# Understanding and Assessing the Usefulness of Present on Admission Indicators as a Predictor of Hospital Readmission

A Dissertation

SUBMITTED TO THE FACULTY
OF
THE UNIVERSITY OF MINNESOTA

James E. Rosenthal

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Stephan T. Parente, PhD, Adviser

August 2014

# Abstract

***Objective***

The objective of this study is to evaluate Present on Admission (POA) indicators as a new data source for which to model hospital readmissions. POA indicators for have been in administrative claims data since 2008. POA indicators' primary purpose is to identify Hospital Acquired Conditions (HACs), which represent 0.14% of overall claims. The remaining non-HAC POA data then falls into a category called "other." This study attempts to gage the secondary usefulness of POA indicators in aiding hospital readmission modeling.

***Methods***

This study used Medicare inpatient 5% Limited Data Sets (LDSs) for the years 2008 through 2011. Patient histories were assembled, index and readmission events were established, and datasets representing the primary diagnosis conditions of Acute Myocardial Infarction (AMI), Heart Failure (HF), and Pneumonia (PN) were extracted. CMS methodologies were followed consistent with the limitations of the source data. A base logistic regression model was created to approximate the CMS hospital readmission models. Three readmission periods were examined: 7 days, 15 days, and 30 days. To this base, three POA variables were developed to address the following research questions: P1) Does the presence of any POA=no indicator (condition occurred after admission to hospital) found on an administrative claim correlate to readmission? P2) Does the number of POA=no indicators found on administrative claim correlate to readmission? P3) Does the hospital-specific POA usage rate per year across all available claims correlate to readmission? These three POA variables were added to the three primary diagnosis datasets, and modeled across the three readmission periods, yielding a total of 27 individual statistical models.

***Results***

For variable P1, all three readmission periods for AMI were statistically significant at the 95% confidence level indicating an increased likelihood of readmission with odds ratios for 7-day: 1.276 (1.051, 1.547); 15-day: 1.269 (1.076, 1.494); 30-day: 1.316 (1.139, 1.520). HF 15-day odds ratio just exceeded statistical significance at 1.061 (1.009, 1.115). For variable P2, results were at the cusp of statistical significance, but probably not clinical significance at all readmission periods. For variable P3, HF and PN were significant, but showed a reduced likelihood of hospital readmission. The data for 2008 showed the widest errors, 2011 the narrowest, indicating an evolution toward more consistent POA use by providers. The odds ratio for 2011 30-day readmission in the HF dataset returned 0.604 (0.476, 0.765), and PN returned 0.730 (0.539, 0.987).

***Conclusions***

POA indicators are not a homogeneous form of data. POA indicators offer an added insight of patient complexity not previously available. POA has personalities based on the primary diagnostic condition. For AMI, there is a link between any POA=no condition during a patient stay and hospital readmission, but this is not true for HF nor PN. When aggregating POA data at the hospital level, HF and PN show a reduced likelihood of hospital readmission, but this does not hold for AMI. This effect could capture the provider's documentation maturity, which is linked to better discharge practices, which in turn reduces readmission.

## Executive Summary

Hospital readmission has gained national attention since the passage of the Patient Protection and Affordable Care Act, which authorized the Centers for Medicare and Medicaid Services (CMS) to identify primary diagnostic conditions having high readmission rates. Under the law, CMS has statutory authority to reduce reimbursements to hospitals with greater than expected readmissions. Penalties are expected to start at ~$280M the first year and climb over the next two, with some estimates predicting them to level off around $750M annually. CMS now has a stick to wield against hospitals as a means of changing their behavior. Individual hospitals and provider networks need to become aware of patient readmission to avoid exceeding annual penalty thresholds. Policy and research agencies such Agency for Healthcare Research and Quality (AHRQ) and the National Quality Forum (NQF) also monitor readmissions. And CMS itself levies the penalty based on calculated expected rates versus actual rates.

Most hospital readmission models at the national level use administrative claims data as their base, not clinical data, resulting in weaker data and more assumptions. At first glance, additional patient data would be valuable to those modeling readmission because it originates directly from the inpatient stay. Hospital Acquired Conditions (HACs) are a subset of International Classification of Disease, 9th Revision, with Clinical Modification (ICD-9-CM) codes which contain the following categories: object left in after surgery, air embolisms, blood incompatibility, catheter-associated urinary tract infection, pressure ulcers, vascular catheter-associated infection, surgical site infections, falls, poor glycemic control, deep vein thrombosis or pulmonary embolism. Present-on-Admission (POA) data is data collected during the patient's hospital stay and added to the administrative claim. A HAC condition that is not POA can be termed a HAC event and will not be reimbursed by CMS. POA data describes the state of secondary diagnosis codes as occurring prior to admission or being acquired at the hospital after admission. The distinction is subtle but offers a glimpse into the patient's stay. POA data is intended for the HAC program in order to identify which claims have conditions which could only be caused by bad care. POA are on all secondary conditions for the majority of Medicare

claims. POA is the means to identifying when a HAC has occurred. It is this "hospital-acquired" aspect which could be valuable to understanding readmission. Think of it as a loose proxy for patient complications during stay.

This thesis attempts to create a base hospital readmission model across 4 years of Medicare, 5% Limited Data Set (LDS) administrative claims data. The intent of the base model is to approximate the model used by CMS given the data limitations. Into this base model, new variables were introduced representing three different views of POA data: two at the individual claims level and one aggregated at the provider level. Three individual datasets were extracted, one for each primary diagnosis condition of interest: Acute Myocardial Infarction (AMI), Heart Failure (HF), and Pneumonia (PN). Three readmission periods were also selected for modeling based on results from the literature search: 7 days, 15 days, and 30 days from index claim. Readmission was considered for all-cause only. Logistic regression captured baseline estimates to which POA variables were then added. A total of 27 models were constructed.

Results indicate that the presence of at least one POA=no indicator on an administrative claim was statistically significant for AMI across all time periods showing an increased likelihood of readmission, when controlling for all other variables. The effect was larger than gender and as large as beneficiary age >= 80 years. This variable was not statistically significant for HF nor PN. The second POA variable approached or just exceeded statistical significant for all conditions and readmission time periods and had small odds ratios producing a small effect. The final variable was not significant for AMI at any time period nor any readmission period, but was statistically significant for HF and PN. The effect was negative, indicating a reduction in readmission odds.

It is curious that claim-based POA variables indicate increased odds for AMI but not HF nor PN. This could be due to AMI being more surgical in treatment compared to HF and PN, which tend toward treatment of chronic or exacerbating conditions. HF and PN may also be associated with other comorbidities requiring greater treatment in general. It could also be that hospital-based POA variables tend to capture hospital documentation practices including discharge-planning, which is associated with reducing hospital

readmission. The effect speaks to the idea that POA data is not homogeneous. POA data has different characteristics depending on primary diagnosis code.

## Acknowledgements

I'd like to take this opportunity to thank the people that made this possible for me: my wife Deborah, my son Adam, and those surrounding my family with assistance to when it really mattered.

I would also like to thank the University of Minnesota Carlson School of Management's Medical Industry Leadership Institute (MILI) for access to the Medicare data used within this study.

Thanks.

## Dedication

I'd like to dedicate this work to my Father and Father-in-law, both passed on now, and to my Mother and Mother-in-Law. Each encouraged me to think for myself and continue my education. They are truly members of the Greatest Generation.

# Table of Contents

## List of Tables

## List of Figures

# 1 Research Problem

## *1.1 Introduction*

"Readmission" refers to a patient's being admitted to a hospital within a certain time period from an initial hospital visit's discharge.[1] Hospital readmission rate for Medicare patients can range as high as 20% of patients per year and cost upwards of $17.4 billion annually[2]. Readmission time period varies from 7, 15, 30, 60, 90 days or more, but the most commonly used definition refers to a 30-day period. Hospital readmissions are a function of the patient's disease state, comorbidities, and demographics, but are also a function of hospital care, principal diagnosis, procedure timeliness and execution. Further still, readmission can be influenced by the community: the type of post-acute care available and general community mores toward health care in general. Many of these factors are not in the hospital's control. Centers for Medicare and Medicaid Services (CMS) is now evaluating hospitals and assessing a penalty for excessive readmissions, which in 2013 amounted to $280M.

Hospital Acquired Conditions (HACs) are situations arising from substandard hospital care resulting in patient harm. As of 2008, CMS is no longer reimbursing providers for HAC conditions. (Note: "provider" throughout this document refers to the institution certified by Medicare to supply services to its beneficiaries. It does not refer to a single physician.) Implementing the HAC program required altering administrative claim forms to include a means of identifying whether or not the patient's secondary diagnosis conditions were Present-on-Admission (POA) or the condition was acquired after the patient was admitted to the hospital.

Is there a relationship between POA indicators and hospital readmission? As a readmission explanatory variable, POA data is interesting because it could increase data richness specific to the patient's stay. This is in contrast to incorporating external data, such as municipal demographics, hospital case mix, economic considerations, or even community views of health care. To date, HAC/POA data has not been incorporated into

many statistical models. The purpose of this study is to understand and assess the usefulness of POA indicators as a potential hospital readmissions predictor.

Understanding this study requires understanding a reasonable background of both 30-day hospital readmission and HAC data. Section 1.2 investigates readmission from the vantage point of national reporting as opposed to state-level or hospital-specific reporting. It takes into account some disease specifics as well as disease progression as a readmission factor. Section 1.3 describes the background of HACs. Of interest here is not so much the conditions themselves, but understanding the administrative claims changes needed to implement HAC reporting via POA indicators.

