Collection and Validation of Patient Self-Reported Race, Ethnicity, and Language (REL) Information In a Postpartum Setting

by

Shivani Mehta

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Honors Bachelor of Science Degree in Neuroscience with Distinction

Spring 2023

Collection and Validation of Patient Self-Reported

Race, Ethnicity, and Language (REL) Information

In a Postpartum Setting

by

Shivani Mehta

Approved: _____

Tania Roth, Ph.D.
Member in charge of thesis on behalf of the Advisory Committee

Approved: _____

Alex Waad, M.A.
Member in charge of thesis on behalf of the Advisory Committee

Approved: Tamara L. Nicol Medina _____

Tamara Medina, Ph.D.
Committee member from the Department of Psychological and Brain Sciences

Approved: Laura Eisenman _____

Laura Eisenman, Ph.D.
Committee member from the Board of Senior Thesis Readers

Approved: _____

Michael Chajes, Ph.D.
Dean, University Honors Program

ACKNOWLEDGMENTS

First and foremost, I would like to thank my entire committee of Alex Waad, Dr. Tania Roth, Dr. Tamara Medina, and Dr. Laura Eisenman for their immense patience with me during this time.

Alex, I would like to thank you for being a truly amazing mentor these past four years. You have taught me more than I ever imagined I would learn and were always there when I needed you. You not only taught me how to conduct research but taught me about public health and its immense importance in the field of medicine. Because of you, I will be a better physician.

Dr. Roth and Dr. Medina, this thesis would not have been possible without your help and willingness to be on my committee. Despite not knowing me in any capacity, you both went above and beyond in supporting this journey and being on my committee. Because of you both, I am able to achieve my aspirations of completing a senior thesis and presenting a project that is extremely heartfelt to me.

Dr. Eisenman, your patience was much appreciated. During a time where my project changed and I scrambled to put something new together, you were there and beyond patient with me. You worked with all my time conflicts and gave very insightful advice during our meetings. Your input will always be appreciated, and I am glad that you were selected to be on this team.

A great part of this project was also conducted through the University of Delaware Summer Scholars Program. I would like to thank the University for giving me the opportunity to pursue my passions. Additionally, the hospital at which this

research was conducted at welcomed me as if I was always a part of their family and readily allowed me to use their facilities. To the team of Health Equity and Cultural Competency at this institution, thank you for bearing with me these last four years and being some of the best and most supportive co-workers I have ever had.

Last but not least, this all would not have been possible without the never-ending support of my family and friends. Mom and dad, thank you for listening to me talk about my thesis even though you most likely had no idea what I was talking about, and still being a source of continuous support—you guys were my biggest cheerleaders. Suhani and Viraj, my amazing siblings and best friends, you guys were always there cheering me on and giving me a source of laughter and release when I needed it. To my friends, especially Daria Willis, words cannot explain how grateful I am to you. Even when my project changed and I was extremely overwhelmed, you stayed by my side and talked me through the stress. Sometimes, the best support are the people closest to you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

As healthcare organizations move toward accountable care agreements, and move away from fee-for-service, there is a greater need for healthcare organizations to have stronger data to support population-based interventions. Moreover, literature has highlighted the discrepancies in collecting identity-based information from patients, including information regarding patient race, ethnicity, and language (REL). REL data has multiple policy and clinical implications as it is utilized to not only determine the allocation of funds for programs but also is used to create evidence-based interventions to decrease health disparities. If the core of this data is incorrect, then resource allocation is futile. More importantly, there is a potential that the resources and interventions that are being created using this data are now not effectively reaching and impacting these communities. Prior research demonstrates that there are large disparities in women's health, especially by race.[1] Given what we know about the nature of flawed data, these disparities are potentially increased or inadequately captured by current interventions. In an effort to assess organizational capacity to collect REL data and identify where discrepancies in the documentation of REL data may occur, this quality and safety improvement project assesses the practices of collecting REL data from patients, as well as concurrence or discrepancy in how REL data is documented within the patient's chart and how they choose to self-identify. These concurrences and discrepancies were measured with a two-pronged approach where one prong involved patient survey of self-identified REL information and the second involved collection and validation from the Electronic Health Record (EHR).

Study results demonstrated an overall concordance between the two data corpuses; however, the discrepancies and variation in certain minority groups were noteworthy. Given the results, the main finding is the need for an EHR with broader fields and/or allowing patients to self-identify their demographic data to allow for the validation of patient identities, create accurate data corpuses, and improve patient health outcomes. Once modified, we expect researchers to have more accurate and credible data to identify health disparities from, driving the eventual closure of the inequities seen within multiple minority populations.

Chapter 1

INTRODUCTION

1.1 Minority Populations and their Vulnerability to Health Disparities

Recent years have brought into the limelight the disparities in healthcare quality given to racial, ethnic, and linguistic minority groups, with a focus given to their vulnerability to poorer health outcomes. These minority groups face multiple barriers to equitable healthcare such as limited social support, lower health literacy, discrimination, poor socioeconomic status, and for those who are Limited English Proficient, language barriers. Inequitable healthcare has further been proven to contribute to poor health outcomes for these patients as indicated by greater patient safety events, which are events that could have or did result in harm to the patient.[2] Given this information, it is clear that minority groups face disproportionate barriers to accessing proper healthcare, driving the widening gap in health disparities between minority and non-minority groups. These disparities are visible in every healthcare field, including maternal-fetal medicine. Maternal and infant mortality rates in America pose a stark contrast to those seen in other developed nations, with the U.S. having one of the highest maternal and infant mortality rates when compared to similarly developed countries. Additionally, the disparities in these rates are drastic when comparing those of racial, ethnic, and linguistic minority groups to the general

population. For example, non-Hispanic Black women have a maternal mortality rate 3.55 times higher than that for non-Hispanic White women, and American Indian/Alaskan Native women have a rate that is 2.3 times higher than that for non-Hispanic White women.[3] When analyzing the data for infant mortality rates, a similar trend is seen. In fact, the infant mortality rate for Black infants is almost double that of White infants. Additionally, other racial/ethnic minorities also have greater infant mortality rates when compared to White infants.[3] These data demonstrate drastic disparities in the field of maternal-fetal medicine due to race and/or ethnicity. The aforementioned risk of patient safety events for linguistic minorities may also contribute to the increased maternal and infant mortality rates observed in these populations. However, more research must be conducted on this topic.

