

Data point: Medical debt and credit scores



Consumer Financial
Protection Bureau

May 2014

- Kenneth P. Brevoort
- Michelle Kambara

This is another in an occasional series of publications from the Consumer Financial Protection Bureau's Office of Research. These publications are intended to further the Bureau's objective of providing an evidence-based perspective on consumer financial markets, consumer behavior, and regulations to inform the public discourse.

Table of contents

Table of contents.....	3
1. Introduction.....	4
2. Data.....	7
3. Empirical approach	9
4. Result I: Medical versus non-medical collections	13
5. Result II: Paid versus unpaid medical collections.....	17
References.....	20

1. Introduction

Information about unpaid medical bills is reported to the nationwide credit reporting agencies (NCRAs) from two sources. It can be reported directly by the medical service provider (e.g., a doctor) or by a third-party debt collection agency that has purchased the debt or been contracted to collect it. The vast majority of medical debt reported to the NCRAs – about 99.4 percent of accounts – is reported by collection agencies. We refer to the medical bills that collection agencies report as “medical collections.”

Once reported, the information about these medical collections is reflected on the credit records that the NCRAs maintain. The information about these medical collections are then used by credit scoring models, such as the FICO and VantageScore models, to derive predictions about the creditworthiness of consumers. These predictions, called “credit scores,” can have a large effect on a consumer’s access to credit.

Credit scoring models generally do not differentiate between medical collections and the other debts that are reported by third-party collection agencies (“non-medical collections”), which include things like unpaid rent or cell phone bills.¹ Traditionally, such models have also not differentiated between collections that have been fully repaid (“paid collections”) and those that remain unpaid (“unpaid collections”).²

¹ The collections evaluated in this Data Point, whether medical or non-medical, reflect debts for non-credit-related bills only. Past due credit card debts, or other credit-related debts, that are in collection are not treated as collections in this Data Point, as these debts are treated differently than the collection accounts that are our focus. For more information about collections, see Avery, *et al.* (2003).

² The most recent version of the VantageScore, version 3.0, excludes paid collections (medical or non-medical). Earlier versions of the VantageScore treated paid and unpaid collections equally.

The use of medical collections in credit scoring models has generated concerns stemming from the unique circumstances under which these debts arise and come to be reported to the NCRAs.³ Among their unique characteristics is that consumers may sometimes be unaware that the medical collections exist. For example, a consumer using the CFPB's complaint system reported a situation like this only last week:

I am an insured individual, yet I have had issues with proper billing practices at hospitals. Twice this has resulted in collections being reported to the credit bureaus on bills that were not properly submitted to my insurer and that were not communicated to me as a patient. Once I tried to resolve these issues, the hospital billing departments are not helpful and the collection agencies are impossible to track down. There's so much focus today on affordable health care and insurance coverage, yet hospitals cannot properly bill those patients who do have insurance.

If consumers are unaware of their medical collections or they view them as illegitimate because they are charges their insurance should have paid, then these debts may provide little information about their creditworthiness.

The CFPB seeks to make decisions in an evidence-based way. In pursuit of that goal and others, Bureau staff have constructed a data resource called the Consumer Credit Panel (CCP), which contains de-identified credit records for a representative sample of consumers. This resource is ideally suited to examine the role of medical collections in credit scoring models and to assess whether the identical treatment of medical and non-medical collections is justified by subsequent debt payment patterns.

Our analysis has two components. First, we examine whether medical and non-medical collections are equally predictive about the subsequent respective credit performance of consumers with these different types of accounts. The answer appears to be “no.” Our results suggest that consumers with more medical than non-medical collections had observed delinquency rates that were comparable to those of consumers with credit scores about 10 points higher.

³ For a summary of the concerns that have been raised about the use of medical collections in credit scoring models, see Rukavina (2014).

Second, we conduct a similar analysis that evaluates whether paid and unpaid medical collections are equally predictive of consumer delinquency rates. Again, the answer appears to be “no.” Consumers with more paid than unpaid medical collections had delinquency rates that were comparable to the rates of consumers whose credit scores were roughly 20 points higher. That is, consumers with paid medical collections were less likely to be delinquent than other consumers with the same credit score.

