Ethical Data Stratification and Analysis to Eliminate Unconscious Bias in Learning Health Systems for Palliative Medicine

VIRGINIA M. MIORI, PHD, KYLE CHALMERS, CISA, DANIEL J. MIORI, RPA-C, MS

vmiori@sju.edu

Abstract: *Learning Health Systems (LHS) produce analyses supporting medical decision making. Though algorithms are chosen to eliminate bias, we are left with unconscious bias present in data, due to lack of representation for marginalized populations. This is especially problematic in palliative care. Medical practitioners lack historical foundations for decision making for patients in these underrepresented populations. This paper first reviews the ethical foundations that drive our approach, then describes data stratification and analysis supporting a data-ethical LHS. Kaplan-Meier curves are generated for each of the stratum, demonstrating the value of palliative care and its impact on survival. Data were collected from government and foundation sources, stratified and analyzed to more appropriately weight medical history and characteristics to improve decision making for all populations.*

"Very well. Sea," cried Canute, "I command you to come no further! Waves, stop your rolling! Surf, stop your pounding! Do not dare touch my feet!" [1]

I. INTRODUCTION

Anticipating that Data Science, as any technology, will advance by its own substantial inertia, a central premise to this work is the belief that just as the sea tide did not cease for King Canute, neither will the tide of technological advance. Data Science in health care will advance, despite the historically sluggish pace of adoption of analytics techniques, especially when such great potential benefit exists. Because of this inexorable advance, discussion and understanding of the appropriate data and analytics techniques, along with the ethical implications of these, require attention and review, particularly within the context of the Four Vs of big data: volume, variety, velocity, and veracity.

---

[1] "The Changing Story of Cnut and the Waves," Medievalists.net, accessed September 7, 2020, http://www.medievalists.net/2015/05/the-changing-story-of-cnut-and-the-waves/

Learning Health Systems are critical to the advancement of any type of care. A Learning Health System is defined by the Agency for Healthcare Research and Quality (AHRQ)[2] as:

> *"a health system in which internal data and experience are systematically integrated with external evidence, and that knowledge is put into practice. As a result, patients get higher quality, safer, more efficient care, and health care delivery organizations become better places to work."*

Palliative Care is best characterized as facilitating the transition from futile, and potentially burdensome and aggressive, curative care for a patient with an irreversible (possibly terminal) condition, to treatment encompassing goals based on quality of life, as agreed on by practitioner and patient. This approach may not necessarily include a wish to limit attempts at resuscitation as is seen in Hospice type care. Frequently, attaining those quality-based goals may involve some continued highly aggressive care to improve function and limit symptoms.

Palliative care is a field with its own unique challenges and philosophic approach which must be considered in discussion of the future of any LHS. According to the World Health Organization[3], only 14% of people (worldwide) who need palliative care, actually receive it. Though the percentage is much higher in the US-, low- and middle-income individuals have significantly less access to palliative care. The WHO also identified key barriers to palliative care, including:

- *"lack of awareness among policy-makers, health professionals and the public about what palliative care is, and the benefits it can offer patients and health systems;*
- *cultural and social barriers, such as beliefs about death and dying;*
- *misconceptions about palliative care, such as that it is only for patients with cancer, or for the last weeks of life; and*
- *misconceptions that improving access to opioid analgesia will lead to increased substance abuse."*

Any discussion of Data Science' integration into an LHS for palliative care must address the unconscious bias due to lack of representation of racial, ethnic and socio-economic groups in treatment data. It is also critical to address the impact of this bias on prognoses and diagnoses rendered for patients. An LHS shares the ethical responsibilities of the clinicians who manage patients, and make recommendations for medical treatment and should consider questions such as:

- What are code survival rates for palliative patients with specific comorbidities?
- What is the likelihood of post-surgical complications for a cancer survivor with a specific treatment profile?
- What is the likelihood of a patient with compromised immunity developing clostridium difficile colitis (c-diff) during a hospital stay?
- What is the likelihood of an elderly patient developing hospital associated delirium?

In palliative care as well as the practice of medicine in general, timely decisions often need to be made on incomplete evidence. Additional time gained by catching the patient earlier in an irreversible disease process allows for more effective clinical treatment to relieve symptoms, and greater time in a therapeutic environment to deal with the emotional trauma of illness and death.

Hospitals and medical practitioners everywhere (rural, suburban, and urban areas), who serve every age group and socio-economic group, should be represented in LHSs and therefore have the ability to provide equal levels of care for all patients. A patient in inner-city hospitals should be able to benefit from knowledge gained across all of these diverse groupings.

To date, the Kaplan-Meier curve has served as a fundamental data analysis tool used to estimate survival times or time-to-event (in this case, the event is death), in the face of missing or incomplete data. In this paper, Kaplan Meier Curves were generated to represent survival among various groups of underrepresented and marginalized populations. The curves served to illustrate the impact of the discrepancies in the delivery of palliative care.

---

[2] "About Learning Health Systems," Agency for Healthcare Research and Quality, accessed July 12, 2020, https://www.ahrq.gov/learning-health-systems/about.html

[3] "Palliative Care Fact Sheet," World Health Organization, accessed September 28, 2020, https://www.who.int/news-room/fact-sheets/detail/palliative-care

II.   LITERATURE REVIEW

A.   Ethics

Ethics surveys have been utilized to understand the role of ethics in clinical settings. Research in clinical ethics support (CES) using surveys (Dauwerse et al., 2013), determined that both implicit and explicit CES are relevant and should be combined to better serve patients. A focus on ethics committees, and identification of roles of the committees (Marcus et al., 2015) found that ethics committees should practice mediation rather than prescription in resolving clinical ethical dilemmas. In addition, they are viewed as a practical means to avoiding medical-legal conflicts.

Ethical considerations of health care go beyond the question of care in individual circumstances. They are also critical when examining disparities in healthcare delivery. Privacy restrictions and lack of consolidated electronic medical records impede the ability of researchers to adequately address inequities.