Disparities in infant mortality rate are not only observed at a national level but can also be identified at a state level. As recently as 2020, data demonstrates that the state of Delaware has an infant mortality rate of 5.4 deaths per 1,000 births, which is equivalent to the national rate. However, since 2010, the state of Delaware has consistently maintained an infant mortality rate higher than the national average with the exception of the 2020 statistic where there was a 30% decrease (See **Appendix A**).[1] Given how alarming the data are for such a small state, the statistics are further broken down by county and race. As indicated by prior literature, there are widespread disparities among racial groups. Black and Hispanic infants consistently have greater infant mortality rates when compared to White, Non-Hispanic infants across all three counties. Within a 5-year period between 2016 and 2020, the average infant mortality

rate across all three counties was 11.6 deaths per 1,000 births for Black infants, 6.3 for Hispanic infants, and 3.8 for White, Non-Hispanic infants (See **Appendix A**).[1] This data is remarkable, indicating an infant mortality rate almost four times higher for Black infants compared to White, Non-Hispanic patients and almost two times higher for Hispanic patients. All counties saw similar trends with Black infants having higher rates of death compared to the other racial groups. Conversely, the city of Wilmington in New Castle County showed that Hispanic infants had a higher mortality rate with 18.8 deaths per 1,000 births; however, Black infants also saw a similar number with a rate of 17.2. Wilmington also saw the greatest disparity among racial groups, with White, Non-Hispanic infants having an infant mortality rate of 7.2 deaths per 1,000 births (See **Appendix A**).[1] These discrepancies are drastic considering the small size of the city. Given these data, it is indicated that there are wide health disparities within Women's and Children's Health in the state of Delaware.

1.2 Barriers to REL Data Collection and their Contributions to Widening Disparities

Despite the United States being considered a leader in modern medicine and international health disparity work, inequities continue to persist. The core of these issues relies upon an increased presence of data collection in hospitals and mandatory requirements with decreased standardization across institutions.[4,5] Additionally, much of the data collection is due to flawed practices that lead to incorrect data corpora.[6] Examples of these flawed practices include visually scrutinizing patients to determine race or ethnicity [4,7,8] as well as categorizing ethnicity based on the patient's last name.[4,7] Studies have determined that using "bad practices" such as the ones detailed

can lead to a patient misclassification rate of up to 50%.[4,7] Visually scrutinizing patients for demographic data is greatly unreliable. In fact, a study has found that visually, most individuals can correctly determine race between Black and White individuals, but often misidentify multiracial and Hispanic patients.[9] Thus, while 78% of American Hospitals collect REL demographics from patients, much of the data is incorrect or poor due to such "bad practices". In addition to poor data collection, hospital systems lack cohesive data collection practices and standards leading to fragmented and incomplete data both within and across healthcare organizations.[8]

Additional studies have been conducted in the United States and Canada to explore whether the barrier to data collection is due to patient reluctance to answering REL questions. Overall, these studies have found that the general public is overall very accepting of answering these questions and experience no discomfort in doing so.[10,11,12] These studies further suggest that the fragmented and incomplete data collection experienced by healthcare systems is due to institutional and systemic barriers rather than those imposed by patients.

There are additional social implications among race and ethnicity that impose further problems in data collection. Race is a social construct that purely exists in the Western world [13,14] and ethnicity contains multiple dimensions.[15] Unfortunately, the United States only offers broad racial categories without sufficient breakdown [8] and the only options for ethnicity data collection are "Hispanic/Latino" or "Non-Hispanic/Latino".

This insufficiency is demonstrated by looking at Cambodians and American-Born Chinese individuals, both of whom would identify as "Asian" and "Non-Hispanic/Latino" on healthcare documentation. Unfortunately, there is a difference of almost 50% in low-birth-weight rates and adequate prenatal care between these groups [16]; however, this difference is difficult to capture during data collection due to the simplified racial and ethnic categories. While we know much about the differences among broad racial categories like "Black" and "White", there is not much known about other racial minorities, and even less is known about disparities within sub-ethnicities within these groups.[17] Additionally, it has been found that collecting data with defined and specific choices not only has limited value for representing patients' identities but can also alienate those whose identities do not match the categories provided.[8] A prime example of this is patients who identify as Arabic when documenting race. Do they indicate that they are "Asian" or "White" when they truly are neither?

With multiple barriers to complete and accurate data collection, it can be summarized that the main concern affecting data quality is the lack of understanding of how to best collect REL information from patients.[8] When comparing patients' self-reported REL data to that in the hospital database, it was further indicated that there were inconsistencies between the two.[18] For instance, a study performed in 13 primary care clinics found that 3% and 6.6% of patients who respectively self-identified as Hispanic and African American, were not documented accurately in the EHR. Additionally,

20% of patients who were indicated as English-speaking in the EHR demonstrated their preferred language as Spanish when taking a survey.[19] Putting this data together, it may be beneficial to both patients and healthcare systems if an overall system of self-reported data with greater racial and ethnic categories was put in place to allow them to identify as they wish.

As discussed, REL data collection allows us to measure health disparities. However, the fine details between racial and ethnic groups are lost, and with incorrect or inconsistent databases, the magnitude of the disparities seen may be entirely different. Hospitals with institutional REL data utilize it to not only look at disparities in care, but also design targeted programs to improve the quality of care, measure the needs of the populations they serve, and provide patient-centered care. REL information additionally improves the quality of care given to patients by ensuring the adequacy of interpreter services, patient information materials, and cultural competency training for care providers.[8] The current data confirm that racial and ethnic minorities not only lack receiving routine medical procedures but also experience a lower quality of care. The quality of care only worsens when the patient speaks a language other than English due to additional communication barriers.[8] A study conducted in Canada exemplifies the notion of poor quality of care and health for patients speaking a language other than English. They found that a greater number of individuals who indicated their ability to speak and understand English as “not well” or “not at all” reported poorer perceived health compared to those who speak and understand English

"very well" or "well".[6] The lack of appropriate medical attention and care for minority populations, especially those who are Limited English Proficient, and their subsequent poorer health amasses to a widening of disparities within these groups. It is imperative to collect accurate language information from patients to accurately reflect the need for interpretation services and medical materials in their preferred languages.

1.3 Study Purpose

If the core of the data that is utilized to craft interventions and allocate resources to racial, ethnic, and lingual minorities is flawed, then there is a potential that these interventions and resources are not effectively impacting and reaching the communities they need to serve. This can potentially increase the pre-existing disparities in these populations. Thus, it is imperative that healthcare provision is culturally and linguistically concordant with patients, and this begins with adequately capturing the patient's language and cultural frame. In an effort to assess organizational capacity to collect REL data and identify where discrepancies in the documentation of REL data may occur, this quality and safety improvement project assesses the practices of collecting REL data from patients, as well as concurrence or discrepancy in how REL data is documented within the patient's chart and how they choose to self-identify. In exploring the collection and documentation of these data, we are able to improve the accuracy and application of patient REL data in assessing the needs of the communities we serve.