## *1.2 30-Day Hospital Readmission Background*

### 1.2.1 Origins of Readmission

Hospital readmission has been the subject of investigation for years. A reasonable point from which to consider contemporary studies is a literature review conducted by Benbassat and Taragin, 2000, titled "Hospital Readmission as a Measure of Quality of Care – Advantages and Limitations."[3] They make the observation that most studies investigated had concluded readmissions were "caused by patient frailty and chronic disease progression." They also identify a wide range of between "9% and 48% of readmissions have been judged to be preventable … and were associated with substandard care." Benbassat and Taragin reviewed articles and books published over the previous 10 years. At that time of publication, no national system for tracking readmission existed. Results were limited to clinical records collected from small documentation reviews and extrapolated upward per condition or per hospital per state.

The Hospital Inpatient Quality Reporting (IQR) program was mandated by Section 501(b) of the "Medicare Prescription Drug, Improvement and Modernization Act (MMA) of 2003"[4], directing hospitals participating in the Inpatient Prospective Payment System (IPPS) to start reporting quality information on core measures. After collection, this data was made available to consumers via CMS's HospitalCompare website, http://www.medicare.gov/hospitalcompare, with the ultimate hope of increasing patient

care by adherence to these basic quality measures. MMA also authorized CMS to reduce reimbursements to hospitals not successfully meeting the quality guidelines. CMS also uses the data to calculate payment rates under the Hospital Value-Based Purchasing (VBP) program.[5] The IQR program was extended to collect additional data via the Section 5001(a) of Pub. 109-171 of the Deficit Reduction Act (DRA) of 2005[6].

Circa mid-2000s, Jencks, Williams, and Coleman cite an overall readmission rate of 20%, meaning 1 in 5 hospitalizations had a readmission within 30 days of a hospital discharge.[7] Hospitals with rates below the baseline readmission rate were praised, while those hospitals with rates above the baseline were shamed. But other than the praise and shame there was no compelling reason for providers to alter behavior.

Medicare Payment Advisory Commission (MedPAC), Congress's Medicare advisory agency, issued "Payment policy for inpatient readmissions" in June 2007, finding Medicare spending on 30-day readmission amounted to $15 billion resulting from an estimated 17.6% all-cause annual readmission rate.[8] Within this report, readmission periods were considered for 7-day, 15-day, and 30-days from the index (initial) discharge. The report also urged readmission definition away from hospital-specific rates and towards specific medical conditions citing "hospitals that concentrate on joint replacements are likely to have lower readmission rates than hospitals that concentrate on cardiac care" when comparing hospital-rate only.

It wasn't until the "Patient Protection and Affordable Care Act of 2010" (ACA) that CMS had the authority under Section 3025 to form the "Hospital Readmission Reduction Program" in order to collect hospital readmission data and withhold reimbursement to hospitals exceeding an expected annual threshold.[9] From a list of top ten readmission conditions, CMS selected three for the first implementation phase: Acute Myocardial Infarction (AMI), Heart Failure (HF), and Pneumonia (PN). These were selected because they are costly, frequent, and were determined to have the highest readmission rates for "preventable" conditions. Data collection started immediately, and reimbursement payments would take effect 2013Q4. Reductions are calculated as shown in Figure 1.[10]

**Formulas to Calculate the Readmission Adjustment Factor**

**Excess readmission ratio** = risk-adjusted predicted readmissions/risk-adjusted expected readmissions

**Aggregate payments for excess readmissions** = [sum of base operating DRG payments for AMI x (excess readmission ratio for AMI-1)] + [sum of base operating DRG payments for HF x (excess readmission ratio for HF-1)] + [sum of base operating DRG payments for PN x (excess readmission ratio for PN-1)] + [sum of base operating DRG payments for COPD x (excess readmission ratio for COPD-1)] + [sum of base operating payments for THA/TKA x (excess readmission ratio for THA/TKA -1)]

*Note, if a hospital's excess readmission ratio for a condition is less than/equal to 1, then there are no aggregate payments for excess readmissions for that condition included in this calculation.

**Aggregate payments for all discharges** = sum of base operating DRG payments for all discharges

**Ratio** = 1 - (Aggregate payments for excess readmissions/ Aggregate payments for all discharges)

**Readmissions Adjustment Factor** =

- For FY 2013, the higher of the Ratio or 0.99 (1% reduction);
- For FY 2014, the higher of the Ratio or 0.98 (2% reduction).
- For FY 2015, the higher of the Ratio or 0.97 (3% reduction).

*Figure 1. CMS readmission adjustment calculation.*

Providers now had a reason to monitor patients for readmission. Using data collected via the IQR, CMS started to publish 30-day hospital readmission rates calculated for each provider using the administrative claims data, all of which was posted on the HospitalCompare website. These rates let consumers know about a hospital's implied quality via the readmission metric.

Between the passage of the ACA and June 2011, CMS had completed the following with respect to the readmission program:

- Defined 30 days as the readmission time period for all-cause readmission
- Limited reimbursement of readmission conditions to AMI, HF and PN
- Established policy consistent with the National Quality Forum (NQF) methodology for calculating excess hospital readmissions for the selected conditions
- Defined readmission methodology consistent with NQF and MedPAC

### 1.2.2 Determining Which Claims Are Readmissions

Understanding readmissions is a tricky business. Early attempts to calculate readmission used all available claims to generate an "overall rate" hospital rate: any patient, any condition, any time. This approach is problematic for a number of reasons. Patients with a progressing disease state return more often and appear to have higher readmission "caused by substandard care." Comorbidities confound readmissions such as kidney problems and heart problems, and meeting different staff within a hospital setting, resulting in treatment for one conditions appears as a readmission for the other. Linking unrelated conditions via readmission inflates readmission rates. A better alternative supported by NQF is to separate claims based upon health condition using sets of ICD-9 diagnostic codes found in the principal diagnosis fields – looking across all patients with AMI, for example. These types of claims have a common set of criteria: patients have similar conditions, the administrative data has similar Diagnostic-Related Groups (DRGs), similar length of stay, similar costs, etc.

Before proceeding too far, it will be necessary to understand the following terms used within this study:

*Administrative claim:* Any claim for service provided by an inpatient service provider containing a summary of services rendered for a patient at a hospital meeting requirements set forth by CMS

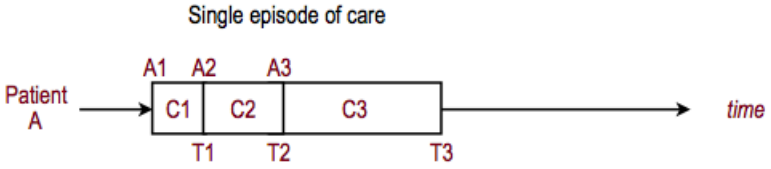
*Potential index claim:* Any administrative claim meeting eligibility requirements for readmission study (ex: claim must be from U.S. state or D.C., patient must be discharged alive, etc.)

*Index claim:* A status given to potential index claims that have an associated readmission claim

*Readmission claim:* A status given to potential index claims that have an associated index claim

The following diagrams illustrate CMS's adopted readmission definitions. Suppose Patient A goes into the hospital for a condition leading to (fill in the blank). Patient A has an operation, then gets moved into another room on a different floor of the hospital.

Several days later, the patient is sent across the street for rehabilitation services to get the patient up on their feet and able to partake of the activities of daily living. Patients stay in the rehab unit for many more days until discharged and sent home. For the average researcher with access to no clinical data but lots of administrative data, this patient's visit to a single provider can instead appear as several events (Figure 2). Providers divide resources for the purposes of billing and reimbursement, making a single visit to the hospital into a chain of disconnected administrative claim records.




Claim C1 - Emergency admission: patient admitted to intensive care and surgery, patient stabilized, transferred to different unit in same hospital. Claim C2 - Recovery in step-down unit, transfer to different hospital for rehabilitation services. Claim C3 - Rehab services, discharge to home/self care.

A*n* = admission date, T*n* = through date

***Figure 2. Administrative claim history for a single episode of care.***

One cannot look at a single administrative claim and determine if it is a readmission. There has to be a context from which to judge. Building the context requires aligning claim records by patient and within patient claims by date. Knowing which claims are upstream events and which are downstream events is still not enough to determine readmission status. CMS's hospital readmission reduction program focuses on short-term acute care hospitals and their care plans for the patient post-discharge. Rehabilitation, psychiatric, long term care, and outpatient facilities are not part of the hospital readmission reduction program. Consequently, these claims are ineligible for this study and must be removed from further consideration. Each claim is subject to screening, the purpose of which is identifying "potential index" claims as the source from which to judge downstream claims as readmissions. The goal is to find the last link in the claims chain responsible for the post-discharge planning (Figure 3).

