

Augmenting Research on South Asian and South Asian Women's Health
in America: The Case for Data Disaggregation

Priyanka Meesa
Thesis Submitted to the Duke Global Health Institute
Duke University
Durham, North Carolina
2023

Abstract

South Asian Americans are a rapidly growing ethnic group in the United States. They experience an array of health disparities that are not well-understood or addressed partially due to a lack of research on the health of this population. There are many factors that contribute to this paucity of research, one is the lack of data disaggregation within the Asian demographic category in U.S. health research datasets. Data disaggregation allows a finding to be analyzed in more granular way. This may help uncover differences that were invisible in the aggregated data. In the context of this paper, data on Asians is often aggregated into one “Asian” category, so it cannot be broken down into East, South, etc. Asian, making the health differences of these populations challenging to identify. This paper provides a deeper understanding of the complex issues surrounding data disaggregation (scientific, technical, regulatory, and ethical), from the perspectives of thought-leaders and subject matter experts informed by their real-world experiences of conducting precision and biomedical research. In this study, participants concurred that data disaggregation is essential to understanding South Asian American health disparities and that aggregation can be harmful. They also noted barriers to and risks of disaggregation. These barriers and risks can start to be addressed with increased funding and researcher awareness of South Asian health disparities. Furthermore, informants highlighted methodological challenges, such as defining what it means to be South Asian and capturing culture in data. This is especially important for understanding the health of women in the community, as South Asians have gendered practices and health behaviors as a result of their cultural and religious beliefs. Participants agreed that country of origin is a good starting point for disaggregation, but more information, such as immigrant status, acculturation, and religion, is important to truly understand health and develop interventions. More research is needed to

understand perspectives of community partners, members of the South Asian American community, policymakers, and research funders on augmenting research on South Asian American health to address health disparities. Increased data disaggregation facilitated by greater funding and awareness among those conducting and participating in research is an important first step to improve the ability of researchers to identify the health needs or outcomes specific to the South Asian population in the United States.

Table of Contents

Abstract	2
Acknowledgements	6
Background	7
South Asian American Health Disparities	7
Stress and Mental Health	10
South Asian American Women Health Disparities	11
Lack of Research Context	14
Data Disaggregation	19
Study Aims	21
Methods	24
Setting	24
Sample	24
Data Collection Procedures	25
IRB Approval	28
Analysis	28
Data Analysis	31
Results	32
Participant Characteristics	32
Challenges for Studying South Asian American Health	40
Benefits of Disaggregation/Harms of Aggregation	41
Risks of Data Disaggregation	41
Barriers to Data Disaggregation	41
How to Disaggregate	42
Impact on South Asian American Women’s Health	43
Discussion	44
Risks, Barriers, and Methodological Challenges	47
Facilitating Disaggregation	50
Impact on South Asian American Women’s Health	52
Limitations	54

Conclusion	55
References	57

Acknowledgements

I have multiple individuals that I would like to thank for their contributions to the development of this thesis project. Firstly, I would like to thank Dr. Subhashini Chandrasekharan, ELSI Team Lead of *All of Us* Policy Office. Dr. Chandrasekharan encouraged my research interests and helped me build my project from the ground up. She offered me countless hours of guidance and invaluable mentorship, empowering me to address my research question and pursue my passions for ethics, genetics research, and women's health. I would also like to thank Dr. Megan Huchko, Hollier Family Associate Professor of Global Health, for serving as my Duke research mentor and facilitating my independent study, and Dr. Dennis Clements, Professor Emeritus of Pediatrics at Duke, for serving as a member of my thesis committee and offering feedback on my project.

Furthermore, I would like to thank Dr. Katherine Blizinsky, Director of Policy for *All of Us*, for her help shaping my project's aims and for taking the time to provide essential feedback. I would also like to thank Dr. Sonya Jooma, Office of Policy NIH *All of Us* Research Program, and Dr. Sheri Schully, Deputy Chief Medical and Scientific Officer of *All of Us*, for sharing their perspectives on the topic of data disaggregation as informational interviewees. Finally, I deeply thank all the key informants who volunteered their time to be interviewed as a part of my project.

Background

As defined by the 2010 U.S. census, an Asian individual is a person who has origins in East Asia, Southeast Asia, or the Indian subcontinent (Hoeffel, 2010). Asian Americans are the fastest growing minority group in the United States (Budiman and Ruiz, 2021). There are more than 23 million Asian Americans from over 20 countries, and the Asian population in the United States is projected to reach 46 million by 2060 (Budiman and Ruiz, 2021). Within the broader Asian racial category, South Asians are an ethnic group of individuals who trace their heritage to Afghanistan, Bhutan, India, Maldives, Nepal, Bangladesh, Sri Lanka, and Pakistan (Cigna, 2020). South Asians are the fastest growing ethnic group in the United States; approximately 5.4 million South Asians live in America, and they compose a diverse community (SAALT, 2009). There are seven major religions practiced in South Asia (*Religions*, n.d.), leading to a host of unique cultural practices and beliefs within this ethnic group. Furthermore, many South Asians Americans speak different languages and eat distinct foods, some are immigrants or refugees, while others were born in the United States. There is wide socioeconomic variation among this population (Cigna, 2020). The diversity of this population is also reflected in the unique health disparities South Asians Americans experience.

South Asian American Health Disparities

South Asians Americans demonstrate differences in their health risks and outcomes compared to the general population of the United States. One health disparity experienced by South Asians is a high risk for cardiovascular disease. South Asians Americans experience four times the risk of cardiovascular disease (CVD) than the general population (Cigna, 2020). They develop heart disease a decade earlier and are twice as likely to have a heart attack (SAALT, 2009). Nearly one in every three South Asians will die from heart disease before the age of 65,

and these coronary events are deadlier than those experienced by any other ethnic group in the United States (Uppal and Gondi, 2021). An estimated 60 percent of the world's cardiovascular patients are of South Asian descent (Jarvis, 2019). The high prevalence of CVD among the South Asian population has been reported in the literature for almost 30 years, yet there is not enough research on the heightened risks to understand its primary causes and how to reduce the disparity (Uppal and Gondi, 2021). For example, the heightened risk of CVD could be a result of lower high-density lipoprotein (HDL) levels or higher levels of ectopic and visceral fat among the population, as these are indicators of CVD among the general population (Volgman et al., 2018). However, obesity does not predict CVD for South Asians in the same way that it does for their white counterparts (Uppal and Gondi, 2021). This is partially because the risk for CVD in South Asians starts at a lower body mass index (BMI) than that in the general population (23 as compared to 25) (Mongraw-Chaffin et al, 2018). One of the reasons understanding the exact risks that predispose this population to higher levels of CVD is challenging is likely due to the lack of research performed on South Asian populations. For example, there is an absence of testing for lipid/inflammatory biomarkers in the population, despite data demonstrating that South Asians have elevated levels of cholesterol and lipoproteins in their blood (Uppal and Gondi, 2021). More research is needed to understand the unique factors that predispose this population to CVD.

Another health disparity associated with South Asian Americans is prevalence of elevated levels of type 2 diabetes. These levels are more than seven times greater than that among white Americans, despite lower reported rates of smoking, obesity, and hypertension (risk factors for diabetes) in the South Asian population in the United States (Uppal and Gondi, 2021). This heightened risk continues to be seen when comparing the South Asian American population

to other ethnic groups in the United States. The prevalence of type 2 Diabetes Mellitus in South Asians is 26% compared to that of African Americans at 18%, Hispanics at 17%, and Chinese Americans at 13% (Flowers et al., 2019). Furthermore, South Asians living with diabetes have a higher mortality rate than other ethnic groups in the U.S. due to micro and macrovascular complications occurring at the time of diagnosis (Cigna, 2020). These high rates may be due to possible genetic factors that contribute to insulin resistance or high rates of hypertension among the population (Asian, 2008). The risks associated with diabetes are further compounded by the high prevalence of CVD in this population, augmenting the need for this group to have targeted diabetes detection and treatment plans that account for South Asian-specific dietary and cultural behaviors that might affect risks for diabetes and health care.

In addition to diabetes, patterns of cancer incidence and prevalence are unique among South Asians. When compared to non-Hispanic white Americans, Asian Americans have a lower cancer mortality rate. In fact, South Asians have one of the lowest rates of lung and colorectal cancer among all ethnic groups in the United States (Nelson, 2016). However, the prevalence of breast cancer, ovarian cancers, and lung cancer among this population is trending upwards. Furthermore, when adjusting cancer prevalence among South Asian Americans to accommodate for BMI and lifestyle factors, such as smoking, the cancer incidence rates become comparable to those found in white Americans (Tran et al., 2018). Considering global data, cancer incidence rates are higher in Southeast Asia and the Republic of Korea than in South Asian countries, including India, Bhutan, and Sri Lanka (Jamal et al., 2019). However, significant differences in cancer outcomes appear when accounting for lifestyle factors, such as sedentary lifestyles and smoking. These behaviors can be common in the Western lifestyle. This suggests that the adoption of the Western lifestyle may affect the cancer trends and outcomes of South Asian

Americans (Nelson, 2016), making a case for the need for more research into cancer among this population to develop early detection and prevention programs. Furthermore, HIV/AIDS significantly affects the South Asian community, as many acquire the disease after immigrating to the United States. However, the disease is highly stigmatized, and information about treatment and transmission is limited among South Asian communities (SAALT, 2009).

Stress and Mental Health

Another potential area for health disparities is in the arena of stress and mental health. There is a lack of research on the prevalence and symptoms of mental health in South Asian American populations; therefore, robust data are not available, and trends are challenging to identify (Lai and Surood, 2013). Research into this area is crucial, as the population is fast-growing and experiences new stressors related to immigration and acculturation to America. Interestingly, studies have found that there exist generational differences in mental illness symptom presentation among younger and older South Asian populations. In younger South Asians, stress was the most common symptom of mental illness, and much of the stress of this population was associated with acculturation (Rastogi et al., 2014). In contrast, in older South Asians, the primary symptoms were psychosis, anxiety disorder, and severe depression (Rastogi et al., 2014). The differences in symptom manifestation between younger and older generations suggests that, even among South Asians, acculturation affects the presentation of mental illness differently within the population, underscoring the need for more research in this area. Furthermore, stress and depression are major risk factors for CVD, which South Asians already experience a higher predisposition to.

The challenges associated with the lack of research into the mental health of South Asian Americans are further compounded by the immense stigma around and denial of many mental

illnesses among the population (Cigna, 2020). Rastogi et al. noted that South Asians are more comfortable with the terms “stress, anxiety, or weakness” as opposed to depression or psychosis (2014). This cultural stigma can often result in the under-diagnosis of mental illness in the population. For example, Lubin and Khandai found that there are smaller numbers of South Asians diagnosed with mood disorders and anxiety conditions than the general population (2017). However, more South Asian women actually meet the DSM-IV criteria for mood and anxiety disorders (Lubin & Khandai 2017). This suggests that there is significant under-reporting of mental illness among the population. Despite lower rates of DSM-IV diagnoses among the population, suicide is the eighth leading cause of death for Asian Americans compared to it being the eleventh leading cause of death for all racial groups in the United States (American, n.d.). This high rate is particularly salient among South Asian women, as past research has found that South Asian female immigrants experience higher rates of suicide compared to their male counterparts and to women native to the country they immigrate to (Patel and Gaw, 1996). This suggests that South Asian women experience mental health in a manner distinct from their male counterparts, highlighting the need for female-specific research.

South Asian American Women Health Disparities

In addition to experiencing higher rates of suicide, South Asian American women also experience unique challenges related to maternal health, domestic violence, osteoporosis, and PCOS that are distinct from the general population in the United States. For example, pregnant South Asian women are at a higher risk for developing gestational diabetes than their white counterparts. Data from the South Asian subcontinent show that women from Bangladesh are the most prone to developing gestational diabetes, followed by women from India, Sri Lanka, and Pakistan (Sanchalika and Theresa, 2015). Along with being more prone to developing gestational

diabetes, the risks associated with type 2 diabetes mellitus in South Asian populations may be augmented for female South Asians due to health behaviors. For example, when examining the role of culture in type 2 diabetes management, Patel et al. found that South Asian women were less likely to participate in physical activities or have the appropriate familial support to manage their T2D (2021). This is because in South Asian cultures, it is often believed that the women's role is to take care of the home, cook, clean and care for dependents. Therefore, it is frowned upon when South Asian women dedicate time to self-care activities, such as physical exercise and taking medication, which would confirm their sick role (Patel, 2021). Not only are South Asian women more likely to suffer from gestational diabetes, but they are also more likely to have worse T2D outcomes than their male-identifying counterparts.