The remainder of this Data Point provides the details of our analysis. In the next section, we discuss the data used in this study. We then discuss the empirical approach used to examine whether medical and non-medical collections, as well as paid and unpaid medical collections, are equally predictive about a consumer’s likelihood of delinquency. We then present the results of our two analyses in Section 4 and 5.

2. Data

The data used in this study come from the CFPB's Consumer Credit Panel (CCP). The CCP is a longitudinal, nationally representative sample of approximately 5 million de-identified credit records from one of the three nationwide credit reporting agencies (NCRAs).

This study uses data from two time periods: September 2011 and September 2013. The entire credit record from both dates was supplied, excluding any direct personally identifying information (such as name, address, Social Security Number). The included information provides account-level information about the debts of consumers in our sample and a commercially-available credit score.

The account-level information that is included in the credit records comprising the CCP allows us to identify which debts reported by third-party collection agencies were from medical or non-medical bills. While we can identify those collections that were from medical bills, nothing in the data reveals anything about the identity of the medical service provider, the type of institution that provided the service, or the nature of the services that were performed.

Using two time periods is a conventional practice to evaluate credit scoring models (Board of Governors, 2007). The two periods allow us to compare the credit record characteristics of consumers at the earlier period with their performance on credit obligations during the ensuing “performance period” from October 2011 to September 2013. We use September 2011 as the start of the performance period because it is the earliest date for which medical collections can be identified in the CCP.⁴ We then use September 2013 as the end of the performance period to

⁴ One of the variables necessary to determine whether a third-party collection was medical or non-medical was not available for earlier dates in the archives from which the CCP data were drawn. As a result, the predictive value of medical collections cannot be evaluated for time periods before September 2011.

provide 24 months of observable performance, the standard length of time used to evaluate credit scoring models.

We evaluate performance using two different but related measures of credit performance (Board of Governors, 2007). The first is the “any account” performance measure, which evaluates performance on both new and existing accounts. In this analysis, a “new account” means one that was opened during the first three months of the performance period (October to December 2011). An “existing account” means one that was already open and in good standing at the start of the performance period. The any account performance measure approximates the type of performance that is used to construct generic credit history scoring models and allows us to evaluate performance for most of our sample.

Our second measure evaluates performance on new accounts only. Focusing on new account performance is an arguably superior approach. When using a performance measure that includes existing accounts, such as the any account performance measure, the characteristics of existing accounts can both directly factor into the calculation of the credit score and affect the performance measure. For example, a mortgage that has been outstanding for a long period of time may both affect a consumer’s credit score and her likelihood of paying on time, since older mortgages likely include more equity. New accounts do not affect the calculation of the credit score, as these accounts did not exist when the score was calculated. As a result, they may provide a better test of a credit scoring model’s predictiveness. The drawback of this performance measure is that only a small portion of the sample, about 16 percent, opened a new account during this window, so we can only evaluate performance for these consumers.

Both performance measures are calculated using the same procedure. Each consumer’s performance is categorized based on that consumer’s worst performing account. The performance measure is assigned a value of 1 for consumers who had at least one account that was 90 or more days past due during the performance period. Otherwise, the performance measure is assigned a value of 0. Consumers whose performance was unobservable are excluded from the analyses.⁵

⁵ For the analysis using new account performance, we cannot observe the performance of consumers who did not open new accounts during the first three months of the performance period. For the analysis using any account