The need for accurate data on race and ethnicity in healthcare settings has been emphasized (Bierman et al., 2002). Collection of data is hindered, resulting in unconscious bias and disparities in care. Based on increased diversity in the US population, improvements must be made to efficiently support the population. Individual data was collected (NRCP on DHHS, 2004) on race, ethnicity, socioeconomic position, acculturation and language. The panel identified the data needed, the effectiveness of data collection systems and critical gaps in the data, along with providing recommendations for improvement completed based on a literature analysis (Cabral and Cuevas, 2020) the conclusion was drawn that unauthorized immigrants within the Hispanic/Latino population have significant disadvantages, even when compared to their documented counterparts. Economic and social conditions such as health care access, health-protective resources and immigration enforcement actions operate differently for these two groups. They provided recommendations for policies and interventions to mitigate the gap.

The COVID-19 pandemic is used as a backdrop (Sivashanker et al., 2020) to show the intersectional (overlapping underserved groups such as being black and a woman) effect of the pandemic. Otherwise, hidden inequities were identified such as: Hispanic non-English speaking patients were dying at higher rates than those who spoke English, and red-lined neighborhoods of color were tested at lower rates, but positive tests were more prevalent than affluent white neighborhoods. These findings led to quality-improvement efforts.

A detailed examination of equitable healthcare (Weinstein et al., 2017) pointed to social, economic, environmental and structural factors in our inability to provide equitable healthcare. Specifically, they pointed to poverty, structural racism and discrimination as limiting factors in obtaining care. They called for community action and governmental programs, policy changes, and partnerships to begin closing the gap.

This formalization of data science using conscious or unconscious decision-making strategies finds a corollary in Belief Ethics. This field "refers to a cluster of questions at the intersection of epistemology, ethics, philosophy of mind, and psychology" (Chignell, 2017) and may be helpful in discussion of Data Science tools to assist with medical decision-making, particularly with regard to prognosis

Previous research has voiced specific concern over indiscriminate use of data dependent prognostic tools in place of the participation of a trained clinician (von Gunten et al., 2016). This seems to be the center of published concerns over ethics within the application of Data Science to the practice of palliative medicine. In addition to voicing concern over the use of data dependent prognostic tools, they also noted current challenges that such tools would address, such as the tendency of physicians to overestimate prognosis at end of life by "300-500%".

The current push for engineering ethics built into algorithms for health care solutions, and the notion of technical neutrality were addressed (McCradden et al., 2020). They noted however that complex causal factors are responsible for difference in medical conditions, with some factors critical to identifying differences, while others do not. In addition, algorithms reflect idealized models rather than real-world models. The solutions are not simple and require careful consideration.

B.   Data Science and Palliative Care

In the case of a Data Science system such as an LHS, to assist with decision-making, a reliable and acceptable method for codifying and incorporating prudential and moral norms would have to be produced. The programmers responsible for the logic of such a system would hold ethical (and potentially legal) responsibility for that system, similarly to the clinicians who use it. Ultimately any LHS would be fallible and must be viewed as additional information for the clinician and patient/surrogate, in much the same way

any other current system, such as MRI or blood test results are now used. Using current testing and standards (as the precedent strongly suggests), the recommendations of an LHS could be acceptable evidence of the improbability of meeting patients' treatment goals. This evidence is an adequate reason to transition to a less aggressive, symptom-control based plan of treatment.

Heuristics allow decisions to be made in the absence of full, possibly unobtainable, information. A phenomenon known in psychology as the illusion of control (Hastie & Dawes, 2009) is one of the understood flaws in medical heuristics. It is the tendency to assume having control in situations where it may not exist.

Heuristic systems are often proposed (Uneno et al., 2017) (Snow et al., 2016) (Smith et al., 2012) (Ying et al., 2015) to assist in prognosis. Existing prognostic tools, The Palliative Performance Scale (PPS) created for use in palliative care, and the Karnofsky Scale that was created for use in oncology, are both based on the patient's overall ability to function, and establish a precedent for the use of formalized grading scales in prognosis. Other, more data driven tools, such as APACHE (APACHE II Score, 2020), are used in a critical care setting and are more complex, involving the processing of lab results and other values to arrive at a score. Additionally, scoring systems for evaluation of frailty in the elderly (Rockwood, et al., 2005) are gaining acceptance, and are considered valid for influencing medical recommendations for treatment.

This judicious use of conscious heuristics can reasonably be extended to the use of data science in medical decision-making. Info markers, such as the use of terms including "Death Anxiety" and "Anticipatory grieving", when used in EMR documentation can trigger consideration for palliative consult (Ying et al., 2015). This and other approaches (Miller, 2016) support the use of prognostic tools and (Smith et al., 2012) (Greer et al., 2013) further develop the concept of data science assisting in a transition from aggressive curative type care to aggressive symptom management.

The Guardian noted that the British National Health Service used their own algorithm, called The Liverpool Pathway to assist with decision-making at end of life. Concerns that a Data Science system should not be substituted for shared decision making between clinician and patient/surrogate informs the British National Health Service's review of the pathway, which directly led to its discontinuation in 2013. Concerns were expressed (von Guten et al., 2016) over efforts to "mine big data for prognostic information…being done by private, proprietary, for-profit companies" which may manipulate data to influence the general public who may not have the technical awareness to discern between valuable recommendations and simple marketing.

The Guardian[4] noted that the British National Health System has now instituted Liverpool's replacement through their National Institute for Health and Care Excellence[5] (NICE). It provides guidance which echoes much of which was offered by The Liverpool Pathway, but in a more patient centered and non-technical format available directly to the healthcare consumer.

Several articles (Pastrana et al., 2010)(Cooper et al., 2018)(Haines, 2021) focused on quality improvement using data, including 1) discussion of documenting the impact of likely benefits of palliative care on quality of healthcare and the associated costs, and 2) accurately defining and identifying quality and success in palliative care. The Center to Advance Palliative Care (CAPC) provides independent data and reference who material on the cost effectiveness of having a palliative service available.

Ultimately, the greatest effect Data Science may have on Palliative care will be through its impact on the practice of medicine in general. Miller discussed views adopting assessment tools like the Risk Readmission Assessment Tool (RRAT), which can evaluate risk of potentially avoidable rehospitalization of patients with advanced heart failure, as a positive feature leading to appropriate and more timely referrals.