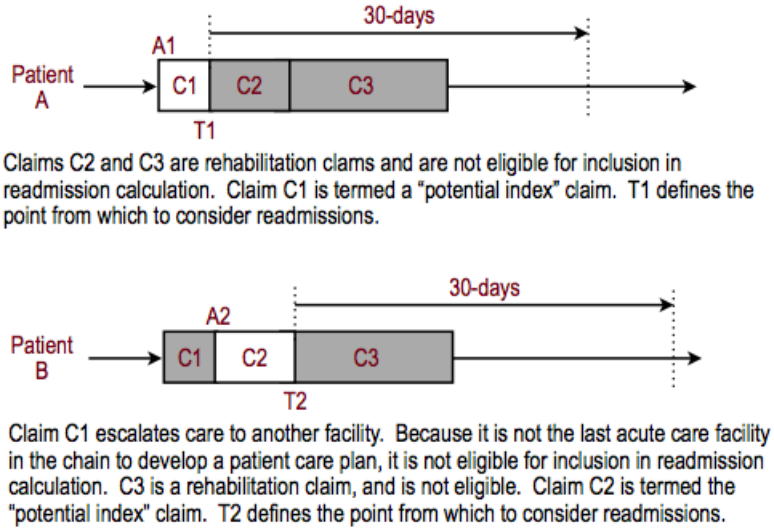



***Figure 3. Identifying "potential index" claims.***

Figure 3 contrasts the cases of Patient A and Patient B. In the case of Patient B, there is a case of escalation from one short-term hospital to another. The administrative claims look similar to that of Patient A except for the point at which the 30-day window starts. Because C1 transferred the patient to a different hospital, the readmission clock starts with the following C2 claim. Again, the purpose of identifying "potential index" claims is to link any readmission with the appropriate hospital, in this case the last hospital in the chain and one responsible for patient discharge planning.

CMS uses a combination of MedPAC- and NQF-derived criteria to determine which administrative claims are eligible and which are not. Some of the exclusion criteria are:

- Claim not from a short-term acute care hospital
- Claim not from a U.S. or D.C. hospital
- Zero-day claims (having a utilization day count of '0', patient didn't spend the night in the facility)
- Transfer claims (claims where the delta between through data and admission date is '0')
- Rehabilitation or psychiatric hospital claims
- Patient <65 years old at time of hospitalization

- In-hospital death claims
- Cancer hospital claims (disease progression in combination with cancer treatment and its side effects confound readmission modeling)
- Patients who leave against medical advice

Claims meeting these conditions would be removed from the patient's claim history. After removing all ineligible claims, actual "index" and "readmission" claims are identified using the earliest through dates from each patient's claims records. There are three considerations illustrated in Figure 4 below.

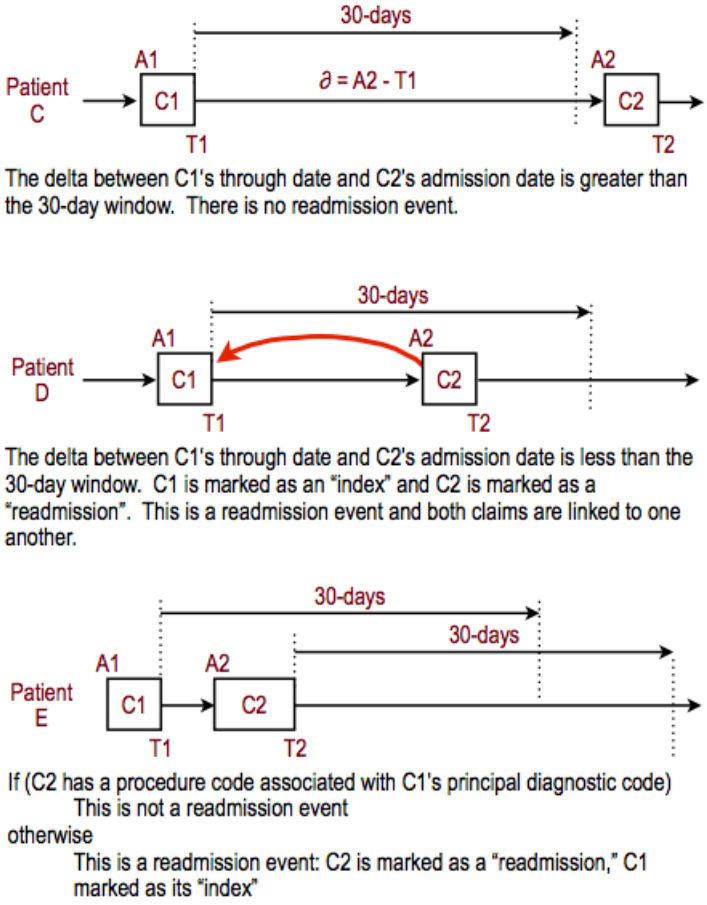



*Figure 4. Identifying single readmission event.*

For Patient C, the distance between the claims exceeds the readmission window and does not count as a readmission event. Patient D, however, has a claim falling within the

readmission window of another claim. This is the normal case, and these two claims become a readmission event. An additional complication arises when patients are diagnosed with a condition on one claim, but are discharged pending scheduled surgery. In the case of Patient E, the second hospitalization appears as a readmission event from the first visit, but it isn't. The question of planned versus unplanned is a function of disease state, urgency, hospital or doctor availability, patient's ability to endure a surgical event, etc. These planned events present an inflated hospital readmission ratio by increasing the numerator counts. These situations can be common at larger institutions negatively affecting the provider. CMS methodology filters out planned events from readmission calculations. Patients with AMI may have staged or planned readmissions, but HF and PN do not.

In addition to determining individual claim eligibility based upon data within a given claim record, some models also add conditions constraining eligibility based upon upstream claims by requiring a 30+ day buffer between "index" claims. The motivation behind this logic addresses the condition when a bad health situation turns worse requiring more hospital visits to stabilize a patient. Under this condition, the claim history gets confused, and filtering out the additional claims produces a better readmission signal. This is true from a provider prospective, but patient advocates point to repeated trips to the hospital as a sign of poor quality. CMS has adopted the second approach as its methodology.

Patients near end-of-life or with chronic disease progression are more likely to need additional care. These patients have many "potential index" claims in their administrative histories and most of these claims have deltas smaller than the 30-days. Figure 5 illustrates the situation for Patient F involving multiple readmission events in close proximity to one another.

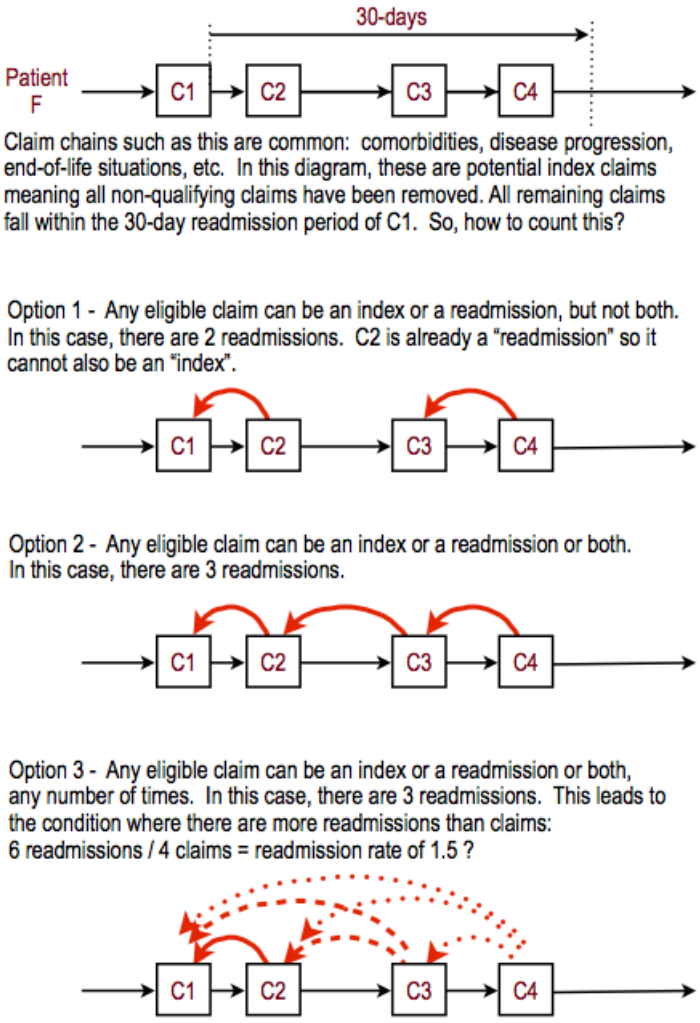



***Figure 5. Identifying multiple readmission events for a single patient.***

When considering the question, "how many readmission events are there?" most methodologies fall into one of three cases. Option 1 adopts the view that a claim can be either an index or a readmission, but cannot be both. This is another method to remove muddled claim histories from the readmission assessment process. Option 2, from the provider perspective, it is plausible to see claim C3 as still somehow involved as correcting the C1 and C2 event chain. From the patient perspective, another trip to the hospital could be seen as substandard care leading to Option 2, where all claims could be either an index or a readmission. In "Working Paper: Simple Methods of Measuring Hospital Readmission Rates" [11], the authors cited another approach of counting *any*

claim falling within the readmission period of a "potential index" claim as a readmission. This is Option 3, and it is included to give perspective on what approaches are under consideration when defining readmissions. Option 3 has a peculiar effect of producing a readmission ration greater than 1.0 caused by an artificially large numerator. It is this reason the America's Health Insurance Plans, Center for Policy and Research list this as a non-preferred method. CMS uses Option 2 in its readmission calculation methodology.

### 1.2.3 Determining Readmission Rates

Given that a repeatable, reliable method for determining readmission events has been defined, the problem now turns toward aggregating results and determining readmission rates. Readmission work of the 1990s attempted to use the individual patients as a point of reference: how many readmissions does a patient and peer group experience? The trouble here is identifying peer groups. At the hospital level? Individual providers don't have access to their peer's data resulting in peer group rates equaling hospital rates. Patients are also problematic as they can travel from one hospital in their region to another for care, thus breaking the chain of events. This method produced wildly varying results and was consequently dropped.