Chapter 2

MATERIALS AND METHODS

This study was conducted in the Postpartum Department of a local Mid-Atlantic Community Hospital utilizing a two-prong approach to data collection (Figure 1.).

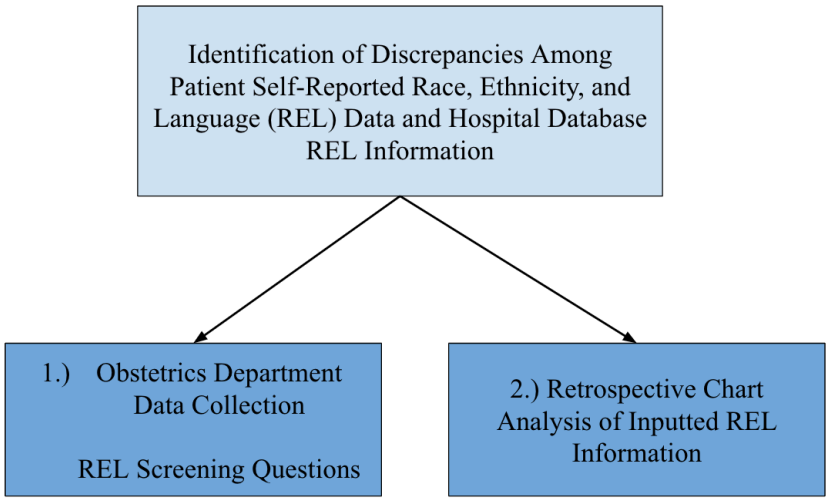

Figure 1. Two-Pronged Approach to Data Collection

During a data collection period of approximately one year between 2021 and 2022, a total of 724 (N=724) postpartum patients were surveyed for their self-reported REL information as a part of a larger study examining barriers and facilitators to language access in the department. All data has been de-identified as per HIPAA regulations.

Hello, my name is [READ NAME], if you have a few moments, I would like to complete the race, ethnicity and language screening that ChristianaCare uses to determine if you need or want language services.

Patient Medical Record Number	_____
In what language would you best understand your doctors or nurses?	◯ English ◯ Spanish ◯ Mandarin ◯ Arabic ◯ Hindi ◯ Hattian Creole ◯ ASL (American Sign Language) ◯ Another language ◯ Decline to answer
Another language, please specify -	_____
Do you identify as Hispanic/Latino or Non-Hispanic Latino?	◯ Non-Hispanic/Latino ◯ Hispanic-Latino ◯ Decline to answer
Which of the following would best describe your race?	◯ American Indian or Alaska Native ◯ Asian ◯ Black or African American ◯ Native Hawaiian or other Pacific Islander ◯ White ◯ More than one race ◯ Other race ◯ Decline to answer
Other Race	_____

LER Screener Results

SCREEN NEGATIVE - NOT ELIGIBLE
Thank you, your responses have been recorded and have a good day.

SCREEN POSITIVE - ELIGIBLE
Thank you, your responses have been recorded, and I will be back in a few minutes with a qualified medical interpreter to discuss your eligibility to participate in a research study at ChristianaCare.

When qualified medical interpreter arrives, they will enter the patient's room with the study team member who will say the following:

Hello again, and as a reminder, my name is [READ NAME]. Because your preferred language is [ler_language], I would like to introduce you to a study with the assistance of an interpreter that is being done at ChristianaCare. The reason we are doing this study is to identify the number of patients who speak a language other than English and Deaf patients within this setting. We hope to better understand your language needs and experience at Christiana Hospital. Are you interested in participating? If you need more information, please let me know.

Figure 2. Sample REL Survey Given to Patients

Patients were verbally surveyed using the questions and options in Figure 2 in order to obtain REL self-identification data. Survey responses demonstrate the categories normally provided to patients during hospital registration, but they were given the option to state "other" and further clarify their demographics. Upon completion of prong 1, prong 2 of the study was initiated utilizing patient medical record numbers (MRNs), which are a unique set of numbers that identify patients within the Electronic Health Record (EHR) System. REL information from the EHR was collected and validated for every patient surveyed. Both data sets were reconciled to obtain a visualization of the degree of variability among patient responses and EHR

information. For visualization, discrepancies were coded as either “0” or “1” where “0” indicated a discrepancy and “1” indicated that the data matched between the EHR and self-identification.

Chapter 3

RESULTS

3.1 Quantitative Data Analysis

When collecting REL information from the EHR, five patients were unable to be found due to incorrect typing of patient MRNs into the survey. Due to this error, the five patients were discarded from final study analysis (N=719). Additionally, patients are provided the autonomy to "decline" answering any demographic both during registration and data collection. Any patients who chose to decline in one or both scenarios were also removed from the sample population. This yielded a final study population of 705 (14 declines), 713 (6 declines), and 718 (1 decline) patients for the race, ethnicity, and language demographics, respectively.

Table 1. *Reconciliation of Self-Reported and EHR REL Demographics*

	Race Reconciliation	Ethnicity Reconciliation	Preferred Language Reconciliation
Demographics Match	601 Patients	701 Patients	697 Patients
Demographics Do Not Match	104 Patients	12 Patients	21 Patients
Total N	705 Patients	713 Patients	718 Patients

Overall, most of the demographic information for patients was in concordance between the two data sets (Table 1). Race had the greatest number of discrepancies

with approximately 17.3% of patients having a self-reported racial identification that did not match with the information in the EHR. However, both ethnicity and preferred language data greatly aligned among the corpora with both having a discrepancy rate of less than 5%. Visual analysis of each demographic; however, demonstrated wide variations in some of the fields provided to patients.