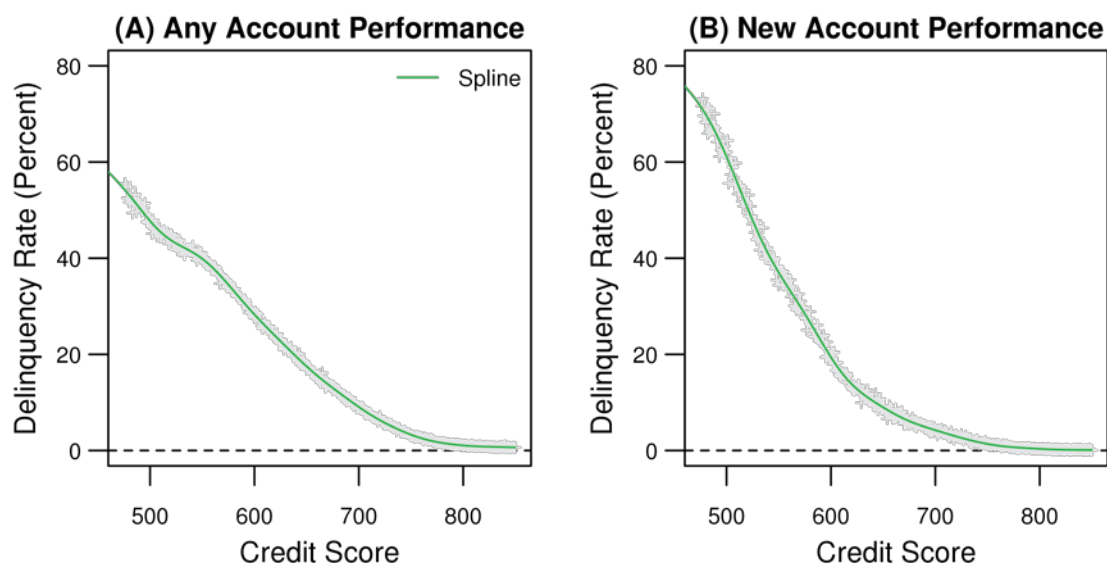
3. Empirical approach

To understand the approach we take, consider two consumers with identical credit records, at the start of the performance period, neither of whom has any collections. Because their credit records are identical, both will have the same credit score, say 780, and would be expected to have the same likelihood of delinquency during the ensuing performance period. Now assume that at the start of the performance period each of the consumers had a debt collection reported on their credit record, one a medical collection and the other a non-medical collection. If the scoring model treats medical and non-medical collections equally, then the scores of both consumers will be decreased by the same amount. Using the estimates published by Johnson (2012), we might expect the scores of these consumers to be decreased by about 115 points relative to the starting assumed credit score of 780. Both consumers would now have scores of 665. Since lower credit scores suggest greater risk, lenders would interpret this as reflecting an increased likelihood of delinquency during the performance period.

But what if medical collections were less predictive about a consumer's credit risk than non-medical collections? In this case, the non-medical collection would increase the expected likelihood of delinquency by more than the medical collection. If the credit scoring model nonetheless treated both types of collections equally, these consumers would both have 665 scores. This means that, if medical collections are truly less predictive about a consumer's creditworthiness than are non-medical collections, consumers with medical collections should perform better (that is, have lower delinquency rates) than other consumers with same credit scores (in this example 665) but with non-medical collections.

performance, performance is unobservable for consumers who either (a) did not have accounts with payments due during the performance period, or (b) did not have accounts that met the definitions for new or existing accounts.

FIGURE 1: PERFORMANCE BY CREDIT SCORE



This is the comparison we make in this Data Point. To facilitate the comparison, we take the portion of our sample with credit scores, regardless of whether or not they have collections, and estimate the relationship between credit scores at the start of the performance period and the delinquency rate of these consumers during the subsequent two year performance period. We do this calculation for both performance measures. These are shown in Figure 1. In both panels, the gray crosses show the average delinquency rate at each credit score and the green lines show the estimated relationship between the credit score and delinquency during the performance period.⁶ As shown, the green lines in both panels are decreasing, suggesting that these credit scores rank-ordered these consumers based on the credit risk they posed.

The next step is to divide the population of consumers who have collections into two groups. In the simple example that we used at the start of this section, consumers had either a medical or a non-medical collection; but in practice, most consumers with medical collections also have non-medical collections. So we divide the population into two groups: consumers with “mostly medical” collections (MM), meaning that they had more medical than non-medical collections, and consumers with “mostly non-medical” collections (MNM), who had more non-medical than

⁶ We estimated the relationships between credit scores and performance in this figure and all the other figures in this Data Point using penalized splines.

medical collections.⁷ Consumers with an equal number of medical and non-medical collections (including consumers without collections) are excluded from these two groups.

For both of these groups, and for both performance measures, we estimate the delinquency rate during the performance period for consumers at each score level as of the start of the performance period. We use these delinquency rates to estimate the extent to which consumers were over- or under-penalized for their medical collections. We do this by taking the delinquency rate calculated at each score level and finding the score level for the general population that exhibited the same delinquency rate. The difference between this score and consumers' actual scores estimates the amount they were over- or under-penalized.