Another approach concentrated on the disease instead of the patient. Hospitals treating AMI or PN should be comparable. This method uses the principal diagnostic code of the index and of the potential-readmission claims to determine if an actual readmission event should be counted. If a patient treated for PN had to go back to the hospital to be treated for PN again, that would count as a readmission. However, this method under counts readmissions and allows for gaming. The provider has access to the clinical data while CMS does not. The temptation to alter a second PN into something that does not count as readmission is too great particularly when peer providers are gaming the system. This approach is further complicated by confounding because the disease condition under investigation would be represented in a statistical model as both an input variable and an output variable: an AMI claim used to predict an AMI readmission. Another set of readmission studies form the 1990s attempted to limit readmissions to those claims relating to the principal diagnosis of the index claim. As a

concept, this approach was abandoned for two reasons. First, every hospital is responsible for treating the whole patient, not just a specific disease, procedure, etc. Second, readmissions occurring at the same hospital are then open for potential gaming; hospitals fudging numbers to appear better than they really are.

Aggregating only at the hospital level also proved unsatisfactory. In this method, all readmissions are counted and divided by the number of claims. MedPAC [7] describes the primary limitation this way, "hospitals that concentrate on joint replacements are likely to have lower readmission rates than hospitals that concentrate on cardiac care." Case mix can allow a statistical model to control for the types and numbers of patients a hospital serves. Service mix can allow a statistical model to control for the types of offerings in which a hospital specializes. Combining both should allow hospital-level comparisons for readmissions. But the results are still unsatisfactory as the hospital is not the correct unit for investigation.

The final recommended solution by MedPAC and NQF came in the mid-2000s with readmission aggregation at the hospital level via specific diagnostic conditions. This method identifies all indexes and readmissions using all administrative claims for each patient. It then identifies index events by principal diagnostic condition (find all AMI-only events, for example). From this pool, all-cause readmission (any reason the patient returns is suspicious) rates are calculated. This method is the current best-practice method for comparing how each hospital handles each condition. Different diagnostic conditions have different readmission characteristics and can mask one another if combined, so removing the confounding aspect of multiple medical conditions creates a better statistical model. CMS uses this approach as the basis for its readmission adjustment factor calculation.

As part of the hospital quality program, CMS monitors risk-standardized readmission rates for three most common or most expensive conditions: AMI, HF, and PN. It publishes those statistics on its HospitalCompare website. CMS created a model using the previous three years' worth of data collected for each hospital to calculate an expected readmission rate. Those hospitals with actual readmission rates exceeding the

expected rates are penalized a proportion of the annual reimbursement for those monitored conditions. The ACA 2010 specifies a maximum penalty of 1% for the first year, 2% for the second, and maxing out at 3% thereafter. Penalties commence with 2013. The first year's assessed penalties for 2013 amount to $280M. Total estimated readmissions are expected to run into the billions of dollars annually. The penalties may seem small in comparison to total reimbursement expenditures, but the purpose of the penalty is to change hospital behavior by redirecting hospital's attention to outcomes and readmission. The purpose is not to recoup all losses and shut facilities down.

## *1.3 HACs/POA Background*

### 1.3.1 Sentinel Events

A reasonable place as any to understand the importance and advent of HACs is with the 1999 Institute of Medicine book, *To Err is Human: Building a Safer Health System.*[12] The authors revealed the lives lost or harmed and lost opportunity cost wasted because of medical errors in the American health care system. There was a certain shock value to the report: "at least 44,000 people, and perhaps as many as 98,000 people, die in hospitals each year as a result of medical errors." How could the U.S. health care system be in this shape? The report's impact raised awareness at an institutional level, a state level, and national level regarding the magnitude of bad medical practice.

As of the late 1990s, quality efforts were ad hoc, provider-based or department-based within a provider. No larger plan existed to identify, track, and report on errors made in the health care system, and few states mandated quality initiatives. Indeed, there existed a culture intolerant of errors and mistakes, preventing the acknowledgement that these types of things might happen in the first place. The authors recommended many things, and among them two stand out: First, standardize medical error reporting mechanisms via the formation of an agency to pursue patient safety and reporting. To accomplish this, the authors recommended secondly to target a limited number of hazardous event types and to concentrate action to reduce or eliminate them. The first would ensure congruent

and consistent reporting across states, and the second would aim for a higher probability of success because the goals would be smaller and more well-defined.

*To Err* also included a summary of state-based adverse event reporting as its Appendix D. It demonstrated how much variability existed within the reporting systems at that time. These 13 states, CA, CO, CT, FL, KS, MA, MS, NJ, NY, OH, PA, RI, and SD, representing approximately 42% of the US, had their own quality reporting systems in place, but those systems varied by what health conditions were tracked, the terms used, consequences if any for not following the standard, and visibility of final reports. Would results be made public or would they be private? The remaining states had no reporting or little state-level reporting. Collecting the data was one issue. Disseminating the results of collection and analysis was another as hospitals were not interested in having any "dirty laundry" aired in public.

A third recommendation from *To Err* was to establishment of what became known as sentinel events – things that were so bad, so egregious, no one could call these legitimate conditions or side-effects of reasonable care. These were the fore-runners of what would become known as HACs.

### 1.3.2 Serious Reportable Events / Never-Events

Despite the reference to *To Err* in nearly all papers defining health care quality, there are other contemporary sources who had arrived at the same conclusion. In 1998 just prior to *To Err* publication, the Quality Interagency Coordination Committee (QuIC) was formed by Presidential directive, 13-Mar-1998 [13]. It was charged "to ensure that all Federal agencies involved in purchasing, providing, studying, or regulating health care services are working in a coordinated way toward the common goal of improving quality of care."[14] The Committee chair is the Secretary of Health and Human Services (HHS) with day-to-day operations directed by another chair, the Director of the Agency for Healthcare Research and Quality (AHRQ). The QuIC Committee quickly recognized the importance and significance of *To Err*'s findings as being in alignment with the viewpoint of driving national health care toward consistent delivery, higher quality and responsibility. The Committee's structure provided backing to the Health Care Finance

Administration, as CMS was called in 2000, to address patient safety concerns brought up by the *To Err* report.

Another organization, the NQF, was created in 1999 at the time of *To Err*'s publication by a coalition of public and private partners under a different but related directive, "to promote and ensure patient protections and health care quality through measurement and public reporting."[15] CMS, at the direction of QuIC, directed NQF to "identify a set of patient safety measurements that should be a basic component of any medical errors reporting system."[16] The logic and virtue in NQF's recommendation is that any health care quality reporting system containing adverse events (however defined) would contain a set of similar conditions because these conditions are so basic that every provider should already be tracking them. NQF would then standardize the definitions and reporting framework.

NQF presented this early framework in 2002. It included events that were "shocking" -- medical errors that appeared to happen regularly and were indicative of such poor health care that those events should just never occur. These became known as "never-events." NQF identified three criteria all potential never-events conditions must meet:

a) The condition is clearly identifiable as an error.

b) The condition is serious in nature, resulting in death or serious injury.

c) The condition is preventable if standard care guidelines are followed.

NQF proposed 28 conditions meeting these criteria, and divided them into 6 categories as seen in below in Table 1.

***Table 1. NQF's 2002 definitions of serious reportable events, aka never-events.***

| ***Category*** | ***Never-events*** |
|---|---|
| Surgical Events | 1.1. Surgery performed on the wrong body part<br>1.2. Surgery performed on the wrong patient<br>1.3. Wrong surgical procedure performed on a patient<br>1.4. Unintended retention of a foreign object in a patient after surgery or other procedure<br>1.5. Intraoperative or immediately postoperative death in an ASA Class I patient |
| Product or Device Events | 2.1. Patient death or serious disability associated with the use of contaminated drugs, devices, or biologics provided by the health care facility<br>2.2. Patient death or serious disability associated with the use or function of a device in patient care in which the device is used or functions other than as intended<br>2.3. Patient death or serious disability associated with intravascular air embolism that occurs while being cared for in a health care facility |

| *Category* | *Never-events* |
|---|---|
| Patient Protection Events | 3.1. Infant discharged to the wrong person<br>3.2. Patient death or serious disability associated with patient elopement (disappearance)<br>3.3. Patient suicide, or attempted suicide resulting in serious disability, while being cared for in a health care facility |
| Care Management Events | 4.1. Patient death or serious disability associated with a medication error (e.g., errors involving the wrong drug, wrong dose, wrong patient, wrong time, wrong rate, wrong preparation, or wrong route of administration)<br>4.2. Patient death or serious disability associated with a hemolytic reaction due to the administration of incompatible blood or blood products<br>4.3. Maternal death or serious disability associated with labor or delivery in a low-risk pregnancy while being cared for in a health care facility<br>4.4. Patient death or serious disability associated with hypoglycemia, the onset of which occurs while the patient is being cared for in a health care facility<br>4.5. Death or serious disability (kernicterus) associated with failure to identify and treat hyperbilirubinemia in neonates<br>4.6. Stage 3 or 4 pressure ulcers acquired after admission to a health care facility<br>4.7. Patient death or serious disability due to spinal manipulative therapy |
| Environmental Events | 5.1. Patient death or serious disability associated with an electric shock while being cared for in a health care facility<br>5.2. Any incident in which a line designated for oxygen or other gas to be delivered to a patient contains the wrong gas or is contaminated by toxic substances<br>5.3. Patient death or serious disability associated with a burn incurred from any source while being cared for in a health care facility<br>5.4. Patient death or serious disability associated with a fall while being cared for in a health care facility<br>5.5. Patient death or serious disability associated with the use of restraints or bedrails while being cared for in a health care facility |
| Criminal Events | 6.1. Any instance of care ordered by or provided by someone impersonating a physician, nurse, pharmacist, or other licensed health care provider<br>6.2. Abduction of a patient of any age<br>6.3. Sexual assault on a patient within or on the grounds of a health care facility<br>6.4. Death or significant injury of a patient or staff member resulting from a physical assault (i.e., battery) that occurs within or on the grounds of a health care facility |

Because these unambiguous events can be distinguished from other conditions, they can be measured and therefore reported. Two recommendation goals stand out: increasing patient safety by reducing medical errors, but also public accountability of clinicians and institutions who have committed these errors. Both are accomplished via public event reporting. These are sound, logical arguments toward the goal of increasing hospital quality. However, as of the mid-2000s, there were no reporting agencies at the federal level collecting this type of data. Nor were there any agencies able to mandate it.