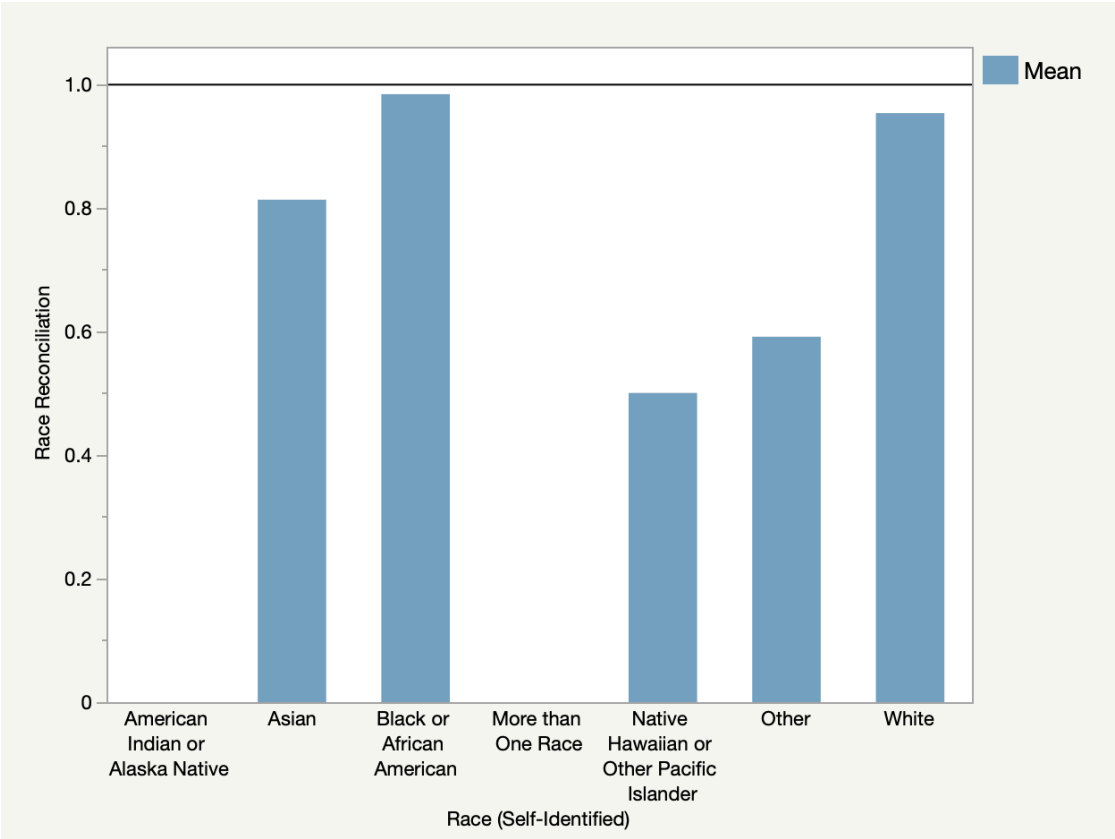

Figure 3. Visual Reconciliation of Race Data (N=705)

While most patients had concordance of race information between the data sets, as seen with means close to 1, indicating a greater percentage of matches, certain racial groups saw immense variation which demonstrates discrepancy. For instance, patients

identifying as Native Hawaiian/Pacific Islander saw great variation as indicated by having the lowest mean matches, further suggesting that many patients had data that did not match between the corpora. It is also noteworthy that most other racial groups had a greater number of matches than no matches as seen with greater mean match percentages. In accord with the literature, most patients were positively identified as Black/African American or White as observed by a relatively high mean match rate. Additionally, two racial groups, American Indian/Alaskan Native and More than One Race obtained no concordance among data sets.

Table 2. *Sample Sizes of Race Self-Identification Data (* = small sample size)*

Self-Identified Race	American Indian or Alaskan native	Asian	Black or African American	More than One Race	Native Hawaiian or Other Pacific Islander	Other	White
Sample Size (N)	4*	64	178	32	2*	89	338

As indicated by Table 2., certain racial groups had low sample sizes. These groups have lower generalizability as even one match or non-match can cause the mean to greatly deviate. Therefore, these groups' match percentages may not be reflective of the overall population.

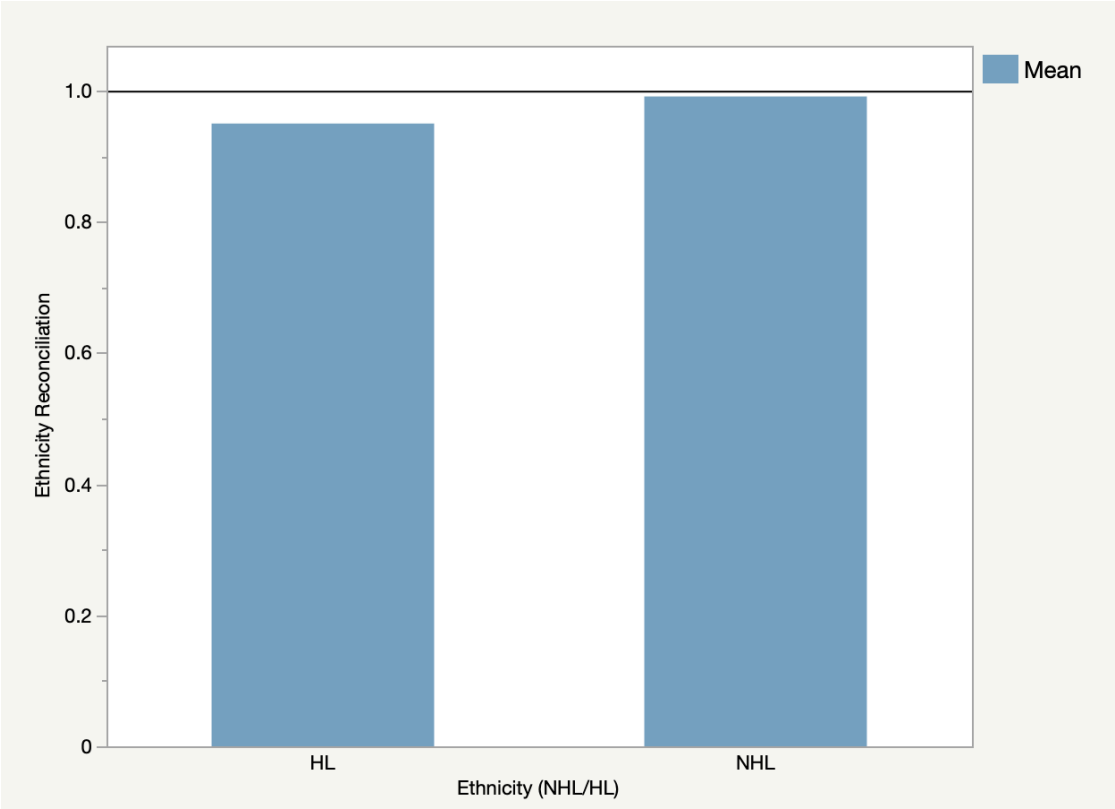

Figure 4. Visual Reconciliation of Ethnicity Data (N=713)

When observing the variation and accuracy of ethnicity data, it is observed that both Hispanic/Latino and non-Hispanic/Latino patients had high data concordance with minimal variation as observed by means close to 1. However, Hispanic/Latino patients on average had greater discrepancies than the latter group.