For example, MM consumers who had a credit score of 620 had an estimated delinquency rate, using new account performance, of 12.8 percent. Using the relationship from Figure 1, we find that this level of performance was comparable to consumers with a credit score of 625, a score difference of 5 points. This implies that MM consumers with 620 credit scores performed like consumers whose scores were 5 points higher. We infer from this positive score differential that they were “over-penalized” by the scoring model.

In contrast, MNM consumers with a 620 credit score had an estimated delinquency rate of 16.2 percent. According to Figure 1, this performance was comparable to consumers with scores of 610, about 10 points lower. Based on their observed performance, the scores of these consumers should have been lower than they were. We infer from this negative score differential that the scores of these consumers were “under-penalized.”

We calculate these score differentials for each MM and MNM consumer. If the credit scoring model's equal treatment of medical and non-medical collections were appropriate, then we would expect the score differentials for both groups to be zero, indicating that both groups exhibited delinquency rates consistent with other consumers with the same credit score. Alternatively, if medical collections are less predictive about creditworthiness than non-medical collections, then we would expect to observe predominantly positive score differentials for MM consumers, reflecting that their scores had been over-penalized, and predominantly negative score differentials for MNM consumers.

⁷ The groups are designated using only collections that are large enough to be factored into the credit scores in our data.

We conduct this analysis for both performance measures and present the results in Section 4. Additionally, we also use this same approach to examine whether the equal treatment of paid and unpaid medical collections was consistent with the subsequent performance of consumers with these collections. In that case, instead of dividing consumers into groups according to whether they had mostly medical or mostly non-medical collections, we divide the population of consumers with medical collections into those who had more paid than unpaid medical collections and those with more unpaid than paid medical collections. We refer to these groups as having “mostly paid medical” (MPM consumers) or “mostly unpaid medical” collections (MUM consumers). Again, consumers who had an equal number of paid and unpaid medical collections, or who had no medical collections, were excluded from these groups. We then calculate the score differentials for both groups as described above. We present the results of this analysis in Section 5.

There are caveats that go along with these analyses. First, we examine only a single performance period that covers October 2011 through September 2013. Due to data limitations, we are unable to analyze other time periods. Credit conditions can and do change over time. While we have no reason to believe that this time period is materially different, analyses for other time periods could produce different estimates.

Second, while we have tried to use measures of credit performance that reflect the types used in industry in constructing and validating credit scoring models, it is possible that other measures of performance could yield different results. It is also possible that medical collections will be more or less informative for specific types of credit (*e.g.*, credit cards, auto loans) or in custom models constructed for individual lenders.

4. Result I: Medical versus non-medical collections

In this section we report the results of our comparison of credit performance of consumers with medical and non-medical collections, using our two performance measures.

