

# Health and Health Care Inequalities: Research Challenges and Considerations for Future Research

May | 2023



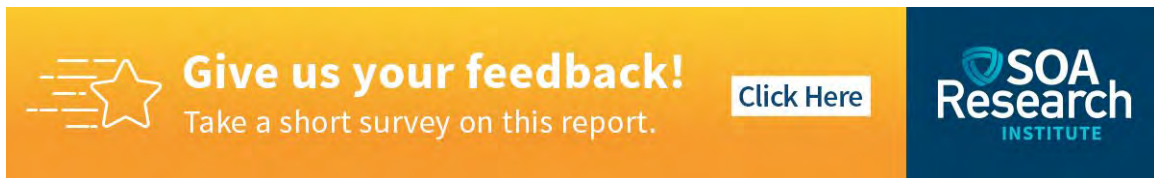


# Health and Health Care Inequalities:

## Research Challenges and Considerations for Future Research

**AUTHORS** Steve Bochanski, FSA, MAAA  
Vicki A. Fendley, ASA, FCAS, MAAA  
Sourabh Prasad  
Jayeeta Sen Mazumder

**SPONSOR** Diversity, Equity and Inclusion Research  
Advisory Council

A horizontal banner with a yellow-to-orange gradient background. On the left is a white star icon with horizontal lines extending from its left side. To the right of the star is the text "Give us your feedback!" in a bold white font, followed by "Take a short survey on this report." in a smaller white font. Further right is a white rectangular button with the text "Click Here" in blue. On the far right of the banner is the SOA Research Institute logo, which consists of a blue shield icon and the text "SOA Research INSTITUTE" in white and blue.

### **Caveat and Disclaimer**

The opinions expressed and conclusions reached by the authors are their own and do not represent any official position or opinion of the Society of Actuaries Research Institute, the Society of Actuaries or its members. The Society of Actuaries Research Institute makes no representation or warranty to the accuracy of the information.

Copyright © 2023 by the Society of Actuaries Research Institute. All rights reserved.

## CONTENTS

<b>Executive Summary .....</b>	<b>4</b>
<b>Introduction.....</b>	<b>6</b>
<b>Section 1: Research Challenges .....</b>	<b>Error! Bookmark not defined.</b>
1.1 Data .....	<b>Error! Bookmark not defined.</b>
1.1.1 Data Availability and Evaluation .....	<b>Error! Bookmark not defined.</b>
1.1.2 Data Granularity.....	<b>Error! Bookmark not defined.</b>
1.2 Selecting a Statistical Approach .....	<b>Error! Bookmark not defined.</b>
1.3 Data Considerations for Modeling.....	<b>Error! Bookmark not defined.</b>
1.3.1 Pre-Processing.....	<b>Error! Bookmark not defined.</b>
1.3.2 Exploratory Data Analysis (EDA) .....	<b>Error! Bookmark not defined.</b>
<b>Section 2: Future Research.....</b>	<b>Error! Bookmark not defined.</b>
2.1 Selecting Data Variables for Research .....	<b>Error! Bookmark not defined.</b>
2.1.1 Race and Ethnicity Data .....	<b>Error! Bookmark not defined.</b>
2.1.2 Health and Health Care Data .....	<b>Error! Bookmark not defined.</b>
2.1.3 Financial Wellness Definitions and Data.....	<b>Error! Bookmark not defined.</b>
2.1.4 Building a Dataset .....	<b>Error! Bookmark not defined.</b>
2.2 Topics to Consider for Future Research.....	<b>Error! Bookmark not defined.</b>
<b>Section 3: Acknowledgments .....</b>	<b>Error! Bookmark not defined.</b>
<b>References.....</b>	<b>25</b>
<b>Appendix A: Specific Considerations for a Linear Regression Approach to Health and Health Care Research.....</b>	<b>Error! Bookmark not defined.</b>
A.1 Selection of Demographic Variables.....	31
A.2 Income Bands as a Proxy for Socioeconomic Groups.....	31
A.3 Base Categories for Regression Analysis.....	32
A.4 Goodness of Fit and Multicollinearity.....	33
<b>Appendix B: Literature Review.....</b>	<b>35</b>
B.1 Financial Wellness Inequalities .....	35
B.2 Health Status/Outcomes and Health Care Quality/Access Inequalities .....	36
B.3 Health care Coverage and Cost Inequalities .....	44
B.4 Chronic Conditions .....	48
<b>About The Society of Actuaries Research Institute .....</b>	<b>49</b>

# Health and Health Care Inequalities:

## Research Challenges and Considerations for Future Research

### Executive Summary

The existence of racial and ethnic disparities in health care is well documented in the United States. Past research has focused on social determinants of health (SDOH), or “non-medical factors that affect health outcomes” (WHO, Health topics), in addition to health and health care access and quality. More recently, the impacts of the COVID-19 pandemic have highlighted the unequal access to the U.S. health care system, which is closely tied to health care coverage.

The Society of Actuaries (“SOA”) Research Institute, an education and research organization dedicated to enabling the provision of expert advice and relevant solutions for financial, business and societal problems involving risk, engaged us to objectively assess the impact of disparities in health care access, cost and quality of care on financial wellness across racial, ethnic, socioeconomic and other demographic groups in the U.S. for reasons other than or in addition to income, wealth or employment status. Better understanding the association and impact of health and health care with financial wellness and other demographic characteristics may provide insights into potential solutions. However, applying statistical analysis to study these topics robustly presents significant challenges that we are unable to overcome at this time. This research paper describes some of those challenges, our learnings, and alternative considerations that researchers may build upon in conducting future research.

Perhaps the greatest challenge is obtaining the necessary data. To understand racial and ethnic disparities in health and health care and facilitate robust interpretation of statistical analysis results, it is important to have individual-level data or, at a minimum, geographic-level data that provides the cross-section of race, ethnicity and other variables being studied (e.g., prevalence of each racial and ethnic group that has health care coverage). However, access to individual-level data is complicated by the highly restrictive Health Insurance Portability and Accountability Act (HIPAA) (CDC, 2018) and a lack of consistent collection of race and ethnicity data by health insurers. In Section 1.1.1: Data Availability and Evaluation, we suggest ten key questions to facilitate researchers’ understanding of the limitations of a dataset prior to undertaking a research project. In particular, it is important to understand the available data variables, their completeness, their level of granularity, and whether the variables contain actual or modeled data.

After identifying one or more datasets and gaining a good understanding of the available variables and their characteristics through exploratory data analysis, researchers will need to select a statistical approach that aligns with the selected research question. Researchers may consider the pros and cons of various approaches, including weighing the use of powerful, modern techniques against the explainability of model results, any limitations of the data, and the amount of time available to conduct the research. It is important to consider possible bias in any model or dataset, particularly when racial and ethnic disparities are being studied.

As we turn our focus to future research, we consider the complexities associated with health and health care—health care coverage, health care costs and affordability, health care access, health care quality, health status, chronic disease, and health outcomes. We also consider financial wellness, using both objective and perception measures, as well as race and ethnicity variables. When studying racial and ethnic disparities, it is imperative to understand the definitions of race and ethnicity in a dataset and how race and ethnicity data is captured (e.g., whether it is self-reported by an individual or based on another person’s perception). It is also important to be cognizant that race is a social construct and to consider the historical context of racial and ethnic experiences and the influence of societal systemic racism in performing this type of research. However, a discussion of the history of

societal systemic racism in the U.S. is beyond the scope of this paper. In this report, we consider various definitions of race and ethnicity and use the 2020 U.S. Census definitions as a framework for discussion. The sensitive nature of racial and ethnic disparities warrants applying careful scrutiny to any research and the associated results.

It is important to consider up-front whether a statistical analysis will yield information on correlation or causality and whether the available data may potentially produce biased results. To study causality (e.g., whether having health insurance impacts financial wellness or vice versa), a dataset must include data at various points in time (i.e., time series data). If a complete dataset—with sufficient sample size and a relevant sample population with a racial, ethnic and other demographic composition aligned with the total U.S. population—is not available for the required variables at a sufficient level of granularity, a new dataset may need to be built. Further, it is important to consider the time span covered by the data used in a statistical analysis and whether one time-frame is relevant to another. A recent example is the COVID-19 pandemic. Data in 2020 and subsequent may not be directly comparable to prior data, and researchers may want to consider these impacts when interpreting the results of a statistical analysis.

We conclude our paper with a list of topics that may serve as a springboard for future research. Future researchers may want to adjust these topics based on the available dataset or identify new topics that are ripe for research. Without the necessary data at a sufficient level of granularity, it will be impossible to obtain robust results from statistical analysis. If the data challenges can be overcome and used to inform the selection of a research question and the type of statistical analysis, future research can provide meaningful insights into racial and ethnic disparities in health and health care.



**Give us your feedback!**

Take a short survey on this report.

[Click Here](#)

**SOA**  
**Research**  
INSTITUTE

## Introduction

Health inequalities have been well documented as a serious problem in the U.S., beginning with the 1985 Malone-Heckler report, which found that Blacks or African-Americans and other marginalized groups had significantly higher mortality rates for leading causes of death (Riley, 2012). Over the last 35 years, research has also focused not only on health, but also health care access and social determinants of health (SDOH), “nonmedical factors that affect health outcomes” (WHO Health topics), including socioeconomic and demographic factors. In the last decade, 38 states and the District of Columbia have expanded Medicaid to adults based on income eligibility alone, which has served to reduce racial and ethnic disparities in health care coverage, although inequalities persist in the remaining states that have not expanded Medicaid eligibility (KFF, June 2022).

More recently, the COVID-19 pandemic has highlighted the unequal access to the U.S. health care system, most notably for American Indians and Alaska Natives, Hispanics and Latinos, Blacks and African-Americans and Native Hawaiians and Pacific Islanders, which exacerbates health inequalities (Hill et al., 2022). As noted in prior research, the lack of access by these groups is closely tied to being uninsured for health care costs and results in “fewer preventive services, poorer health outcomes, higher mortality and disability rates, lower annual earnings because of sickness and disease, and the advanced stage of illness (i.e., many are “sicker” when diagnosed),” which ultimately contributes to higher health care costs (Riley, 2012).

Racial and ethnic disparities in health outcomes, health care access and chronic conditions are intimately related to health care coverage and affordability of health care. Understanding the relationship between these items and socioeconomic factors, such as income level and overall financial wellness, may help identify the underlying causes. With an estimated annual cost of \$35 billion related to the excess health care costs associated with racial health disparities (Ayanian, 2015), finding solutions to address the underlying causes of these racial and ethnic disparities may reduce overall health care system costs.

Significant challenges exist in conducting research regarding these relationships and, in particular, disparities across racial and ethnic groups. Interrelationships with demographic factors and other SDOH further confound the research efforts. In this paper, we discuss some of the challenges in performing research around inequalities in health, health care access and quality, and health care coverage and costs, and we provide suggestions for future research. When designing a statistical approach, selecting data, conducting health and health care research and presenting results, it is important to be cognizant of the sensitivities that exist around health and health care topics and the associated racial and ethnic disparities.

## Section 1: Research Challenges

Studying racial and ethnic inequalities in health status, outcomes, chronic disease, health care coverage and costs, and health care access and quality presents a number of challenges. From defining the relevant measures to identifying a dataset and designing a modeling approach, we highlight many of the challenges researchers may face. Throughout this paper, unless otherwise noted, we use “health and health care” to refer to some or all of the following: health care coverage, health care costs, health status, health outcomes, chronic disease, health care access and health care quality.

### 1.1 DATA

It is often said that the three most important considerations in real estate are “location, location, location.” With respect to health and health care inequalities research, we suggest the three most important considerations are “data, data, data.” In statistical analysis, it is not uncommon to make concessions due to availability, multicollinearity, granularity of available data, and the ability to validate the data against publicly available sources. However, even seemingly insignificant compromises throughout a research project can impact an analysis, and in particular, the interpretation and understanding of observed results. In the health and health care space, these considerations can be significant due to limitations on available health care data due to the Health Insurance Portability and Accountability Act (HIPAA) (CDC, 2018) and a lack of consistency in the collection of race and ethnicity data by health insurers.

Understanding the relationship between health and health care and financial wellness provides additional areas for research. However, coupled with restrictions on the availability of personal financial data, obtaining data related to the intersection of health and health care topics and financial wellness is difficult.

Because of these limitations, undertaking meaningful, new research regarding racial and ethnic disparities in health and health care and the relationship with financial wellness requires a substantial effort to identify data at a sufficient level of granularity. Further, understanding the potential challenges, pitfalls and limitations of one or more datasets, including whether multiple datasets can be joined together is critical in undertaking this type of research.

#### 1.1.1 DATA AVAILABILITY AND EVALUATION

Oftentimes, researchers will identify a question they want to analyze and then set out to find relevant data to perform this research. However, the severe limitations in health and health care and financial data may warrant reversing the order, i.e., identifying one or more robust data sets and then determining meaningful research topics that the data may be used to study. In fact, identifying privately available datasets with a variety of key variables at sufficient granularity would be invaluable as a foundation for much important health and health care research. While public datasets often lack sufficient granularity and may not include all of the relevant variables, private datasets may require licensing and contracting, which can extend the research timeline, and may only be available at a cost.

Table 1 includes a representative list of readily available public data sources that can be leveraged to perform health and health care research. While easily accessible, gaining an understanding of the data included in these sources and the potential shortfalls of the data prior to the beginning of a project will facilitate the design of a research project. As an example, the health care cost data from the U.S. Census Bureau—Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS)—excludes most rural and some suburban areas, which can produce potentially misleading or biased results when studying inequalities, because individuals in rural areas tend to have lower income and access health care less frequently (Rural Health Information Hub, Rural Health Topics).

**Table 1**  
**READILY AVAILABLE, PUBLIC DATA SOURCES**

Data Source	Description	Availability/ Accessibility
<b>U.S. Census Bureau</b> <ul style="list-style-type: none"> <li>• Decennial Census</li> <li>• 2019 American Community Survey (ACS)</li> </ul>	The 2019 ACS includes demographic estimates from the 2010 U.S. Census projected to 2019 through the use of interim, smaller sample surveys. The U.S. Census data includes modeling to roll forward the more complete decile information for each decade using sample survey data. This data is available at a census block group (CBG) level or lesser granularity. (U.S. Census Bureau, ACS).	Publicly available application programming interface (API)
<b>U.S. Census Bureau</b> Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS)	Compiled by the U.S. Census Bureau, the CPS provides information on the U.S. labor force. The ASEC, a survey of more than 75,000 households, includes detailed social and economic characteristics of each person. This database contains out-of-pocket health care costs and premiums at a core-based statistical area (CBSA) level, which includes most urban areas and some suburban areas. (U.S. Census Bureau, ASEC).	Publicly available application programming interface (API)
<b>Centers for Disease Control &amp; Prevention (CDC)</b> PLACES	PLACES, Local Data for Better Health, contains the modeled prevalence for several chronic conditions at various geographic levels as well as a number of other health and health care data variables. This data is available at a U.S. census tract or lesser level of granularity (not available at a CBG level). (CDC, PLACES)	Publicly available application programming interface (API)

Whether using a public or private data source, reviewing the data dictionary is a good first step prior to beginning a research project. However, because data dictionaries and other detailed information may not be readily available, researchers may need to inquire about the data components and characteristics prior to contracting for or purchasing full access to the data. Some important questions to consider include:

1. What variables are included in the dataset?
2. Is individual-level data available? If so, what data field(s) are provided that will allow the data to be joined together with other datasets if needed?
3. If individual-level data is not available, at what level of granularity are these variables available (household, census block group, census tract, ZIP-11, ZIP-9, ZIP-5, county, etc.)? Is the available data free from geographic sampling bias?
4. Is actual data provided for each variable or is it modeled or imputed from available data? If modeled or imputed, what is the algorithm? What data is considered in determining the imputed amounts?
5. What size is the population considered in the dataset? Was the data obtained from a sample of a larger population? Does the population include adults over age 18 only, adults over age 65, children, etc.? Are there other limitations to the population, e.g., limited to a particular state's population?
6. How complete is the data, i.e., is data missing for a portion of the population? If so, what data is missing and what percent of the population does this represent?
7. Is there sufficient distribution of the responses as demonstrated in univariate distributions (i.e., most responses should not be the same value)?
8. Are the variables discrete or continuous? Are the variables qualitative or quantitative? What are the minimum and maximum values? How many possible values are available for each bucket?
9. Has the data been validated against external sources, e.g., if the dataset is a sample of the U.S. population, has the publisher confirmed that the racial and ethnic distribution and other demographic characteristics of the dataset population are similar to the overall U.S. population?
10. For which years is the dataset available? How do the population and individuals in the dataset compare at each point in time? Are the same data variables available for each point in time?

