. Hotz, V. J. & Salvo, J. A Chronicle of the Application of Differential Privacy to the 2020 Census. *Harv. Data Sci. Rev.* (2022) doi:10.1162/99608f92.ff891fe5.
13. McKenna, L. Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing. Work. Pap. (2018).
14. Abowd, J. M. et al. The 2020 Census Disclosure Avoidance System TopDown Algorithm. *Harv. Data Sci. Rev.* (2022) doi:10.1162/99608f92.529e3cb9.
15. Sasa, S. M. & Yellow Horse, A. J. Just data representation for Native Hawaiians and Pacific Islanders: A critical review of systemic Indigenous erasure in census and recommendations for psychologists. *Am. J. Community Psychol.* 69, 343–354 (2022).
16. United States Census Bureau. American Community Survey Data Suppression. [https://www2.census.gov/programs-surveys/acs/tech\\_docs/data\\_suppression/ACSO\\_Data\\_Suppression.pdf](https://www2.census.gov/programs-surveys/acs/tech_docs/data_suppression/ACSO_Data_Suppression.pdf) (2016).
17. Klein, R. J., Proctor, S. E., Boudreault, M. A. & Turczyn, K. M. Healthy People 2010 criteria for data suppression. *Healthy People 2010 Stat. Notes Cent. Dis. Control Prev. Cent. Health Stat.* 1–12 (2002).
18. Patki, N., Wedge, R. & Veeramachaneni, K. The Synthetic Data Vault. in 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA) 399–410 (2016). doi:10.1109/DSAA.2016.49.
19. Soria-Comas, J. & Domingo-Ferrer, J. Big Data Privacy: Challenges to Privacy Principles and Models. *Data Sci. Eng.* 1, 21–28 (2016).
20. Abowd, J. M. & Schmutte, I. M. An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices. *Am. Econ. Rev.* 109, 171–202 (2019).
21. Machanavajjhala, A., Kifer, D., Gehrke, J. & Venkitasubramaniam, M. L-diversity: Privacy beyond k-anonymity. *ACM Trans Knowl Discov Data* 1, 3-es (2007).
22. Dwork, C., McSherry, F., Nissim, K. & Smith, A. Calibrating Noise to Sensitivity in Private Data Analysis. *J. Priv. Confidentiality* 7, 17–51 (2016).
23. Makov, E., Fienberg, S. E. & Steele, R. J. Disclosure limitation, perturbation and related methods for categorical Data. *J. Off. Stat.* 14, 485–502 (1998).
24. Garfinkel, S. Differential Privacy and the 2020 US Census. *MIT Case Stud. Soc. Ethical Responsib. Comput.* (2022) doi:10.21428/2c646de5.7ec6ab93.
25. Slavković, A. & Seeman, J. Statistical Data Privacy: A Song of Privacy and Utility. *Annu. Rev. Stat. Its Appl.* 10, 189–218 (2023).
26. Vaughan, A. S. et al. Applying a Bayesian Spatiotemporal Model to Examine Excess County-Level Cardiovascular Disease Death Rates During the COVID-19 Pandemic. *Am. J. Epidemiol.* kwae330 (2024) doi:10.1093/aje/kwae330.
27. Fadel, A. C., Ochi, L. S., Brito, J. A. de M. & Semaan, G. S. Microaggregation heuristic