FIGURE 2: PERFORMANCE BY CREDIT SCORE, MEDICAL VS. NON-MEDICAL COLLECTIONS

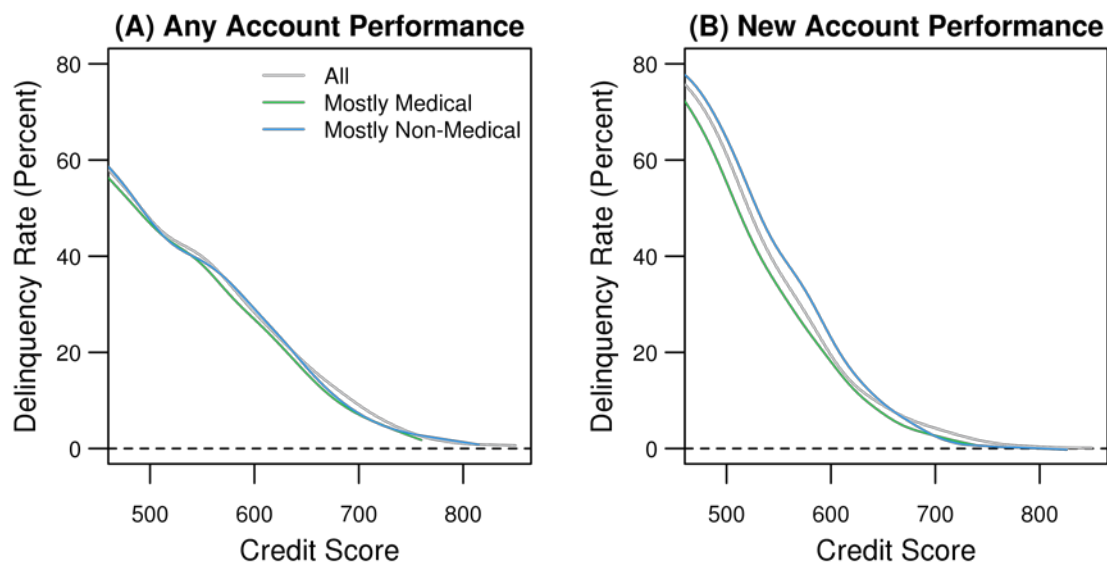


Figure 2 shows the estimated relationship between credit score and the delinquency rate during the performance period. The left panel shows the relationship using any account performance and the right panel using performance on new accounts. In both panels, we show three lines. The gray line is the relationship between credit score and performance for all of the consumers in our sample. These are the same lines that were presented earlier in the two panels of Figure 1. The green and blue lines show the relationship for MM and MNM consumers, respectively.

Because consumers with collections were unlikely to have very high scores, neither the green nor the blue line covers the entire credit score range. For example, none of the consumers in our data with a medical collection had a credit score above 780. As a result, we can only estimate the relationship between credit score and performance for MM consumers up to this score.

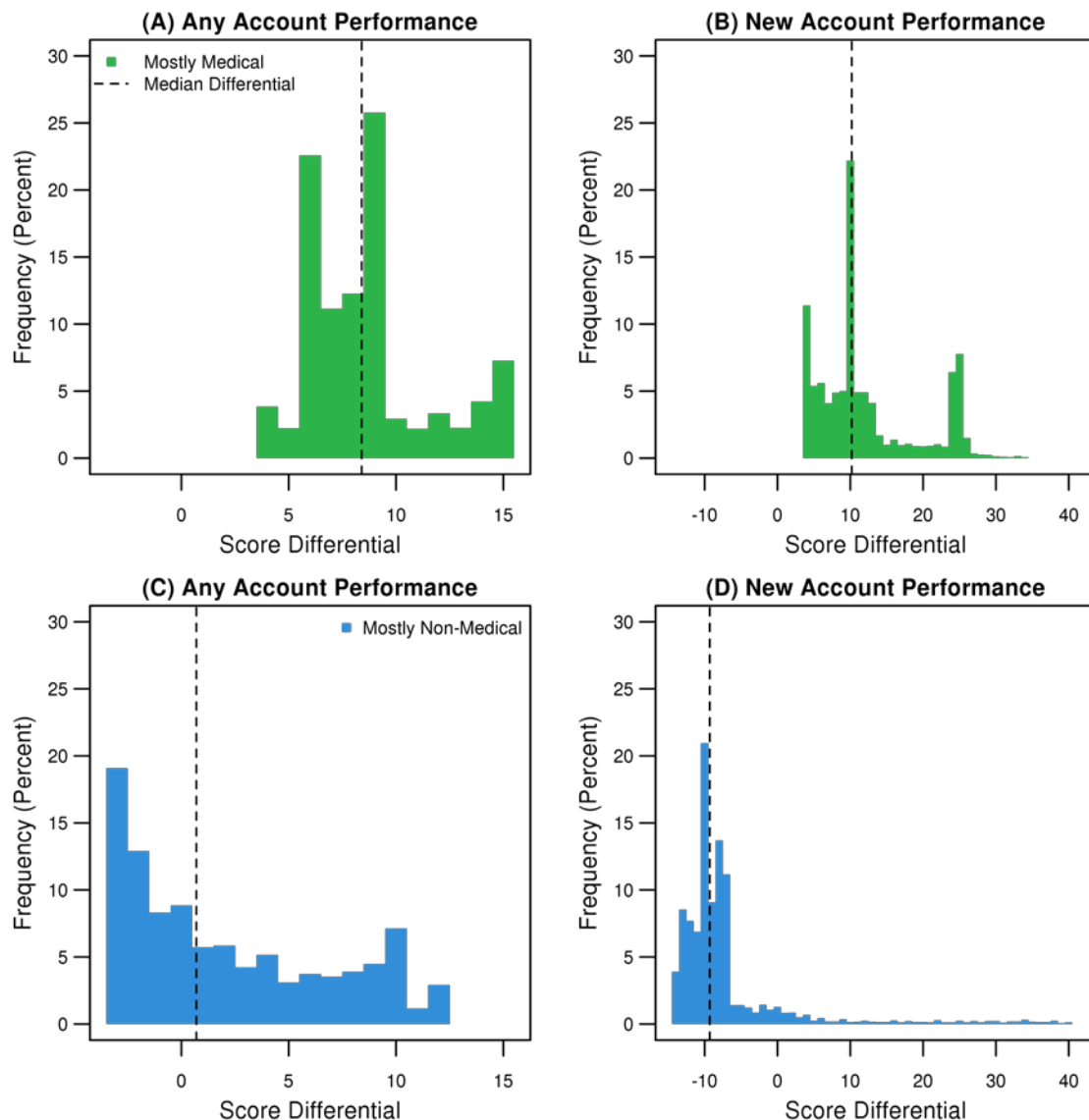
For the green line, both panels exhibit a similar pattern in that the green line is consistently below the gray line. This means that for any given score level, MM consumers had lower delinquency rates than were observed for the population as a whole.