To glean insights into causality, time-series data is needed to conduct a longitudinal study. For example, to study the impact of having a high deductible health plan (HDHP) on health care access or health outcomes, a largely similar



database would need to be available across multiple time periods. Without such data, only relationships or associations between variables may be studied. Time-series data would enable research to determine, for example, whether having an HDHP causes decreased financial wellness, or whether financial wellness causes individuals to elect an HDHP because it has lower premiums. Without such longitudinal data, using existing literature to supplement statistical analysis may help shed light on causality.

When looking at recent data or data over time, it will be critical to consider the impact of the global COVID-19 pandemic, including potentially disparate impacts of the pandemic on different demographic groups. The pandemic has had a significant impact on health, mortality, health care, employment and health care coverage, which will add complexities to research in the foreseeable future.

### 1.1.2 DATA GRANULARITY

Publicly available data typically is aggregated to a level that protects the privacy of the underlying population, which limits the ability to study racial and ethnic disparities in the health care space. Data aggregation occurs largely to comply with HIPAA, federal legislation enacted in 1996 that addresses a variety of health insurance and health care issues, including the privacy, security and electronic exchange of individual health information or protected health information (PHI) (CDC, 2018). Private datasets provide an alternative, although any single source of data, whether public or private, is unlikely to contain all of the relevant information needed to conduct in-depth research in this space.

As implied by the considerations set forth in Section 1.1.1, individual-level data may be preferred for health care-related research. But why? Let's consider U.S. Census data, which is publicly available at a census block group (CBG) level at its most granular. CBG data provides the prevalence of, for example, health care coverage, race and ethnicity, and completion of a bachelor's degree in each CBG. However, CBG data does not enable conclusions to be drawn at an individual level. To study these relationships in a robust manner, individual-level data would be required to facilitate the researcher's understanding of the characteristics of individuals that have an HDHP, for example, including their level of education, income, race and ethnicity, sex, age, marital status and other demographic factors.

Tying health and health care to financial wellness would provide interesting topics for research. However, similar challenges exist with respect to obtaining data on financial wellness, for which data generally is unavailable at an individual-level, but instead as medians or means across geographic groups, which tends to mask heterogeneity within the geographic unit. For example, the median household income in a CBG and the proportion of Hispanics or Latinos in the CBG are known, but the median household income of Hispanics or Latinos in a CBG is not known. The issues are exacerbated when studying health status, outcomes, chronic diseases, health behaviors or attitudes, health care access and quality, and health care coverage and costs, which are individual experiences. Individual-level data can be aggregated or grouped in a more appropriate way before being studied, which may differ from a geographic unit.

Household-level data can be meaningful for use in research, particularly for financial measures or health care coverage, but it raises questions when studying health or demographics. Assigning race and ethnicity to a household where every member of the household is Asian or Asian American is simple, but how can race and ethnicity be assigned to a household that includes individuals of different races and ethnicities? Similar complexities exist around demographic factors such as limited English proficiency, sex, age, educational attainment level and other factors.

While individual data may be preferred for the reasons cited above, it comes with its own limitations. Specifically, actual data is often not known for every individual in a sample population. As such, values must be modeled or imputed from other known information using modeling algorithms to complete the dataset. In addition, individuals self-report much of their own data, potentially subjecting the data to inaccurate reporting stemming from

incomplete or incorrect memories. Depending on the source(s) of any individual data, it may be biased toward a particular segment of the population, e.g., wealthier individuals or particular races and ethnicities and not representative of the U.S. population. As such, validating the data with aggregated external sources is critical as discussed previously.

One significant constraint in performing this type of research is the availability of minimally modeled, individual-level data. The collection of data at an individual-level that does not compromise privacy requirements will enable robust statistical analysis. Currently, race and ethnicity information is not consistently gathered by insurers, although more insurers have begun gathering this data. In addition, while the U.S. Census Bureau and other government entities gather race and ethnicity data, individual-level data is not publicly available. If the data were cleansed of personally identifiable information, it could be made accessible to researchers to enable future research focused on racial and ethnic disparities in health and health care.

Ultimately, research findings will be limited by the granularity level of the least granular data variable. If a dataset is identified that contains individual-level data, research may still be limited if some data variables must be pulled from another dataset and those variables are only available at a lesser level of granularity, e.g., CBG. For example, income, educational attainment, and other demographic variables may be available at an individual-level, but the type of health care coverage may only be available at a CBG level. Analysis using such a dataset can be robust only if the proportion of each racial and ethnic group who has each type of health care coverage is available by CBG. Without that information, the results of statistical analysis will be weaker at best and misleading at worst.

Alternatively, if a significant portion of the individual data has been modeled or imputed, aggregating the data to a geographic level can reduce the high margin of error associated with modeled and imputed variables at an individual level.

Where data may not be available for the U.S. population at an individual level (approximately 335 million) or household level (approximately 120 million) (US Census Bureau, Quick Facts), geographical-level data may be available at the following granularities (from most to least granular): ZIP9 (approximately 45 million), U.S. Census block (220,000), U.S. Census tract (71,000), and ZIP (43,000) (U.S. Census Bureau, Tallies). Generally, statistical analysis that uses more granular variables, while minimizing the amount of modeled data will product more meaningful insights.

## 1.2 SELECTING A STATISTICAL APPROACH

When performing research related to racial and ethnic inequalities in health and health care, it may be prudent to apply additional scrutiny in the selection of a statistical approach and the underlying data because of the complexities associated with the definitions of race and ethnicity and the various health and health care variables coupled with the controversial nature of these topics. This section includes a pedagogical approach to selecting a statistical model, and similarly, Section 1.3 discusses data considerations for statistical modeling. These discussions are in the context of performing research around racial and ethnic inequalities, particularly related to health and health care.

Once the data is selected, including consideration of the source, granularity, time-period (cross-sectional or time-series), and sample size, researchers may define the objective of the analysis, refine the research question, and select the type of statistical or machine learning (ML) model to employ. The research objective may involve but is not limited to: quantifying relationships between variables; estimating if measured relationships are statistically significant; measuring the degree of association; finding patterns or look-alikes using predictive analytics; and forecasting. Depending on the research objective, most common algorithms may be broadly classified as supervised, unsupervised, semi-supervised and reinforcement learning, as outlined in Table 2.

**Table 2**  
**COMMON ALGORITHMS FOR CONSIDERATION IN HEALTH AND HEALTH CARE RESEARCH**

Algorithm	Description
Supervised Learning	<p>As the name indicates, supervised learning algorithms refers to models that identify patterns, learning from observations and making predictions, while being “supervised” or “guided” by the researcher regarding the correct outputs (“labels”). The “supervision” helps the model learn the correct responses and make better predictions, until it produces results with the desired level of accuracy or performance. Supervised learning algorithms further comprise generalized linear models and classification models.</p> <p><i>Generalized Linear Models (GLM):</i> GLMs, including regressions, involve estimating a relationship between one or more predictor variables and an explained or dependent variable. These may be linear or non-linear when the dependent variable is continuous, or logistic if binary<sup>1</sup>. Through hypothesis testing, regressions enable one to test whether the relationship between two variables is statistically significant or not, and to quantify the change in the dependent variable resulting from a change in one or more independent variables. Regressions can also be used for estimating, i.e., predicting the output values for new data based on the estimated relationships built on existing observations.</p> <p><i>Classification Models:</i> These models try to predict to which class or category new observations belong based on observed values of existing features. Some common types of classification models include logistic regression, Naive Bayes Classifier, Support Vector Machines (SVM), Random Forest, Decision Trees, and Neural Networks.</p>
Unsupervised Learning	In scenarios where the output or the “label” is not known, there is no way to “supervise” a function. In unsupervised learning, the model is allowed to work on its own to discover patterns, mine for rules, summarize and group the data points and derive information that was previously undetected. Examples of unsupervised learning include clustering, principal component analysis (PCA), and rules of association.
Semi-supervised Learning	Semi-supervised learning uses both labeled and unlabeled data. Since unlabeled data can be relatively less expensive to acquire than validated labeled data, scenarios can exist where the data can contain a very small amount of labeled data and a very large amount of unlabeled data. While labeled data helps the algorithm to understand the data, unlabeled data also provides meaningful information to the algorithm, which can help make predictions on the unlabeled data. Applications of semi-supervised learning include Natural Language Processing (NLP), text or image classification, clustering and speech analysis.
Reinforcement Learning	Reinforcement learning algorithms continuously learn from the environment, by taking actions that would maximize the reward (“reinforcement signal”) or minimize the risk, until they explore all possible states. Common reinforcement learning algorithms include Artificial Neural Networks (ANN) and Q-Learning.

In addition to the importance of the size and nature of the data as previously discussed, when choosing a model, researchers may want to consider the explainability (interpretation), performance,<sup>2</sup> and complexity of the model as well as training time and dimensionality of data. In general, there is a trade-off between explainability and model complexity (Shmueli, 2010). In fact, explainability may be considered to have an inverse relationship with model complexity—the more complex the model is, the harder it will be to explain its results. For example, random forests and neural networks are complex models involving many hyperparameters that produce results that may not be intuitive and are difficult to explain. When model explainability is important, regression models and decision trees are good candidates. In addition, while the inclusion of more features (high dimensionality) may produce a model that provides a higher level of accuracy, i.e., increased model fit or adjusted  $R^2$ , it will increase the complexity of a model; the researcher may want to consider whether the benefits outweigh the costs.

<sup>1</sup> Other kinds of regression models, including, but not limited to, ridge, LASSO, elastic net, polynomial regressions, may also be considered. An in-depth discussion of these approaches is beyond the scope of this report; these are mentioned for completeness.

<sup>2</sup> A high performing model may refer to a model that predicts well, fits well to the data, trains quickly, finds a robust solution, or any combination of these. Determining whether a model performs “well” is based on consideration of model results relative to the research goals, and other models trained on the same data; as such, model evaluation metrics differ from algorithm to algorithm. For example, a determination that a model performs “well” may be based on achieving a lower prediction error for a classification or regression model, while for clustering, a higher level of dissimilarity between clusters produced by one algorithm relative to alternative clustering algorithms.

Regression analysis is an important statistical tool when studying the relationship between a set of independent variables and a dependent variable, providing results that may be interpreted and explained. Where it is important to study the impact of individual features, such as race and ethnicity or other demographic factors, using a multiple linear regression approach may be appropriate as it describes and quantifies linear relationships, while also ensuring model explainability (see Appendix A for additional considerations regarding regression modeling). In addition, time series models such as vector autoregressive (VAR) models may use time-series data to test possible causality. Unlike multivariate regression models, many advanced techniques involving machine learning and artificial intelligence (AI), for example, are complex “black box” models that make it difficult to understand why certain relationships exist.

Some alternatives to regression models may provide useful insights for health and health care related topics. One example is clustering, which has been used to provide meaningful results around SDOH using geographical clusters (Barrett and Entzminger, 2022). Advanced modeling techniques may be used to determine particular categories to cluster, and once those groupings are determined, clustering provides meaningful results. Clustering may be used to group data variables in other ways as well. For example, clusters may be used to determine characteristics of individuals with different types of health care coverage. However, for Medicaid and Affordable Care Act (ACA) marketplace coverage, differences in eligibility rules (Medicaid) and incentives (ACA) by state render geography a primary consideration.

Many other types of statistical analysis approaches may be useful in researching health and health care topics, although a complete review of possible approaches is beyond the scope of this paper. Consideration of the benefits and shortcomings of each approach in light of the available data can facilitate the selection of an approach deemed to provide the greatest net benefit. It may be helpful for researchers to disclose any shortcomings and discuss the impact on results. In particular, when studying racial and ethnic disparities, it is important to disclose whether the disparities may be understated or overstated as a result of the limitations of the modeling approach and the data.

### 1.3 DATA CONSIDERATIONS FOR MODELING

This section includes a pedagogical approach to evaluating data for use in a statistical model designed to study racial and ethnic inequalities, particularly related to health and health care. As discussed above, many statistical analysis approaches may be used to conduct research regarding racial and ethnic disparities in health and health care. After identifying key data sources available for modeling, the first step is to review the data to identify issues of completeness, collinearity and correlation across variables and select the final data to include in a statistical analysis. In particular, researchers may perform preliminary data analysis, including pre-processing and exploratory data analysis (EDA).

#### 1.3.1 PRE-PROCESSING

Researchers may consider the following initial steps to pre-process the data:

1. *Scoping the analysis:* Identify a set of available variables from various data sources at different levels of granularity that could be used to study the specific research questions.
2. *Identifying data gaps:* Review the preliminary list of available variables to identify additional data needed for the statistical analysis.
3. *Creating a data fabric:* Create a master data model by joining data sources together. For example, for a CBG-level regression, census tract-level variables would be assigned to all CBGs within the census tract.

### 1.3.2 EXPLORATORY DATA ANALYSIS (EDA)

Exploratory data analysis is an important step in better understanding the extracted data; it may involve some or all of the following:

**Summarizing the extracted data:** Summarize the distribution of each variable by the possible values and calculate weighted averages, where appropriate, to validate the data against externally published sources.

1. *Feature engineering:* Based on the study objectives and preliminary data analysis, researchers may transform raw data into desired input features. Feature engineering compatible with the model's requirements can improve the model's performance and efficiency and as such, can play a big role in determining the outcome of the model. Common feature engineering techniques include dealing with missing values (e.g., through imputation or dropping observations); handling outliers; binning (i.e., creating groups that capture outliers that would otherwise skew results); log transformations; handling categorical variables (e.g., one hot encoding to add a dummy, binary variable for all categorical variables in a regression) and feature scaling (e.g., normalization or standardization). For example, it is well-known that income is a highly-skewed variable. A log transformation of income may be used in a regression analysis to remove this skew and improve model fit. However, because regression coefficients on log-transformed variables are complex to interpret, it may be useful to transform the resulting regression coefficients using  $e^{-1}$  to reflect their percentage impact on the dependent variable. Modern datasets increasingly tend to have a large number of features, and this high "dimensionality" can invite its own set of issues. For example, there could be significant correlations among variables or variables can be difficult to inspect individually. It then becomes necessary to reduce the feature set to a more manageable size, or to identify latent driving factors, while trying to retain the information carried by the attributes. In such cases, techniques such as principal component analysis (PCA), exploratory factor analysis, and confirmatory factor analysis are useful tools employed by data scientists for feature reduction or feature aggregation. However, feature aggregation methods transform the actual variables into factors or principal components, and as such, the variables are not easy to interpret since the original underlying features cannot be isolated.
2. *Running correlation analysis:* Derive a correlation matrix to provide quantitative information on collinearity among independent variables, among dependent variables, and between independent and dependent variables. This information may be used to inform the selection of the final variables included in a statistical analysis. Selecting appropriate features reduces the dimensionality of data, which results in lower complexity of the model and a reduced chance of overfitting.
3. *Univariate and Bivariate analyses:* Using univariate and bivariate analysis, researchers may review the data variables for collinearity and correlation to further refine the final variables for inclusion in the statistical analysis. For example, this step may include looking at the minimum and maximum values and quantiles in univariate analysis of the variables to demonstrate sufficient variation between geographic units and align with those from a published, publicly available source such as the U.S. Census. The benefits of pursuing a thorough analysis of model variables are manifold. For instance, many prevalent algorithms are data-hungry, and "feeding" variables identified in the univariate analysis as being highly skewed or not demonstrating meaningful variation will likely hinder the performance of the model. Reviewing bivariate analysis of pairings of the remaining variables can help identify those with high collinearity.

## Section 2: Future Research

Despite the significant challenges, performing meaningful health and health care research is critical to identify relationships with and/or causes of racial and ethnic disparities and address the significant costs to the overall health care system. In pursuing research around these topics, it is important to be mindful of the key issues and determine whether it is acceptable to proceed in light of any limitations in the data or statistical analysis. Ultimately, it may be necessary to develop a database of individual-level data to enable robust health and health care research.

