Optimizing Big Data to Inform the Plausibility of Patient Debt

Sebastián R. Díaz and Vikrant Bhatnagar


Abstract

In 2014, states approving the expansion of Medicaid, a tenet of the Affordable Care Act, received monies to expand health care coverage. Using publicly available national datasets, we mined data to explore the effects of expanded health care coverage on medical debt, and also to inform continued optimization of these data for health policy research. Between 2012 and 2015, there was a decrease in the percentage of individuals claiming medical debt. States adopting Medicaid expansion showed a 23.31% decrease, while states NOT adopting had a 19.01% decrease. Expansion of health coverage showed no statistical increase in the proportion of respondents with regards to the number of nights spent in a hospital in a year and the number of times visited a health professional in the past two weeks. Additionally, data illustrated significant findings in the proportion of individuals who responded in the affirmative to the questions associated with cost of care and delay in care, trouble paying for medical bills, and losing coverage after pregnancy. While the data revealed substantive findings, our exploration detected two ways in which these publicly available datasets can be optimized for health policy research: include respondent’s state of residence, and specify codification of data with empty or null values. By including these two variables, health policy researchers can further leverage these invaluable datasets for informing impacts of Medicaid expansion, and possibly informing the plausibility of patient debt as a significant determinant of health. However, publicly available datasets were able to detect the positive ripples of Medicaid expansion within the first year of implementation. These findings can potentially impact public policy by strengthening the argument for expansion of coverage for individuals with scarce resources without drastically increasing the total financial cost of coverage, impacting the workflow of hospitals/clinics, and and limiting uncompensated care provided by hospitals/clinicians.

Introduction

A 2018 recommendation by the National Academies of Sciences, Engineering and Medicine highlights the increasing importance of Data Science: To prepare their graduates for this new data-driven era, academic institutions should encourage the development of a basic understanding of data science in all undergraduates.15 Note this recommendation is not limited solely to STEM careers, as all disciplines are impacted. Data science’s ubiquity is due in part to more publicly available datasets offering new approaches to informing health policy.

Given their size, many of these Big Data sources expand our traditional notions of minimally adequate samples, allowing us to inform relationships among variables with increased confidence. Yet these data repositories are far from ideal, since they were not designed specifically for health policy research. Instead, they result from the Data Revolution, a convergence of advances in computation resulting in data being codified digitally on a massive scale. Big Data for health policy are here to stay, yet to optimize their utility for informing policy, nascent research is needed to evaluate how these repositories can be improved. Toward that aim, this paper leverages the National Financial Capability Study (NFCS) and National Health Interview Survey (NHIS) as examples of publicly available datasets for exploring the relationship between medical debt, out-of-pocket expenses, and Medicaid expansion to illuminate medical debt’s role as a potential social determinant of health. Exploration of this issue, in turn, highlights opportunities for enhancing publicly available datasets like these that inform health policy.
In 2014, states approving the expansion of Medicaid, a tenet of the Patient Protection & Affordable Care Act\textsuperscript{[4]}, received monies to expand Medicaid coverage from a 100% federal poverty level cutoff to 138\%\textsuperscript{[5]}. This expansion in coverage saw 8,067,744 newly eligible enrollees within the 2014 calendar year\textsuperscript{[6]}. By the end of 2016, thirty states and the District of Columbia had expanded their coverage, increasing the number of newly eligible enrollees to 11,996,598, a 49\% increase\textsuperscript{[7]}. To place this in perspective, Medicaid expansion translated to a $17 billion increase in expenditures from 2014 to the end of 2016\textsuperscript{[8]}, or approximately $1,400 per new Medicaid enrollee. Prior research examining the impacts of Medicaid expansion found that expansion leads to an increase in health coverage improving access and utilization of medical care, prescription drugs, and mental health care that otherwise would have been delayed due to reasons such as cost\textsuperscript{[9]}. Subsequently, Medicaid expansion significantly decreased the uninsured rates in “poor and near-poor nonelderly individuals”\textsuperscript{[10]}, a segment of the population most likely to be affected by burdensome costs of medical bills.