South Asian women also experience health disparities within the realm of maternal health. Studies have found that South Asians, Africans, and South Americans have a higher pooled prevalence of anemia than that found among East Asians (Rahman et al., 2016, p. 495). More specifically, South Asian American women are more likely to have anemia than the general population of the United States (South, 2002). Anemia during pregnancy is associated with low birth weight among newborns, preterm birth, and higher levels of perinatal/neonatal mortality (Engidaw et al., 2022). This trend is particularly salient for mothers who originate from lower-income countries or families, where malaria and poor nutrition is more common (Rahman et al., 2016, p. 495). For example, the risk of maternal anemia-related low birth weight is highest among mothers from Pakistan and Bangladesh. The risk of maternal anemia-related perinatal mortality is most prominent among mothers from India and Pakistan (Rahman et al., 2016, p. 495). Women in the United States who immigrated from India are more likely to deliver low birth weight infants than White women or women from other ethnic groups (South, 2002).

Furthermore, Drysdale et al. found that women born in South Asia had a higher risk of stillbirth than women born in Australia (2012). This trend could be due to a variety of factors, such as culture and circumstance. War, internal conflict, and lack of health financing all contribute to poorer maternal outcomes in lower income countries. Additionally, there has largely been a culture of silence around issues of sexuality and reproductive health in South Asian cultures paired with limited access to contraceptives (South, 2002). More research is needed to understand the factors that lead to anemia and poorer maternal outcomes among South Asian women. Importantly it is also not well-understood how these trends change when South Asian women immigrate and acculturate to the United States.

Another risk that South Asian women face at a disproportionately high rate is osteoporosis and polycystic ovary syndrome (PCOS). Osteoporosis results in bones becoming weak and brittle, causing even small falls or mild stressors to become fractures (Mayo, 2021). Generally, women are four times more likely than men to develop osteoporosis, and individuals of Asian origin are at a higher risk of developing osteoporosis (South, 2002). One in five Asian American and Pacific Islander women develop osteoporosis. Indian women over the age of 50 have a 40% chance of having an osteoporotic fracture at some point in their life, and it is estimated that about 25% of Indian women will have osteoporosis in their lifetimes. South Asian women likely experience a higher risk of osteoporosis due to both their diet and the high number of pregnancies they experience (South, 2002). Having multiple pregnancies leads to a loss of calcium due to lactation frequency, affecting bone recuperation. Additionally, many South Asians consume a vegetarian diet, and the vegetarian options provided by the American diet might not provide them with the necessary levels of calcium to maintain strong bone health (South, 2002).

South Asian women also have a high prevalence of PCOS. PCOS is the most common endocrine disorder in women, and it leads to a range of life and emotional stressors, such as obesity, depression, infertility, insulin resistance, etc. (Factors, n.d.). South Asian women are at a higher risk of developing PCOS than other ethnic groups in the United States. Studies have shown that South Asian women also typically develop PCOS earlier than their white counterparts. The reasons for this increased risk are not well-understood, and, again, more research needs to be performed to understand how acculturation to the United States lifestyle affects the incidence of PCOS in South Asian women. Understanding the risk factors for higher PCOS prevalence is especially important, as South Asian women are also at an increased risk of developing cardiometabolic diseases. Therefore, early detection is imperative (Factors, n.d.).

Finally, South Asian American women experience intimate partner violence at high rates (IPV). For example, Raj and Silverman found that domestic violence was a serious problem and highly prevalent among a cohort of South Asian women living in Boston (2002). Many women in their study did not know what services existed to seek help, and victim-blaming attitudes were highly prevalent among the community (2002). This phenomenon is likely true for many South Asian women in the United States, suggesting that South Asian women often suffer from domestic violence but don't receive the help they need. This highlights the need for more research into intimate partner violence and for developing culturally tailored interventions to prevent/mitigate IPV.

Lack of Research Context

It is evident that South Asian Americans in the United States and, more specifically, South Asian American women, experience health disparities compared to the general population. Although disparities have been identified, the causes of these disparities are not well-understood.

South Asians comprise almost a quarter of the world's population, yet research about the health outcomes, biological, environmental, and social determinants of the health of this population is sparse (*Southern*, n.d.). The reasons for this lack of research are multifaceted and include, but are not limited to, the model minority myth, lack of research participation, stigma, barriers to health care accessibility, lack of funding for research into South Asian health, and lack of data disaggregation within the Asian category in U.S. research datasets.

The model minority myth argues that Asian Americans are a minority group in the United States that are not underprivileged or at a disadvantage. It relies on stereotypes about Asians and perpetuates the narrative that Asian Americans face few health problems, are highly successful, and are wealthy (Blackburn, 2019). The myth was created in the 1960s by white supremacists to oppose the Black Power Movement and deny the existence of racism in the United States by demonstrating how the United States does allow ethnic groups to be well-off (Holland and Palaniappan, 2012). Although decades have passed since its creation, the myth persists, homogenizing all Asian Americans into one category and obscuring the unique health conditions experienced by South Asian Americans. As a consequence, it is perceived that Asian Americans are not at risk for health problems. This results in late detection and treatment among many Asian American populations (South, 2002).

Furthermore, research efforts to better health often extend to those defined as underserved. Due to the model minority myth, Asian Americans fall outside of the classification of underserved, resulting in a lack of health disparities research on this population (Blackburn, 2019). For example, a study performed in 2021 found that Asians are underrepresented in high-impact medical research studies and funding (Nguyen et al., 2021). Despite this, in their recent analysis of grants and postdoctoral trainees supported by the NIH, Asians were not considered by

the NIH to be an underrepresented minority in biomedical research. For example, the NIH website does not list Asians as one of the groups that are underrepresented in research (U.S., n.d., “*When,*” 2021). This narrative is perpetuated for South Asians too. In 2018, the proportion of United States physicians identifying as Indian or Pakistani (4.8%) was greater than the proportion of individuals in the United States identifying as South Asian (1.2%) (Uppal and Gondi, 2021). The overrepresentation of South Asians in the healthcare workforce lends itself to the false narrative that South Asians are well-off and do not experience health disparities. However, looking closer at the gaps in the Asian American population, it is evident that the model minority myth is not accurate. For example, income inequality among Asian Americans is rising. Out of all minority groups in the United States, Asian Americans have the highest level of income inequality (Blackburn, 2019). In fact, in 2016, the richest Asian Americans earned almost eleven times more than the poorest Asian Americans. The average income of Asian Americans obscures this disparity, as those making the most skew the average for those not economically well (Hoeffel et al., 2010). Overall, the model minority myth perpetuates a false narrative of the success of Asian Americans, contributing to the perception of a diminished need for research on the health of these populations.

In addition to this erroneous perception, South Asians in the U.S. are also underrepresented in health research due to a lack of research participation. The lack of participation of South Asian Americans in research is nuanced, and due to a variety of factors, including lack of education, disinterest, fear, logistical barriers, and language barriers. For example, Quay et al. found that South Asians often did not participate in health research because they had a poor understanding of the intentions of researchers or the consent process (2017). Quay et. al. also found that, due to their status as immigrants or outsiders, some South Asians did

not see how participating in research in the United States would benefit them. Many South Asians did not understand the importance or benefit of participating in the research in general, especially after already having received treatment for their health conditions (Quay et al., 2017).

Another factor that affects research participation of South Asians is fear and stigma. Mukherjea et al. found that individuals stated they did not want to participate in research for fear of re-identification (2018). South Asians were also afraid of being reported to immigration, concerned about potential adverse effects of the study on their health with respect to clinical trials, worried about finding updates about their health status, and afraid of being labeled with the stigma associated with a health condition (Quay et al., 2017). For example, a male focus group participant in Mukherjea et al.'s study stated that certain sections of the South Asian community in America would treat him as an outcast the moment they found out he had cancer (2018). Some South Asian women were apprehensive about participating in studies or answering questions without the presence of their preferred male counterpart (husband, son, father, etc.) or were hesitant about revealing sensitive information in front of strangers (Mukherjea et al. 2018). Women stated that they would not participate in gynecological or breast examination studies if the researcher was of the opposite gender (Hussain-Gambles, 2004). Additionally, South Asians, and especially South Asian women, were less likely to participate in research focusing on topics not discussed in South Asian communities, such as mental health and reproductive health.

South Asian Americans also experienced logistical barriers to research participation. Often the travel distance, frequency of research follow-ups, time needed to participate, and duration of the study serves to dissuade South Asians from participating (Mukherjea et al., 2018). Another contributing factor is language barriers. Significant proportions of South Asian American populations do not speak English, including 50% of Bangladeshis and over 30% of

Pakistanis. However, hospitals and research study leaders often do not speak the native languages of the populations or have adequately trained interpreters. This results in family members having to serve as translators, facilitating mistrust or communication errors (SAALT, 2009). Furthermore, recruitment by researchers is often too passive, the study fliers and materials are not linguistically appropriate, or researchers have not invested enough time in building trust with the target communities (Mukherjea et al., 2018).

Another factor contributing to the lack of research on South Asian health is South Asian Americans' under-utilization of the healthcare system, despite the overrepresentation of South Asians physicians in the system. A 2009 review performed by SAALT found that approximately one in five South Asian Americans lack health insurance. Approximately 40% of South Asian Americans under the age of 65 reported lacking a regular source of healthcare, and among the South Asians with health insurance, 15% reported not seeing a primary care physician (2009). A potential contributor to this phenomenon is the culture of valuing privacy and honor, and individuals may perceive that visiting a healthcare provider can disrupt privacy and potentially cause embarrassment (South, 2002).

Additionally, there exists a lack of funding for research on South Asian health. Although Asian Americans compose 6-7 percent of the United States population, only 0.17% of NIH-funded research goes towards Asian American health (Tachibaba and Weiner, 2022). This lack of funding prevents more studies focusing on the health of South Asian Americans from being performed.

Finally, another key issue for the underrepresentation of South Asian Americans in health research is the lack of standardized practices of data disaggregation by meaningful population descriptors in health data. When data on Asian Americans is collected and analyzed, it is often

not disaggregated by ethnic group and sometimes includes data for Native Hawaiian and Pacific Islanders (Holland and Palaniappan, 2012). The United States Census only began separating Asian and Pacific Islanders as a category in 2000, and, in 2003, Congress approved for the census to add Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, and Other Asian to the identification section (Lee et al., 2020). When disaggregated data is collected, it is often difficult for researchers to access and/or understand and/or costs extra. As a result of these barriers, it is exceedingly challenging for researchers to examine health differences among different Asian and Pacific Islander subgroups. This is problematic given the unique experiences, culture, and, as detailed earlier, health conditions of different Asian ethnicities, resulting in the masking of important health disparities.

Data Disaggregation

A study published in 2020 in the American Journal of Public Health examined the impact of data disaggregation on the health of Chinese, Japanese, Korean, Vietnamese, and Filipino adults living in California. The study found that, as an aggregate group, the Asian Americans in the study appeared healthier than non-Hispanic Whites on most metrics. However, every Asian subgroup experienced at least one health disparity when looking at disaggregated data. Filipinos had the most health disparities, scoring higher on obesity, diabetes, high blood pressure, and asthma levels than their white counterparts (Adia et al., 2020). This disparity would not have been recognized if Asian groups were not disaggregated, resulting in a lack of attention or care for the health outcomes of Filipino Americans. Another example occurs when looking at cancer trends. For example, looking at aggregate data, Asian Americans have the lowest cancer mortality rates of all ethnic groups, including Whites. However, when looking at disaggregated data, the incidence rate of cervical cancer in Vietnamese American women is twice that of non-

Hispanic white women (Le and Zeng, 2021). Another example of the misrepresentation caused by a lack of data disaggregation centers around C-reactive protein levels. C-reactive protein is produced by a liver, and the level of the protein increases when there is increased inflammation in the body, which may be attributed to cancer, infection, heart attack, inflammatory bowel disease, lupus, and other diseases (*C-reactive*, n.d.). Researchers found that, in aggregate, Asian American women have the lowest C-reactive protein levels of all groups in the United States (Asian, Hispanic, Black, and non-Hispanic White). However, a similar study conducted in the United Kingdom found that Asian Indians have higher levels of C-Reactive protein than their European counterparts (Forouhi et al., 2001). This suggests the existence of differences of C-reactive protein levels between Asian subgroups and demonstrates how aggregation can lead to potentially false conclusions. Furthermore, looking at data on chronic health conditions, Filipinos, South Asians, and Pacific Islanders tended to have a meaningfully higher prevalence of chronic conditions than the aggregate Asian group, while Chinese, Korean, Japanese, and Southeast Asian groups had lower prevalence (Gordon et al., 2019). In summary, Asian Americans are not a homogenous group, and aggregating their health data masks meaningful health disparities.