The pattern for the blue line, which represents MNM consumers, is more nuanced. Using the any account measure, these consumers performed comparably to other consumers with the same score, with the exception around scores of 700, and to a lesser extent around 550, where these consumers appear to have over-performed their scores. The pattern is less nuanced for the new account measure as these consumers under-performed their scores over most of the score range, again with the exception being scores around 700 where these consumers over-performed their scores.

If we directly compare the MM and MNM consumers, however, there is a clear pattern. Across almost all of the score range, MM consumers had lower delinquency rates than MNM consumers. For the new account performance, these differences appear substantial.

The comparison of these lines, however, may be misleading, as the distribution of scores for consumers with collections is concentrated at lower score levels. For example, only about 16 percent of MNM consumers have credit scores that are between 637 and 755, the range where we observe over-performance, which covers almost one-third of the score range shown. It is also difficult to interpret the scale of these performance differences in terms of the amount of over- or under-performance implied.

FIGURE 3: DISTRIBUTION OF SCORE DIFFERENTIALS, MEDICAL VS. NON-MEDICAL COLLECTIONS



To provide a clearer comparison of the differences between these groups, we use the score differentials that were described in Section 3. Figure 3 shows the distribution of score differentials that were calculated based on the performance relationships shown in Figure 2. The top two panels, labelled Panels (A) and (B), show the score differentials for MM consumers using the any and new account performance measures, respectively. The dashed black line in each panel shows the median score differential. These data, which show consistently positive score differentials, reinforce the consistent pattern of over-performance that was apparent from Figure 2. The median MM consumer over-performed her score by about 8 points using the any account performance measure and 10 points using new account performance.

The pattern for MNM consumers is shown on the bottom in Panels (C) and (D). The score differentials using any account performance spanned a wide range, but were centered close to zero (the median score differential was less than 1 point). Score differentials using new account performance were predominantly negative. This suggests that the score penalty that these consumers received for their collections were either roughly consistent with their subsequent performance or too low given their performance.

A comparison of the distribution of score differentials across these two populations clearly indicates that MM consumers were much more likely to over-perform their scores. These differentials suggest that medical and non-medical collections are not equally predictive of delinquency. Instead, the equal treatment of medical and non-medical collections in the credit scoring model we examine appears to have over-penalized MM consumers.

5. Result II: Paid versus unpaid medical collections

This section presents the results of our analysis comparing the performance of consumers with paid and unpaid medical collections.