In a 2016 Kaiser Family Foundation/New York Times Medical Bills Survey, 73\% of individuals reported having problems paying medical bills for their own medical care\textsuperscript{[11]}. This same survey found that emergency room visits and hospitalizations each comprised over twenty percent of the share of the medical bills\textsuperscript{[12]}. Broken down into ranges, 10\% of individuals who reported trouble paying medical bills in the past twelve months had bills less than $500, 14\% of respondents were between $500 and $1000, 19\% of respondents were between $1000 and $2500, and 24\% of respondents were between $2500 and $5000\textsuperscript{[13]}. To pay off their medical debt, 44\% of individuals used a majority of their savings\textsuperscript{[14]}, draining limited resources needed for food, prescriptions, etc.\textsuperscript{[15]}

Deductibles, coinsurance, copayment for medical services, and costs of services not covered by health insurance collectively define out-of-pocket expenditures\textsuperscript{[16]}. These expenditures have steadily increased for many Americans\textsuperscript{[17]}. Between 2005 and 2015, the average payment toward insurance deductibles increased by 229\%, coinsurance increased by 89\%, copayment decreased by 36\% and collectively, out-of-pocket expenditures increased by 56\% to $4,563\textsuperscript{[18]}. Athena Health sampled n=775,000 patient visits per month from 2012-15 and found that with the removal of inflation, out-of-pocket costs increased by 8\% for pediatric services, 9\% for primary care services, and 13\% for surgical services\textsuperscript{[19]}.

While Medicaid expansion has not resolved medical debt entirely, the rates of medical debt have decreased significantly, as evidenced by NFCS. In December 2017, National Health Interview Survey Early Release Program published that there was a 5.3\% decrease (21.3\% ÷ 15.0\%) in individuals under the age of 65 in families that identified problems paying medical bills\textsuperscript{[20]}. This positive impact is significant, as it represents 13.2 million people whose families are no longer encumbered by the cost of their medical bills\textsuperscript{[21]}. However, this study focused on the impact of medical debt on families rather than individuals, a key distinction from the aforementioned study.

**Materials & Methods**

Given that this research utilizes publicly available, de-identified survey data, no institutional review was required to conduct this study. All analyses were conducted with SPSS statistical software version 24.0 (Chicago, IL: SPSS Inc.) and Tableau visualization software (Seattle, WA: Tableau Software Inc.).

NFCS data were downloaded from the US Financial Capability website\textsuperscript{[22]}, while NHIS data were downloaded from the CDC website\textsuperscript{[23]}. Neither survey ties responses for individual respondents across multiple years’ administrations. Despite this limitation preventing repeated measures comparisons for individual respondents, outcome metrics were compared by year of data collection. Analysis of variance (ANOVA) was utilized to compare mean scores on outcome metrics across the years examined. Given the ordinal nature of the categorical independent variable (i.e. year of administration), the Scheffé test was used to explore post hoc multiple comparisons. Categorical data, which included the survey questions pertaining to medical debt, provided by NHIS was interpreted using the Chi-square test.

One limitation of the NHIS data is that they are not always uniform in scale of measurement, necessitating choosing between the benefit of increased variance versus inclusion of the entire sample. For example, variables related to cost were recorded as specific dollar amounts up to less than $26,000. However, NHIS codifies survey responses involving amounts greater than or equal to $20,000 as a single category. For that reason, outcome metrics related to costs in this study were limited to values below this amount to take advantage, when possible, of the subset of data retaining a continuous scale of measurement.

**Results**
Respondents in the NFCS survey were asked in 2012 and 2015, “Do you currently have any unpaid bills from a health care or medical service provider (e.g., a hospital, a doctor’s office, or a testing lab) that are past due?” With n=27,564 and n=25,509 respondents in 2012 and 2015, respectively, NFCS reported the proportion of affirmative responses to this question for each state. Nationally, reporting of medical debt decreased by 5.67%. The relative percentage decrease per state is illustrated in Figure 1. States that adopted Medicaid expansion had a greater percentage decrease than states not adopting (23.31% vs 19.01%). The NFCS dataset, therefore, allowed us to have a better understanding of the reduction of medical debt given this survey item. To explore further the effects of out-of-pocket premiums and Medicaid expansion, we examined the NHIS dataset for alternative and/or complementary perspectives.