Another factor to consider when thinking about the aggregation of Asians into one group is perceived racism and how that affects the life experience of different Asian subgroups. The standard Asian phenotype in America is that of East Asian. South Asians and Filipinos have darker skin tones than their East Asian counterparts as a result of the ancestral population's geographical location closer to the equator with a higher UV index (Maddodi et al., 2012). As a result, they often have too much melanin to be considered Asian and are not dark-skinned enough to be considered black, so they experience an uncertain racial positioning in America

(Selod, 2018). The racism that an individual experiences influences their health, as exposure to discrimination can lead to poor mental and physical health as well as higher stress (Yom and Lor, 2022). Racism is a result of an individual's perceived racial identity and allows people access to certain privileges as well as propagates discrimination. East Asians are distinct from being identified as black/brown, whereas South Asians and Filipinos fall into this classification. Brown Asians often fall outside of the public's classification of "Asian" and may be racialized as something distinct. Therefore, the privileges and experiences of East Asian groups are not accessible for South Asians or Filipinos. Furthermore, South Asians experience unique racial stressors, such as anti-Indian sentiment and anti-Muslim sentiment. For example, in the late 80s and early 90s, the Dot Busters in New Jersey emerged as a violent group targeting Hindus. This group used violence to force any Hindu they saw on the street out of the city (ArZan, 2020). Another example is the anti-Muslim sentiment post 9/11. Bakhtiari found that exposure to Islamophobia was associated with poorer health outcomes and reduced help-seeking behaviors among individuals perceived as Muslim, including Asian Indians (2020). In contrast, East Asians experience the legacies of the Chinese exclusion act as well as unique racial sentiments as a result of COVID-19. The Asian groups that are traditionally aggregated into one category in health research have distinct historical and socio-political factors that cause them to experience racism and discrimination differently, once again illustrating the pitfalls of homogenizing the group through data aggregation.

Study Aims

Recently, more and more health data are collected about Asian ethnic groups, and research programs like *All of Us* and *Kaiser Permanente* are beginning to focus on the inclusion of diverse groups for biomedical and precision medicine research. The programs have invested in

building “big data” sets with longitudinal and multimodal data collection to capture a holistic picture of the biological, genetic, environmental, and social factors affecting the health risks, disease prevalence, and health outcomes of diverse populations in the United States. The efforts of these programs present a tangible opportunity to advance research on the health disparities experienced by South Asian Americans and South Asian American women by increasing the representation of South Asians in health data. However, some of the barriers to representation of South Asians in this data, as described above, continue to persist despite robust attempts to recruit South Asian participants and collect more data on the population. Furthermore, data repositories across the United States do not have a standardized method of collecting and disaggregating the data on these groups, making it difficult to compare across datasets. Importantly, due to lack of data collection in adequate sample sizes and lack of disaggregation, South Asian health data and disparities may remain invisible in these large data repositories. As a first step to developing a deeper understanding of the complex issues surrounding data disaggregation (scientific, technical, regulatory, and ethical), it is important to understand the perspectives of thought-leaders and subject matter experts informed by their real-world experiences of conducting precision and biomedical research.

The aim of this project is to better understand how precision medicine research can potentially benefit and/or harm the health of South Asian Americans, more specifically South Asian American women. In this paper, a woman is defined as any individual who identifies their gender as a woman, including those whose sex assigned at birth is female. As detailed above, South Asian Americans experience health disparities; however, there is a lack of research on this population to better understand the reasons for and factors contributing to these disparities, and how to mitigate them. The lack of research may be due to reasons such as the lack of

participation of South Asians in health research, and lack of data disaggregation by the Asian subcontinent in most research databases in the U.S., among others. This project focuses on one facet of the issue - lack of resolution within health datasets, which limits South Asians from being examined as a distinct population group within the broader general Asian category in the context of health disparities research. We would like to explore how well-equipped precision medicine research programs are to address health inequities affecting South Asian Americans and South Asian American Women. Specifically, we hope to understand the role lack of resolution (by disaggregating race and ethnicity in the Asian category) within datasets plays in limiting such research to adequately address the health needs of South Asian Americans. Furthermore, we hope to investigate the implications for the design, implementation, and policies of precision medicine programs with regards to data collection and data sharing so precision medicine programs can meaningfully include this population and make a positive impact on South Asian women's health. It is our hope that the findings of the study will inform policies and practices of precision medicine research programs to facilitate the study of health disparities among South Asian communities, especially for women. More broadly, learning how to appropriately disaggregate health data in research databases is important to understand the unique health experiences of all groups in the United States, ensuring health disparities are visible and interventions can be developed for all groups and communities, and especially minority groups

To understand these questions, a series of semi-structured interviews with key implementers of large-scale precision medicine programs and biomedical research initiatives focused on South Asians Americans as well as individuals studying the health of this group were performed. The interview transcripts were qualitatively analyzed to surface themes and

recommendations to better understand barriers to disaggregating South Asians' data for health research and how to ensure appropriate representation of South Asian Americans and South Asian American women in health research data.

Methods

Setting

This study was composed of a series of zoom interviews of participants. All participants resided in the United States. Participants represented four settings: South Asian American research initiatives, biomedical/precision medicine research programs, publications about South Asian Health, or were referred to participate by key informants.

Sample

The study sample consisted of 10 key informants. In order to meet the eligibility requirements, individuals had to work in or perform research in the United States and had to either participate in South Asian American research initiatives [such as SATHI (South Asian Total Health Initiative), SAHI (South Asian Health Initiative), SAAHI (South Asian American Health Initiative), or SAHARA (South Asian Health and Research Group)], biomedical/precision medicine research programs (such as *All of Us*, *Kaiser Permanente*, *Geisinger My Code*, etc.), have published articles about South Asian American Health, or were referred to me as potential subject matter experts by interviewees I reached out to. Informants participating in South Asian American research initiatives worked in roles of program coordinator, assistant clinical director, and director. Participants working for biomedical/precision medicine research programs represented roles of Director of Health Equity; Chair on Committee on Access, Privacy, and Security; Senior Investigator; and Precision Medicine Researcher. Informants who published

articles about South Asian Health performed research on life expectancy and differential racialization as well as South Asian women's health.

Data Collection Procedures

To recruit study participants, individuals who fell into the categories listed above were identified through Google searches of precision medicine initiative program and South Asian research initiative websites, publications on South Asian American health or through citations in publications about South Asian American health, and by asking participants if they could share the contact information of any subject matter experts they would recommend for inclusion in this study. A list of 52 candidates was created, and their email address was found through a Google or Pubmed search, if their contact information was made publicly available. All 52 candidates were sent a recruitment email that described the purpose of the study and the study procedures. Out of the 52 candidates contacted, 17 responded to the recruitment email. Five said they could not participate, two said yes but did not respond to the follow-up email to set a date and time, and ten responded yes and were interviewed. Once responding to the initial email, participants were sent an informed consent form and an interview invitation. Verbal consent was received and recorded at the start of every interview. In the same email, participants were asked for their availability and were sent a zoom calendar invitation. Interviews were scheduled and conducted over a three-week period from January 27th to February 24th, depending on participant availability.

Before interviewing participants, a list of 22 interview questions was developed. Questions fell into the following broad categories: background information and experiences of key informants, perspectives on South Asian health research, data resolution and use, and recommendations for additional information resources/subject matter experts (Table 1).

Interviews lasted approximately 45 minutes and were entirely confidential. Participants were told they could go off the record anytime or not answer questions they did not want to. Interviews were tailored using the interview guide (Table 1) according to the expertise of the key informants. Out of the total of 22 questions, all participants were asked the questions that applied to their background and research experiences. Participants directly studying South Asian American health were asked the questions listed under the *South Asian Health Research* question category (Table 1), whereas this category of questions was omitted for some participants working on biomedical medicine research programs with no specific population group focus. All interviews were recorded, and the recordings were stored in a secured DukeBox folder to be deleted after the completion of this project. All identifying information was removed from transcripts for analysis. The interviews were transcribed for analysis, and the transcripts were stored in the same secure DukeBox folder to be deleted after completion of the project.

Table 1

Interview Questions.

Questions
Background information and experiences of key informant
<ul style="list-style-type: none"> ● Tell me about your role in your program/organization?
South Asian Health Research (Only for those conducting South Asian Health research)
<ul style="list-style-type: none"> ● What does your research focus on?
<ul style="list-style-type: none"> ● What particular health or health disparity issues related to South Asian women do you study?
<ul style="list-style-type: none"> ● What are the challenges you encounter in conducting this research? <ul style="list-style-type: none"> ○ Specifically with regards to research on South Asian women's health?
<ul style="list-style-type: none"> ● From your experience, what are some of the specific challenges that researchers encounter when studying South Asian health? <ul style="list-style-type: none"> ○ And South Asian women's health?

<ul style="list-style-type: none"> • From your perspective, are there gaps in data collection about South Asians, more specifically about South Asian women, for example, in electronic health records or other standardized health data collection tools?
Data Resolution and Use
<ul style="list-style-type: none"> • In general, in biomedical research datasets, are there challenges for focusing on or studying South Asian health? <ul style="list-style-type: none"> ○ What are some of the challenges you have encountered? Specifically what challenges around data collection, resolution, and/or use have you encountered (have any involved lack of data disaggregation?)
<ul style="list-style-type: none"> • Does your organization/program disaggregate the Asian category for research purposes?
<ul style="list-style-type: none"> • Why was this choice made?
<ul style="list-style-type: none"> • What are the barriers to and constraints or challenges of disaggregating data based on [race, ethnicity, ancestry], and specifically disaggregation that would allow for the resolution of South Asian populations?
<ul style="list-style-type: none"> • How might data disaggregation affect research on health outcomes for South Asian Women?
<ul style="list-style-type: none"> • What are the potential benefits or harms of disaggregating South Asian participants' data in biomedical research?
<ul style="list-style-type: none"> • Are there examples you think are especially illustrative of where data disaggregation has been helpful or harmful to South Asians, specifically South Asian Women?
<ul style="list-style-type: none"> • Are there any potential ethical implications to disaggregating data or not disaggregating the data in these ways?
<ul style="list-style-type: none"> • How would someone go about achieving more granular resolution within Asian participant populations (based on racial/ethnic/ancestral criteria) that would improve the ability to address subpopulation-specific health concerns without causing harm?
<ul style="list-style-type: none"> • What do you think might be other solutions to achieving data disaggregation without causing harm? <i>Ex. technological, community, research participant engagement, research oversight</i>
<ul style="list-style-type: none"> • What types of data disaggregation do you think would be most beneficial to study South Asian women's health? <ul style="list-style-type: none"> ○ For example, by geographic location, country of birth, parental country of origin, immigrant status?
<ul style="list-style-type: none"> • What might serve as drivers (such as demographic, political, societal trends) to facilitate disaggregation of the Asian demographic category in U.S. research datasets?
Closing
<ul style="list-style-type: none"> • Are there any other thoughts or concerns you would like to share with us?

<ul style="list-style-type: none"> • Who else would you recommend we talk to, and would you be able to connect me with them?
<ul style="list-style-type: none"> • Are there resources, publications, white papers, or other documents that might help our research efforts that you would be willing to share with us?
<ul style="list-style-type: none"> • Would you be open to a quick follow-up by email if we have any questions?

IRB Approval

All study procedures were approved by the Duke University Institutional Review Board.