- applied to statistical disclosure control. *Inf. Sci.* 548, 37–55 (2021).
28. Yan, Y. et al. Privacy preserving dynamic data release against synonymous linkage based on microaggregation. *Sci. Rep.* 12, 2352 (2022).
  29. Soria-Comas, J. & Domingo-Ferrer, J. Probabilistic k-anonymity through microaggregation and data swapping. in 2012 IEEE International Conference on Fuzzy Systems 1–8 (2012). doi:10.1109/FUZZ-IEEE.2012.6251280.
  30. Soria-Comas, J., Domingo-Ferrer, J., Sánchez, D. & Martínez, S. Improving the Utility of Differentially Private Data Releases via k-Anonymity. in 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications 372–379 (2013). doi:10.1109/TrustCom.2013.47.
  31. Venugopal, R. et al. Privacy preserving Generative Adversarial Networks to model Electronic Health Records. *Neural Netw. Off. J. Int. Neural Netw. Soc.* 153, 339–348 (2022).
  32. Liu, Y., Peng, J., Yu, J. J. Q. & Wu, Y. PPGAN: Privacy-Preserving Generative Adversarial Network. in 2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS) 985–989 (2019). doi:10.1109/ICPADS47876.2019.00150.
  33. Quick, H. Generating Poisson-Distributed Differentially Private Synthetic Data. *J. R. Stat. Soc. Ser. A Stat. Soc.* 184, 1093–1108 (2021).
  34. Tao, F., Xiao, B., Qi, Q., Cheng, J. & Ji, P. Digital twin modeling. *J. Manuf. Syst.* 64, 372–389 (2022).
  35. Walter, M., Kukutai, T., Carroll, S. R. & Rodriguez-Lonebear, D. Indigenous Data Sovereignty and Policy. in *Indigenous Data Sovereignty and Policy* (Routledge, London, 2020). doi:10.4324/9780429273957.
  36. Fang, H. S. A., Tan, T. H., Tan, Y. F. C. & Tan, C. J. M. Blockchain Personal Health Records: Systematic Review. *J. Med. Internet Res.* 23, e25094 (2021).
  37. Elvas, L. B., Serrão, C. & Ferreira, J. C. Sharing Health Information Using a Blockchain. *Healthcare* 11, 170 (2023).
  38. Haleem, A., Javaid, M., Singh, R. P., Suman, R. & Rab, S. Blockchain technology applications in healthcare: An overview. *Int. J. Intell. Netw.* 2, 130–139 (2021).
  39. Australian Institute of Health and Welfare. Data Sources. Indigenous Mental Health & Suicide Prevention <https://www.indigenouismhspc.gov.au/resources/data-resources/data-sources> (2023).
  40. Commonwealth of Australia. Framework for Governance of Indigenous Data | NIAA. <https://www.niaa.gov.au/resource-centre/framework-governance-indigenous-data> (2024).
  41. Walter, M. et al. Indigenous Data Sovereignty in the Era of Big Data and Open Data. *Aust. J. Soc. Issues* 56, 143–156 (2021).
  42. Sporle, A., Hudson, M. & West, K. Indigenous data and policy in Aotearoa New Zealand. in *Indigenous Data Sovereignty and Policy* (Routledge, London, 2020).
  43. Jansen, R. Indigenous data sovereignty: a Māori health perspective. in *Indigenous Data Sovereignty* (eds. Kukutai, T. & Taylor, J.) vol. 38 193–212 (ANU Press, Canberra, 2016).
  44. U.S. Department of the Interior. Compacts of Free Association. <https://www.doi.gov/oia/compacts-of-free-association> (2024).
  45. U.S. Centers for Disease Control and Prevention. Global Health Center. Global Health <https://www.cdc.gov/global-health/index.html> (2024).
  46. Shimkus, J. H.R.4013 - 107th Congress (2001-2002): Rare Diseases Act of 2002. <https://www.congress.gov/bill/107th-congress/house-bill/4013> (2002).
  47. Orwell, G. *An Age Like This, 1920-1940.* (Harcourt, Boston, MA, 1971).
  48. Stivers, R. Ethical Individualism and Moral Collectivism in America. *Humanitas* 16, 56–73 (2003).
  49. Newman, C. Big Data Analytics Shows How America’s Individualism Complicates Coronavirus Response UVA Today. <https://news.virginia.edu/content/big-data-analytics-shows-how-americas-individualism-complicates-coronavirus-response> (2020).
  50. Brey, P. et al. International Differences in Ethical Standards and in the Interpretation of Legal Frameworks. 1–133 <https://satoriproject.eu/media/D3.2-Int-differences-in-ethical-standards.pdf> (2015).
  51. Kukutai, T. & Taylor, J. Indigenous Data Sovereignty: Toward an agenda. in