NHIS data collected 2011 to 2016 (n=627,537) were categorized as either pre-Medicaid expansion (2011-13; n=314,526) and post-Medicaid expansion (2014-16; n=313,011). When examining pre- and post-Medicaid expansion data combined, the proportion of females (51.7%), Black/African Americans (14.5%), and geographic sampling from the South (35.4%) are not markedly dissimilar from national trends. NHIS survey respondents reported out-of-pocket premium cost in dollars, using a combined rating scale employing: 1) a continuous scale ranging from $1 up to $19,999; 2) a collapsed category of $20,000 or more, and; 3) options for responding “Refused” or “Don’t Know.” A significant proportion of respondents (n=338,508; 53.9%) had missing values (n=175,374; 55.8% in pre-Medicaid expansion years vs. n=163,134; 52.1% in post-Medicaid expansion years). It is unclear whether this is the result of respondents reporting zero values, as is addressed below in the Discussion section. When selecting only those values for which amounts less than $20,000 were reported, Analysis of Variance yielded a statistically significant omnibus test (F=150.75; df=5; p<0.001) when comparing out-of-pocket expenditures among all the years from 2011-2016 for which data were analyzed. Given the ordinal nature of the categorical independent variable, the Scheffé test was used for post-hoc comparisons (Figure 2). 2014-2015 was the only two-year transition period during which there was no statistically significant difference in out-of-pocket expenditures (Figure 3). The year to year analysis of money spent for individual medical care saw significant increases in the percentage of individuals claiming zero dollars spent in 2014 and 2015, while 2016 had the highest percentage of individuals in the $2,000 – $2,999, $3,000 – $4,999, and $5,000 or more, categories compared to the previous five years.

Respondents of NHIS were asked to indicate the number of nights within a given year that they spent in a hospital. Possible responses included: 1) a continuous scale ranging from 1 to 365; 2) “Refused,” and; 3) “Don’t Know.” When considering only those individuals who reported at least one night’s hospital stay (n=46,880), Analysis of Variance detected no statistically significant differences (F=0.526; df=5; p=0.757) between pre- and post-Medicaid expansion. When these respondents were asked to indicate the number of times in the past two weeks that they visited a health professional with the same response options (with the exception that the continuous scale ranged from 1 to 50) and only individuals who reported at least one visit are considered (n=104,476), Analysis of Variance detected no statistically significant differences (F=0.840; df=5; p=0.521) among pre- and post-Medicaid. The statistical non-significant of these Analyses of Variance tests hint at an intriguing policy implication for expansion of coverage, as addressed below in the Discussion section.

Categorical data analyses of NHIS data comparing response patterns among the years illuminate significant findings. Respondents were asked if medical care was delayed for cost in the past year with the following response options: 1) “Yes”; 2) “No”; 3) “Refused,” and; 4) “Don’t Know.” When considering only those individuals who answered “Yes” (n=47,418), Chi-Square test detected a statistically significant difference in proportions of affirmative responses among years of the survey (df=15; p<0.001). Given the same four response options as above with consideration given only to individuals who answered “Yes”, Chi-Square test detected statistically significant differences among years for participants who indicated in the past year they needed and did NOT get medical care due to the cost of care (n=35,240; p<0.001), being unable to pay medical bills (n=60,860; p<0.001), problems paying medical bills (n=111,228; p<0.001), medical bills being paid off over time (n=154,899; p<0.001), claiming cost of care is too high (n=14,381; p<0.001), and losing Medicaid/Mediplan after pregnancy (n=3,279; p<0.001). Figure 4 illustrates the percentages of individuals responding “Yes” to the aforementioned questions from 2011 to 2016. These results illustrate that implementation of Medicaid expansion in 2014 saw a decrease in percentages for the seven questions listed that were most closely associated with medical debt, potentially indicating Medicaid expansion’s positive impact in easing the burden of medical debt on individuals.