Analysis

To analyze interview responses, thematic analysis qualitative analysis methodology was performed (*How*, 2022). Out of the ten total transcripts, three diverse transcripts (transcripts where participants had large variety in their responses to encompass a range of codes) were selected to build the codebook. One of my research mentors (Dr. Subhashini Chandrasekharan) and I read the three transcripts independently and proposed codes for the themes we saw. After looking over all three transcripts, a list of potential codes was compiled into a codebook, and these codes were grouped for analysis into categories. 52 codes were created and organized into 10 main thematic categories: current barriers to data disaggregation, risks of data disaggregation, harms of data aggregation, methodological challenges of data disaggregation, benefits of disaggregation, how to facilitate disaggregation, how to disaggregate, impact of research on South Asian women, challenges for research on disaggregated groups, and challenges for studying South Asian health (Table 2).

Table 2

Codebook.

Codebook
1. Current Barriers to Data Disaggregation
1.1 Small sample sizes

1.2 Convention
1.3 Loss of statistical power
1.4 Lack of research funding
1.5 Lack of awareness about importance/lack of researchers' interest
1.6 Lack of collection of granular information
1.7 Participant reticence to provide information
1.8 Lack of participant ability/desire to participate
2. Risks of Data Disaggregation
2.1 Risk of re-identification
2.2 Targeting small populations
2.3 Potential for discrimination/stigmatization
2.4 Misconstruing social phenomenon for biological (disproportionate focus on genetic ancestry of the groups → can miss social components of health)
2.5 Improper Inference
3. Harms of Data Aggregation
3.1 Drawing Incorrect Conclusions
3.2 Missing disparities
3.2a Disparities in access to services
4. Methodological Challenges of Data disaggregation
4.1 Defining populations
4.2 How to disaggregate
4.3 Missing confounding variables (diet, geography, local environment)
4.4 Capturing culture (need for qualitative and other methodological approaches to understand culture and acculturation)
5. Benefits of disaggregation
5.1 Reduce health disparities
5.2 Understanding behavioral differences between populations

6. How to Facilitate Disaggregation
6.1 Larger numbers
6.2 Community engagement (provide more data, participate in qualitative studies, understanding culture in study design)
6.3 Funding
6.4 Education of researchers and increasing awareness
6.5 Partnering with community organizations (building trust, etc).
6.6 Language representation
6.7 Increase researcher diversity
7. How to Disaggregate
7.1 Country of origin
7.2 Immigrant status
7.3 Number of years in the U.S.
7.4 Racial and ethnic background
7.5 Geographic location
7.6 Self-identification
7.7 Religion
7.8 Language
8. Impact of Research on South Asian Women
8.1 Health differences between men and women (develop targeted interventions, identify targeted risk factors)
8.2 Augment South Asian female participation (they have unique barriers to participation)
8.2a Increased awareness about research participation needs
8.3 Understanding South Asian woman experience to develop interventions (dependency, family-based interventions)
8.4 Address misconceptions
9. Challenges for Research on disaggregated Groups
9.1 Challenges accessing disaggregated data

9.2 Lack of disaggregated longitudinal data collection
10. Challenges for studying South Asian Health
10.1 lack of funding
10.2 Community outreach
10.3 Lack of trust
10.4 Longitudinal participation
10.5 Disseminating results impactfully
10.6 Lacking female participants
10.6a Mistrust
10.6b less perceived risk
10.6c Health Taboos Reducing Perceived Need for Research
10.7 Lack of diverse South Asian participant representation (in different SES, different languages, etc.)
10.8 Lack of data disaggregation
10.8a Inability to trust disaggregated data

Data Analysis

To analyze the data, the ten transcripts were imported into Dedoose, a qualitative coding software (*Home*, n.d.). Once imported, all transcripts were read and codes from the previously developed codebook were applied. The codebook was continually reviewed and revised to ensure all participant perspectives and points were accurately captured. Once codes were applied, the codes were examined and grouped into broad themes for analysis.

Results

Participant Characteristics

The ten participants in this study represented a diverse set of perspectives providing viewpoints from the collection of health data in biomedical research programs to the practice of performing research on South Asian American health. Two of ten participants are affiliated with the *All of Us* Research Program and three are affiliated with Kaiser Permanente. One participant is affiliated with the MASALA study, one with the National Cancer Institute, and one with the SAHARA study. The remaining two individuals performed research on South Asian American health, calling for greater need for data disaggregation in their publications/scholarly work. Four participants are men, and six are women. Three are post-doctoral students, and seven have PhDs.

Several dominant themes were identified primarily related to the barriers of, benefits and harms of, and best practices for data disaggregation within the Asian racial/ethnic category to allow for the resolution for South Asians in health research datasets. Illustrative examples for each code were recognized and listed (Table 3). Perspectives commonly shared by participants corresponding to the dominant theses are described below. It is important to note that the results section does not encompass the range of pertinent points and perspectives shared by participants. Instead, it highlights themes often brought up by multiple informants that seemed to resonate among participants as important for consideration when understanding data disaggregation.

Table 3

Illustrative Examples of Codes in Codebook.

Code	Illustrative Example
1. Current Barriers to Data Disaggregation	
1.1 Small sample sizes	“Yeah, I mean, part of it is you have to run a fine line between disaggregating to

	<p>have meaningful representation and understanding of what the data show you, and also having enough numbers in each of the cells. And as you divide people into smaller and smaller ethnic categories, or self-identified categories, you get tinier and tinier cells. And then you really can't make much in terms of talking about characteristics or disease prevalence or risk factor prevalence.”</p>
1.2 Convention	<p>“I think it could also be a partial racist issue. If you ask me, I mean, like you know, this is just the way that we thought about racial types in the United States, dating back almost a 100 years ago. And so, it just really hasn't been revised since then.”</p>
1.3 Loss of statistical power	<p>“I think one of the biggest is statistical power considerations. And so, whether you're looking at race as a primary exposure of interest, or whether you're using it as an adjustment variable. The more categories you add the lower your power gets. So as an adjustment variable, you're increasing those degrees of freedom in your model and so that makes it harder to identify associations. And so, part of it is just statistical considerations around power. And I think when race and ethnicity is not a main exposure of interest, lumping them up together is one of these things where you say, well, I can only have so much, you know I need a parsimonious model.”</p>
1.4 Lack of research funding	<p>“And unfortunately, it comes down to funding. So, the things that change the direction and research that we go is often the National Institute of Health, which is the biggest funder of research even for someone like me who works with the health care system, most of my money comes from the National Institute of Health. When the National Institute of Health puts out specific funding announcements that say this is an important area of research, and we're willing to set aside X amount of money so that you can do this properly, then the research occurs. Without the funding, because this type of research takes more resources - whether it's, you know, doing the in-person recruitment to make sure you get hard to reach populations, or whether it's getting big numbers or whatever it is. And so I'd say research funding.”</p>
1.5 Lack of awareness about importance/lack of researchers' interest	<p>“The other issue is getting people excited about this, because not everyone cares or understands the importance of disaggregating. Sometimes it's like, oh, well if you disagree with this, why don't you disaggregate like Latinos or black women? That's all very valid, but that doesn't make my research any less important. So, getting people to understand the importance of my research is the only goal that I'm trying to get to at this point, because I can't speak to other populations that I'm not as experienced in. So that's a huge struggle - getting people to care.”</p>
1.6 Lack of collection of granular information	<p>“These populations are not well-represented in these datasets, which is a problem from a scientific and business perspective. If you don't have data from certain populations, it can cost millions of dollars. Also, the US is continuously getting more diverse, and it will not get away from this, so we need to collect this data”</p>
1.7 Participant reticence to provide information	<p>“I think some of it touches upon the same reasons that I mentioned - people not wanting to share their perspective, for you know, for you know, very valid reasons, like concerns, and you know, having to build trust.”</p>
1.8 Lack of participant ability/desire to participate	<p>“It's either on the side of the researchers where they're not including Asian groups, or society, participants who didn't want to be in the research because they don't trust, or they don't have the time or energy or financial resources to be able to join research. So, there's many layers of complexity why this happens.”</p>
2. Risks of Data Disaggregation	

2.1 Risk of re-identification	“And then I guess the other thing is identifiability in your population. So when you start to disaggregate, and especially in an area like Colorado, where we don't have as many Asian members in our health system, If you start desegregating down to women, women of a certain age, and women who are Filipino. There might be such a small sample size that you could identify people in research reports, and we don't want that either breaking confidentiality concerns”
2.2 Targeting small populations	“That choice [of not disaggregating data] was made because there were some concerns of targeting smaller populations within the United States. “
2.3 Potential for discrimination/stigmatization	“Another is that there is the potential for stigmatizing research to be performed. However, we have policies against stigmatizing research.”
2.4 Misconstruing social phenomenon for biological (disproportionate focus on genetic ancestry of the groups → can miss social components of health)	“You have to really rule out the social and environmental determinants that are causing the problem. Because the question is, is it a biological component associated with the South Asian population? Do they just have, like a biological predisposition to an inability to clear plaque out of the brain? Then is it a situation in which you have a population where people don't go to get routine checkups at the same rate as other groups such that things Aren't discovered at the same rate. It's unclear right now. Granted you could subdivide and say, like there is a pocket of the population that seems to have disparity, right? And then, and that's fine. But then you have to start really digging apart and picking into that question, say so what is the potential factor? So that we can now potentially make a change like? Do we need a different type of drug or a different type of intervention? Or do we just need to have a different type of public policy with how we go about bringing health care to certain communities? Yeah. So, all relevant questions, all worthy of investigation. You can't begin to ask that question, though, until you actually know the sub-population that you're dealing with, and how they compare to other sub populations. So not disagreeing with you, just pointing this out”
2.5 Improper Inference	“The harms are the small sample size could lead to improper inference. And so, you know, disaggregating groups, and if it's a small sample size that it could really lead to improper inference.”
3. Harms of Data Aggregation	
3.1 Drawing Incorrect Conclusions	“The problem is that you get the answer wrong sometimes, right? So, you may make a broad statement about Asians that only applies to the majority Asian group. So, you're making a broad statement, saying you know health outcomes in Asians are just as strong or just as good as a non-Hispanic white population. But it's because you're really looking at a population that's predominantly Chinese American. And so, when we kind of aggregate, we get it wrong.”
3.2 Missing disparities	“But when you look at Asian groups, and There's either no Asians included, or they'll just have an Asian group that contains everyone. And so that's what I mean when we're talking about how Asians are just not included in these large data sets, and so we just don't know what's happening in their health outcomes.”
3.2a Disparities in access to services	“Well, I'm sure there's a lot of harm to it, too. But I think there's definitely a lot of benefits to disaggregating data, especially when you're thinking about developing interventions, right? A lot of the interventions that are sort of implemented you know, often based on sort of these really like monolithic groups, or they're often based on sort of white populations. And from what we know about thinking about the social factors that influence health, we're not really sort of taking account of

	the unique factors that might be influencing South Asian women, for example, that might sort of hinder the effectiveness of certain interventions. So, data disaggregation is really important when we're thinking about how to address the particular issues that are shaping these certain groups' health.”
4. Methodological Challenges of Data disaggregation	
4.1 Defining populations	“There are a lot of tricks to this problem though. I think you have to be really careful about how you define South Asian. There are many different regions within Asia, right? So, you can imagine the Asian Pacific Region. And so, you do have to be careful about where exactly are you drawing your boundary condition. More importantly, though I I think you run into this question of when you're saying South Asian, are you saying South Asian by biological definition. Or are you saying South Asian by self-assessment? I don't know which one you're talking about - Because you run into this exact same problem with every racial question.”
4.2 How to disaggregate	“Just ensuring that when you disaggregate, you do so in a way that is able to – What does it mean to disaggregate? For example, we always say what does it mean to be Indian right? What does it mean to be Pakistani? It is sometimes difficult to know how far to go with disaggregation. This is the ultimate dilemma for public Health Survey developers, like how much do you need to go and what way?”
4.3 Missing confounding variables (diet, geography, local environment)	“But, you know, it's sort of combined with understanding, why is it that they don't want to come in, you know, for screening. Is it cultural, are there other socio-cultural factors?”
4.4 Capturing culture (need for qualitative and other methodological approaches to understand culture and acculturation)	“Researchers could also benefit from not just relying on quantitative data. So, using qualitative information as well. I think that there's a power to like people's stories and narratives that some like you can't fully capture when you're just doing surveys. And I think especially if we're trying to understand why certain health disparities are happening in certain communities, I think there should also be sort of a mixed methods approach to research that could be really useful.”
5. Benefits of disaggregation	
5.1 Reduce health disparities	“We have seen that if you break it down into various sub populations, you can detect issues that you may not have if you didn't break it down.”
5.2 Understanding behavioral differences between populations	“We've seen it not only in that study, but in some of our research and in our local population screening for other things like colorectal cancer or anything like that certain Asian populations are more compliant, for whatever reason, probably culture, you know, some will just listen to what their doctor tells them. Others will be less or won't come in. Right. So, I think we've seen that in our quality initiatives that we've done locally, we do see differences in certain populations, as well as that lung cancer screening paper, we did see trends, because the numbers were small, we couldn't make it out if they were significant, but we did see trends.”
6. How to Facilitate Disaggregation	
6.1 Larger numbers	“You definitely want to have as many diverse researchers as possible. You also want as many of the diverse participants that you can get. The greater the number, the more disaggregation you can have”