- Indigenous Data Sovereignty: Toward an agenda vol. 38 (ANU Press, Canberra, 2016).
52. Walter, M. & Suina, M. Indigenous data, indigenous methodologies and indigenous data sovereignty. *Int. J. Soc. Res. Methodol.* 22, 233–243 (2019).
  53. Smith, L. T. *Decolonizing Methodologies: Research and Indigenous Peoples.* (Zed Books, London, 2012).
  54. Kukutai, T. & Cormack, D. “Pushing the space”: Data sovereignty and self-determination in Aotearoa NZ. in *Indigenous Data Sovereignty and Policy* (Routledge, London, 2020).
  55. Moutselos, K. & Maglogiannis, I. Evidence-based Public Health Policy Models Development and Evaluation using Big Data Analytics and Web Technologies. *Med. Arch. Sarajevo Bosnia Herzeg.* 74, 47–53 (2020).
  56. Tecun, A., Hafoka, I., ‘Ulu’ave, L. & ‘Ulu’ave-Hafoka, M. Talanoa: Tongan epistemology and Indigenous research method. *Altern. Int. J. Indig. Peoples* 14, 156–163 (2018).
  57. Quanchi, M. Indigenous epistemology, wisdom and tradition; changing and challenging dominant paradigms in Oceania. *Cent. Soc. Change Res. Qld. Univ. Technol.* 1–13 (2004).
  58. Meyer, M. A. Our Own Liberation: Reflections on Hawaiian Epistemology. *Contemp. Pac.* 13, 124–148 (2001).
  59. Arthur, C. Tech giants may be huge, but nothing matches big data. *The Guardian* <https://www.theguardian.com/technology/2013/aug/23/tech-giants-data> (2013).
  60. Aporosa, S. A. & Fa’avae, D. T. M. Grounding Pacific practice: Fono at the fale and veiqaraqaravi vakavanua. *Waikato J. Educ.* 26, 35–43 (2021).
  61. Moana Research. *Pacific Data Sovereignty. The 2019 Moana Research Seminar Series Report.* [https://moanaconnect.co.nz/wp-content/uploads/2021/11/Pacific-Data-Sovereignty-Report\\_FINAL\\_2019.pdf](https://moanaconnect.co.nz/wp-content/uploads/2021/11/Pacific-Data-Sovereignty-Report_FINAL_2019.pdf) (2019).
  62. Noble, S. U. *Algorithms of Oppression: How Search Engines Reinforce Racism.* xv, 229 (New York University Press, New York, NY, 2018).
  63. Boyd, D. & Crawford, K. Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Inf. Commun. Soc.* 15, 662–679 (2012).
  64. Eubanks, V. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.* (St. Martin’s Press, New York, NY, 2018).
  65. Couldry, N. & Mejias, U. A. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism.* (Stanford University Press, Stanford, CA, 2019).
  66. Gillborn, D., Warmington, P. & Demack, S. QuantCrit: education, policy, ‘Big Data’ and principles for a critical race theory of statistics. *Race Ethn. Educ.* 21, 158–179 (2018).
  67. Iliadis, A. & Russo, F. Critical data studies: An introduction. *Big Data Soc.* 3, 2053951716674238 (2016).
  68. Richterich, A. *The Big Data Agenda: Data Ethics and Critical Data Studies.* vol. 6 (University of Westminster Press, London, 2018).
  69. Scotto, C. Digital Identities and Epistemic Injustices. *HUMANA MENTE J. Philos. Stud.* 13, 151–180 (2020).
  70. Kay, J., Kasirzadeh, A. & Mohamed, S. Epistemic Injustice in Generative AI. *Proc. AAAIACM Conf. AI Ethics Soc.* 7, 684–697 (2024).
  71. Symons, J. & Alvarado, R. Epistemic Injustice and Data Science Technologies. *Synthese* 200, 1–26 (2022).
  72. Chen, T., Li, W., Zambarano, B. & Klompas, M. Small-area estimation for public health surveillance using electronic health record data: reducing the impact of underrepresentation. *BMC Public Health* 22, 1515 (2022).
  73. Quick, H., Holan, S. H. & Wikle, C. K. Generating Partially Synthetic Geocoded Public Use Data with Decreased Disclosure Risk by Using Differential Smoothing. *J. R. Stat. Soc. Ser. A Stat. Soc.* 181, 649–661 (2018).
  74. Janmey, V. & Elkin, P. L. Re-Identification Risk in HIPAA De-Identified Datasets: The MVA Attack. *AMIA. Annu. Symp. Proc.* 2018, 1329–1337 (2018).