Discussion
The contentious politicization of healthcare coverage policy may obfuscate Medicaid expansion’s successes in decreasing the proportion of Americans reporting difficulty with medical debt. However, our intent here is not to advocate for or against a policy as complex as Medicaid expansion, but instead to explore objectively how Big Data can better inform its outcomes and challenges. Using two publicly available datasets – NFCS and NHIS – our study explored the effects of Medicaid expansion on medical debt and out-of-pocket costs, and its potential role in enhancing access to care, as a case study for informing the improvement of publicly available repositories. Our study also helps catalyze discussions among researchers regarding the limitations of Big Data, that if alleviated, could better inform public policy and clinical decision-making.

The NHIS dataset was useful for illuminating Medicaid expansion’s impact. There was a significant increase in the percentage of respondents mentioning Medicaid in the NHIS survey from 13.7% in 2013 to 15.0% in 2014, reflective of Medicaid expansion and increased coverage for many individuals. Yet loss of Medicaid coverage was a result for many. Medicaid expansion from 100% to 138% of the Federal Poverty Line (FPL) can potentially explain why there was a significant increase in the percentage of individuals that also claimed loss of Medicaid services due to increases in income or new job. As the expansion increased the pool of eligible enrollees, it also increased the probability that individuals between 100% and 138% of FPL could become ineligible as soon as there was an increase in income or new employment.

Two of our findings have important public policy implications. When only considering variables such as: number of nights spent in a hospital and number of times visited a health professional in the past two weeks, our study found them to be statistically non-significant when comparing among pre- and post-Medicaid eras. These results are inconsistent with perceptions that if more individuals are given insurance at no-cost then those individuals are more likely to seek a healthcare provider when not necessary. Even though eleven million more Americans have insurance from Medicaid expansion, there was no statistically significant difference in the number of visits. Furthermore, categorical data analyses revealed a plateau in the percentages of individuals who had been in a hospital overnight, received home care by health professional, visited a health professional in the office, and received care over ten times in the past year. These findings can potentially impact public policy by strengthening the argument for expansion of coverage for individuals with scarce resources without drastically increasing the total financial cost of coverage, impacting the workflow of hospitals/clinics, and and limiting uncompensated care provided by hospitals/clinicians.

The utility of these datasets to inform potential impacts of Medicaid expansion do present limitations. Our attempts to leverage these datasets to compare and contrast pre- and post-Medicaid expansion outcomes were challenging. One of the complicating factors for studying Medicaid expansion is states’ adoption varied not only with respect to whether or not they did so, but also when. Utilization of the January 1, 2014 date as a cutoff posed a unique problem. Some states never adopted Medicaid expansion, while others adopted at a later date beyond the cutoff. Ideally, therefore, datasets like NHIS could include the respondent’s state of residence to allow researchers to statistically control for the actual adoption date for Medicaid. Addition of this single variable would be invaluable to explore how Medicaid expansion impacted rural states that are more likely to have a populace that experiences a disproportionate amount of medical debt and diminished access to care relative to other states.

Unfortunately, publicly available NHIS data specify only the broader geographic region in which the respondent resides. Our goal here is not to criticize the data with respect to our unique, individual research needs. The NHIS data made publicly available by the CDC provide an excellent resource for investigators wishing to research health outcomes at the broader national level. Regardless, we wish to draw attention to the need for policy researchers’ and others’ input on how these datasets may be designed to optimize their informatics needs.

We hope to raise awareness about another issue we encountered in data exploration. There were certain cells within the data that were empty or null values. When answering questions, respondents were given the opportunity to either provide a categorical (Yes or No) or numerical (e.g. 1 to 50) response and additional options such as “Refused” and “Don’t Know” were also provided. A null, empty response by individuals can impact both the reliability and validity of results. Even though our use of casewise deletion in analyses of these data (versus imputation) did not significantly compromise statistical power, appropriate use of publicly available data repositories requires understanding the specific contexts for how they are codified. Consistent with calls for interprofessional approaches to healthcare, we also need interprofessional approaches to data governance, and to ensure that diverse stakeholders who will increasingly utilize these informatics resources have the opportunity to inform the design of these datasets.
Future research can further inform Medicaid expansion phenomena, yet these same studies can help evaluate formatively how publicly available datasets can be further optimized to inform health policy. By replicating these year-over-year comparisons in health outcomes with alternative datasets, we can extract the advantages and disadvantages of each, and thus begin to develop best practices for how to structure data repositories accordingly. In particular, by subjecting their own repositories of data to replications of this type of study, large provider systems can not only inform the phenomenon explored, they can also evaluate their own data systems’ utility for informing internal research. Moreover, by confining a dataset to a single provider system, researchers could explore outcome metrics longitudinally for a given patient. Second, by limiting data to patients served within a particular Medicaid option state by that provider system, all subjects in the study will have been subjected to Medicaid expansion within the same timeframe.