<p>6.2 Community engagement (provide more data, participate in qualitative studies, understanding culture in study design)</p>	<p>“Well, another path that we've been trying to go down is, like you mentioned, is participant or community engagement. Because, you know, in the EMR, you can get information about the ethnicity, you probably can't get the culture as much, but then, if you can reach out directly to the community and get contacts that way, I think that's another avenue that could be helpful in getting data that we don't already have. But it also, I think, helps to make the community or the culture of an area more engaged with healthcare, and therefore more willing, maybe to one come in, you know, to be seen, because it's not like, they're going to like a foreign place, right? It's inviting because it makes them feel like the medical field is listening to them.”</p>
<p>6.3 Funding</p>	<p>“And unfortunately, it comes down to funding. So, the things that change the direction and research that we do is often the National Institute of health, which is the biggest funder of research even for someone like me who works with the Health care system, most of my money comes from the National Institute of Health. When the National Institute of Health puts out specific funding announcements that say this is an important area of research, and we're willing to set aside X amount of money so that you can do this properly, then the research occurs. Without the funding, because this type of research takes more resources - whether it's, you know, doing the in-person recruitment to make sure you get hard to reach populations, or whether it's getting big numbers or whatever it is. And so I'd say research funding.”</p>
<p>6.4 Education of researchers and increasing awareness</p>	<p>“I think the more people understand the importance of this and just more the education of researchers, of healthcare, or wherever, I think then, you know, the changes can be made in how data is gathered for one thing. I think that's a lot of the issue like with the areas where it's not as granular, right? People tried to make it simpler, which makes sense. On one hand, but on the other hand, when you're really doing research, and you really want to delve into the issues, you can't do that unless you have the data. So, it's about educating, and making people aware that it's important and really trying to influence either whether it's the data gathering applications or whatever it may be to be more granular, in that sense. Yeah, it really is the awareness. It's gonna make people motivated to be able to enable whatever it may be, whether it's the data programming, whether it's the applications.”</p>
<p>6.5 Partnering with community organizations (building trust, etc).</p>	<p>“I think a good way to do that would be to focus, for example, on like local organizations. Right? I think there should be more sort of collaboration between researchers, academic institutions, and sort of community organizations. So, you know, there's a lot of local social justice organizations that work specifically, for example, with South Asians or any other ethnic group. And I think, working together with those communities, some organizations and activists consider. I think it gives researchers an opportunity to sort of be welcomed into those communities and to gain the trust and community members.”</p>
<p>6.6 Language representation</p>	<p>“So South Asians, as you probably know, are a very diverse community that encompasses varieties of religious, cultural, and linguistic diversity, and I think that's the key barrier oftentimes. If one wants to work in older adult populations, particularly first-generation populations, you need to have the linguistic capacity and you need to have those linguistic resources”</p>
<p>6.7 Increase researcher diversity</p>	<p>“You definitely want to have as many diverse researchers as possible. You also want as many of the diverse participants that you can get. The greater the number, the more disaggregation you can have”</p>

7. How to Disaggregate	
7.1 Country of origin	“But I still think country of origin maybe is the most important of those things because you know that changes in health that occur and populations the longer they're in the United States, whether it's increasing in obesity, or whether it's access to screening services. If there is change, the longer people are in the United States. But there still is that link to the country of origin, where we see differences by different groups depending which country they came from.”
7.2 Immigrant status	“South Asians, for example, are like some of the fastest growing like undocumented populations in the Us. So, thinking about the health of South Asians as a whole. It's important to think about immigration status”
7.3 Number of years in the U.S.	“So how long have you been in the Us, immigration status, even citizenship - like are they legal? Which is obviously an ethical issue to collect. But understanding that part because it could really affect their ability to get Jobs, ability to access food or feel comfortable attending public places. So, a lot of all of that would be very important to collect.”
7.4 Racial and ethnic background	“I think, certainly getting the basic ethnic background in the data is important. Because you gotta start from that, right, that's the main thing”
7.5 Geographic location	“Geographic location is important, because again, it's about the culture, right? And even though you have the same populations in two different geographic locations, it's possible that their culture could be a little different.”
7.6 Self-identification	“I think asking people how they like to be identified. A lot of people, you know, have very strong feelings about what their ethnic identity is. And so, asking those questions open-ended is important too.”
7.7 Religion	“South Asian religions are similarly very, very gendered. The pillars of Hinduism are very gendered. So, asking those, particularly in women's health, religiosity, things like that. Not just religion, religiosity, right?”
7.8 Language	“A very innovative way that is currently starting to be done in EHR Data disaggregation is using a language as a means to potentially identify South Asians in the first place. Oftentimes, race or ethnicity may not explicitly be asked, and if it is you get the same issues of disaggregation. One way that's being done now, in trying to work with complex EHR data is through preferred language. Like it is the preferred language in the EHR Bangla, or Urdu.”
8. Impact of Research on South Asian Women	
8.1 Health differences between men and women (develop targeted interventions, identify targeted risk factors)	“I think depending on what the findings are, I think we've seen something definitely between male and female, women and men, we do see quite a lot of differences in different disease processes. So, I think it's even more important to separate, because then you can really identify an issue that you may not be able to if you lump everything together.”
8.2 Augment South Asian female participation (they have unique barriers to participation)	“The first are issues of ensuring there are the cultural considerations that are required. The cultural and sensitivity considerations required to engage South Asian women are often unique in our context, we must have, for many women, we must have a female data collector in order to survey. Sometimes the questions are very sensitive. They're about sexual health. They are about sexual violence. They

	<p>are about reproductive health or just sensitive issues lie in general, and oftentimes it's required for women that we need to have a female staff member to do that. I think this is not the same for men, and there are various reasons for that, but I think that is ultimately a reality that people don't realize in the same extent that is the case with white populations, or even populations in other Asian ethnic subgroups in East Asian or South East Asian, you don't get the same level of gender based unique barriers across gendered lines when it comes to comfortability, sensitivity, what it means to be in a safe space to do research. So there's that one level of really ensuring that the resources, the human capital, the social capital, are all there to be able to have women be more comfortable, to share their perspectives in an ingenuous manner.”</p>
8.2a Increased awareness about research participation needs	<p>“The second issue oftentimes, and this again goes more for I think older, particularly older first generation South Asian women, is dependency and gatekeeping. We find that a big issue, unfortunately, is the husband not allowing women or a wife to be able to participate in research. You know, and or being very dependent on either the husband, the son, the daughter, or other family members, to participate in that research. I cannot come to you because I don't have a car. I don't know how to drive. I, in some cases, don't know how to read or write. Some of our participants in our South Asian health research don't know how to read and write, so ensuring that you're able to have resources and considerations for that as well. So, the barriers that some women may face are unique to those of men, and I think it's then the requirement of the researcher to be able to work through those barriers because we need that perspective in order to get a full picture right of a particular health issue.”</p>
8.3 Understanding South Asian woman experience to develop interventions (dependency, family-based interventions)	<p>‘I see a lot of benefits to doing specific research with regards to the ethnicity of South Asian women is that we would be able to better understand the social, cultural and the geographical context of the diseases that they have. And so, instead of having a very one size fits all approach like giving the same treatment to white women and a South Asian women, for example, like I know that at least the women in my family, they do not want to complete the full course of medication they're like, “Oh, I'm feeling fine. Let's me just stop it.” But as compared to any other people from other regions, they might be incentivized to complete the full course of the medication, so understanding the cultural context of things would be very much important. And then some people, some of the cultures heavily, are very proactive in terms of seeking healthcare, but when it comes to the South Asian population they only want to go to the hospital when things are super bad. So, thinking in terms of that, what are the cultural and social context within which the health of South Asian women operates would be really helpful in biomedical research.’</p>
8.4 Address misconceptions	<p>“It's very much necessary that we start having targeted action about it, and also because we have our own social understanding of diseases, which tends to interfere with the scientific evidence that we have. So how to bring them forward about the misconceptions that they have and try to give them proper information on things rather than relying on the social knowledge that they get about diseases.”</p>
9. Challenges for Research on disaggregated Groups	
9.1 Challenges accessing disaggregated data	<p>“This is sort of related to data disaggregation. But even when I've used a lot of data from the United States census, for example, or like data from the CDC. Even when they disaggregate, it's still really just difficult to access them. You have to go through multiple levels of getting IRB approval. So, it makes it a lot harder for</p>

	researchers to just sort of access those things”
9.2 Lack of disaggregated longitudinal data collection	“You will see Asian Indian/Pakistani, for so many data sets. and then at some time, I think it's like in 2000, they'll start separating Asian Indians from Pakistani, which is good. But like now you're stuck with all this data from prior to that that's not been separated and can't make any generalizations about these ethnicities beyond that, which is very difficult to do.”
10. Challenges for studying South Asian Health	
10.1 lack of funding	“Then also funding because we are trying to collect data of more and more people. Usually, the people who collect data on field are not very highly trained, or they do not get enough compensation for the work that they do. They're not very incentivized to kind of ensure the quality and data collection, because most of the funding that goes in a project is usually on implementation and like distribution of kits or any resources available. So that's a big thing, there isn't enough funding to train people and to provide them with resources”
10.2 Community outreach	“Funding, community outreach and trust and you know, having people stay in the study long term, having people engage authentically in a study both from the participation side as well as from the research side and you know, really disseminating results to the community where it matters and building on results. So, there's challenges on a whole spectrum of research to implementation to community”
10.3 Lack of trust	“And I know there you know there's a long history, and there's a reason right where some people are sort of hesitant to sort of take part in research, right because of sort of the history of how. So, the ethics behind research that has been done in the United States for such a long time. But I think sort of gaining the trust of the populations, especially minority groups. It's really sort of difficult for researchers, but it's really important just to build those relationships.”
10.4 Longitudinal participation	“You really need temporality and temporality requires longitudinal assessments. That's where you have studies like Masala and Mesa. I think Mesa is actually launched as well. I'm not 100% sure, but things like that. There, there's other studies like N-HANES. The issue with those is that they are repeated, cross-sectional studies, which is a little different. It's not that those same people are watched or assessed over time. It's like you get a sample of different people in different year.”
10.5 Disseminating results impactfully	“And I think right now there's sort of a disconnect between what researchers do and sort of what lay-people, what people sort of in the community are experiencing. Researchers need to do a lot more work in, and actually sort of interacting with the community and making their work sort of more legible, more approachable for community members.”
10.6 Lacking female participants	“South Asian women are very difficult to engage in research studies. In our experience, it is easier to engage men to join the study.”
10.6a Mistrust	“So, we have to do a lot of purposeful, meaningful trust building and community outreach, using community organizations and champions and leaders who have a lot more trust in South Asian women.”