### 2.1 SELECTING DATA VARIABLES FOR RESEARCH

Selection of data variables to include in a statistical analysis is an important step. In particular, selecting relevant variables for researching racial and ethnic inequalities in health and health care and the relationship between these variables and financial wellness is rampant with challenges. We touched upon many of the data challenges in Section 1.1, particularly regarding the granularity of the data, but additional considerations around race and ethnicity, health and health care, and financial wellness variables warrant further discussion.

#### 2.1.1 RACE AND ETHNICITY DATA

Some people may perceive race as being clearly defined, but it is far from it. As such, studying societal racial and ethnic disparities introduces many complexities in research. To illustrate the range of perspectives on the definition, meaning and usage of race, we highlight definitions from several sources:

- According to Merriam-Webster, “the term race is understood today as primarily a sociological designation that identifies a group sharing some outward physical characteristics and some commonalities of culture and history, while ethnicity is a word for something you acquire based on where your family is from and the group which you share cultural, traditional, and familial bonds and experiences with” (Merriam-Webster, usage notes).
- The National Human Genome Research Institute of the National Institute of Health (NIH) provides another definition: “Race is a social construct used to group people. Race was constructed as a hierarchical human-grouping system, generating racial classifications to identify, distinguish and marginalize some groups across nations, regions and the world. Race divides human populations into groups often based on physical appearance, social factors and cultural backgrounds.” (NIH, Race, July 2022).
- The Smithsonian provides yet another view: “Although race has no genetic or scientific basis, the concept of race is important and consequential. Societies use race to establish and justify systems of power, privilege, disenfranchisement, and oppression. The notion of race is a social construct designed to divide people into groups ranked as superior and inferior. The scientific consensus is that race, in this sense, has no biological basis—we are all one race, the human race. Racial identity, however, is very real. And, in a racialized society like the United States, everyone is assigned a racial identity whether they are aware of it or not.” (NMAAHC | Smithsonian, Race & Racial Identity).

The U.S. Census is completed for the U.S. population by a householder<sup>1</sup> every 10 years, and race/ethnicity is collected “based on self-identification” (U.S. Census, About the Topic of Race). A subset of the population completes the annual American Community Survey (ACS), which is used to roll-forward the decennial data. As such, U.S. Census data includes imputed values to roll-forward the decennial Census data using the ACS and to address missing information. The data is aggregated to a geographical level for public access; as such, valuable information for research regarding racial and ethnic disparities is lost (U.S. Census Bureau, ACS and Decennial Census).

- Nina Jablovski of The Pennsylvania State University distinguishes between race and ethnicity. "Race is understood by most people as a mixture of physical, behavioral and cultural attributes. Ethnicity recognizes differences between people mostly on the basis of language and shared culture." (Bryce, 2022).
- In the U.S. Census, race is self-reported. According to [uscensus.gov](https://www.uscensus.gov), "The racial categories included in the census questionnaire generally reflect a social definition of race recognized in this country and not an attempt to define race biologically, anthropologically, or genetically. In addition, it is recognized that the categories of the race item include racial and national origin or sociocultural groups. People may choose to report more than one race to indicate their racial mixture, such as "American Indian" and "White." People who identify their origin as Hispanic, Latino, or Spanish may be of any race." (U.S. Census Bureau, About the Topic of Race, March 2022).

The terms "race" and "ethnicity" also tend to be fluid and change over time. Since 2000, the Census allows individuals to select more than one race. Prior to 2000, only one race could be selected in the U.S. Census. Further, as of June 2022, the U.S. government is considering changing the definition of race again, which would be reflected in the 2030 U.S. Census (Wang, 2022). This lack of consistency in the definition of race and ethnicity over time may be problematic for longitudinal studies where consistency in reporting is important.

Datasets that impute race or ethnicity at an individual-level may not reflect different races and ethnicities of multiple people in a household. For example, if a four-person family household includes a father of one racial or ethnic group and a mother of another race or ethnicity who took her husband's surname upon marriage, all four individuals in the household may be mapped to the same racial or ethnic group, i.e., the imputed race or ethnicity of the father. Using ethnicity to impute race may also be problematic. For example, a modeling algorithm may incorrectly assign a Black, French individual as white.

In addition, using unbiased, individual-level data in health and health care research may be ideal, but it is difficult, if not impossible, for most researchers to obtain. Individual-level data, whether used directly or aggregated to a geographic level, is unlikely to be completely free of imputation or modeling or other potential bias. As such, understanding the collection or derivation of race and ethnicity in a dataset is paramount to the interpretation of results and acknowledgement of any potential bias in the statistical model. For example, does the dataset consist of data aggregated largely from self-reported individual-level data (e.g., the U.S. Census) or individual-level data largely imputed based on a modeling algorithm (e.g., some large, private datasets)? It is important to consider how the dataset compares with the racial and ethnic distribution of known, trusted sources, such as aggregated U.S. Census statistics. If the dataset does not closely match the racial and ethnic distribution of a broadly accepted source, sampling may be an option that could alleviate discrepancies, but this approach may be complex.

In addition, an individual's self-reported race may differ from others' perception of that individual's race. This raises a fundamental question: Is one's own or others' perception of race and ethnicity more closely associated with disparities in health and health care? If others' perception of race is determined to be the driver, how can the perceptions of race by others be captured or modeled for use in future research around racial and ethnic disparities?

The U.S. Census reports statistics on several racial and ethnic groups, including Blacks or African-Americans, Hispanics or Latinos, Asians or Asian Americans, and whites as well as smaller populations, such as Alaska Natives or Native Americans, Native Hawaiians and Pacific Islanders, and multi-ethnic groups. Research results regarding disparities for the smaller groups may not be credible or statistically significant because of the small volume. However, combining these groups does not provide a valid alternative because the heterogeneity of the populations would mask disparities, not reflect their unique experiences, and severely limit the insights that could be gleaned from statistical analysis results.

White et al. discuss the importance of incorporating socially-assigned race into future research around health inequalities. They reviewed 18 studies that looked at race using two approaches: asking individuals how they think others perceive their race, i.e., self-identification, and asking interviewers to assign race or provide characteristics that could be used to assign race, i.e., socially-assigned race. Some of the studies looked at the "level of agreement between self-identified and socially-assigned race" and the relationship with self-rated health. White et al. note two studies that compare differences in health screenings and income inequality, respectively, which show larger disparities in individual health outcomes when socially-assigned race is used. However, self-reporting of others' perception of an individual's race showed larger differences for other measures.

The self-identified and socially-assigned race approaches may be practical for a limited sample size survey, but collecting this information for a large dataset, such as the U.S. population, could be difficult if not collected from a thorough source, such as the U.S. Census. White et al. highlight that while the Behavioral Risk Factor Surveillance System (BRFSS) and other large surveys collect socially-assigned race data, the data may be incomplete, e.g., states are not required to participate, and thus the data may not be representative of the broader U.S. population. To the extent that "an individual's idea of self is in part derived from social interactions with others," existing reporting of self-identification of race may still provide a valid basis for research in the absence of socially-assigned race data (White., et al. 2020).

These examples highlight the complexities in studying racial and ethnic disparities in health and health care. In performing research around racial and ethnic inequalities, it is important to consider whether the definition, collection and imputation of race information, model design or sampling techniques employed create bias in the underlying data. The sensitive nature of these topics warrants applying careful scrutiny to any research and the associated results.

### 2.1.2 HEALTH AND HEALTH CARE DATA

Additional complexities in performing health and health care research result from the difficulties faced when measuring qualitative items and the interrelationships among those items. The following paragraphs outline some possible measures for health and health care related topics and associated complexities.

- **Health Care Coverage**

The type of health care coverage alone may not be sufficient to perform robust research around inequalities. For example, geography is an important consideration for:

- *Medicaid*. Eligibility rules for Medicaid coverage vary by state, which renders geography a primary consideration. It may be helpful to consider results for states that have expanded Medicaid coverage separate from those that have not, where racial and ethnic group disparities are more pronounced.
- *Affordable Care Act (ACA) direct purchase*. Direct purchase ACA marketplace incentives or subsidies vary by state.

Geography is less important for other types of coverage:

- *Medicare*. Eligibility rules apply nationally. However, separate consideration of disability related coverage from retiree coverage may facilitate interpretation of results.
- *Employer-based coverage*. Coverage does not vary geographically, but the type of employer-based coverage—HDHP, narrow network, health maintenance organization (HMO), preferred provider organization (PPO), point-of-service (POS), low or no deductible—must be considered for meaningful research.



- **Health Care Costs and Affordability**

The cost of health care services can dissuade individuals from seeking health care apart from acute illness, particularly if the service is not covered by health insurance. Health insurance coverage—specifically, the type and quality of health insurance coverage—significantly impacts costs and access to care. While the introduction of the ACA in 2010 led to a significant reduction in health insurance disparities, policy changes in 2017 brought an increase in uninsured rates among Blacks, African-Americans, Hispanics or Latinos. Racial and ethnic disparities in health care coverage continue to be greater in states that have not expanded Medicaid, and COVID-19-related impacts on jobs and income have further increased racial and ethnic disparities (Artiga, 2021).

Health care cost data is not readily available from public sources and, where available, may not represent all segments of the population. For instance, the U.S. ASEC data provides premium, out-of-pocket and over-the-counter health care costs by CBSA, which by definition include nearly all urban and some suburban areas (U.S. Census Bureau, ASEC).

When studying inequalities, it is important to understand the broader context of the data. Data that indicates a willingness to pay for prescription medications not covered by insurance could also provide insights. However, it would be necessary to understand the context, including the cost of the medication, as the insights would be different if the prescription costs \$4 versus \$400. As such, it may be difficult to research health care cost and affordability without individual-level data.

- **Health Care Access**

As mandated in the Healthcare Research and Quality Act of 1999, the Agency for Healthcare Research and Quality (AHRQ) prepares an annual “National Healthcare Quality and Disparities report to Congress.” A May 2016 presentation of the report includes the following “Elements of Access to Health Care:”

1. “[T]he timely use of personal health services to achieve the best possible health outcomes” from a 1993 paper by the Institute of Medicine (US), now The National Academy of Medicine, Committee on Monitoring Access to Personal Health Care Services; and
2. Four components of access to health care from Healthy People 2020: insurance coverage, services, timeliness and workforce (AHRQ, Elements of Access to Health Care).

Though dated, the Institute of Medicine (IOM) paper continues to provide a relevant, general framework of health care access, stating succinctly that “Access is a shorthand term used for a broad set of concerns that center on the degree to which individuals and groups are able to obtain needed services from the medical care system.” In addition, “access to services is not an end in and of itself. The purpose of gaining access to the personal health care system is to achieve one or more of an array of possible health outcomes” (Millman, 1993).

In a 2018 publication, the National Academy of Sciences, Engineering and Medicine provides an updated discussion of access to health care, including excerpts from several updated publications. Among those is the AHRQ 2009 “National Healthcare Quality and Disparities Reports,” which considers patient access to sites of care and providers that can meet their medical needs. Another publication cited is from the National Center for Health Statistics that discusses the importance of timely health care access “to prevent illness, control acute episodes, or manage chronic conditions.” In addition, this publication highlights a 2013 paper by Levesque that presents five dimensions of accessibility: “approachability, acceptability, availability and accommodation, affordability, and appropriateness” (NASEM, 2018).

It is easy to see the overlap of indicators of health care access with other related topics.

- **Health Care Quality**

The “2019 Healthcare Quality and Disparities Report” published by the AHRQ considers three types of health care quality measures:

- *Processes of Care*—“percentages of people receiving care that they need or percentages of people receiving care that they should not receive.”
- *Outcomes of Care*—“rates of adverse events or deaths” related to a particular process of care.
- *Patient Perceptions of Care*—“percentages of people who perceived problems with aspects of their care.” (AHRQ, 2020)

Currently, on its website, AHRQ includes structural, process and outcome measures as the three classifications of health care quality (AHRQ, Types of Health Care Quality Measures). Regardless of the version, these categories clearly overlap other health and health care topics, including health outcomes and health care access, which illustrates the complexities of research in these areas. For specific measures for each of these categories, see above.

- **Health Status**

Determining appropriate measures of health status is more complex than it may seem. While a variety of factors may provide possible measures, a person’s self-assessment of health status (poor, good, excellent, etc.) is considered a good indicator (Cislaghi et al, 2019). The Office of Disease Prevention and Health Promotion (ODPHP) “Healthy People 2020 Mid-Course Review” breaks out leading health indicators into 12 topics: access to health services; clinical preventive services; environmental quality; injury and violence; maternal, infant and child health; mental health; nutrition, physical activity and obesity; oral health; reproductive and sexual health; social determinants; substance abuse; and tobacco (National Center for Health Statistics (NCHS) Leading Health Indicators, 2016). Research related to health status may incorporate some or all of these measures, depending on the availability of sufficiently granular data.

- **Chronic Disease**

Great variation exists in the U.S. and internationally over the use of the phrases “chronic disease” or “chronic condition.” The CDC compiles chronic disease prevalence by geographic region based on a broad definition, “conditions that last 1 year or more and require ongoing medical attention or limit activities of daily living or both” (CDC, 2022). The top eight chronic conditions driving costs are: heart disease and stroke, cancer, diabetes, obesity, arthritis, Alzheimer’s disease, epilepsy, and tooth decay. The CDC notes that such conditions drive “90% of the nation’s \$4.1 trillion in annual health care expenditures” (CDC, 2022).

Racial and ethnic group disparities in the ongoing, significant health care costs to treat high-cost chronic conditions is well documented. Glied and Zhu note that, relative to the privately insured, the uninsured tend to use cheaper, less effective forms of insulin, spend twice as much on all prescriptions, and are more likely to report being unable to afford prescriptions (Glied and Zhu, 2020). Many research opportunities exist related to disparities in diabetes diagnosis, health care and associated costs.

- **Health Outcomes**

The WHO defines health as “the state of complete physical, mental and social well-being and not merely the absence of disease or infirmity” (WHO preamble to the constitution). In addition, Parrish says that “Positive health outcomes for people include being alive; functioning well mentally, physically, and socially; and having a sense of well-being” (Parrish, 2010).

As noted under Health Care Quality above, AHRQ considers health outcomes as a key component of health care quality, and in fact, it is difficult to find examples of health outcome measures separate from health care quality

measures. For example, The Centers for Medicare and Medicaid Services (CMS) includes the following categories of health outcome measures to score hospitals as part of its 2022 Overall Hospital Quality Star Rating: mortality, safety, readmission, patient experience, and timely and effective care (CMS, Overall Hospital Quality Star Rating). These measures, which overlap health care quality measures, are representative of the types of data that may be useful in studying health outcomes.

High health care costs have created a burden to employers, leading to growth in HDHPs. However, the high deductibles associated with HDHPs may equate to a lack of coverage, particularly for individuals with low income or a lack of financial resources. Galbraith et al. state that “Early data on individual adults in high-deductible plans suggest that financial burden may be more prevalent in these plans compared to traditional plans without deductibles.” (Galbraith et al.) Coupled with the significant growth in the proportion of U.S. workers covered by private sector employer-based coverage under HDHPs, from 39.4% in 2015 to 52.9% in 2020 (Cattanach, 2022), the association between HDHPs and health care access is ripe for research.

Similarly, insurers created narrow network plans (NNPs) to help manage increasing premium costs. However, NNPs may create the need for policyholders to travel long distances to receive in-network care or pay higher copays or deductibles if out-of-network care must be accessed for acute illness. As such, studying the impact of NNPs on health care access and costs could provide an opportunity for meaningful research.

As the market share of HDHPs has grown, health care costs have shifted from employers and insurers to insureds (HDHP Task Force, 2020). However, this shift may not have impacted all racial and ethnic and other groups equally.

Researchers may select some of these measures or others to include in their statistical analysis. However, it may be difficult to study health care access, for example, apart from health care quality or health outcomes because of the significant overlap in the related measures. In addition, setting forth clear definitions of the variables included in a research project and acknowledging any overlap may be prudent. Alternatively, researchers can study multiple health care variables simultaneously and interpret results accordingly.