C. Methodology

The discussion of the Ethics of Belief has created the foundation for data collection, stratification and heuristic creation. The divergent voices in data collection include the US Census Bureau, the US Department of Housing and Urban Development (HUD), the Homeless Research Institute, the Migration Policy Institute (MPI), the Refugee Processing Center (RPC), the Centers for Medicare and Medicaid Services (CMS), the Center to Advance Palliative Care (CAPC), the Centers for Disease Control (CDC), the National Health Interview Survey (NHIS) and the World Health Organization (WHO).

The first five of these sources were used to form a foundational understanding of the demographics of the US population. This data was then used to describe the population, with specific attention paid to characterizing population centers of different sizes, differing urban characteristics and geographic regions. The remaining sources (CMS, CAPC, CDC, NHIS and WHO) were used to characterize the demographics of the population of palliative patients.

---

[4] https://www.theguardian.com/society/2013/jul/15/liverpool-care-pathway-what-went-wrong
[5] "End of Life Care for Adults," National Institute for Health and Care Excellence," accessed October 22,
2020, https://www.nice.org.uk/guidance/qs13

Age, sex, race, educational attainment, ethnicity, income (poverty status), health insurance status, urban vs rural, homeless status, documentation status for immigration, and refugee status were all used to characterize the US population data, and then determine specific factors that led to underrepresentation in the palliative population. Understanding these differences between the overall population and the palliative population allows an LHS and healthcare professionals to begin to remediate the unconscious bias that exists in palliative data.

To illustrate the importance of palliative care in marginalized and underserved populations, we used the NHIS data to identify patient profiles across all strata. These profiles specifically reflect comorbidities, insurance status, and overall propensity to seek medical care. This approach allowed potential patients to be grouped in non-standard ways, ensuring that medical, economic and cultural factors all would be taken into consideration. We then simulated the time of survival for each of the strata and created Kaplan-Meier curves for each stratum, demonstrating the importance of palliative care and demonstrating the need to make palliative care more accessible.

This approach creates a structure for collection of private data in individual practice allows decision makers to map their own patient populations to the strata.

III.   DATA COLLECTION

A.   Total Population

The first goal in data collection was finding and aggregating publicly available data that would serve as an accurate representation of the entire US population. Within the US Census, we focused data collection to American Community Survey (ACS) 5-Year Estimates[6]. The survey is ongoing and provides estimates that each represent a period of time. The 5-year estimates (in this case from 2014-2018) increase statistical reliability, especially for less populated areas and underrepresented populations. To ensure representation, further collection included sources that adequately reflected the homeless population, the unauthorized immigrant population and the refugee population. This aggregated data provides a more accurate representation of the population of the United States. The Migration Policy Institute[7] website has used the same 5-year period estimates, and a unique methodology to assign legal status to the census data and effectively approximate the population of unauthorized immigrants in the United States.

The HUD Exchange maintains a Homeless Data Exchange[8] (HDX) and generates the Annual Homeless Assessment Report presented to Congress from homeless population and subpopulation data submitted by the Continuum of Care Programs (CoCs). These point-in-time estimates are generated annually and include population sizes and demographics by state. The homeless population's characteristics are clarified through the examination of publications of the National Alliance to End Homelessness[9]. In particular, the Demographic Data Project Part IV: The Role of Geography, and the State of Homelessness publication provided the foundation for estimating the homeless population racially and geographically.

B.   Palliative Population

Characterizing the palliative patient population required more depth and evaluation of publicly available data sets. The international reports of the WHO and the Palliative Care Fact Sheet, were used to identify an initial set of chronic conditions that lead to the need for palliative care. These conditions include cardiovascular diseases, cancer, chronic respiratory diseases, AIDS, and diabetes. The CDC completes surveys on risk factors on an annual basis[10], the resulting data directly addresses these and several other chronic conditions. The aggregated and adjusted survey data helps to identify the prevalence of the chronic conditions across demographic intersections.

[6] "American Community Survey 5-Year Data (2009-2019), United States Census Bureau, accessed July 12, 2020-January 4, 2021, https://www.census.gov/data/developers/data-sets/acs-5year.html

[7] "Migration Data Hub," Migration Policy Institute, accessed July 18, 2020- November 1, 2020, https://www.migrationpolicy.org/programs/migration-data-hub?qt-data_hub_tabs=0#datahub-tabs

[8] "PIT Count and HIC Data and Reports," HUD Exchange, accessed July 18, 2020-November 1, 2020, https://www.hudexchange.info/programs/hdx/pit-hic/data-reports/

[9] "The State of Homelessness," National Alliance to End Homelessness, accessed July 18, 2020-November 1, 2020, https://endhomelessness.org/

[10] "Data & Statistic," Centers for Disease Control and Prevention, accessed July 29,2020-January 18, 2021, https://www.cdc.gov/DataStatistics/

If a condition is chronic, yet considered to be reversible, it is not an appropriate predictor of palliative care needs. The CDC data also addresses hospice care, and though palliative patients may eventually require hospice care, the hospice population is not considered to provide an effective estimate of the likely palliative population.

The NHIS data provides an opportunity to use data at a more detailed level of granularity than other sources. The survey data is provided at an individual level, which is the foundation for creation of profiles to generate Kaplan-Meier survival curves.

C. Quantification

Given the wealth of data, it was important to begin the quantitative work descriptively. Data aggregations and visualizations played a pivotal role in understanding the data and formulating appropriate analyses for the data. It was insufficient to create strata based on the intersection of demographic characteristics, rather the strata had the express purpose of representing groups that had access to varying levels of quality healthcare. For example, a homeless individual and an unauthorized immigrant come from very different circumstances, yet face almost exactly the same healthcare challenges. For this reason, we view the response variable as healthcare access. For each data type, a dashboard has been created to provide a visual story. The links to these dashboards are provided at the start of each section.