10.6b less perceived risk	“South Asian women self-perceive themselves as being of less risk than their male counterparts.”
10.6c Health Taboos Reducing Perceived Need for Research	“Secondly the women, because reproductive health and menstrual health is not very much talked about, so they are very hesitant to speak about it because it's considered a taboo to talk about all these things. They just considered that these are like godly phenomena that are just happening. So, they were very shy to communicate with.”
10.7 Lack of diverse South Asian participant representation (in different SES, different languages, etc)	“Yeah, representation, you know, across socioeconomic gradients, that's a really big challenge, especially in the United States, because of having pockets of more recent immigrants who maybe have lower SES and, you know, are busier, have less time and trust in research and researchers to participate. So, getting good universalization and representation for the newer immigrant communities is a real challenge, having a lot of, you know, structures where they feel supported.”
10.8 Lack of data disaggregation	“But often Asians just get lumped together as Asians. We also have trouble, at least within my system, of teasing out native Hawaiian Pacific Islander groups.”
10.8a Inability to trust disaggregated data	“Yeah, often they are just aggregated with all Asians so there isn't a South Asian category necessarily, or even more refined categories in terms of country of origin or an identity of origin. So, it's all Asian. Yes, no, checkbox. And that's completely ridiculous. And, you know, people also may not want to self-disclose, or people do not self-identify as Asian if there's an option. So, they often select “other” or “white.” And so, there's multiple layers of that data collection when it comes to electronic health records, and you can't trust most of them in terms of aggregated data.”

Challenges for Studying South Asian American Health

The participants performing research on South Asian American health were asked to describe the challenges they faced when conducting their research. Among the many challenges described (Codes 10.1-10.8a, Table 3), frequently mentioned included the lack of data disaggregation (code 10.8, Table 2), lack of diverse South Asian participant representation (in different SES, different languages, etc.) (code 10.7, Table 2), and lack of trust (code 10.3, Table 2). Looking more specifically at challenges related to studying South Asian Women's health, participants most raised the lack of female participation in studies pertaining to South Asian health. The lack of female participation was explained by three different codes - mistrust (10.6a),

less perceived risk (10.6b), and health taboos reducing perceived need for research (10.6c) (Table 2).

Benefits of Disaggregation/Harms of Aggregation

To better understand why lack of data disaggregation was a significant challenge for studying South Asian American health, participants were asked about the utility of data disaggregation. The two dominant reasons participants stated were that data disaggregation would be useful to reduce health disparities (code 5.1) and to understand behavioral differences between populations (code 5.2) (Table 2). Participants also explained that a lack of data disaggregation could be harmful primarily because health disparities would be missed (code 3.2) (Table 3). Underneath that category, participants stated that by not disaggregating data, South Asian Americans could experience disparities in access to services (code 3.2a) (Table 2).

Risks of Data Disaggregation

Although participants made a strong case for the utility of data disaggregation, they also mentioned a variety of risks. The dominant risks mentioned include the risk of re-identification (code 2.1), the potential for discrimination/stigmatization, and the potential for misconstruing social phenomena for biological phenomena (codes 2.4 and 2.5) (Table 2).

Barriers to Data Disaggregation

Taking a closer look at the lack of data disaggregation aspect of the challenge of studying South Asian Health, participants were asked about barriers preventing data disaggregation from occurring in current datasets. The barrier that seemed to resonate the most among participants was small sample sizes (code 1.1, Table 2). Participants also commonly mentioned lack of awareness about importance/lack of research interest (code 1.5) and loss of statistical power (code 1.3) as barriers to data disaggregation within the Asian category (Table 2).

In addition to the current barriers to disaggregation, participants highlighted a range of methodological challenges. The two seemingly most pertinent methodological challenges commented by participants were difficulties defining populations (code 4.1) as well as difficulties capturing culture, indicating the need for qualitative and other methodological approaches to understand culture and acculturation (code 4.4) (Table 2).

Furthermore, when using datasets where data is currently disaggregated, participants continued to face nuanced challenges. Specifically, participants experienced difficulties accessing disaggregated data, such as financial costs or hardships locating disaggregated datasets (Table 3). If the data in the datasets being used is disaggregated, participants mentioned challenges drawing meaningful conclusions because the data has not been disaggregated for enough time to provide longitudinal health information (Table 3).

How to Disaggregate

As described above, one of the methodological challenges of data disaggregation frequently described by participants is the challenge of defining populations. To better understand what granular information should be collected and what would be useful in disaggregating populations, participants were asked which types of data disaggregation would be most beneficial to study South Asian American health (Table 1). Participants offered seven different suggestions for disaggregation metrics (7 codes, Table 2), including immigrant status (code 7.2), number of years in the U.S (code 7.3), self-identification (code 7.6), religion (code 7.7), country of origin (code 7.1), racial and ethnic background (code 7.4), and geographic location (Code 7.5) (Table 3).

In addition to what information would be important to collect to separate out Asian populations, participants were also asked how to facilitate data disaggregation in the current

political and research environment of the United States. Participants offered 7 different suggestions (Codes 6.1-6.7, Table 3). Methodology often raised by participants included partnering with community organizations to engage in activities, such as building trust, to augment South Asian participation in research studies (Code 6.5) (Table 3), community engagement, specifically referring to the need to collect more data on South Asian populations, especially qualitative data to understand culture in study design (Code 6.2) (Table 2), and the education of researchers and need to increase awareness about the unique disparities experienced by South Asian Americans (Code 6.4) (Table 3). Funding, specifically the lack of funding for research on South Asian health in America, manifested as both a barrier to research on South Asian health (Code 10.1) and as a method of facilitating data disaggregation (via increased funding) (code 6.3) (Table 3).

Impact on South Asian American Women's Health

One of the goals of this project is to understand how data disaggregation might impact the health of South Asian women in America. Throughout the interview, participants were asked how their responses might apply to South Asian women (Table 1). Participants offered unique perspectives on the topic, as encompassed by codes 8.1-8.4 (Table 3). Frequently mentioned by participants was that data disaggregation would allow researchers to identify health differences between men and women in order to develop targeted interventions for both genders and identify risk factors unique to both groups (Code 8.1) (Table 3). Additionally, another perspective shared by participants was the utility of disaggregation in understanding the South Asian female experience, such as dependency and family situations, to develop interventions (Code 8.3) (Table 3).

Discussion

Despite the health disparities South Asians experience in the United States, there is a lack of research on this population to understand the causes of these disparities or how to develop targeted interventions. One contributing factor to the paucity of research is lack of data disaggregation of the Asian racial/ethnic category within U.S. research datasets. The reasons for the lack of disaggregation are nuanced. The purpose of this study is to begin to understand the perspectives of biomedical and precision medicine program leaders and South Asian health researchers on data disaggregation, specifically the utility of data disaggregation in mitigating South Asian American health disparities, the barriers to/risks of data disaggregation, and how to meaningfully disaggregate the Asian ethnic category. The main findings of this study include participant agreement that a lack of data disaggregation of datasets is a key challenge of studying South Asian health in America. Furthermore, primary benefits of disaggregation include the ability to reduce health disparities and understand population-specific health behaviors. Risks include potential re-identification of participants and stigmatization of individuals and communities. These risks are related to the barrier to disaggregation of small sample sizes, which results in a lack of statistical power. Participants find it challenging to define ethnic/racial categories when collecting granular information and offered numerous suggestions for data to collect in order to meaningfully disaggregate populations. Participants also highlighted community outreach and engagement as methods of facilitating disaggregation to increase sample sizes and capture qualitative experiences of South Asian Americans. Additionally, they noted that data disaggregation would help researchers better understand existing health differences between South Asian men and women as well as develop interventions that account for factors uniquely experienced by South Asian women, such as dependency or health-related

taboos. Essentially, disaggregation is necessary to adequately address the health of South Asian American populations. However, it is not possible without increased research participation of this population through trust building, community engagement, and qualitative studies. These practices require funding and researcher interest. These perspectives are important to inform future policy makers and thought leaders how to adequately disaggregate health data and the steps they can take to facilitate the study of South Asian health to identify and mitigate disparities that are obscured by data aggregation.

The current challenges for studying South Asian health in America identified through this project can be grouped into two categories - lack of data disaggregation and lack of research participation (due to a variety of factors, such as mistrust, lack of community engagement, etc.). When asked about challenges for studying South Asian American health, participant's frequent citation of lack of data disaggregation as a challenge indicates that greater disaggregation within the Asian ethnic/racial category in datasets is currently impeding research. For example, a participant stated, "*I think mainly a lot of times when we're doing these multi-site studies across the US, you would see a table and you see [Asians] lumped together... It's probably good to try to take them apart and disaggregate them because disease burden is different.*" This quote illustrates how aggregating all Asian subgroups poses challenges in understanding sub-group specific disease burden, which is essential for addressing health disparities. Participants also stated that a lack of diverse South Asian participant representation posed a challenge to research. As a participant explained, "*South Asians are a very diverse community that encompasses varieties of religious, cultural, and linguistic diversity.*" South Asians make up a diverse group in the United States, and the nuanced identities of these individuals need to be captured in order for researchers to accurately understand the health experiences of South Asians in America and

develop interventions. Agreeing with this sentiment, a participant stated, *“there's very limited data that really speaks to the entire diaspora of South Asian Americans.”* Finally, lack of trust was frequently cited by participants as a challenge to studying South Asian Health. As a participant explained, *“there's a reason why some people are sort of hesitant to take part in research, because the ethics behind how research has been done in the United States for such a long time.”* All three challenges are important and highlight how increasing research on South Asian participants is nuanced and has multiple barriers.

Looking closer at the focus of this study - lack of data disaggregation, participants concurred that disaggregation is imperative because it helps identify and reduce health disparities in the South Asian American population and ensure that the population has access to the health services they need. They also brought up the harms of aggregation, as disparities can be obscured, or incorrect statements can be made when data on Asian subgroups is aggregated. As a participant stated, *“to my theory we're not even there to the point that we know what risks are prominent among South Asians, because we just don't have enough data and people to study it.”* With the current lack of data disaggregation, it is challenging for researchers to even identify disparities. Continuing this line of thought, another participant explained how disaggregated data is essential to developing interventions. A participant said, *“I think there's definitely a lot of benefits to disaggregating data, especially when you're thinking about developing interventions. A lot of the interventions that are implemented are based on these really monolithic groups, or... white populations. And from what we know from thinking about the social factors that influence health, we're not really taking account of the unique factors that might be influencing South Asian women, for example, that might sort of hinder the effectiveness of certain interventions. So, data disaggregation is really important when we're thinking about how to address the particular*

issues that are shaping these certain groups' health.” They also highlighted that aggregating data can cause the harm of drawing incorrect conclusions. A participant said, *“and so when we kind of aggregate, we get it wrong, you know, in Asian populations that could benefit from services and underserved populations might not be getting the services they need then, or getting the focus that they need because we say, oh, well, Asian outcomes in health are actually quite good, so we really need to focus on black populations and Hispanic populations that have lower, which we do... But there are Asian subgroups that probably really need that focus that are getting overlooked.”* Data aggregation can obscure the health differences of Asian ethnic groups, perpetuating the false perception that this population does not experience negative health outcomes. This is harmful to Asian populations, preventing their health disparities from being identified and addressed.

Risks, Barriers, and Methodological Challenges

Although the benefits of data disaggregation are clear, there are risks, barriers, and methodological challenges to data disaggregation that explain why it’s not currently being practiced in all datasets. These challenges need to be considered when pursuing greater disaggregation within the Asian ethnic category. It is important to note that none of the risks participants described are particularly unique to South Asian populations. Instead, they apply to individuals performing any type of research using bio samples or sensitive information from immigrant or minority populations. The primary risk participants described was the risk of re-identification of participants, which persists when collecting and disseminating any type of health information. Participants explained that this risk can be mitigated by having larger sample sizes and being cautious about the information shared with researchers. They also noted the potential for discrimination/stigmatization of the population. As a participant stated, *“being an*

immigrant community, [South Asians] come with their own vulnerabilities. And so, if we are not conscious and are not using data properly, that could result in somehow the marginalization of these communities.” Participants went on to state their programs either have policies against stigmatizing research or that researchers should partner with “*community liaisons or people from the community who deeply care about this topic... and understand the history behind South Asia*” to mitigate this risk. The last risk frequently discussed by participants was that of misconstruing social phenomena for biological phenomena. As a participant explained, “*you have to really rule out the social and environmental determinants that are causing the problem. Because the question is, is it a biological component associated with the South Asian population? Do they just have, like a biological predisposition to an inability to clear plaque out of the brain?*” When performing research, it is important not to attribute a health outcome to the biology of a group if it is caused or influenced by the social and environmental conditions. This risk highlights the importance of collecting and sharing both genetic data and information about social determinants of health in data repositories to allow repository users to stratify analyses by genetic and/or social factors. While these risks are important, they are not unique to South Asian populations and apply when studying any minority population. They can be mitigated by the solutions proposed above.