### 2.1.3 FINANCIAL WELLNESS DEFINITIONS AND DATA

To study disparities in health and health care, it may be important to consider the relationship with financial wellness. Relevant literature generally defines financial wellness and financial insecurity through two different perspectives: either objective measures and circumstances or perceptions of one’s circumstances. However, a lack of consensus exists on the specific definitions of financial wellness. Table 3 provides definitions of each type of measure.

**Table 3**  
**DEFINITION OF FINANCIAL WELLNESS AND INSECURITY**

Type of measure	Definition
Objective measures	Scholars subscribing to the "objective" perspective define financial wellness as "the comprehensive, multidimensional concept incorporating financial satisfaction, objective status of financial situation, financial attitudes, and behavior that cannot be assessed through a single variable" (Joo, 2008). In U.S. health care, other authors (Ketterman and Pippidis, 2020; Klein and Patel, 2021) also regard one's ability to pay for large, unexpected health care expenses as highly relevant. In addition, the Consumer Financial Protection Bureau (CFPB) has noted that a level of liquid savings sufficient to cover unexpected expenses is an important indicator of financial wellness (CFPB, September 2017).
Perception measures	From the "perception of circumstances" viewpoint, scholars focus on an individual's subjective perception of their circumstances and their emotional response to that perception. Some scholars define financial wellness as a condition in which "a person can fully meet current and ongoing financial obligations, can feel secure in their financial future, and is able to make choices that allow them to enjoy life" (CFPB, May 2017). A lack of financial well-being, or financial insecurity, may include "an individual's perception of the risk of economic misfortune" (Scheve and Slaughter, 2004, 663) and anxiety produced by "possible exposure to adverse economic events and by anticipation of difficulty in recovering from these events" (Bossert & D'Ambrosio, 2013, 1018).
Combined definitions	Combining these formulations, financial insecurity may be defined as "an intersection between "perceived" and "actual" downside risk" (Jacobs, 2007).

Intuitively, liquid assets may be a good indicator of one's ability to pay for unexpected health care costs, and as such, may be a good measure of financial wellness. Alternative objective measures of financial wellness are net worth and income. However, similar patterns of disparities may exist across racial groups for liquid assets, income and net worth because of high correlations among these variables. While income is available in U.S. Census data, income producing assets (IPA) and net worth are not, and none of these variables are likely to be available for the entire U.S. population at an individual level unless modeled or imputed. However, other sources may provide this information for a smaller population. Regardless of the source, it is important to consider any potential bias that may result from the racial and ethnic distribution of the individual dataset.

Perception measures do not lend themselves to regression or some other types of statistical analysis. However, clustering or other techniques could be used. For example, if individual data were available with a data variable for a self-assessment of health status (poor, good, excellent, etc.), insights could be gleaned by gaining an understanding of the characteristics of individuals in each category. Data that indicates a willingness to pay for prescription medications not covered by insurance could also provide insights. However, it would be necessary to understand the context, including the cost of the medication, as the insights would be different if the prescription costs \$4 versus \$400.

#### 2.1.4 BUILDING A DATASET

While U.S. Census data is collected at an individual or household level, individual-level data is not available to the public because of privacy reasons. As such, it may be necessary to develop a database of individual-level data with minimal modeled variables to enable robust health and health care research. Determining the proper sampling method for distribution of a survey will be critical so that the resulting database closely resembles the racial and ethnic and demographic characteristics of the U.S. Ideally, such a dataset would include a wide variety of individual data regarding health status; health care quality (what measures); health care access (what measures); health care costs; specific health diagnoses, including cancer and chronic diseases; access to and appropriateness of preventative care; and health care coverage (including details on the type of coverage, provider and terms such as copays or deductibles). To enable statistical analysis of longitudinal data so that causal effects may be studied, this data would need to be collected on an annual or other periodic basis. Such a dataset could significantly enable future health care research.

## 2.2 TOPICS TO CONSIDER FOR FUTURE RESEARCH

Section 1 provided a discussion of the challenges in performing health and health care research related to data granularity, statistical modeling approaches and other data considerations. Section 2.1 provided a discussion of considerations for researching racial and ethnic inequalities in health and health care as well as the intersection with financial wellness. Throughout that discussion, we suggested various alternative approaches, methodologies and variables that could be used to overcome the significant challenges.

In this section, we suggest narrowly focused topics pertaining to the relationship between health care access, utilization, quality and cost with financial wellness that may be ripe for research. In doing so, we offer topics that researchers may consider alongside the available data to select a statistical approach that addresses the specific research topic. Because data may be a limiting factor regarding the types of statistical analysis that may be used, researchers may want to adjust the research topics after selecting a dataset and reviewing any associated limitations.

1. For a given level of resources, e.g., income, IPA, etc., how does the quality of health care received differ by racial, ethnic and socioeconomic group?
2. For a given level of resources, e.g., income, IPA, etc., how do health outcomes differ by racial, ethnic and socioeconomic group?
3. Rural populations tend to have lower health care utilization compared to suburban and urban populations, although access to health care may not be the primary factor (Rural Health Information Hub). In addition to proximity and transportation and the ability to pay (related to income, savings, IPA and health care coverage), what are the drivers of health care access in different types of geographical areas, i.e. urban, suburban and rural, and how do they differ by race and ethnicity? How do differences in health care utilization impact total health care costs related to the resulting delays in diagnosis of chronic or other health conditions, and how do they differ by race, ethnicity and socioeconomic group?
4. Controlling for differences in income, savings and health care coverage, certain racial and ethnic groups, e.g., Blacks or African-Americans and Hispanics or Latinos, tend to have lower-than-average health care utilization, possibly a result of mistrust of the medical community (CDC, 2021). What drives this mistrust for each race and ethnicity? How can the underlying causes be addressed for each racial and ethnic group to increase health care utilization and drive more relevant and timely preventative care and management of chronic conditions to improve mortality and other health outcomes?
5. How does the type of health care coverage relate to health outcomes, and how does that relationship vary by racial, ethnic and socioeconomic group? Types of health insurance plans may include some of the following:
  - a. Employer-based coverage: HDHPs versus low-deductible plans, small group versus large group plans
  - b. Health maintenance organization (HMO), preferred provider organization (PPO) or point-of-service (POS) plans
  - c. Medicaid
  - d. Medicare
  - e. Narrow networks
  - f. Types of direct purchase: ACA, concierge plans or short-term plans
6. The proportion of U.S. workers covered by private sector employer-based coverage under HDHPs has grown significantly in recent years, from 39.4% in 2015 to 52.9% in 2020 (Cattanach, 2022), and Galbraith et al. found that families with chronic conditions that are covered under HDHPs have an increased financial burden relative to those covered under traditional plans (Galbraith, 2011). How has the growth in HDHPs impacted health care utilization? How does this vary by racial, ethnic and socioeconomic group? How are

other groups, including different racial, ethnic or socioeconomic groups, including families without chronic conditions, impacted?

7. Narrow network plans have more restrictive networks that may reduce the feasibility of obtaining in-network health care and can drive up the costs of out-of-network services, mostly for conditions requiring immediate attention. Compared with preferred provider organizations (PPOs), for example, how does this restricted access impact the timeliness of diagnoses and treatments, health outcomes and costs? Are there racial and ethnic inequalities associated with these networks?
8. How do health outcomes vary by race, ethnicity and socioeconomic group? Why? Consider other factors that may explain the variation by race and ethnicity, including quality of health care received, access to appropriate health care treatments, level of financial wellness, socioeconomic status (e.g., income, education level), gender, age, poverty indicator, limited English proficiency and/or immigration or citizenship status.
9. How is health care spending by racial and ethnic and socioeconomic group related to financial wellness; does health care spending drive changes in the level of financial wellness or does the level of financial wellness drive the amount of health care spending?
10. What racial and ethnic inequalities exist regarding the type of health care coverage, socioeconomic group and:
  - a. the prevalence of diabetes, separately for type 1 and type 2?
  - b. preventative care and treatment after diagnosis, including treatment by an endocrinologist?
  - c. the gap between having a diabetes diagnosis and whether insulin and diabetes medications are purchased? Note that off-label use of certain diabetes medications, e.g., Metformin, adds complexity to this question.
  - d. diagnosis with pre-diabetes and appropriate care or lifestyle changes to delay or prevent the onset of diabetes?
  - e. Percentage of insulin costs covered by the individual versus health care coverage? Glied and Zhu note that relative to the privately insured, the uninsured tend to use cheaper prescription drugs and less effective forms of insulin (Glied and Zhu, 2020).
  - f. quality of health care received?

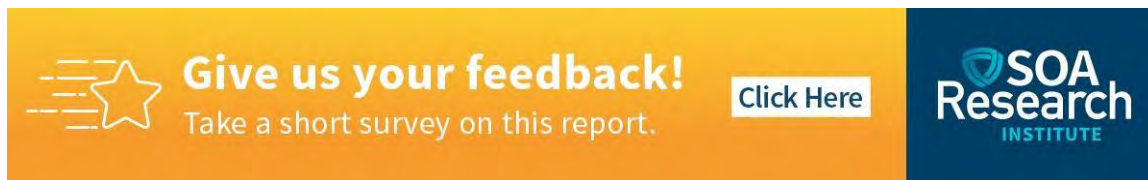
This question could be adapted for other chronic illnesses, such as end-stage renal disease (which may qualify an individual for Medicare coverage) or other chronic diseases, including those presented in Table 3.

11. In addition, Hostetter and Klein discuss cancer as being a significant contributor to health care costs (Hostetter and Klein, 2018). How are the associated health care costs affected by the stage of diagnosis? How does this vary by race, ethnicity and socioeconomic group?
12. Mental health issues are a growing concern, exacerbated by the COVID-19 pandemic, and it is well known that mental health issues are common comorbidities with other chronic illnesses and financial insecurity.
  - a. How do these relationships vary by racial, ethnic and socioeconomic group?
  - b. Do mental health issues cause other chronic diseases as a result of, for example, worse self-care, or does dealing with chronic disease create mental health issues?
  - c. Do mental health issues increase the likelihood of financial insecurity or vice versa?

What changes could be made to reduce the rate of mental health issues or, if deemed a cause, chronic disease or financial insecurity?
13. How do racial and ethnic inequalities vary by state and, in particular, between Medicaid expansion states and states that have not expanded Medicaid? Are marginalized groups impacted more or less in non-expansion states?

14. It has been suggested that COVID has exacerbated the racial and ethnic inequalities in health and health care that exist in this country. How have racial and ethnic inequalities in health and health care access and quality and health care coverage changed pre- and post-COVID pandemic?
15. North Carolina Mutual Life Insurance Company (NCMLC) was founded in Durham, NC in 1898 to improve access to affordable life insurance for Blacks or African Americans (NC History Encyclopedia, 2022). How has the December 2018 rehabilitation order impacted access to life insurance by Blacks or African-Americans and other underserved/marginalized groups? What can be applied to help reduce racial and ethnic disparities in other contexts?

Researchers may want to refine these research topics in light of the available dataset or identify additional questions that may provide insights into racial and ethnic disparities in health and health care.



The banner is split into two color sections: orange on the left and dark blue on the right. On the orange background, there is a white star icon with motion lines to its left. To the right of the star, the text 'Give us your feedback!' is written in a bold, white, sans-serif font. Below this, in a smaller white font, is the text 'Take a short survey on this report.' To the right of this text is a white rectangular button with the text 'Click Here' in a dark blue font. On the dark blue background, the SOA Research Institute logo is displayed in white, featuring a shield icon to the left of the text 'SOA Research INSTITUTE'.

## Section 3: Acknowledgments

The researchers' deepest gratitude goes to those without whose efforts this project could not have come to fruition: the Project Oversight Group for their diligent work overseeing and editing this report for accuracy and relevance.

Project Oversight Group members:

Mischelle Schweickert, FSA, MAAA, Chair

Paul Houchens, FSA, MAAA

Rebecca Johnson, MBA

Roy Machamire, FSA, MAAA, CERA

Rebecca Owen, FSA, MAAA, FCA

Colin Ramsay, ASA, MAAA

Rui Zhong, FSA, MAAA

At the Society of Actuaries:

Lisa Schilling, FSA, EA, FCA, MAAA, Senior Research Actuary

Other researchers:

Stella Bleikhman

Mathias Gleichgevicht

Hong Beng Lim

Collin Nill

Kazuko Prock

Ram Narain Y. Shukla

Radha Srinivasan



## References

- Agency for Healthcare Research and Quality, Rockville, MD (Content last reviewed June 2021). 2019 National Healthcare Quality and Disparities Report. *AHRQ.gov*, <https://www.ahrq.gov/research/findings/nhqrdr/nhqrdr19/index.html>.
- Agency for Healthcare Research and Quality, Rockville, MD (December 2020). National Healthcare Quality and Disparities Report Introduction and Methods. *AHRQ.gov*, AHRQ Publication No.20(21)-0045-EF. Available from: <https://www.ahrq.gov/sites/default/files/wysiwyg/research/findings/nhqrdr/2019qdr-intro-methods-cx061721.pdf>.
- Agency for Healthcare Research and Quality, Rockville, MD. Topic: Access to Care. <https://www.ahrq.gov/topics/access-care.html>.
- Agency for Healthcare Research and Quality, Rockville, MD. Types of Health Care Quality Measures. Content last reviewed July 2015. <https://www.ahrq.gov/talkingquality/measures/types.html>.
- Alom, S., and Marjanovic, S. (July 26, 2021). Fintech's Role in Reducing Disparities and Financial Barriers to Health Care Access. *The RAND Blog*, RAND Corporation, <https://www.rand.org/blog/2021/07/fintechs-role-in-reducing-disparities-and-financial.html>.
- Artiga, S., et al. (16 July 2021) Health Coverage by Race and Ethnicity, 2010-2019. *KFF*. <https://www.kff.org/racial-equity-and-health-policy/issue-brief/health-coverage-by-race-and-ethnicity/>.
- Ayanian, J.Z. (October 1, 2015). The Costs of Racial Disparities in Health Care. *Harvard Business Review*. <https://hbr.org/2015/10/the-costs-of-racial-disparities-in-health-care#:~:text=Racial%20health%20disparities%20are%20associated,%24200%20billion%20in%20premature%20deaths>.
- Barrett, J. and Entzminger, S. (2022). How Does Where You Live Impact Your Health? *Society of Actuaries*. <https://www.soa.org/resources/research-reports/2022/live-impact-health/>.
- Boschert, J., et al. (January 2020). Understanding the connections between health and financial wellness. *Milliman.com*, Milliman. [https://us.milliman.com/-/media/milliman/pdfs/articles/understanding\\_the\\_connections\\_between\\_health\\_and\\_financial\\_wellness.ashx](https://us.milliman.com/-/media/milliman/pdfs/articles/understanding_the_connections_between_health_and_financial_wellness.ashx).
- Bossert, W. and D'Ambrosio, C. (2013). Measuring Economic Insecurity. WIDER Working Paper. Helsinki: UNI-WIDER. <https://www.wider.unu.edu/sites/default/files/wp2013-008.pdf>.
- Bryce, Emma. What's the difference between race and ethnicity? *Live Science*. April 9, 2022. <https://www.livescience.com/difference-between-race-ethnicity.html>.
- Cattanach, J. (24 Jan 2022). High-Deductible Health Plans Continue to Grow in Popularity, but Are They Right for You? *Value Penguin | Lending Tree*. <https://www.valuepenguin.com/high-deductible-health-plan-study>.

Centers for Disease Control and Prevention (CDC) (19 April 2021). Health Equity Considerations and Racial and Ethnic Minority Groups. *CDC.gov*. <https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html>.

Centers for Disease Control and Prevention (CDC) (Last reviewed 14 September 2018). Publications and resources: Health Insurance Portability and Accountability Act of 1996 (HIPAA). <https://www.cdc.gov/phlp/publications/topic/hipaa.html#:~:text=The%20Health%20Insurance%20Portability%20and,the%20patient's%20consent%20or%20knowledge.>

Centers for Disease Control and Prevention (CDC) (Last reviewed 25 May 2022) National Center for Chronic Disease Prevention and Health Promotion/About Chronic Disease: Health and Economic Cost of Chronic Disease. <https://www.cdc.gov/chronicdisease/about/costs/index.htm>.

Centers for Disease Control and Prevention (CDC). PLACES: Local Data for Better Health. <https://www.cdc.gov/places/measure-definitions/health-outcomes/index.html#mental-health>.