## Appendix

### Tables and Figures

Table 1

Region	ZIP codes	Counties <i>Population est.,</i>	Districts/municipalities <i>Population est.,</i>	Census tracts <i>Population est.,</i>	BRFSS units.
<b>Hawai'i neighbour islands</b>					
Maui	30	1 County <i>168,000</i>	1 <i>164,754</i>	36 <i>3300-7600/tract</i>	1 <i>168,000</i>
Lanai			1 <i>7,345</i>		
Molokai			1 <i>7,345</i>		
Kauai	14	1 County <i>73,000</i>	2 Kauai <i>72,800</i> Niihau <i>200</i>	16	1 <i>73,000</i>
<b>USAPI</b>					
Guam	16	1 Territory: <i>170,000</i>	19 districts: Dededo: <i>44,943</i> Yigo: <i>20,539</i> Tamuning: <i>18,972</i> Mangilao: <i>15,191</i> Barrigada: <i>9,676</i> Chalan Pago-Ordot: <i>6,822</i> Agat: <i>4,915</i> Santa Rita: <i>6,084</i> Sinajana: <i>2,592</i> Hagatna: <i>1,051</i> Asan-Maina: <i>2,137</i> Piti: <i>1,454</i> Yona: <i>6,484</i> Merizo: <i>1,850</i> Umatac: <i>647</i> Inarajan: <i>2,273</i>		1 <i>170,000</i>

Region	ZIP codes	Counties <i>Population est.,</i>	Districts/municipalities <i>Population est.,</i>	Census tracts <i>Population est.,</i>	BRFSS units.
			Talofofo: 3,050 Mongmong-Toto-Maite: 6,825 Agana Heights: 3,496		
Northern Mariana Islands	5	1 Territory: 47,000	4 municipalities: Tinian: 3,540 Saipan: 43,385 Rota: 2,527 Northern Islands: 377		1 47,000
American Samoa	5	1 Territory: 55,000	5 districts: Eastern District Western District, Manu'a District Swains Island, Rose Atoll	18 Tracts: Tract 9501: 1,487 Tract 9502: 1,158 Tract 9503: 2,415 Tract 9504: 1,376 Tract 9505: 2,623 Tract 9506: 3,218 Tract 9507: 1,762 Tract 9508: 2,081 Tract 9509: 3,031 Tract 9510: 1,446 Tract 9511: 6,252 Tract 9512: 3,418 Tract 9513: 2,932 Tract 9514: 2,132 Tract 9515: 1,698 Tract 9516: 4,293 Tract 9517: 1,342	1 55,000
Region	ZIP codes	Counties <i>Population est.,</i>	Districts/municipalities <i>Population est.,</i>	Census tracts <i>Population est.,</i>	BRFSS units.
<b>Compact of Free Association States</b>					
Federated States of Micronesia	4 1 each for Yap Pohnpei Chuuk	Total: 105,000 4 states Yap: 12,000 Pohnpei: 36,000 Chuuk: 49,000			

	Kosrae	Kosrae: 8,000		
Republic of the Marshall Islands	1	1 State: 65,000	24 inhabited atolls and islands, though no formal county divisions: Majuro: 27,800 Ebeye (Kwajalein Atoll): 15,000 Other islands and atolls: ~200-2000	
Palau	1	1 State: 18,000	16 states, which function similarly to counties or municipalities Koror: 11,000 Other states: ~200-4000	

Table 2

<i>NS= Not suppressed</i>	Population	Prevalence threshold for suppression at 10 cases 1 in X	Common disorder (1 in 100)	Common disorder (1 in 500)	Common disorder (1 in 1000)	Uncommon disorder (1 in 1,500)	Very uncommon disorder (1 in 10,000)	OrphanNet rare disorder (1 in 200,000)	
<b>Kauai</b>	73298	7329.8	<i>NS</i>					<b>Possible suppression</b>	
<b>Maui County</b>	168307	16830.7	<i>NS</i>					<b>Possible suppression</b>	
Maui Census Tract 301	4500	450.0	<i>NS</i>	<b>Possible suppression</b>					
Maui Census Tract 302	6800	680.0	<i>NS</i>		<b>Possible suppression</b>				
Maui Census Tract 303	5200	520.0	<i>NS</i>		<b>Possible suppression</b>				
Maui Census Tract 304	6100	610.0	<i>NS</i>		<b>Possible suppression</b>				
Maui Census Tract 305	3400	340.0	<i>NS</i>	<b>Possible suppression</b>					