In addition to the risks, there are nuanced barriers to data disaggregation, including small sample sizes, loss of statistical power, and lack of research interest. These barriers are interconnected, and addressing one works towards addressing the others. The most frequently mentioned challenge was small sample sizes, which exacerbates the risk of identification. The small sample sizes of South Asian research participants brings up another barrier of loss of statistical power. As a participant explained, “*the problem with that is the small numbers. You*

know, when you talk about population research in general, when to be able to get large enough numbers to make a significant finding becomes an issue. We recently published a paper looking at lung cancer screening. And we did break it down to the sub populations. And while we could see trends, we couldn't really make absolute significant statements, because the numbers were so small.” South Asian research participation in many research programs is too low to allow for disaggregation while maintaining the ability for researchers to draw significant conclusions. Increasing sample sizes to allow for disaggregation while maintaining statistical power would take targeted community engagement and meaningful trust building. This cannot be done without awareness about the health needs of South Asians as well as research interest - another barrier frequently mentioned by participants. As a participant stated, *“It’s just policies and interests, like do people care about it? How can we make people care about it? And what can we do to make people care about it so that we can get money to pay for the disaggregation of data? It’s the combination of it all. People don’t understand why – people think all Asians look alike, they just don’t care, if that makes sense. So, taking time to gather all of those pieces separately, and then making those decisions to be able to disaggregate would be really important.”*

Essentially, getting people to care about the issue and garnering researcher interest is one of the first steps to facilitating disaggregation. This is followed by funding to finance community outreach, which would ideally increase participant sample sizes in datasets, as participants explained.

Even if the barriers and risks to disaggregation are ameliorated, methodological challenges to disaggregation exist. Participants highlighted the challenge of defining populations as well as the need for qualitative data to understand culture to adequately capture the South Asian experience. For example, a participant stated, *“I think you have to be really careful about*

how you define South Asian. There are many different regions within Asia right? So, you can imagine the Asian Pacific Region. And so, you do have to be careful about where exactly you are drawing your boundary condition. More importantly, though I think you run into this question of, when you're saying South Asian, are you saying South Asian by biological definition, or are you saying South Asian by self-assessment? I don't know which one you're talking about - because you run into this exact same problem with every racial question." Defining what it means to be South Asian is challenging; however, this challenge is not unique to the South Asian ethnic group. It applies to all racial/ethnic classifications. Participants also mentioned the challenge of defining mixed race individuals, which will persist as a challenge for all racial/ethnic categories in datasets as the world becomes more interconnected and intermixed. Furthermore, qualitative data is needed to understand the health experiences of all populations and immigrant communities as they acclimate to America, not just South Asian Americans. As a participant stated, *"the culture of South Asian may have a big impact.... I think that's important, especially here in the way we see that a lot of people adapt to the culture. So, I think you know, that's where lifestyle factors become more of an issue. So really, when it comes to lifestyle factors, it's more about what culture do you practice? What culture do you live in?"* Defining any ethnic and racial categories is a murky practice. While definitions are important, in cases where no one definition is truly adequate, standardization across datasets is likely the most important practice. Once these steps are taken, researchers can focus on gathering more qualitative experiences to learn more about participants and further stratify analyses.

Facilitating Disaggregation

One of the key methodological challenges of data disaggregation is how to define populations. As discussed in the results section, participants offered seven unique suggestions for

the type of data disaggregation that would be most beneficial to study South Asian American health. The diversity of the methods suggested by participants indicates the complexity of defining South Asian American as an ethnic group and how to meaningfully allow for disaggregation. It also demonstrates how nuanced an individual's identity is and the multitude of factors that affect health. As a participant stated, *“you could start by asking about country of origin, parent country of origin, grandparent country of origin, a little bit more about immigration factors, acculturation is really important - how long they've lived here - language fluency, etc. Acculturation and immigration are important to understand differences by group. And ancestry is really important.”* The participant continued on to mention the importance of self-identification and geographic location. There are a large number of factors that affect the health of immigrant populations in the United States. Despite the vast number of responses, many agreed that ethnic background/country of origin are good starting points. As a participant stated, *“I think, certainly getting the basic ethnic background in the data is important. Because you gotta start from that, right, that's the main thing.”*

Interestingly, in terms of facilitating data disaggregation, the recommendations participants gave were directly linked to the barriers mentioned. Participants highlighted the importance of partnering with community organizations to build trust and augment awareness about the importance of the research both among South Asian American communities and for researchers. As a participant stated, *“a lot of research now is really showing right that there's a big need for data disaggregation. But I think, in order to get community buy-in, you really need to work with local organizations – building community members' trust and showing them that the value of this work isn't just for researchers. It's not just producing entrepreneurship, but it's meant to improve the lives of the community members.”* In addition to showing community

members the importance of the work, participants also frequently suggested educating researchers. As a participant mentioned, *“I think the more people understand the importance of this and just more education of researchers, of healthcare, or wherever, I think then, you know, the changes can be made in how data is gathered.”* Employing this tactic would address the barrier participants mentioned of lack of researcher interest/awareness about South Asian health disparities. Finally, participants emphasized the need for community engagement to collect qualitative data and understand culture. Collecting this data would help researchers understand biological as well as social phenomena and the methodological challenge of the need to understand culture. Furthermore, employing all these methods of targeted community engagement would also work towards increasing research participation and consequently sample sizes and statistical power. However, as participants also highlighted, it all ultimately comes down to funding as well - funding is necessary to perform community engagement.

Impact on South Asian American Women’s Health

Data disaggregation is beneficial to better the health of all members of South Asian American communities. Participants pointed out that it might be especially beneficial for South Asian American women because this group experiences unique, gendered health outcomes. South Asian women can have vastly different roles and freedoms than men and often are less likely to perceive themselves as at risk. As a participant explained, *“the cultural and sensitivity considerations required to engage South Asian women are often unique in our context.... The second issue oftentimes, particularly for older first generation South Asian women, is dependency and gatekeeping. We find that a big issue, unfortunately, is the husband not allowing women or a wife to be able to participate in research... so the barriers that some women may face are unique to those of men.”* A lack of data disaggregation would homogenize the

experiences of all Asian women into one, making it impossible to begin to understand the uniquely gendered experiences of South Asian women and develop targeted interventions for this group.

The participants interviewed for this study make a strong case for the importance of data disaggregation. Most of the barriers to disaggregation participants discussed are not uniquely experienced by South Asian Americans but rather are pertinent to research on all minority groups in the United States. This suggests that they should not serve as barriers hindering the disaggregation of the Asian ethnic category if disaggregated research on other minority groups continues to be performed despite the challenges mentioned. All in all, this study demonstrates how the lack of disaggregation is tied to many different factors, and the challenges of disaggregation are interconnected. The absence of disaggregation is related to a lack of funding and researcher interest/awareness in the topic. For example, if interest is augmented via a greater understanding of the tangible health disparities in the population, more researchers will advocate for funding and more funding will be provided. Greater funding will allow for more community engagement efforts, augmenting South Asian research participation in the United States, and foster building trust among these communities, resulting in greater sample sizes and adequate statistical power. In terms of defining populations, while this is a challenging matter, participants agreed that the most basic and necessary metric is country of origin. Standardization across all datasets is essential as well. Once basic data is captured and disaggregation is allowed, researchers can take next steps to further segment and understand South Asian American health by performing qualitative studies and capturing more nuanced data, such as language and years in the U.S.

In the future efforts must be taken to augment research funding and awareness about these issues allow for disaggregation and facilitate research on this community. Furthermore, researchers need to be cognizant of the data they collect and share on subgroups and how it might shape understanding of health risks. For example, by failing to collect information on years in the United States, religion, and/or geographic location, they might not be able to truly understand the lifestyle factors causing health disparities. They also need to be aware of how aggregation might unintentionally cause harm by obscuring disparities. Understanding these implications is important when performing research on any minority group in the United States.

Limitations

This study has some important limitations. One limitation is the small sample size. Only ten participants were interviewed for this study. To gain a better understanding of the barriers to and need for data disaggregation, a larger sample size would be beneficial to reach saturation of themes. Furthermore, there was a lack of diversity in the participants represented. Only two biomedical research programs were represented in the participant cohort and having greater representation of research initiatives would garner a more diverse and comprehensive set of perspectives on the topic. Furthermore, no experts on South Asian American women's health were interviewed in this study, as those with such expertise were non-responsive to interview requests. The perspectives of these researchers would be especially beneficial to understanding how data disaggregation might specifically affect the health of South Asian women. Finally, this study has a limited scope. It only looks at one aspect of the problem of the lack of research on South Asian American health through the lens of precision medicine/ biomedical researchers' perspectives. To truly augment research on this population to address health disparities, it is imperative to understand all aspects of the problem, not just a lack of data disaggregation.

Furthermore, it is important to hear the perspectives from a more diverse set of individuals, including South Asian American community partners, research participants, funders, and patient advocates.

Conclusion

South Asian Americans are the fastest growing ethnic group in the United States (Budiman and Ruiz, 2021). They experience health disparities compared to the general population of the United States, such as increased CVD and diabetes incidence as well as higher rates of PCOS and osteoporosis in women (Cigna, 2020). The reasons for these disparities are not well understood partially due to a lack of research on South Asian health. Many factors contribute to the lack of research, including lack of South Asian research participation, lack of funding, and mistrust. One of the key contributing factors is a lack of data disaggregation within the Asian category in U.S. research datasets to allow for the resolution of South Asian populations. Researchers and thought leaders highlighted how data disaggregation is imperative to understanding health disparities and to develop interventions to effectively better the health of this population. However, there is low awareness, lack of funding, and lack of large enough sample sizes to meaningfully disaggregate within all U.S. research datasets. Furthermore, there is not a standardized method of collecting granular information or disaggregating data, as there are numerous factors that affect the health of immigrant populations. Increased funding and awareness to augment community engagement are first steps to working towards data disaggregation, as these practices could foster increased numbers of South Asian participants. Additionally, standardization and disaggregation by country of origin are starting points to understand the unique health risks of this ethnic group.

Building on this study, it is important to understand perspectives of South Asian American community members and health advocates on inclusion/representation of South Asians and South Asian women in health research. Additional research is needed to understand their perspectives on how to build trust and augment research participation. The perspectives of policy makers, South Asian American women's health researchers, and research funders are also important to determine how to augment funding for research on this group and how to advance awareness about the health disparities of South Asian Americans.

There is a paucity of research on many minority groups in the United States. South Asians are one of these groups, and they should not be left out of the research advancements made possible by interventions created from and disparities identified using granular data in U.S. research datasets. All populations deserve the chance to improve their health and access the services they need to attain wellbeing. Data disaggregation within the Asian ethnic category is the first step to ensuring South Asian Americans have a chance for health equity in America and globally.

References

- Adia, A. C., Nazareno, J., Operario, D., & Ponce, N. A. (2020). Health conditions, outcomes, and service access among Filipino, Vietnamese, Chinese, Japanese, and Korean adults in California, 2011–2017. *American Journal of Public Health, 110*(4), 520–526.
<https://doi.org/10.2105/ajph.2019.305523>
- American Psychological Association. (n.d.). *Suicide among Asian Americans*. American Psychological Association. Retrieved April 19, 2023, from <https://www.apa.org/pi/oema/resources/ethnicity-health/asian-american/suicide>
- arZan. (2020, June 2). *Remembering Navroze Mody: A zarathushti victim of hate crime in the United States*. Parsi Khabar. Retrieved April 7, 2023, from <https://parsikhabar.net/news/remembering-navroze-mody-a-zarathushti-victim-of-hate-crime-in-the-united-states/23086/>
- Asian American Health Initiative. (2008). *Asian American Health Priorities - Strengths, Needs, and Opportunities for Action*.
- Bakhtiari, E. (2020). Health effects of Muslim racialization: Evidence from birth outcomes in California before and after September 11, 2001. *SSM - Population Health, 12*, 100703.
<https://doi.org/10.1016/j.ssmph.2020.100703>
- Blackburn, S.-S. (2019, March 21). *What is the model minority myth?* Learning for Justice. Retrieved April 7, 2023, from <https://www.learningforjustice.org/magazine/what-is-the-model-minority-myth>

Budiman, A., & Ruiz, N. G. (2021, May 16). *Asian Americans are the fastest-growing racial or ethnic group in the U.S.* Pew Research Center. Retrieved April 7, 2023, from

<https://www.pewresearch.org/fact-tank/2021/04/09/asian-americans-are-the-fastest-growing-racial-or-ethnic-group-in-the-u-s/>

C-reactive protein. Mount Sinai Health System. (n.d.). Retrieved April 7, 2023, from

<https://www.mountsinai.org/health-library/tests/c-reactive-protein>

Cigna. (2020, September). WHITE PAPER: SOUTH ASIAN HEALTH DISPARITIES.