Cislaghi, B., & Cislaghi, C. (2019). Self-rated health as a valid indicator for health-equity analyses: evidence from the Italian health interview survey. *BMC public health*, 19(1), 533. <https://doi.org/10.1186/s12889-019-6839-5>.

Consumer Financial Protection Bureau (CFPB) (September 2017). “Financial well-being in America.” *CFPB*. [https://files.consumerfinance.gov/f/documents/201709\\_cfpb\\_financial-well-being-in-America.pdf](https://files.consumerfinance.gov/f/documents/201709_cfpb_financial-well-being-in-America.pdf).

Consumer Financial Protection Bureau (CFPB) (May 2017). CFPB Financial Well-Being Scale: Scale Development and Technical Report. *CFPB*. <https://www.consumerfinance.gov/data-research/research-reports/financial-well-being-technical-report/>

Fidelity Investments (15 September 2020). “Racial Inequalities in Financial Wellness.” PlanSponsor. <https://www.plansponsor.com/thought-leadership/fid-racial-inequalities-financial-wellness/>.

Galbraith, A. A., Ross-Degnan, D., Soumerai, S. B., Rosenthal, M. B., Gay, C., & Lieu, T. A. (2011). Nearly half of families in high-deductible health plans whose members have chronic conditions face substantial financial burden. *Health affairs (Project Hope)*, 30(2), 322–331. <https://doi.org/10.1377/hlthaff.2010.0584>.

Gillen, E.M., et al. (2017). “The Effect of Narrow Network Plans on Out-of-Pocket Cost.” *American Journal of Managed Care*, vol. 23, no. 9, 2017, pp. 540-545. *AJMC*. <https://www.ajmc.com/view/the-effect-of-narrow-network-plans-on-out-of-pocket-cost>.

Glied, S.A., and Zhu, B. (25 September 2020). Not So Sweet: Insulin Affordability over Time. *The Commonwealth Fund*. <https://www.commonwealthfund.org/publications/issue-briefs/2020/sep/not-so-sweet-insulin-affordability-over-time>.

Hayes, S.L., et al. (29 August 2016). High-Need, High-Cost Patients: Who Are They and How Do They Use Health Care? A Population-Based Comparison of Demographics, Health Care Use, and Expenditures. *The Commonwealth Fund*. <https://www.commonwealthfund.org/publications/issue-briefs/2016/aug/high-need-high-cost-patients-who-are-they-and-how-do-they-use>.

High Deductible Health Plan Task Force.

[https://www.cga.ct.gov/ins/tfs/20190822\\_High%20Deductible%20Health%20Plan%20Task%20Force/final%20report%20Feb%202020/High%20Deductible%20Health%20Plan%20Task%20Force%20Final%20Report%20pgs%201-149.pdf](https://www.cga.ct.gov/ins/tfs/20190822_High%20Deductible%20Health%20Plan%20Task%20Force/final%20report%20Feb%202020/High%20Deductible%20Health%20Plan%20Task%20Force%20Final%20Report%20pgs%201-149.pdf). February 24, 2020.

Hill, Latoya, et al.. Key Facts on Health and Health Care by Race and Ethnicity. KFF. <https://www.kff.org/racial-equity-and-health-policy/report/key-facts-on-health-and-health-care-by-race-and-ethnicity/>. January 26, 2022.

Holle, M, et al. (2021). Trends in the Concentration and Distribution of Health Care Expenditures in the US, 2001-2018. *JAMA Network Open*. 4(9):e2125179. doi:10.1001/jamanetworkopen.2021.25179. <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2784058>.

Hostetter, M, and Klein, S (27 September 2018). In Focus: Reducing Racial Disparities in Health Care by Confronting Racism. *The Commonwealth Fund*. <https://www.commonwealthfund.org/publications/2018/sep/focus-reducing-racial-disparities-health-care-confronting-racism>.

Jacobs, E. (September 2007). The Politics of Economic Insecurity. *The Brookings Institution*: Washington DC. Issues in Governance Studies. Number 10. <https://www.brookings.edu/wp-content/uploads/2016/06/jacobs20070913.pdf>

Joo S. (2008). Personal Financial Wellness. In: Xiao J.J. (eds) Handbook of Consumer Finance Research. Springer, New York, NY. [https://doi.org/10.1007/978-0-387-75734-6\\_2](https://doi.org/10.1007/978-0-387-75734-6_2).

Kaiser Family Foundation. Status of State Medicaid Expansion Decisions: Interactive Map. June 29, 2022. <https://www.kff.org/medicaid/issue-brief/status-of-state-medicaid-expansion-decisions-interactive-map/>

Ketterman, J, and Pippidis, M. (January 2020). Bridging the Gap Between Financial Wellness and Physical Wellness – FSA, HSA, and HRAs. *The Standard Newsletter*, AFCPE, <https://www.afcpe.org/news-and-publications/the-standard/1st-quarter-2020/bridging-the-gap-between-financial-wellness-and-physical-wellness-fsa-hsa-and-hras/>.

Kirby, J. and Muhuri, P. (May 2018). Insurance and Access to Care in Urban and Rural Areas, 2014-2015. Statistical Brief #512, *Agency for Healthcare Research and Quality*. [https://meps.ahrq.gov/data\\_files/publications/st512/stat512.pdf](https://meps.ahrq.gov/data_files/publications/st512/stat512.pdf).

Klein, A. and Patel, K (1 June 2021). How financial volatility impacts health care decisions: What we don't know is hurting us. *Brookings Institution*. <https://www.brookings.edu/blog/up-front/2021/06/01/how-financial-volatility-impacts-health-care-decisions-what-we-dont-know-is-hurting-us/>.

Lechner, A. (2011). Forgoing Medical Care Due to Costs: The Causal Effect of Health Care Burdens on Health. Georgetown University, 2011. Master's Thesis. <https://repository.library.georgetown.edu/handle/10822/558168>

Lee, De Chih, et al. (2021). The convergence of racial and income disparities in health insurance coverage in the United States. *International Journal for Equity in Health*, vol. 20, no. 96, 2021, pp. 1-8. *BioMed Central*. <https://equityhealthj.biomedcentral.com/articles/10.1186/s12939-021-01436-z>.

Lewis, T. T., Cogburn, C. D., & Williams, D. R. (2015). Self-reported experiences of discrimination and health: scientific advances, ongoing controversies, and emerging issues. *Annual review of clinical psychology*, 11, 407–440. @2015 Annual Review, <http://www.annualreviews.org>. <https://doi.org/10.1146/annurev-clinpsy-032814-112728>.

Lichtenstein, R. (2015). The Vexing Problem of Health Inequalities in the United States. *Medical Care*, 53(3) 215-217. doi: 10.1097/MLR.0000000000000332. [https://journals.lww.com/lww-medicalcare/Citation/2015/03000/The\\_Vexing\\_Problem\\_of\\_Health\\_Inequalities\\_in\\_the.1.aspx](https://journals.lww.com/lww-medicalcare/Citation/2015/03000/The_Vexing_Problem_of_Health_Inequalities_in_the.1.aspx).

The Difference between “Race” and “Ethnicity”: How they differ and overlap. Merriam-Webster.com. <https://www.merriam-webster.com/words-at-play/difference-between-race-and-ethnicity>.

Millman, M. (Ed.). (1993). Access to Healthcare in America: 2, A Model for Monitoring Access. *Institute of Medicine (US) Committee on Monitoring Access to Personal Health Care Services; National Academies Press (US)*. <https://www.ncbi.nlm.nih.gov/books/NBK235891/>.

Mitchell, E.M. (January 2021). Concentration of Healthcare Expenditures and Selected Characteristics of High Spenders, U.S. Civilian Noninstitutionalized Population, 2018. *Statistical Brief #533. Agency for Healthcare Research and Quality, Rockville, MD*. [https://meps.ahrq.gov/data\\_files/publications/st533/stat533.pdf](https://meps.ahrq.gov/data_files/publications/st533/stat533.pdf).

National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Health Care Services; Committee on Health Care Utilization and Adults with Disabilities. Washington (DC): [National Academies Press \(US\)](https://www.ncbi.nlm.nih.gov/books/NBK500097/); 2018 Mar 1. <https://www.ncbi.nlm.nih.gov/books/NBK500097/>.

National Center for Health Statistics (2016). Chapter IV: Leading Health Indicators. *Healthy People 2020 Midcourse Review*. Hyattsville, MD. [https://www.cdc.gov/nchs/data/hpdata2020/CHIV\\_LHI.pdf](https://www.cdc.gov/nchs/data/hpdata2020/CHIV_LHI.pdf).

National Human Genome Research Institute. Race. <https://www.genome.gov/genetics-glossary/Race>. Updated July 1, 2022.

National Museum of African American History & Culture | Smithsonian. Talking About Race: Race and Racial Identity. <https://nmaahc.si.edu/learn/talking-about-race/topics/race-and-racial-identity>.

Ndugga, N, and Artiga, S. (11 May 2021). Disparities in Health and Health Care: 5 Key Questions and Answers. *KFF*. <https://www.kff.org/racial-equity-and-health-policy/issue-brief/disparities-in-health-and-health-care-5-key-questions-and-answers/>. (note that this article uses the terminology, “people of color”)

NorthCarolinahistory.org: An Online Encyclopedia, “North Carolina Mutual Life” <https://northcarolinahistory.org/encyclopedia/north-carolina-mutual-life/> (accessed August 4, 2022).

Ortaliza, Jared, et al. (14 January 2022). How does cost affect access to care? *Health System Tracker*, Peterson-KFF. <https://www.healthsystemtracker.org/chart-collection/cost-affect-access-care/#item-start>.

Parrish, R.G. (July 2010). Measuring Population Health Outcomes. *Center for Disease Control and Prevention Special Topic*. [https://www.cdc.gov/pcd/issues/2010/jul/10\\_0005.htm](https://www.cdc.gov/pcd/issues/2010/jul/10_0005.htm)

Riley, W.J. (2012) Health disparities: gaps in access, quality and affordability of medical care. *Transactions of the American Clinical and Climatological Association* vol. 123 (2012): 167-72; discussion 172-4. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540621/>

Rural Health Information Hub. Healthcare Access in Rural Communities. *Rural Health/Topics & States/Topics: Healthcare Access in Rural Communities*. <https://www.ruralhealthinfo.org/topics/healthcare-access#:~:text=A%20shortage%20of%20healthcare%20professionals,were%20located%20in%20rural%20areas>.

Scheve, K. and Slaughter, M. (1 October 2004). Economic Insecurity and the Globalization of Production. *American Journal of Political Science* 48(4):662 DOI:[10.2307/1519926](https://doi.org/10.2307/1519926).

Scott, KW, Scott, JW, Sabbatini AK, et al. (2021). Assessing Catastrophic Health Expenditures Among Uninsured People Who Seek Care in U.S. Hospital-Based Emergency Departments. *JAMA Health Forum*,2(12):e214359. doi:10.1001/jamahealthforum.2021.4359. <https://jamanetwork.com/journals/jama-health-forum/fullarticle/2787660>.

Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A Practical Guide to Calculating Cohen's  $f(2)$ , a Measure of Local Effect Size, from PROC MIXED. *Frontiers in psychology*, 3, 111. <https://doi.org/10.3389/fpsyg.2012.00111>.

Shmueli, G. (August 2010). To Explain or to Predict? *Statist. Sci.* 25 (3) 289 - 310, <https://doi.org/10.1214/10-STS330>.

United States Census Bureau. Annual Social and Economic Supplements. <https://www.census.gov/data/datasets/time-series/demo/cps/cps-asec.html>.

United States Census Bureau. American Community Survey: [https://www2.census.gov/programs-surveys/acs/tech\\_docs/subject\\_definitions/2019\\_ACSSubjectDefinitions.pdf](https://www2.census.gov/programs-surveys/acs/tech_docs/subject_definitions/2019_ACSSubjectDefinitions.pdf).

United States Census Bureau. About the topic of race. <https://www.census.gov/topics/population/race/about.html>. Last revised March 1, 2022.

United States Census Bureau. The Importance of the American Community Survey and the Decennial Census. <https://www.census.gov/programs-surveys/acs/about/acs-and-census.html>.

United States Census Bureau. Questions and Answers for Census 2000 Data on Race. March 14, 2001. [https://www2.census.gov/census\\_2000/census2000/raceqandas.html](https://www2.census.gov/census_2000/census2000/raceqandas.html)

United States Census Bureau. Quick Facts, <https://www.census.gov/quickfacts/fact/table/US/PST045221>.

United States Census Bureau. Tallies, 2020. <https://www.census.gov/geographies/reference-files/time-series/geo/tallies.html>.

U.S. Centers for Medicare & Medicaid Services. Overall Hospital Quality Star Rating. July 2022. <https://data.cms.gov/provider-data/topics/hospitals/overall-hospital-quality-star-rating/>.

Vacha-Haase, T. and Thompson, Bruce. October 2004. How to Estimate and Interpret Various Effect Sizes. *Journal of Counseling Psychology*. 51(4):473-481. DOI:[10.1037/0022-0167.51.4.473](https://doi.org/10.1037/0022-0167.51.4.473). <https://www.researchgate.net/publication/232499875> How to Estimate and Interpret Various Effect Sizes.

Viswanath, K, and Ackerson, L. (January 2011). Race, Ethnicity, Language, Social Class, and Health Communication Inequalities: A Nationally-Representative Cross-Sectional Study. *PLoS ONE*, 6(1), e14550. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0014550>.

Wang, Hansi Lo. Biden officials may change how the U.S. defines racial and ethnic groups by 2024. National Public Radio. June 15, 2022. <https://www.npr.org/2022/06/15/1105104863/racial-ethnic-categories-omb-directive-15>.

Weida, E.B., et al. (18 May 2020). Financial health as a measurable social determinant of health." *PloS one* vol. 15, 5 e0233359., doi:10.1371/journal.pone.0233359.  
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0233359>.

White, Kellee, et al. Socially-assigned race and health: a scoping review with global implications for population health equity. *International Journal for Equity in Health*. 2020. <https://doi.org/10.1186/s12939-020-1137-5>.

World Health Organization (WHO) Health topics. Social determinants of health: Overview.  
[https://www.who.int/health-topics/social-determinants-of-health#tab=tab\\_1](https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1).

World Health Organization (WHO). Preamble to the Constitution of WHO.  
<https://www.who.int/about/governance/constitution#:~:text=Health%20is%20a%20state%20of,belief%2C%20economic%20or%20social%20condition>.

## Appendix A: Specific Considerations for a Linear Regression Approach to Health and Health Care Research

In this appendix, we discuss several considerations relevant to research around racial and ethnic disparities in health, health care and health care coverage, including differences by level of financial wellness or socioeconomic group. Based on preliminary data analysis previously discussed, researchers can design a model using the selected data. Regression models have specific advantages and disadvantages in conducting this type of research, and we present herein some specific considerations for use of this type of model. Note, however, that this discussion is not a recommendation for the use of regression modeling.

In this appendix, consider a Weighted Least Squares (WLS) regression model studying the relationship with a dependent, health and health care variable,  $Y_i$  :

$$Y_i := \alpha_i + \sum_{j=1}^J \beta_j \text{RaceVariable}_{ji} + \sum_{m=1}^M \theta_m \text{DemographicVariable}_{mi} + \epsilon_i, \\ i = 1, 2, \dots, N$$

where,

$N$  is the total number of observations included in a study

$\beta_j$  is the coefficient of the  $j^{\text{th}}$  racial and ethnic group.

$\theta_m$  is the coefficient of the  $m^{\text{th}}$  demographic group.

$\text{RaceVariable}_j$  is the variable corresponding to the  $j^{\text{th}}$  race

$\text{DemographicVariable}_m$  is the variable corresponding to the  $m^{\text{th}}$  demographic group

### A.1 SELECTION OF DEMOGRAPHIC VARIABLES

The choice of which demographic variables to include in a regression model will depend on the specific variables available at the desired level of granularity. Professional judgment may be used to select among the available variables, but this approach could introduce unintended, unintentional bias. As an alternative, objective approach, a forward regression approach may be used to determine the set of demographic or other variables to be used as controls for the regressions. For example, the variable (or group of categorical variables) which leads to the largest improvement in model fit may be added first. For this type of approach, various measures of model fit, including Bayesian Information Criterion (BIC), where a lower value indicates a better model fit, or adjusted  $R^2$ , where a higher value indicates a better model fit, could be used. For example, a variable that produces marginal increase in adjusted  $R^2$  of a specified threshold, e.g., at least 0.25%, could be considered for inclusion in the regression model. In addition to BIC and adjusted  $R^2$ , Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE)\* could also be incorporated into a forward regression approach so that the fitted models progressively have lower average errors in prediction, i.e., better prediction accuracy.