FIGURE 4: PERFORMANCE BY CREDIT SCORE, PAID VS. UNPAID MEDICAL COLLECTIONS

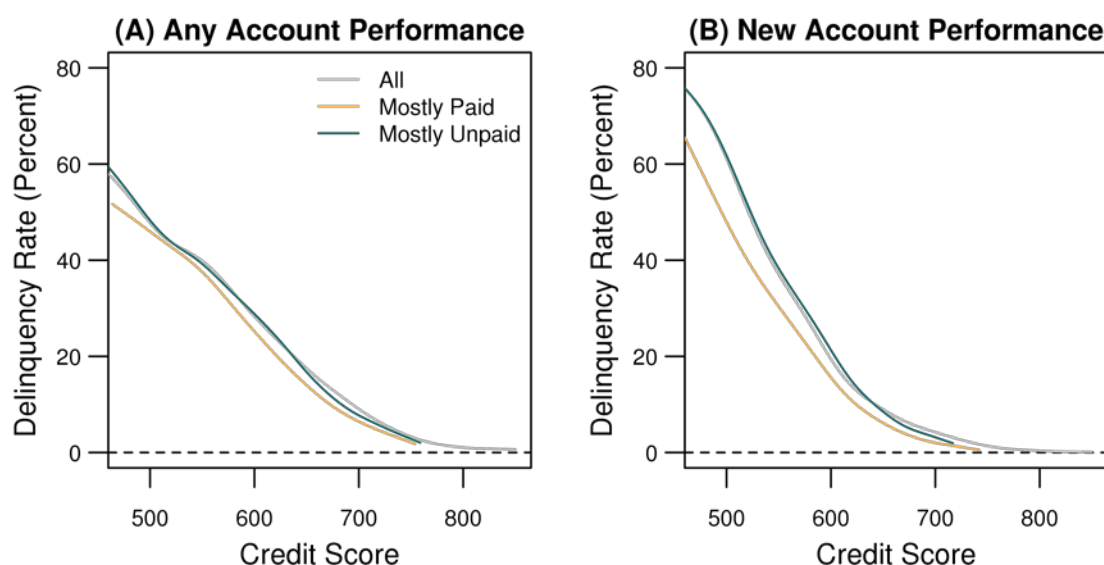


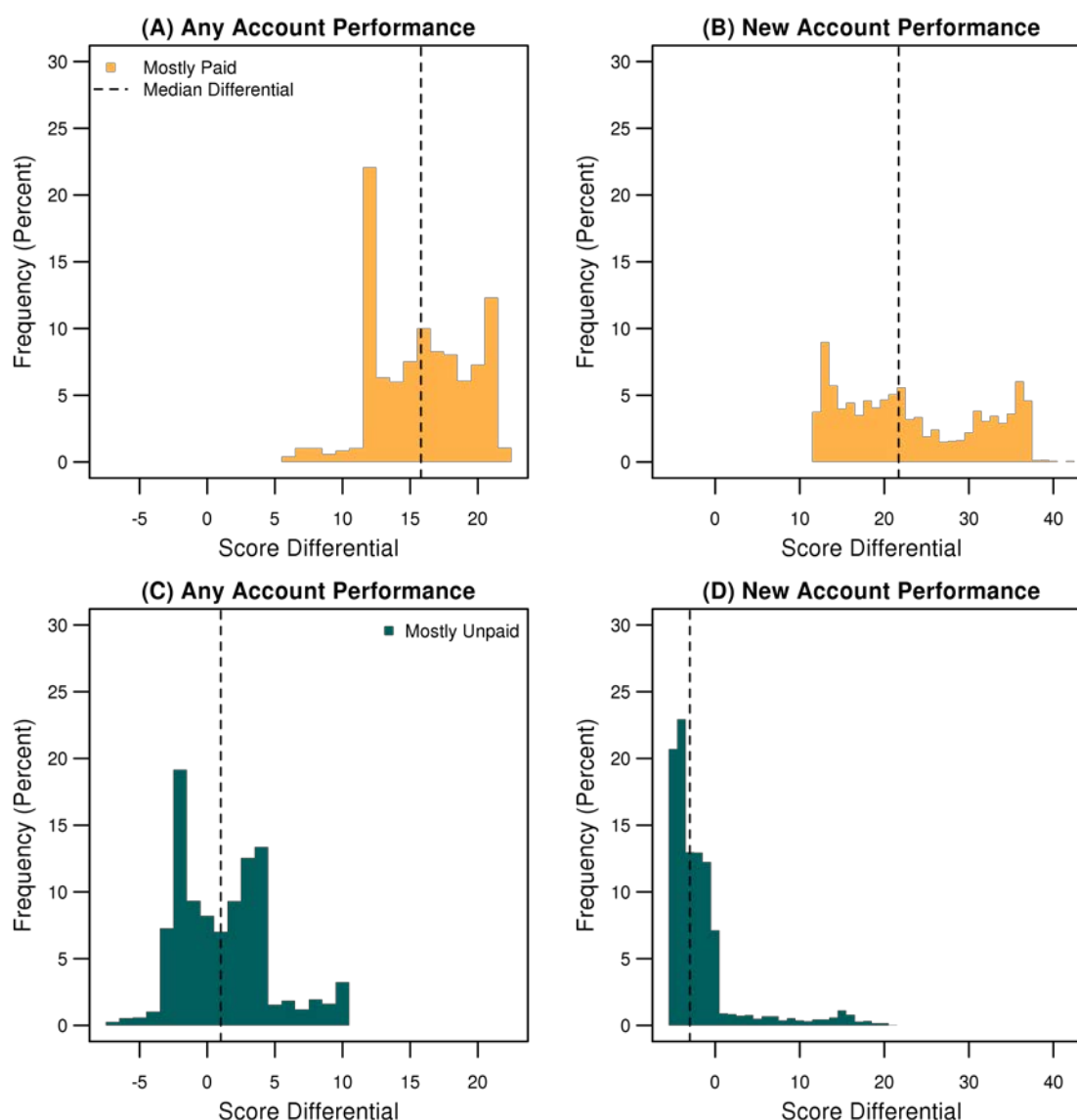
Figure 4 presents the estimated relationship between credit score and delinquency during the performance period. Panel (A) shows the relationship using any account performance and Panel (B) using new account performance. As in the last section, the gray lines show the relationship using our entire sample. In this figure, the gold line represents MPM consumers and the teal line represents MUM consumers.