Things were different at the state level. By 2007, a report on state-level patient safety[17], identified 10 states, CA, CT, IL, IN, MN, NJ, OR, VT, WA, and WY, that had voluntarily adopted some or all of the NQF never event guidelines into their own

departments of health. Minnesota is one example that led the way by legislated reporting starting in 2003[18], resulting in its first patient safety report dated Jan-2005.

### 1.3.3 HACs

With the passage of the DRA of 2005 Section 5001(c)[19], Congress directed CMS to alter DRG payments in order to limit inpatient hospital reimbursements for certain conditions that could have reasonably been prevented had the hospital followed standard-of-care guidelines. The "Quality Adjustment in DRG Payments for Certain Hospital Acquired Infections" section of the DRA Section 5001(c) gave CMS direction to identify conditions meeting these three criteria:

- Condition has high cost or high volume, or both.
- Condition results in higher DRG payment when present as secondary diagnosis.
- Condition could reasonably have been prevented, as determined by application of evidence-based guidelines.

Prior to the DRA Section 5001(c), Medicare would reimburse a hospital more for treating a patient whose inpatient stay cost more as a result of the hospital's own faulty actions or inactions. A New York Times article "Not Paying for Medical Errors", explains that a patient getting an infection after being admitted to a facility: "even if the infection is caused by sloppy sanitary practices in the hospital itself … [results in] rewarding incompetence rather than penalizing it."[20] Wald and Kramer[21] add to the sentiment by focusing on one condition, Catheter-Associated Urinary Tract Infections (CAUTIs), showing $400M annually spent treating infections, most of which, in their opinions, could have been prevented. They also note providers may be reimbursed 60% more for the same underlying conditions because of complications caused by the infection the provider failed to prevent. By way of example, they cite a Colorado hospital's costs to treat a patient with AMI (Table 2).

They further estimate an annual expenditure of $400M for just to treat complications for this one condition. With the DRA Section 5001(c), CMS gained the ability to restrict reimbursement to those patient-centric conditions and not to those that are hospital

related. In the previous example, the hospital would have to absorb the $3,468 difference from the standard of care compared to the cost to treat the original condition. In addition, the law prevents hospitals from billing the patient for the difference. These legal actions attempt to focus hospitals on the quality of care they provide by removing a financial incentive to tolerate poor quality.

***Table 2. Additional reimbursement for CO hospitals cited by Wald and Kramer.***

| ***AMI Patient*** | ***Hospital Costs*** |
|---|---|
| Patient with no complications | $5437 using current standard-of-care |
| Patient with a UTI complications | $6721, ~23% more than standard |
| Patient with major UTI complications | $8905, ~63% more than standard |

To implement a program limiting reimbursement, CMS would need to do three things.

1. Publish a HAC list consisting of International Classification of Disease, 9th Revision, with Clinical Modification (ICD-9-CM) codes so providers could establish programs targeting these codes. [22]
2. Alter existing DRG payments and codes to segregate conditions by their severity: a) condition is baseline normal case, b) condition has Complications or Comorbidities (CC), or c) condition has Major CC (MCC). This is called the Medicare Severity-DRG (MS-DRG) and offers additional reimbursement for additional care when conditions are warranted.
3. Define a POA indicator as "yes" or "no" for each secondary diagnosis code submitted on each claim containing a HAC condition to enable CMS to detect HAC/POA conditions.

Payments would be held to the base MS-DRG severity rates if a HAC condition exists on the administrative claim, and that HAC happened to the patient after admission to the hospital. HAC data was phased in starting as a voluntary system in 2007Q4, becoming mandatory in 2008Q4.

#### 1.3.3.1 Publish a HACs List

Word spread through the health care community in the early 2000s to the public about U.S. hospitals and never-events, raising the outrage surrounding Medicare *paying* for bad care under the fee-for-service agreements. Awareness eventually moved to Congress. With the passage of the DRA of 2005, signed into law in 2006, Congress granted authority to CMS to define conditions such as never-events and to create a reporting structure capable of processing claims limiting reimbursement in cases of bad health care-caused conditions. It was from NQF guidelines which CMS adopted its first list of HACs. The two lists are not congruent because they target different populations. The audience for Serious Reportable Errors from NQF are U.S health care patients whose age could be anywhere from infant to elder. CMS conversely needed conditions targeting patients whose ages are 65+ years old. Per the DRA section 5001(c), CMS was called to identify at least two conditions meeting the criteria, and to publish those no later than 2007Q3 such that providers could start to report on them. CMS chose eight conditions for their initial implementation, and expanded the list to ten conditions effective 2009Q4.[23] (See Table 3.)

***Table 3. CMS's definition of HAC categories.***

| *Hospital Acquired Conditions* | *ICD-9-CM Diagnosis Code* |
|---|---|
| Object left in surgery | 998.4, 998.7 |
| Air embolism | 999.1 |
| Blood incompatibility | 999.60, 999.61, 999.62, 999.63, 999.69 |
| Catheter-associated urinary tract infection | 996.64 |
| Pressure ulcers | 707.23, 707.24 |
| Vascular catheter-associated infection | 999.31-999.33 (new 2007) |
| Surgical site infection (mediastinitis) | 519.2 plus 36.10-36.19 procedure code |
| Falls | 800-839, 850-854, 925-929, 940-949, 991-994 |
| Manifestations of poor glycemic control | 250.10–250.13, 250.20–250.23, 251.0, 249.10–249.11, 249.20–249.21 |
| Deep vein thrombosis or pulmonary embolism following certain orthopedic procedures | 415.11, 415.19, 453.40–453.42 plus one of the following procedure codes: 00.85–00.87, 81.51–81.52, or 81.54 |

#### 1.3.3.2 Alter Existing DRG Payments

The DRA Section 5001(c) gave CMS the ability to modify the DRG payment definitions and the administrative claims data upon which the DRG is based. DRGs were first proposed in 1980 to address the rising cost of Medicare spending by proposing a

reimbursement system for providers to control their own costs through common payment for a given medical condition. The idea behind DRGs is to gather clinically cohesive groups of medical treatments together and provide one base reimbursement amount. For example, under a simple fee-for-service model, all treatments, drugs, equipment, therapy, etc., for a Coronary Arterial Bypass Graft (CABG) procedure would be billed separately. All hospitals would be reimbursed for what they billed. Under the DRG system, an average cost for the CABG and associated costs is reimbursed. The resulting intermediate reimbursement amount is then altered based on hospital status, teaching institution status, patient outlier conditions, etc. The base amount varies based on the age of beneficiary, gender, etc. The provider has a financial incentive to treat the patient more efficiently, a cost below the reimbursed amount, resulting in a financial saving or income to the hospital. Providers exceeding the reimbursement cost would then realize a financial loss to the institution for that claim. Hospitals contracting with CMS to provide Medicare service for seniors agree to reimbursement under the IPPS utilizing DRGs.

CMS replaced the existing DRG (Version 25) definition with an extended DRG definition (Version 26) which incorporated a relative complexity or severity measure. CMS called it the MS-DRGs. Prior to Version 26, conditions which complicate treatment were not represented well. Hospitals serving more complicated patients were not reimbursed more for the additional services needed for the complication.

Starting with version 26, the Grouper application, could select between multiple MS-DRG codes for the same base condition which are differentiated only by the severity of the patient's overall conditions: base condition equals base reimbursement. Grouper would select the code 'CC' to designate patients with "complications or comorbidities," (also confusingly defined as "complicating conditions" depending upon the source), reflecting the providers efforts to treat patients above the normal case because the patient required a longer hospital stay, additional drugs and monitoring, more lab work, more rehab, etc. When these conditions become excessive, Grouper would then select the "major complications or comorbidities" or "MCC" code. Not all MS-DRG codes have all levels of CC/MCC. Some MS-DRGs differentiate only the most complicated

conditions from the rest, and others only the base condition and the rest. These are accomplished with the "w/o CC", "w/o MCC", and "w/o CC/MCC" codes.

Table 4 below shows how CC and MCC modify the MS-DRG. Complicated peptic ulcers have three flavors: MCC, CC, or neither. Uncomplicated peptic ulcers only come in two flavors: MCC or not MCC, which results in the question, how can an uncomplicated peptic ulcer have major complicating conditions? The answer is that the ulcer itself is uncomplicated, but the patient's other conditions would impact treatment enough to a) increase the amount of care needed as denoted by "with MCC" severity level, or b) not interfere with the ulcer's treatment denoted by "without MCC" severity level.

***Table 4. Sample MS-DRG definitions illustrating how the severity designator works.***

| ***MS-DRG Number*** | ***Description*** | ***Severity*** |
|---|---|---|
| 380 | Complicated Peptic Ulcer | With MCC |
| 381 | Complicated Peptic Ulcer | With CC |
| 382 | Complicated Peptic Ulcer | Without CC/MCC |
| 383 | Uncomplicated Peptic Ulcer | With MCC |
| 384 | Uncomplicated Peptic Ulcer | Without MCC |

CMS reimburses providers at higher rates for difficult-to-treat patients compared to the standard cases. All incoming claims are assigned an MS-DRG code using a program called Grouper, a CMS software product available for purchase designed for hospital billing departments to help prepare administrative data for submission to CMS[24]. It reads all of the diagnosis and all of the procedures plus additional information to determine the appropriate MS-DRG, severity, and ultimately the reimbursement dollar amount for that claim.