### A.2 INCOME BANDS AS A PROXY FOR SOCIOECONOMIC GROUPS

Apart from estimating the relationship between the explained variable and independent variables (usually termed the "main" effects), researchers are often interested in studying differences in model results for variables under study, (e.g., racial and ethnic disparities) across different groups (e.g., socioeconomic groups using income levels, education level or occupation). These are termed interaction effects, where the association with a group can "interact" with how the independent variables relate to the dependent variable. In regression models, the introduction of dummy variables, which are usually binary indicators denoting whether an observation belongs to a particular group, and their interaction with other variables can help identify if differences in effects exist between groups. However, this approach increases model complexity when applied to models with multiple predictive

variables and may also introduce collinearity among predictors and their interaction terms, which invites its own set of issues.

A practical concession in such cases could be to split observations by the groups, e.g., income bands, and run individual regression models on each. Depending on the level of granularity of the variables, it may be important to consider any loss of information that may result from mapping from one level of granularity to another. For example, if a particular variable, such as diabetes prevalence is available at a census tract level, it will be assigned to all census block groups (CBGs) or individuals within the tract. Because the socioeconomic group related to having or not having a diabetes diagnosis is not known, results for middle income bands may be largely similar.

For a regression that considers the relationship between health and health care variables, financial wellness and socioeconomic groups, additional complexities may be introduced as many of the variables that could be used to measure financial wellness—income, net worth and income producing assets (IPA)—are correlated with socioeconomic status. As such, we do not recommend defining financial wellness as the dependent variable in an income band regression model. The confounding effects of the variables add unnecessary complexity to the models and the interpretation of results. Instead, considering disparities by socioeconomic group in health status, outcome, chronic disease, health care coverage and costs, or health care quality and access may result in a better model design, facilitate interpretations and produce better model fits (e.g., higher adjusted  $R^2$ ).

### A.3 BASE CATEGORIES FOR REGRESSION ANALYSIS

For categorical variables such as race, statisticians and data scientists traditionally run regressions by choosing one category—typically the largest category—as the baseline, and then fitting regression coefficients for each remaining category. Statisticians then interpret the regression coefficients as the average difference in the dependent variable for individuals belonging to a given category relative to the baseline category. This approach may have several disadvantages, especially when studying racial and ethnic group disparities:

1. While this is a standard statistical approach, comparisons to the baseline category may seem to imply that the base category is the standard. With respect to race and ethnicity, for example, the standard statistical approach would use whites as the base category (as whites are the most populous group in the U.S.), and regression results would show all other racial categories relative to whites. This approach can create unintended connotations because whites may be interpreted as the standard to which all comparisons are made.
2. The lack of a coefficient for the base category may not be intuitive to non-statisticians and may also add complexity to the interpretation of results.

It may be possible to construct a national average baseline for presentation of regression results using the standard regression approach that drops the single, largest category, which is used as the baseline. This approach would likely employ a weighted sum-to-zero constraint on the coefficients that result from regressions run using the standard approach, and the resulting coefficients would represent the average difference between an observation belonging to category  $j$  and the national average. However, we are not aware of such an approach in statistical literature. Another alternative approach could be to use indices instead of absolute values, where the indexing would be performed relative to national average values.

The primary disadvantage to this approach is two-fold: the interpretation of regression results is more difficult or complex, and this adds a step which could introduce an opportunity for technical errors. A robust technical review process can easily overcome the latter. We note that the level of complexity increases with each additional national baseline category created.



#### A.4 GOODNESS OF FIT AND MULTICOLLINEARITY

While considerations regarding the goodness-of-fit of a model and multicollinearity of the data variables are not specific to health and health care research, we include herein a high-level discussion of some common measures and considerations that may be used to address these issues to facilitate discussions between actuaries and data scientists.

The goodness-of-fit of a model refers to measuring how well a statistical (assumed) model fits to the observed data. Adjusted  $R^2$  may be used to determine goodness of fit. A measure of statistical significance of regression coefficients may also be considered either for dropping variables if the sample size is small or for a large sample size in the context of interpreting results. Specifically, p-values (taking alpha = 5%) in the final model results indicate which variables turn insignificant at the alpha-threshold.

Another statistical mean, effect size (ES), is a value assessing the strength of findings in research studies, typically by measuring the magnitude and direction of the difference between two groups or the relationship between two variables. An ES can be a difference between means, a percentage, or a correlation (Vacha-Hasse & Thompson, 2004). However, no straightforward relationship exists between the size of an effect and its practical value. For example, depending on the context, a lower effect size could potentially be more important for one outcome than a higher effect size for another outcome. When an effect or difference is statistically significant, especially in the wake of large sample sizes, it does not necessarily mean that it is big, important, or helpful in decision-making. To understand whether effects of differences are also meaningful, effect sizes are often said to be an important complement to be reported.

The appropriate effect size measure can vary depending on the data and nature of the research study (Selya et al. 2012). Some of the common measures are summarized in Table 4.

**Table 4**  
**COMMON STATISTICAL MEASURES OF EFFECT SIZE**

Name of Measure	When is it used
Cohen's d (or d-family of standardized mean differences)	Effect size is the difference in means between two groups divided by the standard deviation of the control group, and thus, requires the independent variable of interest to be categorical. Standardized mean differences also cannot determine local effect sizes, i.e., individual effect sizes of particular variables within a multivariate model that includes other categorical and continuous independent variables.
Cramer's $\phi$ or Cramer's V method of effect size	Chi-square is the best statistic to measure the effect size for nominal data. In nominal data, when a variable has two categories, then Cramer's phi ( $\phi$ ) is the best statistical measure to use. When there are more than two categories of nominal data, then Cramer's V statistics is preferred.
Pearson r correlation (or r-family measures of strength of association)	The r family effect sizes describe the proportion of variance that is explained by group membership [e.g., a correlation (r) of 0.5 indicates 25% (R-Squared) of the variance is explained by the difference between groups].
Cohen's $f^2$	In a multiple regression model where both independent and dependent variables are continuous, one of the most common methods for calculating the effect size of each of the variables or constructs is Cohen's $f^2$ .

We would be remiss if we did not address the consideration of multicollinearity necessary when running regression models. Multicollinearity occurs when two or more independent variables are highly correlated with one another. While most statistical studies include independent variables that have some degree of correlation, multicollinearity in a regression model can create uncertainty in attributing the effect that the interrelated variables have, or, more specifically, which interrelated variable has an effect on the dependent variable.

While it is difficult, if not impossible, to eliminate multicollinearity completely—especially when using independent variables that are interrelated (e.g., the proportion of Hispanics or Latinos and the proportion of limited English-

speaking households in a CBG)—statistical measures may be used to minimize the effect of multicollinearity. One such measure is Variance Inflation Factor (VIF), which is a standard measure used to detect multicollinearity, with values ranging from one (not correlated) to infinity (perfectly correlated). While considerable debate exists over the specific threshold to use, a VIF of five or larger is generally considered problematic. To deal with high VIF, there are two common approaches: (1) drop independent variables with a high VIF (greater than five, for example), beginning with the variables added in the later steps, until all independent variables have a VIF less than the selected threshold, or (2) transform multicollinear independent variables using principal component analysis (PCA) or factor analysis. Details regarding these approaches are beyond the scope of this paper.

While a regression model cannot eliminate multicollinearity completely, the use of VIFs largely addresses high multicollinearity in regression models, and significant distortion in the variance of estimates from remaining collinearity would not be expected. In addition, a large dataset, such as data for 335 million individuals that represent the total U.S. population significantly overcomes concerns of multicollinearity and likely will produce results similar to more advanced modeling techniques that can be difficult to explain and interpret.

As discussed in the previous section, consideration of p-values when interpreting regression model results can also be used to address these concerns. Specifically, the uncertainty created by multicollinearity can inflate p-values and yield unreliable results (i.e., the interpretation of the regression coefficients and their significance becomes unreliable) while minimally affecting the overall model fit. Generally, not relying on regression coefficients associated with high p-values (i.e., greater than or equal to 0.05) will largely address any remaining concerns regarding multicollinearity.

## Appendix B: Literature Review

We reviewed existing literature on financial wellness, health and health care quality inequalities as well as the impact of disparities in health care coverage, costs and access on financial wellness and its impact on health outcomes. We specifically sought articles that examined the disparities in these relationships across key demographic characteristics, including race and ethnicity. In addition, we identified studies that assessed the impact that differences in the prevalence of diabetes across these key demographic groups has on financial wellness.

### B.1 FINANCIAL WELLNESS INEQUALITIES

Consumer Financial Protection Bureau (CFPB) (September 2017). “Financial well-being in America .” CFPB. [https://files.consumerfinance.gov/f/documents/201709\\_cfpb\\_financial-well-being-in-America.pdf](https://files.consumerfinance.gov/f/documents/201709_cfpb_financial-well-being-in-America.pdf).

The Consumer Financial Protection Bureau’s goal is to help consumers “achieve their own financial and life goals” through education. The CFPB focuses on “defining, measuring and studying what contributes to financial well-being” and has developed a scale to measure financial well-being. This report presents the results of the CFPB’s 2016 nationwide U.S. survey, the first to use the financial well-being scale along with a variety of demographic and financial data and attitudes. Below is a summary of the key findings:

- The perception of one’s own financial well-being varies greatly throughout the U.S. and is likely affected by multiple factors. The average score was 54, with one third falling at or below 50, between 51 and 60, and above 60. The CFPB found that the scores were consistent with the relative ease or difficulty with which an individual can make necessary payments.
- Financial well-being scores vary across income levels. A more important, less variable, indicator is the amount of liquid savings sufficient to cover unexpected expenses as well as specific circumstances and perceptions.
- Individuals who have had negative experiences with debt, such as collections or denial of credit, had lower scores; whereas individuals who feel confident in their financial abilities had higher scores.
- Income, occupation, education level and older age, particularly individuals older than 65, are associated strongly with financial well-being. Geographic region and sex are not. Even differences in average financial well-being among racial and ethnic groups are small, with the exception of non-Hispanic, non-Latino white adults who have higher scores than all other groups.

Fidelity Investments (15 September 2020). “Racial Inequalities in Financial Wellness.” *PlanSponsor*. <https://www.plansponsor.com/thought-leadership/fid-racial-inequalities-financial-wellness/>.

This article looks at differences in financial wellness across groups, with a focus on race and ethnicity of employees, recognizing the influence of personal traits and circumstances beyond traditional debt, income and savings measures. Specifically, financial literacy and confidence as well as personal and financial responsibilities can influence financial wellness or one’s view of financial wellness. The authors offer an example that providing support to extended family is less common among white employees than their Black or African American, Hispanic or Latino and Asian or Asian-American colleagues. Blacks or African Americans, in particular, are more likely to sacrifice their own financial wellness to provide community support. Some racial and ethnic groups are challenged to achieve financial wellness as a result of historical and continuing wealth and income inequality, possibly because of a lack of sufficient educational, employment, borrowing, and employer retirement savings plan opportunities. In this article, Fidelity focuses on four “domains” of financial wellness: budgeting, debt, savings, and protection, and acknowledges that many short-term and long-term factors impact financial wellness.

The results of this study indicate that the number of Black or African-American and Hispanic or Latino employees eligible to participate in an employer retirement savings plan who have low financial wellness scores is two to two and one-half times higher than the average, even when controlled for income. Fidelity notes that the various racial/ethnic groups score comparably on budgeting, but Blacks or African-Americans and Hispanics or Latinos fall behind on debt, savings and protection measures. The study also found higher levels of financial stress around emergency savings for Blacks or African-Americans and Hispanics or Latinos.

Fidelity also explored groups that identify with more than one demographic group, e.g., race or ethnicity and gender. While men had more financial confidence than women, Black or African-American and Hispanic or Latino women and Black or African-American and Hispanic or Latino men had higher confidence than their female and male white counterparts, respectively, despite the fact that Blacks or African-Americans and Hispanics or Latinos have a lower retirement savings rate.

## B.2 HEALTH STATUS/OUTCOMES AND HEALTH CARE QUALITY/ACCESS INEQUALITIES

Lewis, T. T., Cogburn, C. D., & Williams, D. R. (2015). Self-reported experiences of discrimination and health: scientific advances, ongoing controversies, and emerging issues. *Annual review of clinical psychology*, 11, 407–440. ©2015 by Annual Reviews, <http://www.annualreviews.org>. <https://doi.org/10.1146/annurev-clinpsy-032814-112728>.

The authors trace the history of studies linking exposure to self-reported racial discrimination and health outcomes. The following adverse effects of self-reported discrimination were identified:

- Higher prevalence of adverse health indicators (hypertension, mortality, incident asthma, breast cancer, and poor mental health), early indicators of clinical disease (inflammation, carotid visceral fat, obesity, coronary artery calcification, shorter telomeres and cortisol dysregulation) and health behaviors (poor sleep quantity and quality, cigarette smoke, and substance abuse).
- Increased cardiovascular response, mental health symptoms, and hypertension.
- Lower levels of seeking health care and reduced adherence to healthy behaviors.
- Lesser access to health-enhancing services as a result of historical racial discrimination around housing and financial accessibility related to residential segregation. That type of segregation has been connected to poorer quality and less affordable housing, increased exposure to environmental toxins and reduced access to resources shown to improve quality of life (education, health care, employment).

A large part of the article also talks about ongoing controversies on how to more effectively conduct such research, as well as interesting areas for future research such as cyber discrimination and the physical and psychological mechanisms through which discrimination can affect health outcomes.

Viswanath, K, and Ackerson, L. (January 2011). Race, Ethnicity, Language, Social Class, and Health Communication Inequalities: A Nationally-Representative Cross-Sectional Study. *PLoS ONE*, 6(1), e14550. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0014550>.

Using a nationally-representative survey of 5,187 U.S. adults, Viswanath and Ackerson assess the propensity for individuals to seek out health information and the degree to which they pay attention to and trust in various forms of media (e.g., television, radio, newspaper, magazines, internet). Viswanath and Ackerson conducted the study within the context of cancer and specifically looked at how these behaviors vary with race, ethnicity, language, and social class. The following significant differences relative to the non-Hispanic, non-Latino white population were found:

- Spanish-speaking Hispanics or Latinos exhibited the lowest propensity to seek out health information.
- Hispanics or Latinos tended to pay more attention to and trust health messages on radio.
- Non-Hispanic, non-Latino Blacks tended to pay more attention to and trust health messages on television.

Further, college graduates paid significantly more attention to newspapers, magazines, and the internet, while exhibiting less trust in radio and television, than individuals who did not graduate from high school.

Centers for Disease Control and Prevention (CDC) (19 April 2021). Health Equity Considerations and Racial and Ethnic Minority Groups. *CDC.gov*. <https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html>.

This article lays out the channels through which COVID-19 exacerbates health inequities. Those include:

- *Discrimination in systems meant to protect well-being or health, (e.g., housing, education, criminal justice, and finance)*: This inequity leads to unequal access to resources for protecting oneself against COVID and can lead to chronic and toxic stress among affected populations.
- *Health care access and use*: Socioeconomic factors, cultural differences between patients and providers, inequities in treatment by health care providers and historically-based mistrust of the government (e.g., Tuskegee syphilis study; sterilization without consent) are components of this disparity.
- *Occupation*: Certain racial and ethnic minority groups are overrepresented in higher risk essential work.
- *Educational, income, and wealth gaps*: Lower education levels make these groups more likely to work in high-risk settings, and their lack of savings means they can ill afford to miss work.
- *Housing*: A higher percentage of such groups live in crowded housing conditions.