Maui Census Tract 306	4200	420.0	NS	Possible suppression
Maui Census Tract 307	7000	700.0	NS	Possible suppression
Maui Census Tract 308	3100	310.0	NS	Possible suppression
Maui Census Tract 309	5600	560.0	NS	Possible suppression
Maui Census Tract 310	7600	760.0	NS	Possible suppression
Maui Census Tract 311	2300	230.0	NS	Possible suppression
Maui Census Tract 312	5900	590.0	NS	Possible suppression
Maui Census Tract 313	4800	480.0	NS	Possible suppression
Maui Census Tract 314	3700	370.0	NS	Possible suppression
Maui Census Tract 315	4500	450.0	NS	Possible suppression
Maui Census Tract 316	3900	390.0	NS	Possible suppression
Maui Census Tract 317	5400	540.0	NS	Possible suppression
Maui Census Tract 318	4200	420.0	NS	Possible suppression
Maui Census Tract 319	3300	330.0	NS	Possible suppression
<b>Republic of Palau</b>	18000	1800.0	NS	Possible suppression
Koror	11000	1100.0	NS	Possible suppression
<b>Federated States of Micronesia</b>	105000	10500.0	NS	Possible suppression
Yap (FSM)	12000	1200.0	NS	Possible suppression
Chuuk (FSM)	49000	4900.0	NS	Possible suppression
Pohnpei (FSM)	36000	3600.0	NS	Possible suppression
Kosrae (FSM)	8000	800.0	NS	Possible suppression
<b>Republic of the Marshall Islands</b>	60000	6000.0	NS	Possible suppression
Majuro (Marshall Islands)	27800	2780.0	NS	Possible suppression
Ebeye (Marshall Islands)	15000	1500.0	NS	Possible suppression

<b>Guam</b>	170179	17017.9	<b>NS</b>	<b>Possible suppression</b>
Dededo (Guam)	44943	4494.3	<b>NS</b>	<b>Possible suppression</b>
Yigo (Guam)	20539	2053.9	<b>NS</b>	<b>Possible suppression</b>
Tamuning (Guam)	18972	1897.2	<b>NS</b>	<b>Possible suppression</b>
Mangilao (Guam)	15191	1519.1	<b>NS</b>	<b>Possible suppression</b>
Barrigada (Guam)	9676	967.6	<b>NS</b>	<b>Possible suppression</b>
Chalan Pago-Ordot (Guam)	6822	682.2	<b>NS</b>	<b>Possible suppression</b>
Agat (Guam)	4915	491.5	<b>NS</b>	<b>Possible suppression</b>
Santa Rita (Guam)	6084	608.4	<b>NS</b>	<b>Possible suppression</b>
Sinajana (Guam)	2592	259.2	<b>NS</b>	<b>Possible suppression</b>
Hagatna (Guam)	1051	105.1	<b>NS</b>	<b>Possible suppression</b>
Asan-Maina (Guam)	2137	213.7	<b>NS</b>	<b>Possible suppression</b>
Piti (Guam)	1454	145.4	<b>NS</b>	<b>Possible suppression</b>
Yona (Guam)	6484	648.4	<b>NS</b>	<b>Possible suppression</b>
Merizo (Guam)	1850	185.0	<b>NS</b>	<b>Possible suppression</b>
Umatac (Guam)	647	64.7	<b>Possible suppression</b>	
Inarajan (Guam)	2273	227.3	<b>NS</b>	<b>Possible suppression</b>
Talofofo (Guam)	3050	305.0	<b>NS</b>	<b>Possible suppression</b>
Mongmong-Toto-Maite (Guam)	6825	682.5	<b>NS</b>	<b>Possible suppression</b>
Agana Heights (Guam)	3496	349.6	<b>NS</b>	<b>Possible suppression</b>
<b>Northern Mariana Islands (NMI)</b>	47329	4732.9	<b>NS</b>	<b>Possible suppression</b>
<b>Saipan (NMI)</b>	43385	4338.5	<b>NS</b>	<b>Possible suppression</b>
Tinian (NMI)	3540	354.0	<b>NS</b>	<b>Possible suppression</b>
Rota (NMI)	2527	252.7	<b>NS</b>	<b>Possible suppression</b>