#### *1.3.3.3 Define POA Indicators for each Secondary Diagnosis*

The administrative claims data format circa 2005 was not adequate to handle the burden of tracking HAC conditions, nor were the proposed DRG modifications. It is this ability to add complications which gives CMS the mechanism by which to react to complicating conditions caused by a hospital's poor care. But, there is one additional piece of information needed to prevent reimbursement for HACs: Did the patient have

the listed diagnosis at the time the patient was admitted to the hospital? If so, then Grouper would classify the claim as it always had. If not, Grouper would limit reimbursement to the base MS-DRG amount only. This distinction was not possible to derive from existing administrative claims data prior to 2007Q4. CMS defined the POA indicator flags as seen in Table 5. On an administrative claim, if a HAC condition exists as one of the diagnosis fields, then each diagnosis must be accompanied by a POA flag. If no HAC condition exists in the diagnosis fields, POA flags need not be reported (POA field may be left blank on administrative claim).

***Table 5. POA indicator flags.***

| ***POA Flag*** | ***Definition*** | ***CMS Payment for CC/MCC DRG if HAC condition present*** |
|---|---|---|
| Y | Yes. Present at the time of inpatient admission. | Yes |
| N | No. Not present at the time of inpatient admission. | No |
| U | Unknown. The documentation is insufficient to determine if the condition was present at the time of inpatient admission. This indicator will be treated the same as 'N'. | No |
| W | Clinically Undetermined. Unable to clinically determine whether the condition was present at the time of inpatient admission. This indicator will be treated the same as 'Y'. | Yes |
| 1 | Unreported/Not used. Exempt from POA reporting. | Yes |
| blank | If no HAC condition present, no POA flags are required. | NA |

The DRA Section 5001(c) mandates HAC/POA changes be phased in beginning with 2007Q4 as the date when hospitals would need to start reporting POA indicators with their administrative claims data, priming the pump as it were to get the kinks worked out of the system. Starting in 2008Q4, CMS would start using this HAC/POA data to limit reimbursements for bad care. Since the initial definition in 2007Q4, '1' and 'blank' indicator flags had been phased out as valid POA options effective as of 01-Jan-2010. [25]

Adding HACs into the MS-DRG system results in reduced reimbursement to the base level for claims containing a HAC not POA. Any HAC complication caused by the provider is not reimbursed, and is therefore the provider's burden. Figure 6 illustrates the nuance of MS-DRG reimbursement flow when HACs are involved [26].

The DRA Section 5001(c) limits CMS in the following ways: First, only acute care hospitals and providers participating in the IPPS need identify secondary diagnosis as

being POA for each secondary condition identified on their submitted administrative claims. This excludes "Critical Access Hospitals, Rehabilitation Hospitals, Psychiatric Hospitals, or any other facility not paid under the Medicare Hospital IPPS"[27] Only these claims will be subject to a reduction in the MS-DRG payment if a HAC condition is not POA. Knowing hospital type and claim payment type will be critical to modeling hospital readmissions using POA data because the data does not include all claims nor all hospitals.

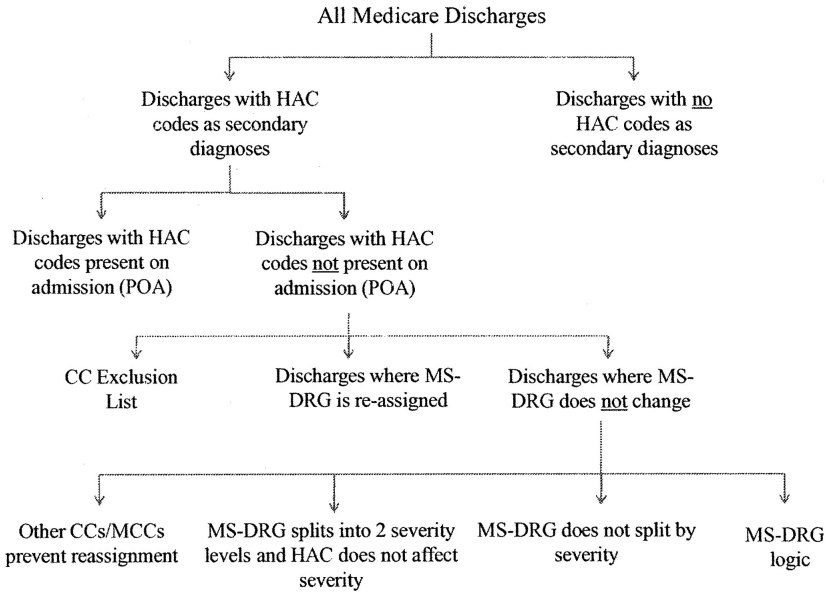



*Figure 6. CMS HAC reimbursement logic.*

### 1.3.4 Significance of POA data

CMS monitors the HAC program producing annual reports containing rates, trends, costs, etc.[28, 29] But POA data is new to health care research. There are 5 years of provider experience, 3 years of which is suitable for study. POA data offers a way to distinguish between patient comorbidities and potentially preventable events.[30] This added patient complexity may be useful when assessing hospital readmission risk, and this data may be available to clinicians before the patient is discharged.

During the initial years of the HAC program, strict interpretation of DRA Section 5001(c) was interpreted by some providers as having to include POA data for all secondary diagnoses found on a patient's administrative claim *only when a HAC condition was included*. Consequently, many hospitals rarely include POA data. Other providers had a different interpretation and included POA data for all reported secondary diagnoses codes even when no HAC condition was included. Still other providers included POA data seemingly at random. CMS clarified POA use by mandating all inpatient providers shall use POA indicators all of the time starting in 2010Q4. This interpretation difference makes early POA data harder to use because of the hit-or-miss quality. Later data can be used to better effect precisely because of the "everyone must use it" directive. Ninety-five percent of all administrative claims are under this guidance.

## 1.4 Research Goals

The purpose of this study is to understand and assess the usefulness of POA indicators as a potential hospital readmissions predictor. Readmissions are a function of the patient's disease state, comorbidities, and demographics. They are also a function of hospital care, principal diagnosis, procedure timeliness and execution. Further still, readmission can be influenced by the community: the type of post-acute care available and general community mores toward health care in general. Many of these factors are not in the hospital's control. POA is data captured at time of care and is currently not used as a predictor by CMS, perhaps because the data is relatively new and unknown, perhaps because of questions about the existing readmission model getting clouded by additional data. This study will focus on POA use to establish whether or not it would be a useful as a predictor for readmission.

A Medicare 5% inpatient Limited Data Set (LDS) for the period between 2007Q4 and 2011Q4 contains approximately 2.8M administrative claims records. One POA flag per secondary diagnosis field on each administrative claim yields ~15,000,000 pieces of data for those years. This is information particular to a patient's hospital visit that can potentially give new insight as to the patient state at hospital admission, information

previously unavailable. That patient state may have some correlation to hospital readmission.

Within that same set of claims, approximately 20% of claims, or 500,000 records, represent 30-day hospital readmissions. Also within that same set of claims <7%, 200,000 records, contain HAC conditions where POA=yes (patient admitted with condition), and of those records only ~3100 claims contains HAC conditions where POA=no (patient acquired condition after admission to hospital). Because the HAC/POA=no segment is so small, HACs will not be used in this study.

## 1.5 Research Question

POA indicators are new to the health research community. To date, there have been relatively few studies using POA indicators not involving HAC codes. Given the large amount of data and the limited POA indicator analysis, a large number of potential research questions could be investigated. The purpose of this study is to understand the relationship, if any, between POA and hospital readmission. I selected one question upon which to focus this effort.

*Would POA indicators be a useful set of data to include when modeling hospital readmissions?*

Hospital readmissions is a significant topic because of new laws. CMS, providers, private insurers, policy researchers, etc. are focusing on readmissions right now. POA data is relatively new and unstudied, it is national in scope, it is readily available within administrative claims data, and it captures patient state during a hospital stay in a way that no other administrative data can. It appears logical to combine POA with readmissions.

## 1.6 Research Database Process Overview

I have developed a data mining study investigating POA indicators for non-HAC administrative claims. To adequately study the POA indicators, it was necessary to understand the data and its relationships at a sufficiently deep level. Complicating this mission was the general form of the raw data itself. Creating a normalized research data

mining database solved the learning issue and the raw data issue as described in section 3.1 Data Mining. Index and readmission claims were then identified. A standard data set was then extracted for three conditions of interest: AMI, HF, PN. Once this step was completed, a base statistical readmission model using logistic regression was developed using extracted data. This base logistic regression model was extended to include several POA variables and compared to the base model.

## *1.7 Rationale and Significance*

The HAC/POA program is a relatively new program requiring providers to submit new data to CMS. This data is national in representing a huge injection of data into the claims system. For 2011, there are 12.4M claim records with ~10 POA indicators per claim or 123M new pieces of data. This study is significant in that:

- This is new data (< 5 years old).
- It is available nationally.
- It is available across most providers (short-term & IPPS).
- It is a significant volume of data.
- < 200 articles refer to HAC/POA data.
- Previous incident rates relied upon estimating harm or complications; now there is a direct way to measure.

Significance: what difference does this study make?