Table 3. *Sample Sizes of Ethnicity Self-Identification Data (* = small sample size)*

Self-Identified Ethnicity	Hispanic/Latino (HL)	Non-Hispanic/Latino (NHL)
Sample Size (N)	140	575

Table 3., indicates that both ethnicity demographics had relatively large sample sizes, suggesting that the mean match rates seen are generalizable to the overall population.

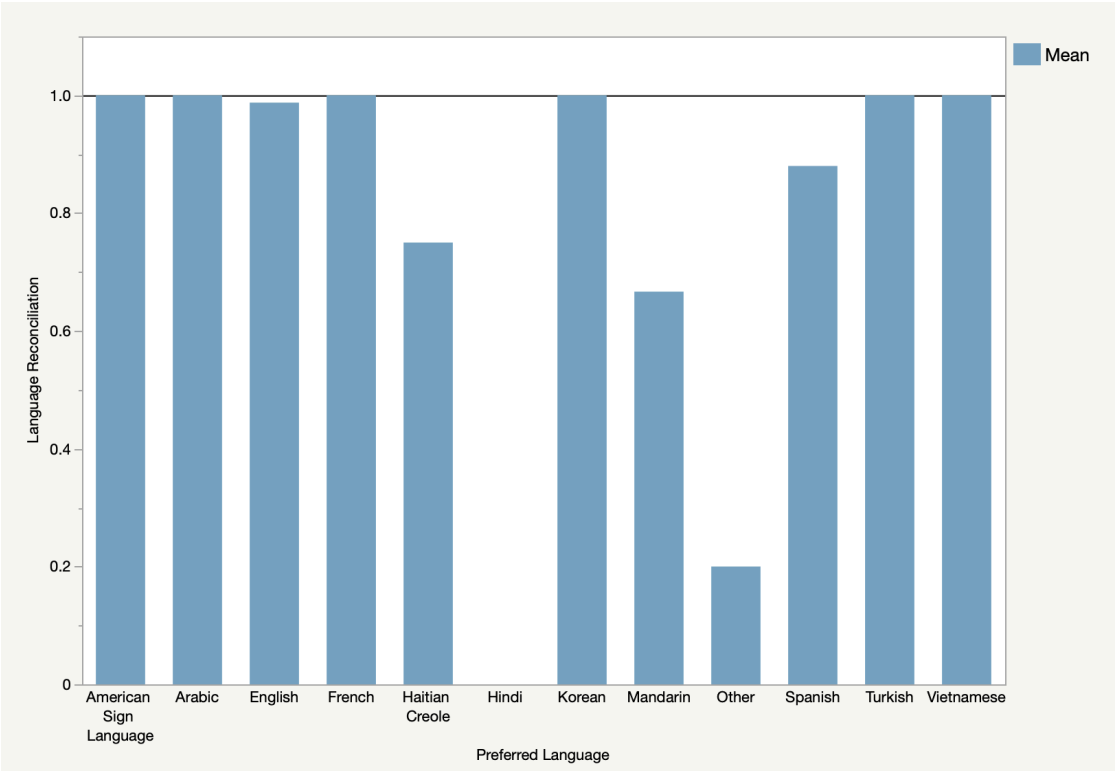

Figure 5. Visual Reconciliation of Preferred Language Data (N=718)

Preferred language data reconciliation demonstrated the greatest variation of the three demographic variables measured. Patients whose preferred languages were American Sign Language (ASL), Arabic, French, Korean, Turkish, and Vietnamese had 100% concordance between the data sets as demonstrated by a match rate of 1. English-speaking patients also had relatively high matches between the data sets as seen by a mean close to 1. Spanish-speaking patients, conversely, observed greater discrepancies as portrayed by a lower match rate which is indicative of greater variation. Patients who speak Mandarin or another language not provided as an option had the greatest discrepancies as demonstrated by the lowest match rates. This suggests that these two languages had a large population of patients whose data that did not match between the EHR and self-identification. It is also important to note the lack of any information for the Hindi language. The data set had one patient who self-identified as preferring Hindi as their language; however, their preference did not match with the EHR information yielding a reconciliation of 0.

Table 4. *Sample Sizes of Preferred Language Self-Identification Data (* = small sample size)*

	American Sign Language (ASL)	Arabic	English	French	Haitian Creole	Hindi	Korean	Mandarin	Other	Spanish	Turkish	Vietnamese
Sample Size (N)	2*	7*	642	1*	4*	1*	1*	3*	5*	50	1*	1*

As indicated by Table 4., several of the preferred languages that patients self-identified with had minimal sample sizes. As aforementioned, these smaller sample sizes have limited generalizability with even the most minute variations having the potential to skew the mean match rate. Therefore, some of the means observed may not be truly indicative of the general population.

3.2 Qualitative Data Analysis

A qualitative analysis of the compiled information from both corpuses provided further insight of the data visualizations performed.

When examining the options for sources of discrepancies between race information, it became clear that the EHR does not support an option for patients to identify as "more than one race". This lack of inclusivity forced many patients to choose one of their multiple races as one "main one" that they chose to identify with. Interestingly, many multi-racial patients whose races included Black/African American chose to identify with that race as opposed to their other identities. Further analysis of patient preferences revealed that Hispanic/Latino patients and Middle Eastern patients provided varying responses for their racial identification, which aligns with prior literature. Middle Eastern patients often wavered between the identification of "White" or "Asian" in the EHR. Conversely, during REL data surveys, these patients often opted to state "Other" and specify their racial identity as "Arabic" or "Middle Eastern". Similarly, in the EHR, many Hispanic/Latino patients often identified as "White"; however, when given the ability to self-identify, they would often state "Other" and further clarify by stating identities such as "Hispanic", "Mexican", "Latina", etc. These identifications under the "Other" category further demonstrate a gap in patient knowledge about the differences between the terms "race" and "ethnicity" as many patients who indicated "Other" further stated their racial identity as an ethnicity. For example, many Hispanic/Latino patients clarified their race as "Latina", which is considered an ethnicity.

Ethnicity qualitative analysis yielded similar insights as Figure 4. and failed to highlight additional information that could describe any sources of discrepancies.

Additionally, discrepancies between patient ethnicity information between the two data sets did not display any identifiable trend nor pattern.