The patterns in this figure are much less nuanced than was observed in the previous section. MUM consumers, shown by the teal line, had delinquency rates that were generally comparable

to those observed for the general population, though delinquency rates appear to have been slightly higher over much of the score range using new account performance.

In stark contrast, MPM consumers had delinquency rates that appear well below the levels experienced by other consumers with the same scores. This pattern of over-performance, or lower delinquency rates, is observed over the entire score range.

FIGURE 5: DISTRIBUTION OF SCORE DIFFERENTIALS, PAID VS. UNPAID MEDICAL COLLECTIONS



Using these estimated relationships between credit scores and performance, we calculate the score differentials. These are presented in Figure 5. The top two panels show the score

differentials for MPM consumers and the bottom two panels the differentials for MUM consumers. Again, the median score differential is labelled by the black dashed line.

These figures show that MPM consumers appear to have been substantially over-penalized for their paid medical collections. The median score differential was approximately 16 points using the any account performance measure or 22 points using new accounts.

For MUM consumers, again there is limited evidence of a similar pattern. Using the any account performance measure, these consumers performed comparably to other consumers with the same scores (the median differential was 1 point). The new account performance measure yields predominantly negative score differentials, suggesting that the scores of these consumers were higher than they should have been given their subsequent performance. That is, they appear to have been under-penalized.

References

- Avery, Robert B., Paul S. Calem, Glenn B. Canner, and Raphael W. Bostic. 2003. “An Overview of Consumer Data and Credit Reporting,” *Federal Reserve Bulletin* 89(2): 47-73.
- Board of Governors of the Federal Reserve System. 2007. *Report to the Congress on Credit Scoring and Its Effects of the Availability and Affordability of Credit*. (Washington, DC: Federal Reserve Board).
- Johnson, Carla K. 2012. “How to Check your Credit for Medical Debt Snags,” *Washington Times*, March 4, 2012. Available online at <http://www.washingtontimes.com/news/2012/mar/4/how-to-check-your-credit-for-medical-debt-snags/> (Last visited, May 19, 2014).
- Rukavina, Mark. 2013. “Medical Debt and Its Relevance When Assessing Creditworthiness,” *Suffolk University Law Review* 46(3): 967-982.