- This is research: adding to knowledge by investigating new data.
- Other researchers should know if these POA flags are useful or not within their models.
- This is a first attempt to categorize POA usefulness.
- Are there any potential HAC conditions present within the data?

This remainder of this research study contains the following chapters:

- Chapter 2 Literature Review contains a discussion of what has been done with POA indicators by researchers thus far. A number of examples were selected, and brief synopses presented as representative of different types of use.

- Chapter 3 Methodology frames the research study by defining data mining stages concluding with sections describing statistical approach and data necessary to test the research hypothesis.
- Chapter 4 Presents results of base hospital readmission model, POA characterization, and hypothesis tests.
- Chapter 5 Conclusions presents the usefulness along with implications of the research.

# 2 Literature Review

*Contains a discussion of what has been done with POA indicators by researchers thus far. A number of examples were selected, and brief synopses presented as representative of different types of use.*

## 2.1 Purpose

The significance for investigating POA indicators is to find alternate uses, understand how well POA indicators worked to fulfill study goals, and develop a set of POA uses this study could implement as statistical variables.

This section presents a systematic analysis of published literature surrounding HACs and POA indicators. As the data is managed by a U.S. government agency, many of the primary sources searched were other U.S. Government agencies, publications, reports, and congressional records, or contractors working for the aforementioned. Search sources include CMS, National Institutes of Health (NIH), PubMed, Google Scholar, Health Information Management Systems Society (HIMSS), and the American Medical Informatics Association.

## 2.2 POA Review

POA data has only existed from 2007 onward, and consequently the effective date range is limited. Hospital readmissions as a topic, however, has been studied for a decades. PubMed has references from 1956. It is impractical and unreasonable to review the entire history of hospital readmission starting from that point. A more reasonable date is the point is the timeframe of the ACA's Medicare Hospital Readmission Reduction Program (ACA passage 23-Mar-2010, HRRP effective 01-Oct-2013). It is during this timeframe when a national methodology was developed. Because this thesis is interested in the modeling aspects, 2010 was selected as the early date cutoff for this literature search. POA data is also U.S. health care system specific, so only U.S. publications were targeted.

The purpose of POA indicators in administrative claims data is to act in concert with ICD-9-CM codes to identify HACs, by definition. Consequently, the search terms "hospital acquired condition" and "HAC" along with "present-on-admission" and "POA" have explicit connotations on their own but also associations between them. "HAC/POA" are often found together under the CMS name for the program, in publications describing never-event evolution, and in hospital quality brochures for patients, which were removed from further consideration. When conducting this literature search, I started my search with "HAC" and expanded from there toward "POA". This process enabled me to make note of how the given study used POA data.

The research question limits the literature search to secondary uses of POA data. Any information in the literature consistent with that goal was included. If a study concentrated upon existing HACs (incident rates, trends, etc.) more so than using POA data in a new manner, I filtered it out as not relevant to this study. Any public-relation-oriented or marketing-oriented material, I filtered out of this study. There were a small number of technical publications targeting Information Technology (IT) personnel which aimed at clarifying data formats, expectations, and timelines, how to prepare and submit administrative claims, etc. These I also filtered out. The HAC/POA program is evolving. In my search, I found minutes from various committees describing pros and cons of splitting an ICD-9-CM code to better track an existing HAC, to track potential HACs, split or join MS-DRG by CC/MCC codes to differentiate complications from potential HACs, etc., and I also filtered these out.

It is also worth noting an absence of studies. Using "POA" and "readmission, "re-hospitalization," "30-day readmission," and a few addition terms resulted in zero search results. The combination of "present-on-admission" and "readmission" resulted in one study using the English phrase "present-on-admission," but not the meaning "administrative data fields called 'present-on-admission'." All told, 68 search results were examined resulting in the immediate culling of half the entries because they focused primarily on HACs and not secondary uses of POA. Of the remainder, I reviewed all

abstracts. I constructed a list of 23 usable search results for full reading resulting in the information presented in this study.

Having read each of the final search result documents, eleven investigations were not useful for this study and were omitted. The remaining twelve studies can be categorized based upon usefulness to this study by grouping them according to how and what data were investigated:

*Hospital POA Documentation Practices:* Five studies tried different approaches to increase POA reporting accuracy.

*The Yes/No Nature of POA Data:* Four studies were found using POA data to track specific disease states within a hospital setting.

*Groups of ICD-9 Codes and POA Data:* Three studies aggregated groups of ICD-9 codes and combined them with POA data to form downstream quality measures.

These categories are presented below, each detailing two studies in order to highlight POA usage within their category. A brief discussion of each category follows this section.

## 2.3 POA Data Studies

What are possible secondary uses of this new data? Literally, secondary use means for some purpose not originally intended. Because HAC and POA are already connected, secondary use here could mean a) HACs without POA (which is not possible given the technical definition of HAC as being an instance of a select ICD-9-CM code coupled with a valid POA indicator), or b) POA without HAC. It is this second option which I follow now. This subsection highlights articles found while conducting this literature search. Each represents a number of similar articles, and each comes at the POA indicator usage from a different aspect. It is these differences which are interesting. Each subsection below answers the question "how do researchers use POA indicators today?"

### 2.3.1 Hospital POA Documentation Practices

The following sections contain two study examples targeting the increased accuracy of POA reporting within a given hospital. Both studies restrict themselves to pressure ulcer documentation for the purpose of ensuring enough documentation gets into the medical record to clearly make the assessment of POA=yes or POA=no. This situation is intriguing as it represents a potential point of differentiation between hospitals: how well different providers document POA data.

#### *2.3.1.1 Using POA Data in a Clinical Decision Support System*

Are providers supplying enough documentation to make a determination of POA=yes within their Electronic Medical Record (EMR)? This question is of interest for two reasons:

- As a point of policy, providers will not be reimbursed if there is insufficient documentation to show POA=yes, the patient had the condition when they were admitted.
- A provider's POA rates may vary from that of it's neighbors because of its documentation practices.

Several studies examine an attempt to reduce HAC rates by adding Clinical Decision Support (CDS) mechanisms utilized by nurses at admission time. In several instances, CDS was inserted into the EMR process and workflow to check for and document conditions found on the patient at admission, providing enough information within the clinical record to make a determination of POA status for the administrative claim record. One such study, by C. Rodgers, titled "Improving Processes to Capture Present-on-Admission Pressure Ulcers" [31], targets pressure ulcer rates. The author notes that after several years in operation, the HAC/POA program and its implications are having an effect on providers.

> "Even if the PrU [pressure ulcer] was present-on-admission (POA) but not documented, it is considered as a hospital-acquired condition (HAC) and not reimbursable. The loss of repayment

for HACs, such as Stages III and IV PrUs, increases the importance of having an accurate process to identify and document PrUs that are POA."

The medical record must contain enough documentation to show POA=yes. HAC/POA is based on the standard-of-care guidelines. Hospital staff may not have training to differentiate "between PrU Stages I and II and maceration-associated skin damage (MASD) or deep tissue injury (DTI)" – different conditions contributing to false positive POA condition when one is not present.

This article clearly shows this provider understands the consequences of not gathering enough information to show POA=yes. The author describes a potential solution to adequately checking for pressure ulcers at admission time via a CDS approach. Within the admission workflow, nurses enter patient information and are prompted to check for pressure ulcers explicitly.

#### *2.3.1.2 Using POA Data in Terminology, Assessment, and Tracking*

"New opportunities to improve pressure ulcer prevention and treatment: implications of the CMS inpatient hospital care Present on Admission (POA) indicators/hospital-acquired conditions (HAC) policy. A consensus paper from the International Expert Wound Care Advisory Panel," [32] is another study of interest, which used POA indicators not as the source of the information, but as the end. In FY07, CMS reported ~260,000 cases of pressure ulcers they consider to be preventable. The authors conduct an informal study to investigate pressure ulcer stage definitions and how those definitions are used by clinicians to document POA. They find raising awareness is not enough to change pressure ulcer EMR documentation. They find with additional education, nurses and admitting physicians could be given a framework for incorporating pressure ulcer detections, along with daily inspection and information capture. This additional attention can significantly lower pressure ulcer incidence rates. This change involves shifting the documentation from the nurses who usually deal with skin issues to the admitting physician who is in charge of the patient's care. This shift raises the likelihood that enough documentation will make it into the medical record for the purpose of indicating

POA=yes. A consequence of this is better attention toward skin management and lowering of hospital acquisition.

This change is different from the pressure ulcer CDS in that it isn't throwing "technology to the rescue" solutions at the problem. It is providing the education to the people doing the work about the quality of care, total cost, and quality-reporting HAC aspects of pressure ulcers. Education in combination with a workflow framework allows initial assessment to flow into continued monitoring under a prevention strategy that results in the observed reduction of preventable pressure ulcers.

### 2.3.2 The Yes/No Nature of POA Data

These next studies illustrate an interesting point: one needs to understand both the nature of the POA indicator and the ICD-9 code for which it goes. The first is a POA success story showing that hospitals are not the source of Adverse Drug Events (ADEs) but are downstream from the event based upon the POA=yes indicators. The second study urges caution when combing POA data with certain day-surgery procedures, as all complications from them will be marked as POA=yes even though the complications occurred after the surgery.