D. Census Data

Dashboard Link: Race/Ethnicity and Income/Poverty Census Data Analysis (Click Tableau Public Link)

The census data examined was primarily concerned with race/ethnicity (RE) and income/poverty (IP) metrics. The RE columns measured gender, age, race, ethnicity, and housing units. The IP columns measured overall household income, other sources of household income, labor force status, health insurance coverage, and poverty level. Note that employment and health insurance data points were not captured until 2012, while all other data points were captured back to 2010. Table 1 includes meaningful changes in demographics while Table 2 includes relevant correlations[11].

In order to accurately represent the percentage change in Social Security, independent of population shifts, the calculation was based on COLA increases from 2010-2018[12]. The inflation rate, based on the consumer Price Index – All Urban Consumers (CPI-U), over the time period from 2010-2018 was 16.34%[13]. The income measures marked with an asterisk (*) have fallen behind inflation during this time. Expanded results available upon request.

Category	Base Year	2018	Percentage Change
Income (2010-2018)			
Mean Household Income	$70,883	$84,938	19.83%
Median Household Income*	$51,914	$60,293	16.14%
Social Security Income*	$15,495	$19,346	*14.38%*
Below Poverty Level	10.1%	11.5%	13.86%
Employment (2012-2018)			
Employment Rate	70.13%	73.18%	4.35%
Unemployment Rate	7.09%	4.41%	-37.80%
Health Insurance (2012-2018)			
Uninsured	14.79%	9.33%	-36.92%
Private Insurance	59.96%	60.37%	0.68%
Public Insurance	26.73%	31.28%	17.02%
Race/Ethnicity (2010-2018)			

---

[11] Correlations with an absolute value of 0.5 or greater are deemed significant

[12] "Cost of Living Adjustments," Social Security Administration, accessed March 15, 2020, https://www.ssa.gov/oact/cola/colaseries.html

[13] "Inflation Rate in the United States of America, 2020," StatBureau, accessed March 15, 2021, https://www.statbureau.org/en/united-states/inflation

African American/Black	13.04%	13.5%	3.53%
Asian	5.13%	6.19%	20.67%
Latinx	16.72%	18.65%	11.54%
White	74.13%	72.88%	-1.69%
Lack of English Fluency	8.7%	8.5%	-0.23%
Age (median: 2010-2018)	36.9	37.9	2.71%

Table 1: Census Data Change Summary

Variable	Significant Negative Correlation	Significant Positive Correlation
Private Health Insurance Rate	Food Stamps Supplemental Social Security < $25,000 Household Income Households Below Poverty Line Unemployment/Not in Labor Force	Mean Income Mean Social Security Income > $50,000 Household Income Employment Rate
Public Health Insurance Rate	Mean Income Mean Social Security Income > $50,000 Household Income Employment Rate	Food Stamps Supplemental Social Security < $25,000 Household Income Households Below Poverty Line Unemployment/Not in Labor Force

Table 2: Relevant Insurance Rate Correlations

E. Unauthorized Immigrant Data

Dashboard Link: Unauthorized Immigrant Data Analysis (Click Tableau Public Link)

The unauthorized immigrant data contains demographic information such as country of birth, region of birth, years of US residence, age, gender, parental status, marital status, school enrollment, educational attainment, English proficiency, languages spoken at home, labor force participation, top employment industries, family income, health insurance coverage, and home ownership.

The number of unauthorized migrants positively correlates with the population size and presence of large metropolitan areas within a state. The top 5 states in order of unauthorized population are CA, TX, NY, FL, and IL. CA has almost 900,000 more migrants than TX, and TX has almost 900,000 more NY.
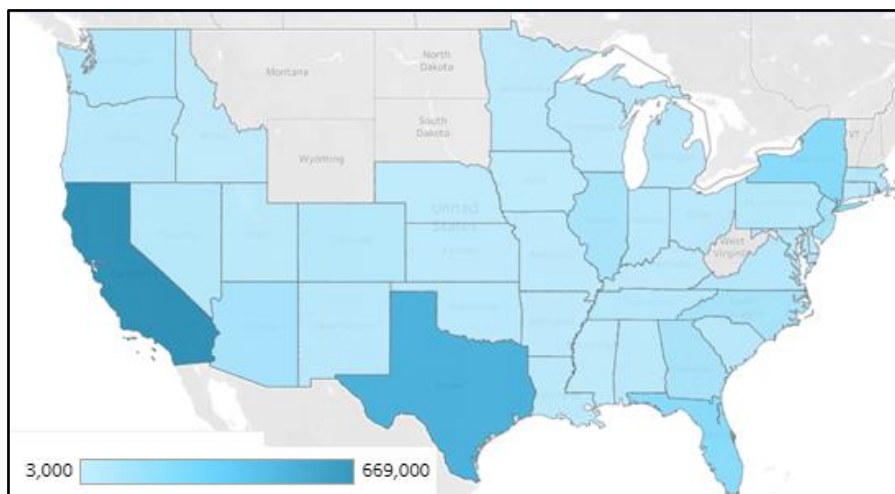

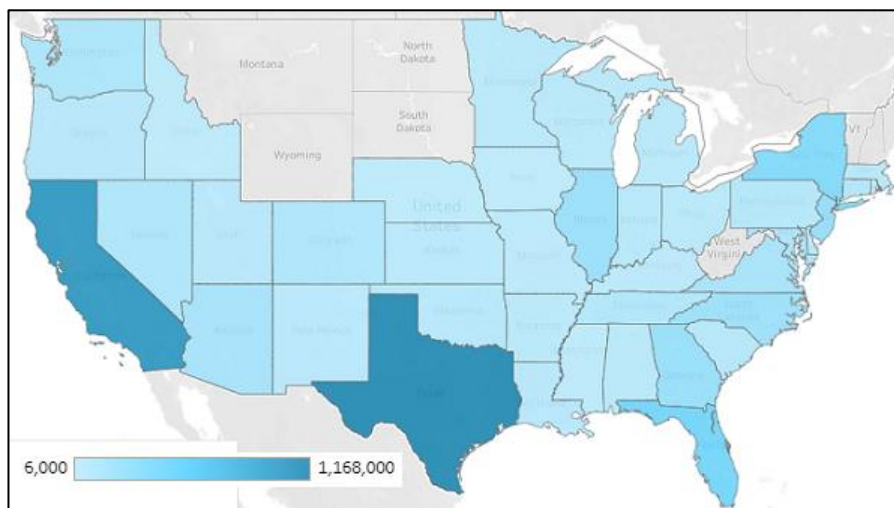*Figure 1: Unauthorized Immigrants below Poverty Level*

Figure 2: Uninsured Unauthorized Immigrants

While many variables can act as predictors of access to quality healthcare, three dominant variables are income below the poverty line, uninsured status and inability to speak English well. The heatmaps in Figures 1 and 2 show that very high concentrations of unauthorized immigrants are marginalized by income and insurance status.