<b>American Samoa (AS)</b>	49710	4971.0	<b>NS</b>	<b>Possible suppression</b>
AS Eastern District	23030	2303.0	<b>NS</b>	<b>Possible suppression</b>
AS Western District	19950	1995.0	<b>NS</b>	<b>Possible suppression</b>
AS Manu'a District	1100	110.0	<b>NS</b>	<b>Possible suppression</b>
AS Swains Island	50	5.0	<b>Possible suppression</b>	
AS Census Tract 9501	1487	148.7	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9502	1158	115.8	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9503	2415	241.5	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9504	1376	137.6	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9505	2623	262.3	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9506	3218	321.8	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9507	1762	176.2	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9508	2081	208.1	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9509	3031	303.1	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9510	1446	144.6	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9511	6252	625.2	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9512)	3418	341.8	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9513	2932	293.2	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9514	2132	213.2	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9515	1698	169.8	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9516	4293	429.3	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9517	1342	134.2	<b>NS</b>	<b>Possible suppression</b>
AS Census Tract 9518	1708	170.8	<b>NS</b>	<b>Possible suppression</b>

**Figures:**

Figure 1: A. Statistical data suppression overview

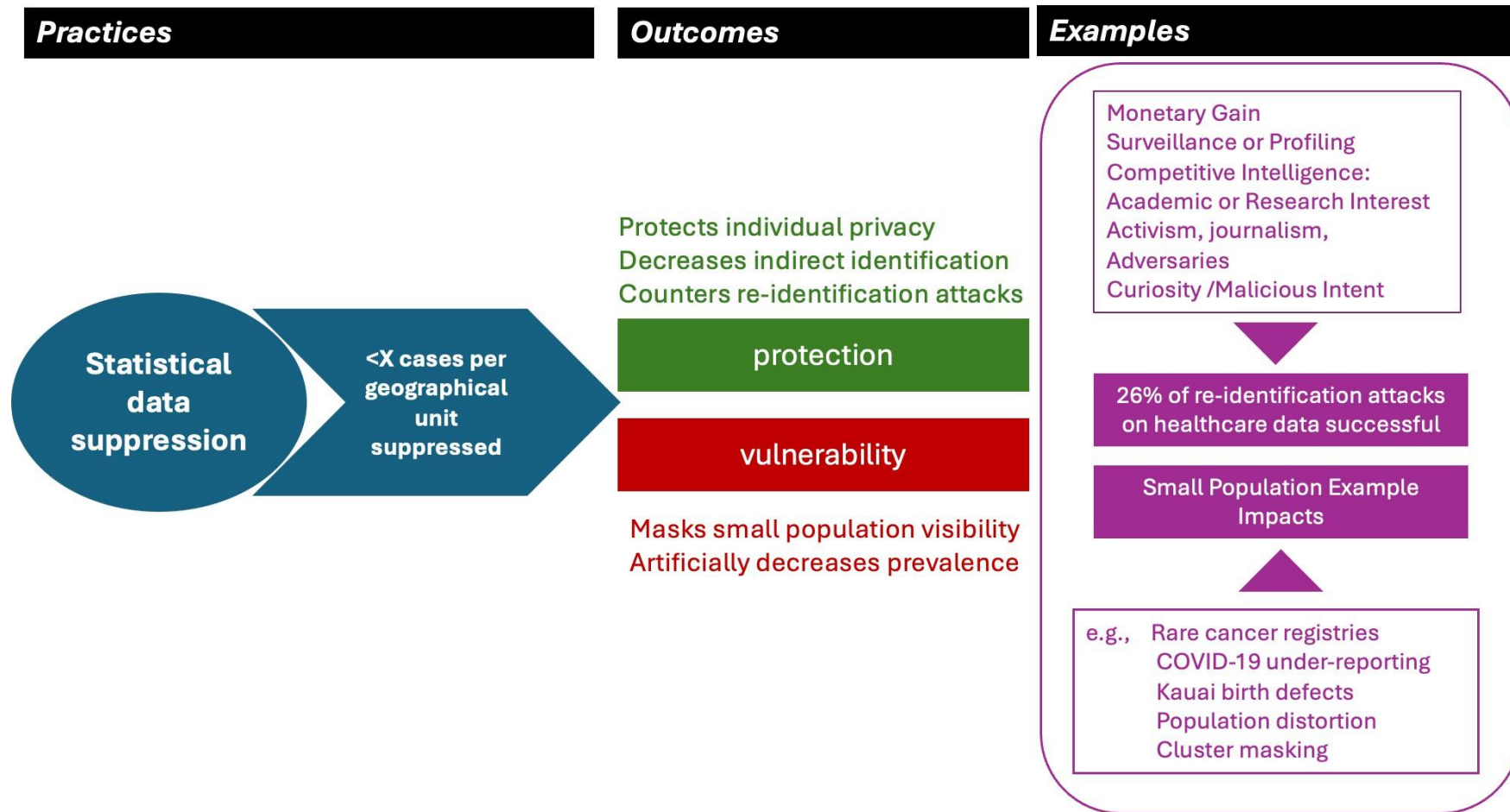
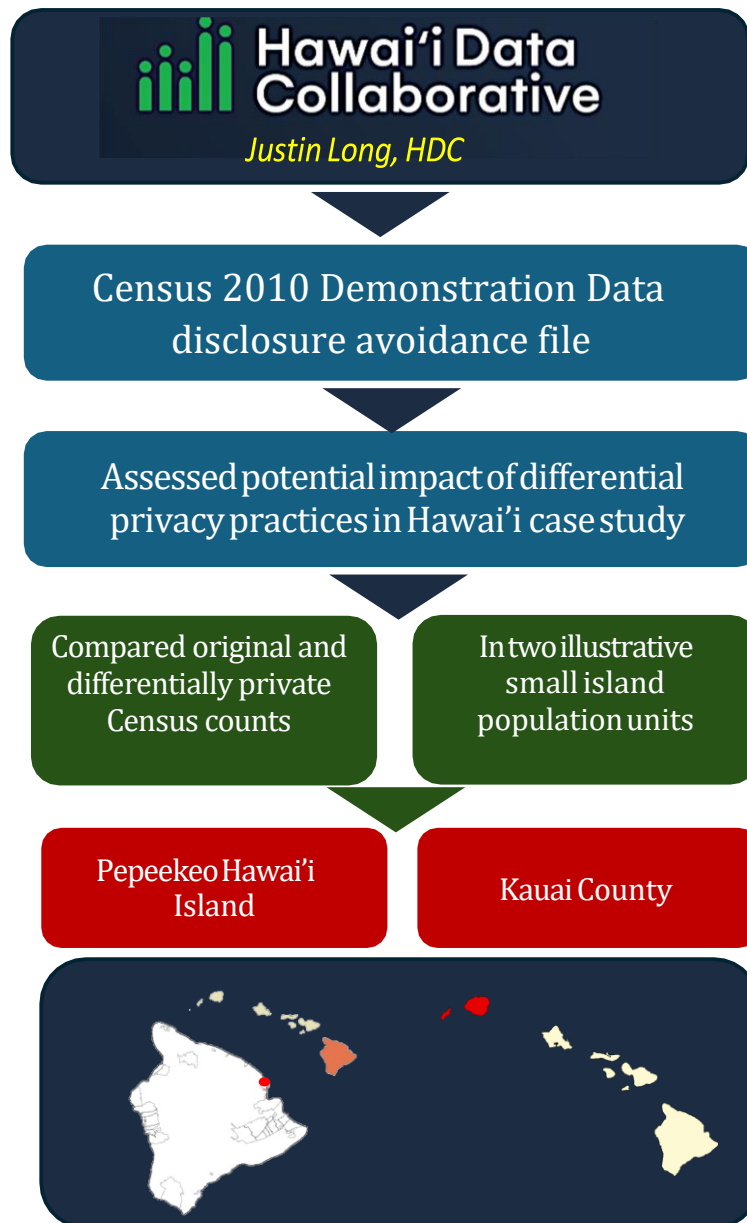


Figure 1: B. Example Hawai'i impacts of differential privacy practices



- **Population of 1,789 people per the 2010 Census**
- **604 persons fewer in the differentially private 'count' using the Census 2020 demonstration data file (1,185 persons)**
- **12 Black/African American Alone and 6 American Indian/Alaskan Native alone persons disappeared, and these categories were reported as zero in differentially private data**
- **NHOPI population pyramid structure was found to be distorted between original and differentially private counts, but total population did not vary significantly**
- **Examples:**
  - **117 males aged 5-9 disappeared in the differentially private counts**
  - **139 males aged 70-74 appeared in the differentially private counts**

Figure 2: Wayfinding through Pacific data suppression – a framework