Evaluating patient preferred language data sets also yielded greater insight into sources of discrepancy, as well as reasoning for trends seen with matched data. As visually portrayed, patients whose preferred languages were American Sign Language (ASL), Arabic, French, Korean, Turkish, and Vietnamese had complete matches among data sets (Figure 5). However, analysis of the patient population size revealed that the number of patients that prefer these languages is very small with less than ten patients per language. Given this information, it is likely that the high level of concordance seen with these languages is due to sample size and if the study included more patients from these linguistic identities, discrepancies would likely be observed. The evaluation also revealed that a large majority of the patient population preferred their medical information be provided in English or Spanish, as expected by the overall demographic makeup of the United States. When exploring sources of discrepancies for trends, it was found that a few patients whose EHR data marked "French" as their preferred language otherwise indicated "Haitian Creole" upon self-identification. Discrepancies were also noted between Asian languages such as Hindi, Telugu, Bengali, etc. where patients self-identified as preferring either of those languages, but the EHR indicated "English" instead. Furthermore, it was discovered that the EHR did not support many Asian languages, besides those clearly described in Figure 5. Spanish-speaking patients also faced discrepancies, with English often being

listed as their preferred language in the EHR; however, the data could not identify specific sources for the inconsistencies. Similar to the Asian languages, it was further found that the EHR also cannot support Spanish sub-dialects and groups all patients speaking any dialect of Spanish under the category of “Spanish, Castilian”, which is only one sub-dialect of the language.

Chapter 4

CONCLUSIONS

4.1 Discussion

Despite the discovery of discrepancies between the hospital EHR system and patient self-identified REL information, overall, the data corpuses had similar concordance (Table 1). Further breakdown of the numeric data via qualitative and visual quantitative analyses; however, identified sources of discrepancies and drastic variations among corpora which, unless corrected, have the capacity to misrepresent health disparities and subsequent interventions.

While race information had the greatest number of discrepancies (Table 1), qualitative analysis revealed the source as the lack of options for patients to identify as "More than One Race" in the EHR. Given this finding, it can be stated that race information is greatly concordant. However, the lack of broad options for racial identification is a noteworthy finding. Patients who identify as multiracial were not uncommon and comprised a sizable part of the study population (Table 2). Additionally, patients of Middle Eastern or Latin American/Spanish origin faced the same problem of reduced and confusing identification options. These findings validate literature stating that hospital racial identification options are not broad and restrict patient autonomy in

choosing how to identify. Racial reconciliation findings are also in accord with literature stating that due to “bad practices”, hospitals have better information and proper identification of patients as Black or White (Figure 3) when compared to patients of other races. While this discovery does not ascertain the use of “bad practices” within the hospital, they further validate the notion that visual observation can be deceiving. Final analysis of race data allows us to say with relatively high certainty that there is a validation of race information between the two data corpora; however, this validation comes with the conclusion of a system lacking in diversity, which must be addressed.

Ethnicity information also seems to be highly concordant between patients who identified as Hispanic/Latino and patients who identified as non-Hispanic/Latino (Figure 4). While any discrepancies were not further delineated by qualitative analysis of the corpora, the restricted nature of the ethnicity options is imperative to discuss. The options of NHL and HL can be considered as simple. One either is of a Hispanic/Latino origin or not. However, as suggested by background research, these simple options fail to capture the wide diversity of patient ethnicities. Within non-Hispanic/Latino ethnicities and even Hispanic/Latino ethnicities, there are multitudes of subgroups that fail to be captured by the simplistic nature of our demographic collection. Thus, the results for ethnicity may be concordant, but this concordance is deceiving given the condensed nature of ethnicity options.

The variation observed between preferred language data corpora (Figure 5) is of utmost importance. Given the volume of patients who qualify for placement in the "Other" language category in the EHR and the wide variation in their mean match rate between corpora, it is important to list what languages belong in those categories. These patients who speak "other" languages clearly lack representation of their linguistic identity in the EHR, leading to lack of resources such as interpretation services and paperwork in that language. In alignment with prior studies, this lack of language access in healthcare only widens disparities within minority populations. As aforementioned, the EHR imposes too many restrictions upon the possible demographic identities that patients can choose from, leading many to fall into some other category or get misclassified. The qualitative analysis also highlighted a trend about sub-dialects and similar languages often getting conflated. For instance, many patients who self-identified their preferred language as "Haitian Creole" often had "French" marked within the EHR. Upon further research, it was found that both languages are similar and could be considered sub-dialects of each other; however, similar to the Spanish sub-dialects, there are marked differences between the two. This conflation has occurred with two sets of languages, indicating some pattern or trend. While for the Spanish language this conflation can be explained by the lack of EHR options for any sub-dialects, the origin of this conflation between French and Haitian Creole is unknown. Further analysis must be conducted to ascertain the relationship between those languages and how they are misidentified within the EHR despite having two distinct options. Another interesting finding that many of our Spanish-

speaking patients faced, as well as a few others of other languages, was English being listed as their preferred language in the EHR despite verbally stating otherwise during the survey. While the data could not suggest a clear reason for the discrepancies, a plausible explanation may be the differences between linguistic ability in a medical setting and a casual setting. Another "bad practice" that is common in healthcare settings is the assumption of proficient medical English abilities from patients who may be proficient in casual conversation. These assumptions mark patients' preferred languages as English in the EHR and may hinder the use of interpreters. However, there is a notable striking difference between medical English and conversational English. While many patients may be proficient or even fluent in conversational English, their proficiency in medical terminology in English may be less than adequate, thus requiring interpreter use. Language, arguably, is one of the most important demographics relevant to patients during healthcare visits and the ensuring of good health outcomes. Without proper documentation of patient preferred language, the one that they would be most comfortable receiving medical information in, patients may face poor health outcomes and the consequences of those almost immediately.

Overall, the study completed its main objective of collecting and validating patient REL information between two distinct sets of data. Additionally, the combination of a qualitative analysis with a quantitative one allowed for the delineation of certain trends and sources of discrepancy. Nevertheless, data trends with certain groups were

not further defined by this study and it is imperative to take a closer look at those trends in future studies. The current data supports the notion of wide variations in patient self-identification aligning with EHR demographics and it is indisputable that certain racial, ethnic, and linguistic groups face greater discrepancies than others. Given these results, it is evident that these groups require further examination to delineate specific trends or causes.