F. Homelessness Data

Dashboard Link: Homeless Data Analysis (Click Tableau Public Link)

The homeless data contains demographic information like age, gender identity, race, and ethnicity about homeless populations by US state and year. In the United States, the overall homeless population was 564,708 people in 2015 and 552,830 people in 2018, a decrease of 11,878 people. In the intermediary years the homeless population was always at least 10,000 less than either of these numbers, but there were no downward fluctuations greater than 15,000. Examining race, white people consistently represent 47-49% of the homeless population across these years. Black people consistently represent 39-41% of the homeless population across these years. It is important to note that black people are disproportionately represented in the homeless population given they only represent around 13% of the population while white people represent over 70% of the population. Additionally, population hubs like NY and CA account for 36.13% of the homeless population in 2015, and by 2018 accounted for 39.71%.

Examining specifically the sheltered vs. unsheltered homeless populations, the sheltered homeless population has declined from 69.32% in 2015 to 64.82% in 2018 while the unsheltered population has increased from 30.68% to 35.18%. Examining race over the years measured, black people consistently represented 45-47% of the sheltered homeless population 26-28% of the unsheltered population. White people consistently represented 43-44% of the sheltered homeless population and 56-59% of the unsheltered homeless population.

G. Refugee Data

Dashboard Link: Refugee Data Analysis (Click Tableau Public Link)

The refugee data has four critical components that were measured from 2010 to 2020: refugee's country of origin, the refugee's destination state, year, and number of refugees. The summary of the largest refugee populations is presented in Table 3.

Rank	Refugee Country of Origin (COE)	Refugee Destination State	Largest Refugee Population (COE/state)
1	Burma (125,135)	Texas (59,273)	Iraq/California (20,446)
2	Iraq (109,135)	California (54,412)	Burma/Texas (17,512)
3	Bhutan (77,409)	New York (35,589)	Iraq/Michigan (15,675)
4	DROC (70,444)	Michigan (30,123)	Iran/California (12,980)
5	Somalia (54,208)	Arizona (25,959)	Iraq/Texas (11,918)

Table 3: Largest Refugee Population Summary

The US accepted a total of 600,832 refugees over these 11 years. The year with the largest number of refugees was 2016 with 84,933 refugees. 2017-2020 were the four lowest years in terms of volume of refugees accepted to the US. The only country for which the US continued to allow large numbers of refugees was the Democratic Republic of Congo (DROC), while Iraq, Burma, Bhutan, Somalia, Syria, and almost every other nation had significantly fewer refugees. The same patterns were replicated at the state level.

H. Chronic Disease Indicators Data

Dashboard Link: *Chronic Disease Indicators Data Analysis (Click Tableau Public Link)*

The chronic disease indicators data contained data on alcohol usage, arthritis, asthma, cancer, cardiovascular disease, chronic kidney disease, chronic obstructive pulmonary disease, diabetes, disability, immunization, mental health, nutrition and physical activity, older adults, oral health, overarching conditions, reproductive health, and tobacco. The data was collected for each state and the overall United States from 2008-2018, though data collection has not been completely consistent during this time. The data was primarily collected for the overall population, by race/ethnicity, and by gender. The data has 193 distinct questions asked in it regarding the different topics listed above, but Table 4 offers the results from the most general questions for the overall population.

The metrics in Table 4 provide a mixed overall picture. While many of the categories do not provide race/ethnicity data, disproportionately adverse outcomes prevail for minorities when examining poverty rate, prevalence of two or more chronic conditions for 65+ Medicare-enrolled people, and disabilities for adults 65+. Whites are consistently the lowest represented group with a poverty rate within the range of 9-11%. Blacks, Hispanic/Latinx, and Native Americans consistently have poverty rates over 20%. The percentage of black Americans (65+) on Medicare with two or more conditions consistently exceeds 71%, while every other race consistently falls below 70%. The percentage of disabled black Americans falls within the 43-39% range. The percentage of disabled Native Americans consistently hovers just below 50%, while the percentage of disabled white and Asian Americans falls below 35%.

Category	Base Year	Final Year	Year Range
Chronic Liver Disease Mortality	0.1 cases per 100,000	0.13 cases per 100,000	2010-2017
Heavy Drinking Among Adults 18+	6.6%	6.4%	2011-2018
Arthritis Among Adults 18+	24.4%	26.3%	2011-2018
Current Asthma Prevalence Adults 18+	9.1%	9.4%	2011-2018
Incidence of Invasive Cancer	5.017 cases per 100,000	5.044 cases per 100,000	2010-2012
Mortality of Invasive Cancer	1.85 cases per 100,000	1.85 cases per 100,000	2010-2012
Mortality from Total Cardiovascular Diseases	2.54 cases per 100,000	2.64 cases per 100,000	2010-2017
Awareness of High Blood Pressure Adults 18+	30.8%	32.2%	2011-2017
Prevalence of Chronic Kidney Disease Adults 18+	2.4%	2.9%	2011-2018
Prevalence of Chronic Obstructive Pulmonary Disease Adults 18+	6.1%	6.6%	2011-2018