Riley, W.J. (2012) Health disparities: gaps in access, quality and affordability of medical care. *Transactions of the American Clinical and Climatological Association* vol. 123 (2012): 167–72; discussion 172–4. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540621/>

This article chronicles advances made in the public understanding of health disparities in the last three decades, starting with the 1985 Malone-Heckler report. In addition to discussing disparities in treatment received, the authors address cost and access issues; specifically noting that being uninsured results in using “fewer preventive services, poorer health outcomes, higher mortality and disability rates, lower annual earnings because of sickness and disease,” and a tendency to be in more advanced stages of a disease when diagnosed. “The uninsured tend to be disproportionately poor, young, and from racial and/or ethnic minority groups.” These disparities create a financial burden on the state, estimated at “\$1.5 trillion over a 3-year period.”

Ndugga, N, and Artiga, S. (11 May 2021). Disparities in Health and Health Care: 5 Key Questions and Answers. KFF. <https://www.kff.org/racial-equity-and-health-policy/issue-brief/disparities-in-health-and-health-care-5-key-question-and-answers/>. (note that this article uses the terminology, “people of color”)

In this paper, the authors address structural and societal systemic racial inequities, noting that recent incidents of police brutality, racial hate crimes and the impacts of COVID-19 have brought these issues to the forefront. The authors look at several definitions of health disparities and health equity, noting that “health and health care disparities refer to differences in health and health care between groups that stem from broader inequities,” and “health equity generally refers to individuals achieving their highest level of health through the elimination of disparities in health and health care.” The authors discuss the importance of several health care systems and other factors that affect health disparities, including economic stability; education; health care coverage, availability and quality; social and environmental factors; and nutrition. Disparities impact various demographic groups and occur over the course of an individual’s lifetime and across subgroups. These disparities also often overlap, e.g., race and age. The authors further mention that the federal government is working to reduce disparities in certain key populations, including “people of color, low-income populations, women, children/adolescents, older adults, individuals with special health care needs, and individuals living in rural and inner-city areas.”

Disparities that had previously existed were exacerbated as a result of the COVID-19 pandemic. In fact, the Surgeon General first brought attention to health disparity issues two decades ago, highlighting differences in mental health care access and tobacco use among certain racial and ethnic groups. Ultimately, these disparities lower life expectancy, with Blacks or African-Americans living, on average, four years less than whites. Adding to the complexity of identifying factors impacting health care access is the higher prevalence of uninsured among Blacks or African-Americans and Hispanics or Latinos. Job losses resulting from the pandemic have made these issues more acute. In addition, more Black or African-American adults reported negative experiences with health care providers compared to white adults. For Asians or Asian-Americans, Blacks or African-Americans, Native Hawaiians, other Pacific Islanders, and Hispanics and Latinos, the rate of sickness, severe illness, and even death from COVID-19 were higher than their proportion of the population, as these groups experienced greater difficulty accessing testing and treatment. The paper also looked at decreases in mental health and increases in financial insecurity that resulted from the pandemic, which negatively impacted “people of color, low-income people, LGBT people and other underserved groups” at a higher rate than suffered by white people. In addition, the paper discussed lower percentages of Black or African-American and Hispanic or Latino communities vaccinated for COVID-19.

The authors highlight the national economic benefit of addressing health and health care disparities, estimating an annual cost of \$135 billion in medical expenses and lost productivity alone, and discussing the federal government’s current efforts to solve health inequities. The efforts focus on “structural racism and racial inequities in biomedical research,” with an emphasis on providing equitable resources. The National Institutes of Health (UNITE Initiative), the Centers for Disease Control and Prevention, and the Department of Health and Human Services support those initiatives. Further, efforts to expand Medicaid and strengthen the Affordable Care Act, among other measures, are aimed at addressing health care coverage inequities. Additional efforts are focused on maternal health and mortality issues.

Hostetter, M, and Klein, S (27 September 2018). In Focus: Reducing Racial Disparities in Health Care by Confronting Racism. *The Commonwealth Fund*.

<https://www.commonwealthfund.org/publications/2018/sep/focus-reducing-racial-disparities-health-care-confronting-racism>.

Hostetter et al. reviewed inequities in health care access and quality, including less preventive care, for various racial and ethnic groups, with a focus on Black or African-American patients. The authors highlight the inequities that remained after normalizing for “income, neighborhood, comorbid illness, and health insurance type” in the Institute of Medicine’s 2003 report, *Unequal Treatment*. While these results prompted the Institute to “add equity to a list of aims for the U.S. health care system,” efforts have focused primarily on reducing costs and on overall health care quality rather than equity (see the 2010 IOM report, [How Far Have We Come in Reducing Health Disparities?](#)). According to the authors, explicit acknowledgement of racial influences on health care is important. They review the influence of implicit bias, or “learned stereotypes and prejudices that operate automatically and unconsciously” and societal structural racism, which is “the many ways societies foster racial discrimination through housing, education, employment, media, health care, criminal justice, and other systems.” They refer to results of a study by Ronald Wyatt et al., *Achieving Health Equity: A Guide for Health Care Organizations*, which showed that while mortality rates related to heart disease improved from 1990 to 2005, the gap between non-Hispanic, non-Latino Black or African-American and non-Hispanic, non-Latino whites grew.

The authors provide context by discussing the role that the Greensboro, NC “sit-ins” and health discrimination incidents played in health care desegregation and the creation of the Greensboro Health Disparities Collaborative in 2003. Some key findings from the Greensboro Cancer Care and Racial Equity study include:

- *Breast cancer mortality*: Although the incidents of breast cancer are slightly lower among Black or African-American women, the five-year survival rate is much worse: 80% for Black or African-American women compared to 91% for white women. Fewer Black or African-American patients (81%) completed treatment compared to white patients (87%), after normalizing for income, age, marital status, comorbidities and health insurance coverage. Following additional training and the addition of patient navigators that improved health care literacy, the gap between Black or African-American (91%) and white patients (89%) who completed treatment narrowed and reversed slightly.
- *Cancer screening*: In 2009, HealthPartners conducted focus groups and found a significant difference in the prevalence of colorectal screening between persons of color (43%) and whites (69.2%). That effort identified concerns over the inconvenience and invasive nature of traditional colonoscopy screening as driving the differences. Three months after sending fecal immunochemical test (FIT) home screening kits for colon cancer to 3,000 persons of color and encouraging physicians to not communicate bias for traditional screening to patients, the gap had narrowed to 70.1% for “patients of color” and 77.7% for white patients.
- *Chronic conditions*: By addressing misconceptions and encouraging self-monitoring, Kaiser Permanente’s efforts to reduce disparities in chronic, serious illnesses, such as high blood pressure, resulted in a more than 14-point increase in controlled hypertension among Blacks or African-Americans. That reduction nearly closed the gap with white patients.
- *Maternal morbidity and mortality*: The authors also note significantly higher mortality rates for Black or African-American women associated with pregnancy and childbirth, cervical cancer, and to a lesser extent, heart disease. Pregnancy-related death rates among Black or African-American women are three to four times higher than white women and have increased from 2000 to 2014.

The article also discusses other efforts aimed at improving health care equity, including several health care systems that have instituted specific programs.

Lichtenstein, R. (2015). The Vexing Problem of Health Inequalities in the United States. *Medical Care*, 53(3) 215–217. doi: 10.1097/MLR.0000000000000332. [https://journals.lww.com/lww-medicalcare/Citation/2015/03000/The\\_Vexing\\_Problem\\_of\\_Health\\_Inequalities\\_in\\_the.1.aspx](https://journals.lww.com/lww-medicalcare/Citation/2015/03000/The_Vexing_Problem_of_Health_Inequalities_in_the.1.aspx).

Richard Lichtenstein’s commentary succinctly captures the state of health inequalities in the U.S. The author presents four interrelated domains of inequalities:

- *Socioeconomic and demographic factors*: the so-called social determinants of health that include the socially disadvantaged—in income, education, and neighborhood facilities, among other components. Those factors tend to yield higher morbidity and mortality rates, adjusted for age.
- *Access factors*: disadvantaged populations are more likely to lack a “medical home” or access to preventive services, and thus are more likely to forgo or delay medical care for treatable conditions.
- *Health system factors*: Blacks or African-Americans, other underrepresented/marginalized racial and ethnic groups, and low-income populations tend to use different health systems with measurably lower quality of care than higher-income or white patients.
- *Psychosocial factors*: it has been found that, even when the same physician treats patients with identical diagnoses and coverage, non-whites generally receive lower-quality care than whites.

Artiga, S., et al. (16 July 2021) Health Coverage by Race and Ethnicity, 2010–2019. *KFF*. <https://www.kff.org/racial-equity-and-health-policy/issue-brief/health-coverage-by-race-and-ethnicity/>.

The Kaiser Family Foundation (KFF) analyzed health care coverage by race and ethnicity from 2010 to 2019 using both KFF and American Community Survey data. The study found that the ACA impacted health care coverage in these ways:

- Even with enactment of the ACA, the proportion of the population that was uninsured remained higher for marginalized groups relative to whites from 2010 to 2016.
- The prevalence of the population that was uninsured dropped the most for Hispanics or Latinos, falling by almost half from 32.6% to 19.1%.
- After policy changes and the rollback of ACA coverage expansions, uninsured rates for Blacks or African-Americans and Hispanics or Latinos increased from 2016 to 2019.
- Based on 2019 data, non-elderly Black or African-American, Hispanic/Latino, Native Americans, Alaska Natives, and Hawaiian/Pacific Islanders are more likely than whites to be uninsured.
- Uninsured, non-elderly Hispanics or Latinos and Asians are less likely than whites to be eligible for Medicaid coverage because of citizenship requirements.
- Despite efforts to fill gaps in insured populations via programs such as Medicaid and the Children’s Health Insurance Program (CHIP), the authors suggest that the U.S. government lacks the support needed to provide full health care coverage. In states that have chosen to roll back or maintain Medicaid as is, disparities are exacerbated. KFF attributes the high uninsured rates among marginalized groups to the high prevalence of poverty among such groups, which leads to economic barriers in accessing health care coverage. The authors recommend reducing disparities by increasing access to health care coverage through increased subsidies and expanded enrollment eligibility from private and public sources.



Weida, E.B., et al. (18 May 2020). Financial health as a measurable social determinant of health." *PloS one* vol. 15, 5 e0233359., doi:10.1371/journal.pone.0233359.  
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0233359>.

While several studies have focused on the relationship between an individual's financial wellness and health outcomes, Weida et al., suggest incorporating concepts such as financial literacy and "opportunities to save and build wealth" into future studies. These authors propose financial health as a measurable social determinant of health. They define financial health, based on an analysis of the responses to a financial wellness survey, as consisting of four interrelated components: the ability to spend, save, borrow and plan. The authors found that each of these components were significantly associated with various physical and health measures, cementing their role as a social determinant of health. Further, unlike many studies of financial wellness focused on the banked population, the data source for this study looks at the unbanked and underbanked population.

Lechner, A. (2011). Forgoing Medical Care Due to Costs: The Causal Effect of Health Care Burdens on Health. Georgetown University, 2011. Master's Thesis.  
<https://repository.library.georgetown.edu/handle/10822/558168>

Lechner's study establishes a direct causal effect of financial wellness on health by investigating the characteristics of the underinsured population, operationally defined as those who had to forgo medical care because of costs. Lechner found that of those who had to forgo medical care because of costs, about one-third were, in fact, insured. Further, the author found that among those who went without medical care, the probability of a bad health outcome—asthma attack, diabetic complication, pneumonia, heart attack, congestive heart failure, stroke, chronic obstructive pulmonary disease or renal failure—was 36%, compared to 4% among those who received care.

Boschert, J., et al. (January 2020). Understanding the connections between health and financial wellness. *Milliman.com*, Milliman, [https://us.milliman.com/-/media/milliman/pdfs/articles/understanding\\_the\\_connections\\_between\\_health\\_and\\_financial\\_wellness.ashx](https://us.milliman.com/-/media/milliman/pdfs/articles/understanding_the_connections_between_health_and_financial_wellness.ashx).

In this paper, Milliman explores the "strong connection between health and wealth," asserting first that wealth impacts health care access and ultimately health outcomes, whether for the richest or those living in poverty. The general relationship between greater wealth and better health is consistent across races and entire country populations. Milliman found that wealth affects health and health affects wealth. Further, the article concludes that individuals who take a disciplined approach to the future apply that to financial health, such as saving for retirement, as well as physical health, such as living healthier. Bloodwork substantiated the correlation between financial and physical health.

Milliman also noted the strain on many to care for family members—for some, that means caring for aging parents and children at the same time. "In particular, the financial impact on caregivers is more severe when they have to take care of family members with dementia." Due to the time required to care for someone with dementia, caregivers may be unable to work or may have to reduce their work hours or pursue a different type of employment. In addition, when the Milliman paper was published, life expectancy was longer than for prior generations. Increasing life expectancy, fewer working years for caregivers, health care inflation in excess of overall inflation, and the insufficiency of Social Security to pay full benefits by 2035 means that many people are not saving enough for retirement.

Milliman also noted results of a National Public Radio (NPR) survey that supports the association of workplace stress and personal stress, such as caring for family members that are ill, with chronic and other health conditions. The authors note that treatment of medical conditions is not necessarily unaffordable if diagnosed early enough. Some can be prevented. Chronic conditions, in particular, are quite costly. And while diabetes is not among the five most expensive medical conditions according to this paper, long-term treatments are costly and the disease itself often is preventable. Specifically, Milliman notes that one in every seven dollars of U.S. health care spending relates to diabetic care, and on an individual level, diabetic care costs are “approximately 2.3 times higher than nondiabetic patients.” Further, diabetics without health insurance do not receive the medical care or prescription drug treatments that their insured counterparts do. Uninsured diabetics also make 168% more visits to the emergency department. In addition, individuals who are generally unhealthy, physically or mentally, are more likely to need long-term care or other assistance, which can quickly drain retirement savings.

Agency for Healthcare Research and Quality, Rockville, MD (Content last reviewed June 2021). 2019 National Healthcare Quality and Disparities Report. *AHRQ.gov*, <https://www.ahrq.gov/research/findings/nhqdr/nhqdr19/index.html>.

The Agency for Healthcare Research and Quality (AHRQ) prepares an annual report, as mandated by Congress, that analyzes the quality of health care and disparities across racial and socioeconomic groups in the US. The AHRQ highlights that many social determinants of health (SDOH) impact health care quality, and in turn, affect health outcomes. The report notes the World Health Organization’s definition of SDOH—“conditions in which people are born, grow, live, work, and age”—and suggests the following contexts for discussing SDOH: social, economic, education, physical infrastructure and health care. AHRQ assesses quality across the following priorities: “patient safety, person-centered care, care coordination, effective treatment, healthy living, and care affordability.” In the latest report, AHRQ presents the results from an analysis of more than 250 measures from 2000 to 2016–18. Key findings of the study include:

- *Health care access:* Of the health care access measures in the study (having health insurance, having a usual source of health care, timeliness of care and health care infrastructure), more than 50% showed a greater penetration of individuals with health care coverage. Penetration was worse for 20% of respondents and largely unchanged for the remaining 25%.
- *Quality:* The study also showed improvements in health care quality over the time period, but the degree varied across measures (person-centered care, patient safety, healthy living, effective treatment, care coordination and care affordability.)
- *Disparities:* While the study showed improvement in some disparities, inequality for low-income and uninsured groups worsened or was unchanged.
- *Racial and ethnic disparities:* The current data compared health care quality measures relative to whites and found that 40% of the measures were worse for Blacks or African-Americans, Native Americans, and Alaska Natives; more than 33% were worse for Hispanics or Latinos and Native Hawaiians and Pacific Islanders. While almost 30% of the metrics were worse for Asians or Asian-Americans, almost 33% were better.
- *Geographic disparities:* AHRQ also compared health care quality measures against “large fringe metropolitan areas” and found that nearly 25% were worse for individuals living in “large central metropolitan areas” and 33% were worse for individuals living in “noncore” and “micropolitan” areas. For individuals in “medium and small metropolitan” areas, quality measures were worse for slightly less than 20% of quality measures.

Klein, A. and Patel, K (1 June 2021). How financial volatility impacts health care decisions: What we don't know is hurting us. *Brookings*. <https://www.brookings.edu/blog/up-front/2021/06/01/how-financial-volatility-impacts-health-care-decisions-what-we-dont-know-is-hurting-us/>.