4.2 Limitations

This study poses no large identifiable limitations. One notable disadvantage from our methods is the use of only one postpartum population. The hospital studied has multiple branches and postpartum departments. The introduction of the extra branches would have amplified the study population and may have allowed for the identification of multiple and more concrete trends. In addition, there was an approximate six-month gap in data collection in which the main person who surveyed the patients changed. This gap and change impaired the consistency of data collection during the study. While we foresee no great impacts from this discrepancy, it was mentionable. The small sample sizes for certain races (Table 2) and preferred languages (Table 4) further limits the overall generalizability of this study. A broader study that captures more patients of these racial and linguistic identities is required in order to ascertain trends.

4.3 Clinical Implications

The overall results and discussion of this study can be summarized to one crucial point: the need for broad patient identification fields and/or the allowance of self-identification for demographic information. The current, restricted options do not allow for the validation of patient identities and can lead to a sense of alienation, which may contribute to poorer health outcomes and increased patient safety incidents. The crucial demographic for safety incidents, as aforementioned, is language. When the preferred language of a patient and the one in the EHR, which care providers utilize for reference, are not in accord, then patients may not completely understand the medical information and care received, leading to poor outcomes. A prime example of this is an incident that occurred in the Spanish language. A patient in the emergency room was provided a prescription to take medication once (one time) a day. Unknown to the provider, however, the patient preferred their medical information in Spanish and in their language, "once" correlates to the number 11. Unfortunately, the patient was later readmitted due to an overdose on that medication having taken 11 pills. Situations such as the one detailed above are completely avoidable if patient demographics match their true self-identifications, which begs the necessity for an improved EHR or better method of demographic data collection.

Patient safety incidents do not only occur in the emergency department, but in all fields of medicine. It is known that there are wide and unwavering disparities in maternal fetal medicine in the United States and it seems that current interventions are

not functioning properly. A potential source of error could be the lack of accurate demographic data, such as REL information, in hospital EHRs. Given the faulty data corpus, researchers may struggle with identifying health disparities, and once identified, the group(s) the disparities are suggested in could be incorrect or not adequately capture their magnitude. Millions of resources, both financial and personnel, are contributed to minimize and close health disparities. However, it seems that in maternal fetal medicine, the efforts have been relatively futile. After careful examination of REL information from postpartum patients and their self-identifications relative to the EHR, it can be stated with great confidence that there are discrepancies that can have drastic implications for the allocation of resources towards health disparities. In the relatively small sample size seen in this study, a minimum 3% of each demographic collected had patients whose data was not in concordance with the EHR (Table 1). When expanded to a large population, 3% can be a significant population. Thus, it is imperative that hospital systems have EHRs that are concordant with patient self-identifications. With this correction, researchers in the future will be able to easily delineate health disparities by group and create meaningful interventions that actually target the correct populations.

In addition to improvement of the EHR and the methodology for REL data collection from patients, there are educational initiatives that must be taken to improve demographic data corpuses. The first intervention is the education of patients and staff members of the differences between race and ethnicity. As observed with the self-

identification of patients' race under the "Other" category, multiple patients incorrectly stated an ethnicity such as "Hispanic" or "Latino" when stating their racial identity. While there is systemic error in the categorization of race and ethnicity in this country, many people do not understand the nuances between the two. Race often simply refers to the physical attributes of a person whether it is skin color (ex. White) or other, sometimes biological, attributes (ex. Asian). Ethnicity, on the other hand, further breaks race down and refers to broad cultural expressions (ex. Hispanic) or places of origin (ex. Latino). While the race categorizations are understandably insufficient to adequately capture patients of Hispanic/Latino origin or Middle Eastern origin, that is a systemic error that must be corrected. From the perspective of the patient and providers, however, education of these two concepts can aid patients in making more consistent and accurate selections for their race/ethnicity. Another area of required education is for hospital staff and patients with respect to "bad practices" and the differences between medical and casual linguistic ability. In the best interest of the patient and their health outcomes as well as keeping concordant REL data in the EHR, it is imperative that both patients and providers understand what "bad practices" for data collection consist of in order to best avoid them. Additionally, education about linguistic ability is imperative so that both patients and providers understand when an interpreter may be necessary during a hospital visit.

The nuances of this study confirm the suggestions from prior literature suggesting discrepancies in the demographic data stored in the EHR as compared to patient self-

identification. Given these findings and the patterns and trends delineated from them, these problems can be addressed to improve the EHR system and collect meaningful data that efficiently and accurately identifies health disparities. This data will be valuable in the future as they will promote the creation of evidence-based interventions to tackle them. Fortunately, a prior study assessing hospital registrars', who are the staff that record patient demographic information upon intake, attitudes to changes in the EHR at the same institution holds promising results. In a study conducted by Mehta et al.[19] to assess hospital registrars' perceptions regarding the addition of gender identity information to the EHR, results indicated that many registrars were welcome to these changes. These findings are promising regarding the cumulative changes required in the EHR in order to create affirming care spaces for patients. Given these outcomes, we hope that appropriate steps are taken forward in the inequities faced by minority groups as a whole within healthcare and within maternal fetal medicine.

REFERENCES

1. Paul, DA. DHMIC Data Update. PowerPoint Presented at: DHMIC Meeting; September 14, 2022; Wilmington, DE
2. Chauhan A, Walton M, Manias E, et al. The safety of health care for ethnic minority patients: a systematic review. *Int J Equity Health*. 2020;19(1):118. Published 2020 Jul 8. doi:10.1186/s12939-020-01223-2
3. Dagher RK, Linares DE. A Critical Review on the Complex Interplay between Social Determinants of Health and Maternal and Infant Mortality. *Children (Basel)*. 2022;9(3):394. Published 2022 Mar 10. doi:10.3390/children9030394
4. Escarce JJ, Carreón R, Veselovskiy G, Lawson EH. Collection of race and ethnicity data by health plans has grown substantially, but opportunities remain to expand efforts. *Health Aff (Millwood)*. 2011;30(10):1984-1991. doi:10.1377/hlthaff.2010.1117
5. Institute of Medicine. (2009). *Race, ethnicity, and language data: standardization for health care quality improvement*. Washington, DC: National Academies Press.
6. *We Ask Because We Care. The Tri-Hospital + TPH Health Equity Data Collection Research Project Report*. Toronto Health Equity; 2013. Accessed April 8, 2023. http://torontohealthequity.ca/wp-content/uploads/2017/05/We-Ask-Because-We-Care-Summary-Report.pdf
7. Andrews RM. Race and ethnicity reporting in statewide hospital data: progress and future challenges in a key resource for local and state monitoring of health

disparities. *J Public Health Manag Pract*. 2011;17(2):167-173. doi:10.1097/PHH.0b013e3181f5426c