Prevalence of Diagnosed Diabetes Adults 18+	9.5%	10.9%	2011-2018
Mortality due to Diabetes Reported as any Cause of Death	0.83 cases per 100,000	0.83 cases per 100,000	2010-2017
Disability Adults 65+	36.7%	34.6%	2010-2017
Mean Recent Mentally Unhealthy Days for Adults 18+	3.8 Days	4 Days	2011-2018
Overweight or Obesity Among Adults 18+	63.8%	66.5%	2011-2018
Healthy Weight Adults 18+	34.5%	31.6%	2011-2018
Prevalence of 2 or More Chronic Conditions for 65+ Medicare-enrolled People	69.8%	67.7%	2010-2015
6 or More Teeth Lost for Adults 65+	39.2%	33.7%	2012-2018
Poverty Rate	15.3%	13.4%	2010-2017
Timeliness of Routine Healthcare Checkups Among Women 18-44	62.4%	74.5%	2011-2018
Current Smoking Among Adults 18+	21.2%	16.1%	2011-2018

Table 4: Chronic Condition Data Change Summary

Disparities in quality of health, based on race, carry weight in palliative care decisions. The leading causes of death among the African American/Black American population and the Hispanic/Latinx population were found to be heart disease, cancer, and accidents (Carratala and Maxwell, 2020). African Americans have the highest mortality rate for all cancers combined among racial/ethnic groups. Table 5 shows percentage incidence of chronic conditions by race, as reported by the CDC in 2018.

Condition	African American/ Black American	Hispanic American/ Latinx American	Non-Hispanic White
Fair or poor health	14.7%	12.3%	8.3%
Overweight/Obese	72.0%	71.1%	65.3%
Hypertension	32.2%	23.7%	23.9%
Diabetes	13.1%	13.2%	8.6%
Heart Disease	10.0%	8.2%	11.5%
Stroke	3.9%	2.5%	2.6%
Cancer	5.1%	4.2%	9.1%
COPD	4.5%	3.7%	5.0%
AIDS	0.01%	0.02%	<0.003%

Table 5: Chronic Conditions Percentage Incidence

A scientific statement for the American Heart Association (Rodriguez et al., 2014), highlighted cardiovascular disease and stroke among Hispanics/Latinx identified personal beliefs and social issues faced by this population. Among the common problems were language skills of patients, providers and translators. In addition, there may be cultural traditions such as folk/alternative medicine, prayer and other rituals that must be respected. Inclusion of family is critical to effective care.

Identification of differences in access to, and utilization of, healthcare by documented and unauthorized Mexican immigrants (Vargas Bustamante et al., 2012) was studied. Unauthorized immigrants from Mexico are 27% less likely to have visited a doctor in the previous year and 35% less likely to have a regular doctor than documented Mexican immigrants.

The increased risk of certain chronic diseases faced by non-white populations has a significant impact on life expectancy and serves to emphasize the importance of eliminating bias from health care data. Palliative patients must be assessed using the most relevant data to ensure intervention that improves quality and duration of life. Kaplan-Meier curves will be used to demonstrate the downstream impact of chronic conditions and other characteristics.

I.   Kaplan Meier Curves

The data gathered for this research, contains as complete a representation of the US population as is possible. From this data, we have completed analyses to identify related variables and factors that impact the acquisition and quality of palliative care. As noted previously, this approach allowed potential patients to be grouped in non-standard ways, ensuring that medical, economic and cultural factors will be taken into consideration.

Health care systems and palliative care practices with patient level data would have the ability to complete multiple regressions to estimate survival times, and logistic regressions (Miori and Miori, 2010) to estimate survival probabilities. In the absence of patient level data, survival times and quality of life may still be estimated using Kaplan-Meier curves. Characteristics more closely identified with each type of patient are defined in Table 6.

Characteristic	Definition
Advocacy	Presence of a family member/social worker, etc. who advocates for the patient, or a patient who is able to advocate for themself
Translation	Ability of a patient or patient advocate to completely understand information communicated by medical professionals
Insurance	Type of insurance coverage (if any) held
Education	Level of educational attainment (relative to secondary school)
Employment	Level of permanent employment
% of Poverty Level	Patient income as a percentage of poverty level, ($25,100 based on a family of 4 in the Federal Poverty Guidelines[14])
Overall Health	Reflects overall health prior to becoming palliative. This factor reflects comorbidities such as Diabetes, Hypertension, COPD, AIDS, etc. This factor also reflects race and ethnicity, noting that the probability of comorbidities changes by race.
Denial of Need for Own Care	An individual patient may deny the need for their own care. This may result from a lack of acceptance of an illness of potentially cultural, social or ethnic norms.

Table 6: Definitions of Patient Characteristics

Three strata of palliative patients and patient care status are identified in Table 7. In the presence of detail level patient data, these discrete categories would be more appropriately represented as a smooth continuum. We present these groups at this time for the purpose of demonstrating procedure. Each Group (I, II and III) represents a likely level of palliative care administration for populations displaying the associated characteristics in Table 6. With the exception of advocacy and fluency, assignment to a care group would be determined by the concentration of characteristics. Patients who lack advocacy or are subject to poor medical fluency, would automatically fall into Group I. An unemployed and uninsured patient, with poor overall health, and less than a secondary level education, would likely fall into Group I as well. As expected, highly educated patients with strong advocacy, fluency and private insurance are most likely to fall into Group III and receive exceptional care.

---

[14] "Federal Poverty Guidelines," FamiliesUSA, accessed March 15, 2021, https://familiesusa.org/resources/federal-poverty-guidelines/

Patient Goal Status/Patient Care Status	Group I: Poorly Defined or Absent Goals/Absence	Group II: Possible Discussion of Goals/Adequate Care	Group III: Well Defined Goals/Exceptional Care
Advocacy (yes or No)	No	Yes	Yes
Translation	Poor	Conversant	Fluency
Insurance	Uninsured	Public Insurance	Private Insurance
Education	Incomplete Secondary	Secondary	Post-Secondary
Employment	Unemployed	Under-employed	Well Employed
% of Poverty Level	0%-150%	150%-250%	> 250%
Overall Health	Poor	Good	Excellent
Denial of Own Need for Care (Yes, No)	Yes	No	No

Table 7: Characteristics of Palliative Patient Care Groups

Most likely survival times, least expected survival times and longest expected survival times were estimated for each type of patient.[15] We chose the triangular distribution to represent survival time for each group. Being bounded and unimodal, it serves as a reasonable estimate in the absence of detail patient data, while also serving to effectively demonstrate the procedure. The most likely survival time is taken as the mode, with the least and most likely survival times being taken as the minimum and the maximum. The probability density function (pdf) of the triangular distribution is stated in Equation 1.