This article mainly reviews and comments on a University of Pennsylvania study by Addo and Servon, which included interviews of “racially diverse” college graduates that have health insurance and earn at least \$60,000 annually. In particular, for Americans that live paycheck-to-paycheck, including “a growing number in the middle class,” income volatility, the rising cost of health care, lack of transparency in health care costs, the “quality of their health insurance coverage,” and the lack of dental and mental health coverage under health insurance policies may impact health care access. The authors note that “delaying health care can worsen health outcomes and drive up total health care costs.”

Fintech (i.e. new financial technology) helps moderate income swings through promoting better budgeting and saving but is insufficient in handling the large and unexpected nature of health care expenses. Aaron and Patel note that existing metrics, which compare average health care spending against average income, do not explicitly account for the unexpected nature of health expenses and income volatility among Americans. The authors recommend formulating new metrics which do so. They also recommend integrating financial counseling services into medical offices as is done currently in oncology. Fintech also could help medical providers streamline their existing, outdated medical billing systems.

Alom, S., and Marjanovic, S. (July 26, 2021). Fintech's Role in Reducing Disparities and Financial Barriers to Health Care Access. *The RAND Blog*, RAND Corporation, <https://www.rand.org/blog/2021/07/fintechs-role-in-reducing-disparities-and-financial.html>.

Other than reducing transaction costs and expediting payments, fintech is being considered to improve underserved communities' access to health care financing. Key components include:

- Fintech can be useful in low-income countries as smartphone ownership is increasing. However, out-of-pocket health care costs are high, people do not necessarily have bank accounts, and good credit scores are hard to attain.
- For high-income countries, fintech can be used to bridge accessibility gaps.

Some examples of fintech applications are:

- Digital lending: pre-approved, point-of-care lending with low interest rates, and digital health care loans for emergency, point-of-care needs.
- Health wallet applications, e.g., on mobile phone platforms.
- Insurtech, e.g., using internet-enabled devices to dynamically inform premiums.
- Crowdfunding, e.g., digitally-enabled risk pools without upfront costs or platforms enabling donors to fund costs in specific clinical areas of interest.

Still, a lack of understanding remains over how scalable a given potential application is and how to structure it to increase community impact. Addressing these issues would require better understanding the readiness of target populations to adopt different fintech technologies or how to integrate these technologies into existing health care delivery systems.

Ortaliza, Jared, et al. (14 January 2022). How does cost affect access to care? *Health System Tracker*, Peterson-KFF. <https://www.healthsystemtracker.org/chart-collection/cost-affect-access-care/#item-start>.

This study examines 2019 data from the Centers for Disease Control (CDC) National Health Interview Survey (NHIS) to explore the association of health care costs with health care access. While this data is prior to the COVID-19 pandemic, the authors recognize that one-third of U.S. adults did not access preventive or other medical services during 2020 as they attempted to mitigate their risk of infection. The study found that 89% of adults in the U.S. have health care coverage, and 86% consider themselves to be in good, very good or excellent health. Even before the pandemic, 9% of U.S. adults surveyed said cost was the reason they did not seek medical care or delayed care. Below are more specifics around this observation:

- A higher percentage of Hispanics or Latinos and non-Hispanic, non-Latino Blacks or African-Americans reported foregoing or delaying care; whereas, fewer non-Hispanic, non-Latino whites and even fewer Asian or Asian-Americans reported cost considerations having an impact on the health care they received.
- Twice as many adults who self-assessed as having worse health reported delaying or foregoing coverage as compared to those in better health. More than half of those in worse health worry about whether they will be able to pay health care costs, and one-quarter of those in worse health are likely to have difficulty paying, versus 10% of those in better health. In addition, slightly more individuals in better health do not use a regular medical care facility compared to those in worse health.
- More than five times as many uninsured adults delayed or did not seek medical care compared to those covered by health insurance. Further, the data showed a three-point uptick in uninsured adults foregoing or delaying medical care from 2016–2018, while the uptick was less than one percent for insured adults. More than 75% of uninsured adults worry about whether they will be able to pay health care costs, compared with 22% who are likely to have difficulty paying. A total of 44% of insured adults worry about paying medical costs compared to 11% who are likely to have difficulty paying. In addition, nearly five times the number of uninsured relative to those with health care coverage did not use a regular medical care facility. Of those uninsured, they are almost twice as likely to seek medical care at an urgent care or retail clinic, nearly four times as likely to obtain medical care at an emergency department, and nearly half as likely to seek care in a doctors' office or health center.
- The study bifurcated the data between adults with income above or below twice the poverty line and found that 15% of those below the line avoided or delayed care, compared to 6% of those above the line. Dental services were the most common services foregone or delayed, followed by prescriptions and medical services. Only 5% of mental health services were delayed or foregone.

### B.3 HEALTH CARE COVERAGE AND COST INEQUALITIES

Several authors also looked at studies linking health care coverage and costs to financial wellness, either actual or perceived.

Ketterman, J, and Pippidis, M. (January 2020). Bridging the Gap Between Financial Wellness and Physical Wellness—FSA, HSA, and HRAs. *The Standard Newsletter*, AFCPE, <https://www.afcpe.org/news-and-publications/the-standard/1st-quarter-2020/bridging-the-gap-between-financial-wellness-and-physical-wellness-fsa-hsa-and-hras/>.

Ketterman and Pippidis focus on the role of flexible spending accounts (FSAs), Health Savings Accounts (HSAs) and health reimbursement accounts (HRAs) as tools allowing one to save pre-tax money to plan for future health and dependent care expenses, bridging the gap between financial and physical wellness. They note that a 2018 Federal Reserve study found that 1 in 4 adults skipped medical care because of cost, and 1 in 5 had a large, unexpected

medical bill—a finding that underscores the importance of planning for medical expenses and thus the role of such accounts in helping do so. Both employee and employer may contribute to these accounts. The principal differences are:

- For FSAs, money unused at the end of the plan year is forfeited. For HSAs and HRAs, unused money can roll over.
- HSAs are available only to those with a qualifying high-deductible health insurance plan (HDHP) and are employee-owned. HRAs are similar to HSAs but are employer-owned.

Lee, De Chih, et al. (2021). The convergence of racial and income disparities in health insurance coverage in the United States. *International Journal for Equity in Health*, vol. 20, no. 96, 2021, pp. 1–8. *BioMed Central*. <https://equityhealthj.biomedcentral.com/articles/10.1186/s12939-021-01436-z>.

Using 2017 data from the Household Component of the Medical Expenditure Panel Survey (MEPS), Lee et al. investigated the combined effect of race and income on health care coverage using a vulnerability framework and logistic regression. The study found that income and, independently, race had a significant impact on coverage. When combined, these two factors had a “devastating” effect. Specifically, the odds of a low-income, non-white person being insured were 68% lower than for a high-income white person.

Galbraith, A. A., Ross-Degnan, D., Soumerai, S. B., Rosenthal, M. B., Gay, C., & Lieu, T. A. (2011). Nearly half of families in high-deductible health plans whose members have chronic conditions face substantial financial burden. *Health affairs (Project Hope)*, 30(2), pp. 322–331. <https://doi.org/10.1377/hlthaff.2010.0584>.

While high-deductible health plans (HDHPs) may provide a low-cost alternative to traditional plans because of more affordable premiums, insured cost sharing is higher for HDHPs. Galbraith et al. conducted a study of the comparative financial burden under traditional versus HDHPs for families with one or more chronically ill family members. The survey found that the percentage of “families with chronic conditions” that reported a “health care-related financial burden” was more than twice as high for the group covered under a HDHP (48%) compared with a traditional health plan (21%). For this study, a family was considered to have a health-related financial burden if the family responded that they experienced one or more of the following in the last year: difficulty paying family medical bills; medical costs made it difficult to pay for basic rent, food or heat; or the family set up a medical payment plan. The study also found that the proportion of families that had health care expenses in excess of three percent of income was almost double for families covered by HDHPs (53%) compared to families covered under traditional plans (29%). While tax-exempt health savings accounts or health reimbursement arrangements may be used to pay for the shared costs, according to the authors, “the majority of high-deductible plan enrollees report that they do not have one.”

Kirby, J. and Muhuri, P. (May 2018). Insurance and Access to Care in Urban and Rural Areas, 2014–2015. Statistical Brief #512, *Agency for Healthcare Research and Quality*. [https://meps.ahrq.gov/data\\_files/publications/st512/stat512.pdf](https://meps.ahrq.gov/data_files/publications/st512/stat512.pdf).

The urban-rural health care gap is well-known and hypothesized to result from disparities in access. Kirby and Muhuri used 2014–2015 data from the Medical Expenditure Panel Survey—Household Component (MEPS-HC) to assess patterns in health care disparities along this dimension. The findings are as follows:

- Among those less than 65 years of age, the percentage of uninsured was the same among those living in metro and non-metro areas, but the latter tended to have higher rates of public insurance coverage and lower rates of private insurance coverage.
- Among those aged greater than 65, the percentage relying exclusively on Medicare (vs. some supplemental private insurance) is higher among those living in non-metro areas.
- The percentage of people lacking a usual source of care was higher in non-metro areas for the under-65 group; for the over-65 group no significant differences existed between metro and non-metro.
- Differences tended to be insignificant in the percentage of people reporting unmet need for medical care, dental care or prescription drugs.

Gillen, E.M., et al. (2017). “The Effect of Narrow Network Plans on Out-of-Pocket Cost.” *American Journal of Managed Care*, vol. 23, no. 9, 2017, pp. 540–545. *AJMC*. <https://www.ajmc.com/view/the-effect-of-narrow-network-plans-on-out-of-pocket-cost>.

Narrow network plans promote cost savings by promising in-network providers a more concentrated volume of business in exchange for a lower rate of reimbursement. This study by Gillen et al. looks at whether joining a narrow network plan significantly decreases the mean number of outpatient visits and out-of-pocket expenditures. The results are negative for the former and affirmative for the latter, suggesting narrow networks do not decrease utilization while enabling cost savings. Notably however, this study is limited in that data originates from a single large insurer in the Southeast, and narrow network plans can vary widely in their implementation.

Mitchell, E.M. (January 2021). Concentration of Healthcare Expenditures and Selected Characteristics of High Spenders, U.S. Civilian Noninstitutionalized Population, 2018. *Statistical Brief #533*. *Agency for Healthcare Research and Quality*, Rockville, MD. [https://meps.ahrq.gov/data\\_files/publications/st533/stat533.pdf](https://meps.ahrq.gov/data_files/publications/st533/stat533.pdf)

Emily Mitchell of The Agency for Healthcare Research and Quality (AHRQ) examined 2018 Medical Expenditure Panel Survey (MEPS) health care spending data and found that only five percent of the population drove half of all health care spending. The author also examined the proportion of total health care expenditures by various groups, including race and ethnicity, age and the highest-ranked spenders. The highest one percent of spenders accounted for one in every five health care dollars, and the bottom half of spenders paid for slightly more than three percent of total health care costs. The top spenders included a disproportionate number of older, non-Hispanic, non-Latino white individuals.

Holle, M, et al. (2021). Trends in the Concentration and Distribution of Health Care Expenditures in the US, 2001–2018. *JAMA Network Open*. 4(9):e2125179. doi:10.1001/jamanetworkopen.2021.25179. <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2784058>.

Holle, Wolff and Herant studied 2001 to 2018 MEPS data, which is based on a survey of household expenses for a “nationally representative sample of the U.S. civilian noninstitutionalized population.” In addition to presenting similar statistics on top spenders and noting consistency over the 18 years, the authors highlighted that only 2.3% of the 2018 population incurred half of all prescription drug costs compared to 6.0% of the 2001 population. The pharmaceutical industry’s greater focus on expensive, specialty medications drove the increase in average costs for top spenders, more than ten-fold the 2001 mean when normalized to 2018 dollars.

The authors further subdivided individuals into groups based on whether they were covered under public insurance (with some finer breakdowns), private insurance or uninsured as of the end of the year. Holle, Wolff and Herant also split individuals into family income bands—specifically, “family income relative to the poverty line.” The authors found that despite changes in U.S. demographics and health care coverage, including the introduction of The Affordable Care Act, the proportion of health care expenses paid by high, medium and low spenders changed very little from 2001 to 2018. However, Holle et al. note changes in the concentration of prescription drug costs over the time period. Specifically, the portion of prescription costs related to individuals with private insurance increased. The reason: further ACA impacts that led to a larger proportion of young, healthy individuals comprising the uninsured population, which grew from 35% to 40% during the time period.

Scott, KW, Scott, JW, Sabbatini AK, et al. (2021). Assessing Catastrophic Health Expenditures Among Uninsured People Who Seek Care in U.S. Hospital-Based Emergency Departments. *JAMA Health Forum*,2(12):e214359. doi:10.1001/jamahealthforum.2021.4359. <https://jamanetwork.com/journals/jama-health-forum/fullarticle/2787660>.

Uninsured persons rely on emergency departments (EDs) predominantly for timely medical care. However, they are liable for up to 100% of the billed expenses; while bills can be negotiated down for uninsured persons, the process is slow and complicated, and could involve aggressive collections practices. This can put them at risk for a catastrophic health expenditure (CHE), defined as an annual out-of-pocket medical expenditure exceeding 40% of one’s income. Previous studies have found ED visits resulting in hospitalization for emergency conditions to carry a 70–90% risk for a CHE. This study focuses on ED visits which do not result in hospitalizations, specifically during the time period 2006–2017. The study found the median ED charge from such visits to have increased from \$842 in 2006 to \$2,033 in 2017. Further, the risk of CHE from such visits is 18%, having increased from 13.6% in 2006 to 22.6% in 2017. Those in the lowest income quartile bore the brunt of this risk, with 1 in 3 resulting in a CHE in 2017.

## B.4 CHRONIC CONDITIONS

Glied, S.A., and Zhu, B. (25 September 2020). Not So Sweet: Insulin Affordability over Time. *The Commonwealth Fund*. <https://www.commonwealthfund.org/publications/issue-briefs/2020/sep/not-so-sweet-insulin-affordability-over-time>.

Glied and Zhu looked at 2014 to 2017 data from the National Health Interview Survey (NHIS) and MEPS to assess the out-of-pocket costs of insulin use among those with various levels of health coverage, aged 18–64. They find:

- While compassionate use or discount programs exist to help the uninsured pay for insulin, most of the individuals end up paying full list price.
- The uninsured tend to use cheaper, less effective forms of insulin.
- The uninsured and those with gaps in coverage constitute 80% of those who paid full list price.
- On average, the uninsured spend more on all prescriptions (\$2,456) compared to the privately insured (\$1,274) and are more likely (38% vs. 10%) to report being unable to afford prescriptions.
- Nevertheless, insulin costs remain a significant portion of total out-of-pocket costs for the privately insured.
- Those enrolled in higher-deductible plans reported greater affordability problems than those in conventional plans.

The issue of rising insulin prices is secondary to the lack of health care coverage, and those in copay-only plans tend to fare better than those with coinsurance. Nevertheless, since deductibles and out-of-pocket maximums apply to all care and not just diabetes, the authors recommend policymakers address cost-sharing for all care, confirming that those with diabetes are financially well-protected.

Hayes, S.L., et al. ( 29 August 2016). High-Need, High-Cost Patients: Who Are They and How Do They Use Health Care? A Population-Based Comparison of Demographics, Health Care Use, and Expenditures. *The Commonwealth Fund*. <https://www.commonwealthfund.org/publications/issue-briefs/2016/aug/high-need-high-cost-patients-who-are-they-and-how-do-they-use>.

Hayes et al. studied the demographic characteristics of patients with “complex and costly care needs.” Using data from the 2009 to 2011 Medical Expenditure Panel Survey, the authors recommend ways to improve health outcomes for such patients. The study found that relative to those with chronic conditions and no functional limitations, “high-need” adults—defined as those with “three or more chronic diseases and a functional limitation in their ability to care for themselves or perform routine daily tasks”—had annual expenditures that were two to three times higher and out-of-pocket expenditures which were more than a third higher, despite having lower income. The authors also found that rates of hospital use were more than twice as high for high-need adults. Rates of doctor visits and the use of home health care were found to be significantly higher as well. In terms of demographics, the data showed a disproportionate prevalence of high-need individuals who were older, female, white, less educated, low-income, publicly insured, and in fair or poor self-reported health.



## About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its [strategic research programs](#): aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of [topical research available](#), including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute  
475 N. Martingale Road, Suite 600  
Schaumburg, Illinois 60173  
[www.SOA.org](http://www.SOA.org)