**Conclusion**

Much more research is needed to inform the impact of Medicaid expansion on medical debt. Regardless, our study found existing publicly available datasets do allow detection on positive impacts on reducing medical debt. Yet to explore these phenomena in more granular fashion and to better control for covariates, these publicly available datasets need to include additional key variables (e.g. state of residence for respondent/subject) that can illuminate potential impacts on patient’s access to and utilization of healthcare services, while simultaneously informing how to optimize these data, more broadly, for healthcare informatics.

**Acknowledgements**

The author wishes to thank the Osteopathic Heritage Foundation for providing a fellowship to pursue a dual-degree (DO/MBA) at Ohio University Heritage College of Osteopathic Medicine.

**Illustrations**

![Figure 1](image1.png)

Figure 1. Relative percentage decrease for the respondents in the NFCS survey answering “Yes” to the question, “Do you currently have any unpaid bills from a health care or medical service provider (e.g., a hospital, a doctor’s office, or a testing lab) that are past due?” States adopting Medicaid expansion as of 2015 appear in blue & non-expansion states are in orange.

![Figure 2](image2.png)

Figure 2. Scheffé test was used for post-hoc comparisons to determine statistical significance between years, with significance detected between every set of years except 2014 & 2015.
Figure 3. Average out-of-pocket costs were calculated from 2011 to 2016. Statistically significant differences were seen when two separate years were compared. However, 2014-2015, represented in orange, was the only set of comparison that were not statistically significant different.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unable to pay medical bills</td>
<td>62.8%</td>
<td>55.0%</td>
<td>54.3%</td>
<td>53.0%</td>
<td>53.5%</td>
<td>51.3%</td>
</tr>
<tr>
<td>Medical care delayed for cost</td>
<td>8.9%</td>
<td>8.5%</td>
<td>7.9%</td>
<td>7.1%</td>
<td>6.5%</td>
<td>6.3%</td>
</tr>
<tr>
<td>Needed and did NOT get care due to cost</td>
<td>11.1%</td>
<td>12.1%</td>
<td>10.8%</td>
<td>10.8%</td>
<td>10.8%</td>
<td>10.8%</td>
</tr>
<tr>
<td>Cost is too high</td>
<td>41.0%</td>
<td>43.1%</td>
<td>43.0%</td>
<td>42.6%</td>
<td>42.6%</td>
<td>42.2%</td>
</tr>
<tr>
<td>Medicaid/Medicaid stopped after pregnancy</td>
<td>5.0%</td>
<td>4.4%</td>
<td>4.4%</td>
<td>4.2%</td>
<td>4.1%</td>
<td>2.7%</td>
</tr>
<tr>
<td>Medical bills being paid off over time</td>
<td>25.3%</td>
<td>25.7%</td>
<td>24.9%</td>
<td>24.1%</td>
<td>23.6%</td>
<td>23.9%</td>
</tr>
<tr>
<td>Problems paying medical bills</td>
<td>20.3%</td>
<td>19.6%</td>
<td>18.8%</td>
<td>17.1%</td>
<td>15.7%</td>
<td>14.8%</td>
</tr>
</tbody>
</table>

Figure 4. Percentages of individuals responding “Yes” to questions associated with medical debt that were asked to NHIS respondents from 2011 to 2016. Implementation of Medicaid Expansion in 2014 saw a decrease in percentages in the seven listed questions. Chi-square test was utilized to determine statistical significance.

About the Authors

Sebastián R. Díaz, PhD, JD is an Associate Professor of Family Medicine, at Ohio University Heritage College of Osteopathic Medicine.

Vikrant Bhatnagar is a fourth-year medical student at Ohio University Heritage College of Osteopathic Medicine.