Let:

$$Let\ a = the\ minimum\ value$$
$$Let\ c = the\ most\ likely\ (peak)\ value$$
$$Let\ b = the\ maximum\ value$$

$$f(x) = \begin{cases} 0 & x < a \\ \dfrac{2(x-a)}{(b-a)(c-a)} & a \le x \le c \\ \dfrac{2(b-x)}{(b-a)(c-a)} & c \le x \le b \\ 0 & x > b \end{cases} \quad (1)$$

Using the RAND() function[16] in Microsoft Excel and the pdf of the triangular distribution, 250 instances of the triangular distribution were generated for each group, for a total of 750 patients. When using private patient level detail, the patient survival time should be modeled using whatever statistical distributions are found to provide the most significant fits.

---

[15] Note that these estimates are intended to provide relativity between the categories and are not intended to be accurate estimates of any single patient's condition.

[16] RAND() generates pseudo random numbers that fall between 0 and 1.

The simulated instances served as survival time estimates for each group, and were used to generate and Kaplan-Meier curves. These curves in Figure 3 clearly demonstrate higher quality of life for longer periods with higher quality palliative care. Populations that are marginalized due to race, ethnicity, poverty, and other statuses fall into Group I with significantly greater frequency. It is clear that accessibility to quality palliative care is extremely important and must be more broadly available.
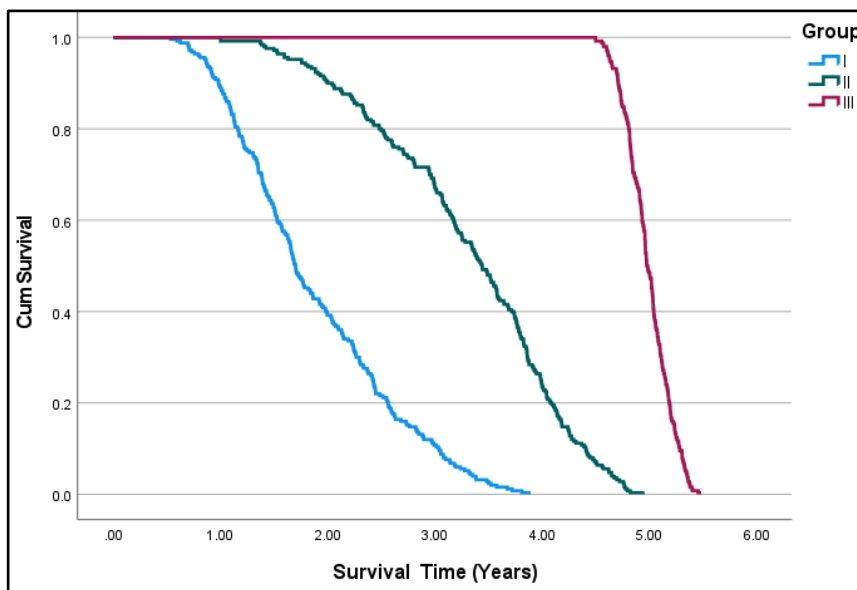

Figure 3: Kaplan-Meier Curves

IV. CONCLUSION

In this research, a process was identified that can be broadly applied in the absence of patient level data. The population data and chronic conditions data were publicly available from primarily the US Census Bureau and the Centers for Disease Control. Though the data collection was manual, Application Programming Interfaces (APIs) are also available to automate this process.

The data and analysis substantiated the need to build an LHS free from unconscious bias. Rather than simply focus on a single factor such as race, this research was particularly focused on the intersections that exist among populations, and the characteristics that serve as indicators of the expected quality of palliative care.

We have demonstrated that identification of marginalized populations requires consideration of intersections of advocacy, quality translation, race, ethnicity, socio-economic status, cultural characteristics (including denial of need for care), education, employment, and type of insurances among others. The use of these modifiers allowed for the creation of strata (groups) that effectively represent marginalized populations across all intersections.

Based on the Kaplan-Meier curves, it is clear that from poor to adequate palliative healthcare has a significant impact on life span in marginalized populations. LHS algorithms must reflect both conscious and unconscious biases through adequate representation of marginalized populations.

A. Future Research

Extensions of this research cross multiple areas. The first and most obvious extension is the inclusion of detail level patient data. Not only would this level of data allow for a further validation of the process, but it would also allow for generation of more representative statistical distributions of survival and Kaplan-Meier curves. A continuous variable characterization of survival would eliminate the discrete division of groups in the current analysis and establish greater nuance in the survival expectations.

Real data will provide a foundation for sizing of the strata. Further, analyses such as multiple regression, logistic regression, and simulation can be used to predict individual survival time, predict survival rates, and identify bottlenecks and inequities in palliative care delivery.

Beyond the examination of data, future research includes algorithmic design to identify and resolve inequities and unconscious bias in the data, ultimately leading to the creation of a Learning Health System focused on palliative care.

V.   REFERENCES

Cabral, Jacqueline and Adolfo G. Cuevas. "Health Inequities Among Latinos/Hispanics: Documentation Status as a Determinant of Health." *Journal of Racial and Ethnic Health Disparities* 7 (2020):874–879. https://doi.org/10.1007/s40615-020-00710-0

Sofia Carratala and Connor Maxwell. "Health Disparities by Race and Ethnicity." *Center for American Progress Fact Sheet* (2020). https://cdn.americanprogress.org/content/uploads/2020/05/06130714/HealthRace-factsheet.pdf

Chignell, Andrew, "The Ethics of Belief", *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edited by Edward N. Zalta, https://plato.stanford.edu/archives/spr2017/entries/ethics-belief/.