Drysdale, H., Ranasinha, S., Kendall, A., Knight, M., & Wallace, E. (2012). Ethnicity and the risk of latepregnancy stillbirth. *The Medical Journal of Australia*, 197(5), 278-281.

Retrieved 8/1/2020 from <https://www.mja.com.au/journal/2012/197/5/ethnicity-and-risk-late-pregnancy-stillbirth>

Engidaw, M. T., Eyayu, T., & Tiruneh, T. (2022). The effect of maternal anaemia on low birth weight among newborns in northwest Ethiopia. *Scientific Reports*, 12(1).

<https://doi.org/10.1038/s41598-022-19726-z>

Factors for increased incidence of polycystic ovarian syndrome in South Asian women. Drexel

University College of Medicine. (n.d.). Retrieved April 7, 2023, from

<https://drexel.edu/medicine/academics/womens-health-and-leadership/womens-health-education-program/whep-scholars/imari-patel/#:~:text=Studies%20have%20demonstrated%20that%20South,can%20be%20misleading%20for%20clinicians.>

- Flowers, E., et al (2019). Body Composition and Diabetes Risk in South Asians: Findings From the MASALA and MESA Studies. *Diabetes Care*; 42(5):946–953.
- Forouhi, N. G., Sattar, N., & McKeigue, P. M. (2001). Relation of C-reactive protein to body fat distribution and features of the metabolic syndrome in Europeans and South Asians. *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity*, 25(9), 1327–1331.
<https://doi.org/10.1038/sj.ijo.0801723>
- Gordon, N. P., Lin, T. Y., Rau, J., & Lo, J. C. (2019). Aggregation of Asian-American subgroups masks meaningful differences in health and health risks among Asian ethnicities: An Electronic Health Record based Cohort Study. *BMC Public Health*, 19(1).
<https://doi.org/10.1186/s12889-019-7683-3>
- Hoeffel EM, Rastogi S, Kim MO, Shahd H. The Asian Population: 2010. 2010 Census Briefs. 2012;24.
- Holland, A. T., & Palaniappan, L. P. (2012). Problems with the collection and interpretation of Asian-American health data: omission, aggregation, and extrapolation. *Annals of epidemiology*, 22(6), 397–405. <https://doi.org/10.1016/j.annepidem.2012.04.001>
- Home | Dedoose. (n.d.). [Www.dedoose.com](https://www.dedoose.com/). <https://www.dedoose.com/>
- How to do thematic analysis — Delve. (2022, May 2). Delve.
<https://delvetool.com/blog/thematicanalysis>
- Hussain-Gambles, M. (2004). South Asian patients' views and experiences of clinical trial participation. *Family Practice*, 21(6), 636–642. <https://doi.org/10.1093/fampra/cmh611>

- Jamal, A., Torre, L., Soerjomataram, I., & Bray, F. (2016). *The Cancer Atlas* (3rd Ed.). American Cancer Society Inc.: Atlanta, Georgia. Retrieved 3.4.2020 from <https://canceratlas.cancer.org/>
- Jarvis, J. (2019, August 29). *Endowed professorship focuses on health disparities faced by South Asians*. Newsroom. Retrieved April 7, 2023, from <https://www.unthsc.edu/newsroom/story/endowed-professorship-will-focus-attention-on-health-disparities-faced-by-south-asians/>
- Lai, D. W., & Surood, S. (2013). Effect of service barriers on health status of aging South Asian immigrants in Calgary, Canada. *Health & social work*, 38(1), 41–50. <https://doi.org/10.1093/hsw/hls065>
- Lee, J. R., Maruthur, N. M., & Yeh, H. C. (2020). Nativity and prevalence of cardiometabolic diseases among U.S. Asian immigrants. *Journal of diabetes and its complications*, 34(12), 107679. <https://doi.org/10.1016/j.jdiacomp.2020.107679>
- Le, T., & Zeng, E. (2021, March 31). *Data disaggregation shows startling health disparities among Asian Americans: Commentary*. Baltimore Sun. Retrieved April 7, 2023, from <https://www.baltimoresun.com/opinion/op-ed/bs-ed-op-0331-asian-health-disparities-20210331-xciw35hutzg2pmbdduv73prdm4-story.html>
- Lubin, M., & Khandai, A.C. (2017) Prevalence and Determinants of Psychiatric Disorders Among South Asians in America. *The American Journal of Psychiatry Residents' Journal*, 1-9. DOI: 10.1016/j.ajp.2013.09.005
- Mayo Foundation for Medical Education and Research. (2021, August 21). *Osteoporosis*. Mayo Clinic. Retrieved April 7, 2023, from <https://www.mayoclinic.org/diseases-conditions/osteoporosis/symptoms-causes/syc-20351968>

- Maddodi, N., Jayanthi, A., & Setaluri, V. (2012). Shining light on skin pigmentation: the darker and the brighter side of effects of UV radiation. *Photochemistry and photobiology*, 88(5), 1075–1082. <https://doi.org/10.1111/j.1751-1097.2012.01138.x>
- Mongraw-Chaffin, M., et al (2018). Relation of Ectopic Fat with Atherosclerotic Cardiovascular Disease Risk Score in South Asians Living in the United States (from the Mediators of Atherosclerosis in South Asians Living in America [MASALA] Study). *The American journal of cardiology*, 121(3), 315–321.
- Mukherjea, A., Ivey, S. L., Shariff-Marco, S., Kapoor, N., & Allen, L. (2018). Overcoming Challenges in Recruitment of South Asians for Health Disparities Research in the USA. *Journal of racial and ethnic health disparities*, 5(1), 195–208. <https://doi.org/10.1007/s40615-017-0357-x>
- Nelson, R. (2016). Asian Americans Show Different Cancer Pattern Than Whites. *Medscape News* (Sept. 30, 2016). Retrieved from <https://www.medscape.com/viewarticle/869545>
- Nguyen, H. T., Zheng, A., Gugel, A., & Kistin, C. J. (2021). Asians and Asian Subgroups are Underrepresented in Medical Research Studies Published in High-Impact Generalist Journals. *Journal of immigrant and minority health*, 23(3), 646–649. <https://doi.org/10.1007/s10903-021-01142-6>
- Patel, S P., and A C. Gaw. "Suicide Among Immigrants From the Indian Subcontinent: a Review." *Psychiatric Services (Washington, D.C.)*, vol. 47, no. 5, 1996, pp. 517-21.
- Patel, T., Umeh, K., Poole, H., Vaja, I., & Newson, L. (2021). Cultural Identity Conflict Informs Engagement with Self-Management Behaviours for South Asian Patients Living with Type-2 Diabetes: A Critical Interpretative Synthesis of Qualitative Research Studies.

- International Journal of Environmental Research and Public Health*, 18(5), 2641.
<https://doi.org/10.3390/ijerph18052641>
- Quay, T. A., Frimer, L., Janssen, P. A., & Lamers, Y. (2017). Barriers and facilitators to recruitment of south asians to health research: A scoping review. *BMJ Open*, 7(5)
 doi:<http://dx.doi.org/10.1136/bmjopen-2016-014889>
- Rahman, M., Abe, S., Rhaman, S., Khanda, M., Narita, S., Bilano, V., Ona, E., Gilmour, S., and Shibuya, K. (2016). Maternal anemia and risk of adverse birth and health outcomes in low- and middle-income countries: systematic review and meta-analysis. *American Journal of Clinical Nursing*, 103(2), 495-504.
- Raj, A., & Silverman, J. G. (2002). Intimate partner violence against South Asian women in greater Boston. *Journal of the American Medical Women's Association* (1972), 57(2), 111–114.
- Rastogi, P., Khushalani, S., Dhawan, S., Goga, J., Hemanth, N., Razia K., Kosi, R. (2014). Understanding clinician perception of common presentations in South Asians seeking mental health treatment and determining barriers and facilitators to treatment, *Asian Journal of Psychiatry*, 7, 15-21. doi.org/10.1016/j.ajp.2013.09.005
- Religions of south asia*. Center for South Asia Outreach. (n.d.). Retrieved April 7, 2023, from <https://southasiaoutreach.wisc.edu/religions/>
- SAALT. (2009, September). Health Care Issues Affecting South Asians in the United States. *Southern Asia population (live)*. Worldometer. (n.d.). Retrieved April 7, 2023, from <https://www.worldometers.info/world-population/southern-asia-population/>

- Sanchalika, A., & Teresa, J. (2015). Risk of Gestational Diabetes Among South Asian Immigrants Living in New Jersey--a Retrospective Data Review. *Journal of racial and ethnic health disparities*, 2(4), 510–516. <https://doi.org/10.1007/s40615-015-0099-6>
- SELOD, S. (2018). *Forever Suspect: Racialized Surveillance of Muslim Americans in the War on Terror*. Rutgers University Press. <https://doi.org/10.2307/j.ctv2n7f77>
- South Asian Public Health Association. (2002). A Brown Paper: The Health of South Asians in the United States . South Asian Public Health Association.
- Tachibana, C., & Weiner, J. (2022, January 6). *Busting myths about the Asian American "model minority"*. Penn LDI. Retrieved April 7, 2023, from <https://ldi.upenn.edu/our-work/research-updates/busting-myths-about-the-asian-american-model-minority/>
- Tran, H., Udaltsova, N., Li, Y., & Klatsky, A. (2018). Low Cancer Risk of South Asians: A Brief Report. *The Permanente Journal*. 22(2018), 17-095. Retrieved 3.4.2020 from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5882184/#b6-17-095>
- U.S. Department of Health and Human Services. (n.d.). *Populations underrepresented in the extramural scientific workforce*. National Institutes of Health. Retrieved April 7, 2023, from <https://diversity.nih.gov/about-us/population-underrepresented>
- Uppal, N., & Gondi, S. (2021, February 22). *The paradox of South Asian disparities in medicine*. HMSR. Retrieved April 7, 2023, from <https://www.hmsreview.org/issue/4/south-asian-disparities>
- Volgman, A. S., Palaniappan, L. S., Aggarwal, N. T., Gupta, M., Khandelwal, A., Krishnan, A. V., Lichtman, J. H., Mehta, L. S., Patel, H. N., Shah, K. S., Shah, S. H., Watson, K. E., & American Heart Association Council on Epidemiology and Prevention; Cardiovascular

Disease and Stroke in Women and Special Populations Committee of the Council on Clinical Cardiology; Council on Cardiovascular and Stroke Nursing; Council on Quality of Care and Outcomes Research; and Stroke Council (2018). Atherosclerotic Cardiovascular Disease in South Asians in the United States: Epidemiology, Risk Factors, and Treatments: A Scientific Statement From the American Heart Association. *Circulation*, 138(1), e1–e34. <https://doi.org/10.1161/CIR.0000000000000580>

“When it comes to applying for NIH grants and scholarships, who qualifies as an underrepresented person?” | NIH: National Institute of Allergy and Infectious Diseases. (2021, January 21). [www.niaid.nih.gov](https://www.niaid.nih.gov/grants-contracts/underrepresented-person-definition). <https://www.niaid.nih.gov/grants-contracts/underrepresented-person-definition>

Yom, S., & Lor, M. (2022). Advancing Health Disparities Research: The Need to Include Asian American Subgroup Populations. *Journal of racial and ethnic health disparities*, 9(6), 2248–2282. <https://doi.org/10.1007/s40615-021-01164-8>