8. Hasnain-Wynia R, Baker DW. Obtaining data on patient race, ethnicity, and primary language in health care organizations: current challenges and proposed solutions. *Health Serv Res*. 2006;41(4 Pt 1):1501-1518. doi:10.1111/j.1475-6773.2006.00552.x
9. Smith T. *Measuring Race by Observation and Self-Identification*.; 1997. Accessed April 8, 2023. https://gss.norc.org/Documents/reports/methodological-reports/MR089.pdf
10. Baker DW, Cameron KA, Feinglass J, et al. Patients' attitudes toward health care providers collecting information about their race and ethnicity. *Journal of General Internal Medicine*. 2005;20(10):895-900. doi:https://doi.org/10.1111/j.1525-1497.2005.0195.x
11. Baker DW, Hasnain-Wynia R, Kandula NR, Thompson JA, Brown ER. Attitudes toward health care providers, collecting information about patients' race, ethnicity, and language. *Med Care*. 2007;45(11):1034-1042. doi:10.1097/MLR.0b013e318127148f
12. Varcoe C, Browne AJ, Wong S, Smye VL. Harms and benefits: collecting ethnicity data in a clinical context. *Soc Sci Med*. 2009;68(9):1659-1666. doi:10.1016/j.socscimed.2009.02.034
13. Krieger N. Does racism harm health? Did child abuse exist before 1962? On explicit questions, critical science, and current controversies: an ecosocial perspective. *Am J Public Health*. 2003;93(2):194-199. doi:10.2105/ajph.93.2.194

14. Williams DR. Miles to go before we sleep: racial inequities in health. *J Health Soc Behav*. 2012;53(3):279-295. doi:10.1177/0022146512455804
15. Statistics Canada. (2003). *Ethnic diversity survey: a portrait of a multicultural society*. Ottawa: Statistics Canada.
16. National Research Council (US) Panel on DHHS Collection of Race and Ethnicity Data, Melnick D, Perrin E, eds. *Improving Racial and Ethnic Data on Health: Report of a Workshop*. Washington (DC): National Academies Press (US); 2003.
17. Agency for Healthcare Quality and Research. 2003. *The National Healthcare Disparities Report.* Washington, DC: AHRQ
18. Moscou S, Anderson MR, Kaplan JB, Valencia L. Validity of racial/ethnic classifications in medical records data: an exploratory study. *Am J Public Health*. 2003;93(7):1084-1086. doi:10.2105/ajph.93.7.1084
    Klinger EV, Carlini SV, Gonzalez I, et al. Accuracy of race, ethnicity, and language preference in an electronic health record. *J Gen Intern Med*. 2015;30(6):719-723. doi:10.1007/s11606-014-3102-8
19. Mehta S, Waad A, Brooks M, Siegel SD. "I Think There's Only Two Fields for That": Hospital Registrar Attitudes and Practices for Collecting Patient Gender Identity Data. *Transgender Health*. Published online April 22, 2022. doi:https://doi.org/10.1089/trgh.2021.0083

Appendix A

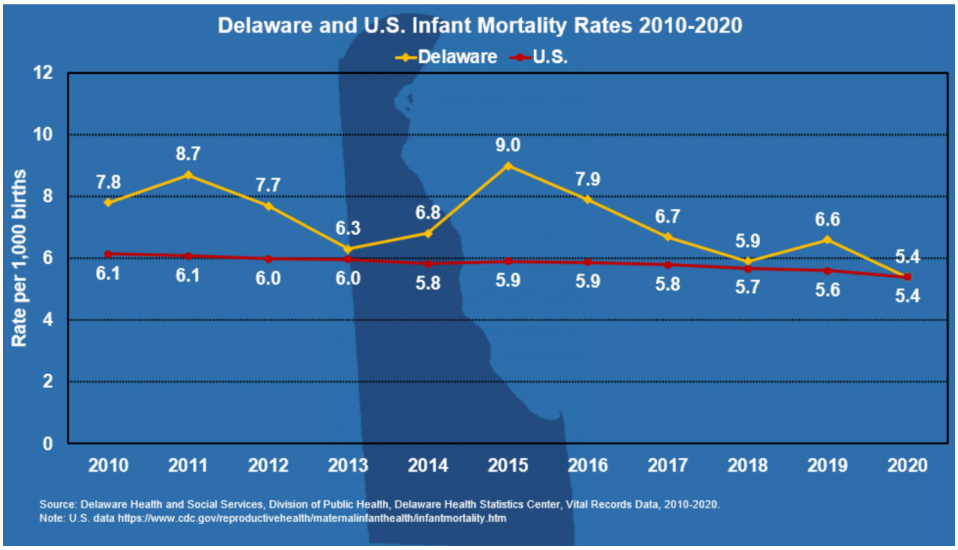

Figure 6. Delaware and U.S. Infant Mortality Rates 2010-2020

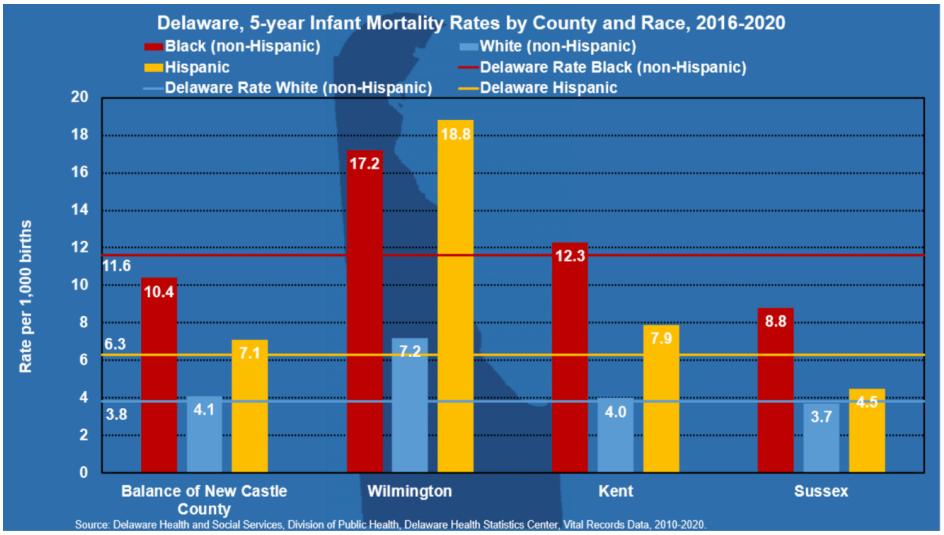

Figure 7. Delaware Five Year Infant Mortality Rates by County and Race, 2016-2020