Cooper, Zack, Stuart V Craig, Martin Gaynor, and John Van Reenen. "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured." *The Quarterly Journal of Economics* 134 no. 1 (2019): 51-107. https://doi.org/10.1093/qje/qjy020

Dauwerse, Linda, Froukje Weidema, Tineke Abma, Bert Molewijk, and Guy Widdershoven. "Implicit and Explicit Clinical Ethics Support in the Netherlands: A Mixed Methods Overview Study." *HEC Forum* 26 (2013): 95-109.

Greer, Joseph A, Vicki A Jackson, Diane E Meier and Jennifer S Temel. "Early Integration of Palliative Care Services with Standard Oncology Care for Patients with Advanced Cancer." *CA; A Cancer Journal for Clinicians* 63 no. 5 (2013): 349-63.

Haines, William. *The Internet Encyclopedia of Philosophy,* s.v. "Consequentialism." Edited by James Fieser and Bradley Dowden, ISSN 2161-0002, https://iep.utm.edu/conseque/, accessed February 6, 2018.

Hastie & Dawes *Rational Choice in an Uncertain World* 2nd ed, 141-143, SAGE Publications, Inc, 2009 Thousand Oaks, CA

Marcus, Brian S, Gary Shank, Jestin N Carlson, and Arvind Venkat. "Qualitative Analysis of Healthcare Professionals' Viewpoints on the Role of Ethics Committees and Hospitals in the Resolution of Clinical Ethical Dilemmas." *HEC Forum* 27 (2015): 11-34.

McCradden, Melissa, Shalmali Joshi, Mjaye Mazwi, and James A Anderson. "Ethical limitations of algorithmic fairness solutions in health care machine learning." *The Lancet Digital Health* 2 no. 5 (2020): e221-e223. https://doi.org/10.1016/S2589-7500(20)30065-0

Miller, Susan. "Providing Palliative Care and Beyond…" *Journal of Palliative Medicine* 19 no. 10 (2016): 1022–1024. doi:10.1089/jpm.2016.0394

Miori, Virginia M and Daniel J Miori. (2010). "Evaluating Survival Likelihoods in Palliative Patients Using Multiple Criteria of Survival Rates and Quality of Life." *Advances in* Business *and Management Forecasting*. 7 (2010): 221-238. https://doi.org/10.1108/S1477-4070(2010)0000007018

Pastrana, Tania, Lukas Radbruch, Friedemann Nauck, Gerhard Höver, Martin Fegg, Martina Pestinger, Josef Roß, et al. "Outcome indicators in palliative care-how to assess quality and success. Focus group and nominal group technique in Germany." *Support Care Cancer* 18 no. 7 (2010):859-868.

Rockwood, Kenneth, Xiaowei Song, Chris MacKnight, David B Hogan, Ian McDowell and Arnold Mitnitski. "A Global Clinical Measure of Fitness and Frailty in Elderly People." *Canadian Medical Association Journal*. 173 no. 5 (2005): 489-495. https://doi.org/10.1503/cmaj.050051

Rodriguez, Carlos J., Matthew Allison, Martha L. Daviglus, Carmen R. Isasi, Colleen Keller, Enrique C. Leira, Latha Palaniappan, et al. "Status of Cardiovascular Disease and Stroke in Hispanics/Latinos in the United States." *A Science Advisory from the American Heart Association.* 130 no. 7 (2014):593-625. DOI: 10.1161/CIR.0000000000000071

Sivashanker, Karthik, Cheri Couillard, Jennifer Goldsmith, Normella Walker and Sunil Eappen. "Addressing the caste system in U.S. healthcare in the era of COVID-19." *International Journal for Equity in Health* 19 no. 184 (2020). https://doi.org/10.1186/s12939-020-01298-x

Smith, Thomas J, Sarah Temin, Erin R Alesi, Amy P Abernathy, Tracy A Balboni, Ethan M Basch, Betty R Ferrell et al. "American Society of Clinical Oncology Provisional Clinical Opinion: The integration of Palliative Care Into Standard Oncology Care." *Journal of Clinical Oncology* 30 no. 8 (2012): 880-887.

Snow, Richard, Karen Vogel, Bruce Vanderoff, Benjamin P. Kelch, and Frank D Ferris. "A prognostic indicator for patients hospitalized with heart failure." *Journal of Palliative Medicine* 19 (2016):1320-1324. https://doi.org/10.1089/jpm.2015.0531

Uneno, Yu, Kei Taneishi, Masashi Kanai, Kazuya Okamoto, Yosuke Yamamoto, Akira Yoshioka, Shuji Hiramoto et al., "Development and validation of a set of six adaptable prognosis prediction (SAP) models based on time-series real-world big data analysis for patients with cancer receiving chemotherapy: a multicenter case crossover study." *PLoS ONE* 12 no. 8 (2017). https://doi.org/10.1371/journal.pone.0183291

Vargas Bustamante, Arturo, Hai Fang, Jeremiah Garza, Olivia Carter-Pokras, Steven P. Wallace, John A. Rizzo, and Alexander N. Ortega. "Variations in Healthcare Access and Utilization Among Mexican Immigrants: The Role of Documentation Status." *Journal of Immigrant Minority Health*. 14 no. 1 (2012):146–155. DOI 10.1007/s10903-010-9406-9

Von Gunten, Charles F., Joan M. Teno, and R. Sean Morrison. "Big Data and End-of-Life Care: Promise and Peril." *Journal of Palliative Medicine* 19 no. 12 (2016):1240. https://doi: 10.1089/jpm.2016.0374

Weinstein, James N, Amy Geller, Yamrot Negussie, and Alina Baciu. *Communities in Action: Pathways to Health Equity.* National Academies of Sciences, Engineer, and Medicine. Washington, DC: The National Academic Press, 2017. DOI: https://doi.org/10.17226/24624

Yao, Yingwei, Janet Stifter, Miriam O Ezenwa, Muhammad Lodhi, Ashfaq Khokhar, Rashid Ansari, Gail M Keenan, and Diana J Wilkie. "Infomarkers for Transition to Goals Consistent with Palliative Care in Dying Patients." *Palliative & Supportive Care* 13 no. 5 (2015): 1427-1434.