#### *2.3.2.1 Using POA Data to Monitor Drug-Usage Events*

An AHRQ study titled "Origin of Adverse Drug Events (ADE) in U.S. Hospitals," [33] used the 2011 Healthcare Cost and Utilization Project (HCUP) data from 32 states. It also used POA indicators along with ICD-9-CM codes to classify within the Medicare system. By FDA regulation, ADE data flows to the FDA via drug manufacturers and not via hospitals. The authors use hospital visit as their unit of study. POA is used directly to tally ADE events. The authors had access to an entire year's worth of data, and report on a census not a sampling. They list actual ADE rates by using a subset of ICD-9-CM codes found in ~20M claim records. Without considering POA, ADE rates could be seen as hospital deficiency: poor oversight, lack of clinician training, etc. Once POA indicators are included however, the results show that "across all causes of adverse drug events, there were three times as many ADEs that were present-on-admission than

originated during the stay." POA data refutes the claim of poor hospital quality when it comes to drug events.

#### *2.3.2.2 Using POA Data to Monitor Hospital Procedural Effectiveness*

Instead of targeting a hospital's effectiveness, another study utilizing POA indicators focused on downstream events from Carotid Angioplasty and Stenting (CAS).[34] The authors propose a study to gage procedure complications by checking for secondary diagnosis conditions using POA indicators as a guide to determining if the conditions existed before or after the surgery. They study data from California for years 2005-2008, New York covering 2008, and New Jersey covering 2008. First, the authors define a set of ICD-9-CM procedure codes as targets, and then define a second set of codes as potential complications resulting from the procedures. They construct three statistical models one of which uses POA as a model parameter. The authors find conflicting outcomes between their models and illustrate with an example of stroke.

One identified side effect of Carotid Endarterectomy (CEA) and CAS is stroke. Half of the patients under study had "stroke POA=yes." To the model not using POA indicators, this presents an underrepresentation of hospital effectiveness by lumping all strokes as either "hospital acquired" or "the patient has a stroke before admission." However, patients could have a stroke with these procedures and have this fact go undocumented because of the administrative claims data limitation in recording such events: "patient had a stroke condition? Yes. Was there a stroke condition present-on-admission? Yes."

The authors make three additional points about the inherent difficulties of capturing POA documentation. First, CEA and CAS procedures are most likely to be same-day admission procedures. Any potential complications are therefore likely to be coded as POA=yes. Second, there is significant miscoding within the POA data the authors examined. It was their conclusion POA data miscoding represented poor classification and no evidence was seen to conclude it was a reimbursement gaming activity. Third, the ICD-9-CM diagnosis codes are not capable of indicating intraoperative or postoperative,

nor where complications such as stroke occurred. This last set of limitations are addressed in part with the release of ICD-10 coding.

### 2.3.3 Groups of ICD-9 Codes and POA Data

Two example studies combine groups of ICD-9 data with POA data to draw forward an inference that POA=no is a sign of some patient complications. In the first study, ICD-9 codes for sepsis were expanded and grouped by POA to track post-surgery infections. The second study is similar to the first mechanically, but significantly expanded to include a wide array of groups derived via the claims data itself. These represent complication to patient stay and may also be linked to readmission.

#### *2.3.3.1 Using POA Data to Monitor Infection Control Programs*

There are several studies which utilize POA codes as part of a statistical modeling effort. "Post-admission Sepsis as a Screen for Quality Problems: A Case-Control Study," [35] from Nov-2013 focuses estimating a hospital's infection control program by examining administrative claims for sepsis. There are five to six ICD-9-CM diagnosis codes for sepsis, depending upon ICD-9 version. The authors identify an additional 20 ICD-9-CM diagnosis codes associated with sepsis. Their case-control study attempts to identify links between hospital quality of care practices (as identified by these 20 secondary conditions) and hospital acquired sepsis cases by utilizing the POA=no indicators within the administrative claims data. The study's premise predicts a significant rise in the secondary sepsis conditions at hospitals having poor infection screening programs. Their method combines secondary conditions with POA indicators to form a quality measure proxy used to screen in-hospital infections. This study used baseline data from New York state hospitals starting with years just prior to the CMS mandate, New York being a leader in the early adoption of POA indicators implemented their program in the 1990s.

#### 2.3.3.2 Using POA Data to Monitor Hospital-wide Patient Complications

A significant study utilized POA indicators in a non-HAC setting attempts to identify Potential Inpatient Conditions (PICs) as the next circle out from HACs. In "Identifying Hospital-Wide Harm: A Set of ICD-9-CM-Coded Conditions Associated With Increased Cost, Length of Stay, and Risk of Mortality,"[36] the authors note the incidence of HAC conditions constitute a small number of claims, and therefore non-HACs contain the remainder of claims essentially in a category called "other." They propose a method to divide these "other" claims into PICs and "other" by defining a methodology leading to a set of PICs. They further describe how PICs could be used to improve health care. The study has two parts: forming PICs, and correlating PICs with other claim attributes.

*Forming PICs:* Central to the methodology is the use of POA indicators. Using a national data set covering 2008Q4 through 2010Q3, the authors partition claims into 32 service lines: gastroenterology, cardiology, etc., based on MS-DRG codes. These groups are then sorted into 4 larger groups. For each group, they create tallies of all ICD-9-CM diagnosis codes having POA=no status. The results are sorted by tally and significant ICD-9-CM codes were added to clusters by manual review.

*Correlating PICs:* 500,000 national claim records between 2008Q4 and 2010Q3 are used to link the cluster with length of stay (LOS), mortality, and cost. The authors identify 86 conditions they consider "high-impact conditions" associated with real phenomena at statistically significant rates. The study does not go into causality: "longer LOS might increase the risk for pressure ulcers rather than the pressure ulcers driving the longer LOS." But, PICs can indicate a hospital's rates above/below a given quality threshold. Within the PIC study, PIC attributes are correlated to a single administrative event. No hospital readmission modeling is conducted.

## 2.4 Summary

I have examined how researchers use POA data for secondary purposes. The previous section characterizes three different categories of POA data usage relevant to this study. Overall, the number of studies conducted using POA data is relatively small

due in part to the relative age of the HAC/POA program and the limited amount of data available, making it a harder subject to study. It is for this reason that a limited number of studies makes this a fresh area for research.

From using CDS to assist with POA accuracy to shifting documentation practices from nurses to doctors when recording HAC/POA data, comes the idea that a provider's POA rates may vary from its peers because of a hospital's documentation practices, training, support technology, or staff training. It may be possible to find a way to create a POA metric based at the hospital level that could be tracked over time.

From the ADE and procedural effectiveness study, it appears POA data has a personality based upon its associated ICD-9 code. Just because POA is used, doesn't mean it represents the same thing for different diseases. There is a nuance to POA data. Defining POA metrics and applying them over a wide array of conditions may not produce consistent results. Disease or procedure categories play a role in POA characteristics.

From the infection control and potential inpatient condition studies, the idea that POA data can cluster around certain diseases, codes, or procedures makes it possible to link more than one code or group together, such as AMI, HF, or PN.

# 3 Methodology

*Frames the research study by defining data mining stages, concluding with sections describing statistical approach and data necessary to test the research hypothesis.*

## *3.1 Data Mining*

Under more normal analytic circumstances, raw administrative claims files would be loaded directly into a statistical engine for analysis. Administrative claims files are usually received as one large file per calendar year, each file containing between 600,000 and 900,000 rows of data for a 5% random sample. In this case, I have graciously been granted access to Medicare's inpatient LDS files covering years 2007 through 2011 via the University Of Minnesota Carlson School Of Management's Medical Industry Leadership Institute (MILI). These files together comprise a total of ~2.9M rows of administrative inpatient claims data. However, each raw data file has a different number of columns, different column names for the same data field, different column order, different data encodings for the a column's year-to-year content, in addition to anywhere from dozens to hundreds of superfluous columns.

Raw data could not simply be loaded into a statistical package directly and consumed. It required additional processing to normalize and standardize column data. Any normalization process is time-consuming and potentially error prone. To prevent generating inferences upon questionable data, I chose to build a data mining database by writing my own custom database build and verification tools. Building a database is the first of three steps needed to build a data mining capability. These steps include:

1. Assembling the Research Database.

   I built a data mining database containing the claims and mashed them with DRG definitions, ICD-9-CM definitions, U.S. census data, Research Data Assistance Center (ResDAC) names and codes with field information, and CMS hospital data. Data cleansing is another activity completed with this

step. I validated all fields against expected values. By completing this step, I gained a base understanding of the exposed relationships within the data.

2. Creating Experimental Data Sets.

   From this source database above, I extracted study sets for analysis. Data mining is a process of learning and knowledge building. It is very much an *a priori* method of learning; often where the ends are not always clear from the start. The central output from this activity gave me a base understanding of the hidden relationships within the data.

3. Defining Statistical Models.

   I utilized logistic regression to explore interactions and relationships among data set variables.

I expand and explain each of these steps in greater detail below.

This approach controls for variation of data formats and ensures consistent definition of data across the study years. This approach also allows for flexibility to explore relationships within the data by using varied tools (spreadsheet, pivot tables, custom visualization, statistical) without committing to a particular format. This approach also allows for learning and extending of relationships to include new data. Finally, this approach allows imposing statistical rigor to questions of interest, ultimately including the research questions.

## 3.2 Assembling the Research Database

I chose MySQL [37] along with its associated tools as my database engine of choice. Within MySQL Workbench, I designed a snowflake schema within which I could maintain the claims records and layer external definitions and additional data (Figure 7). Administrative claims act as the central fact table. Surrounding the fact table are dimension tables, which are normalized and indexed data from the central fact table. These appear as veins of the snowflake radiating away from the center. A dimension example is the DRG table. The database build process loads all known MS-DRGs and crosslinks them to the central fact table. By this method, all central fact table DRGs are

validated against the known list. In addition, query time in either direction for example, DRG-to-central fact or central fact-to-DRG, is reduced by removing DRG information and replacing it with database keys.

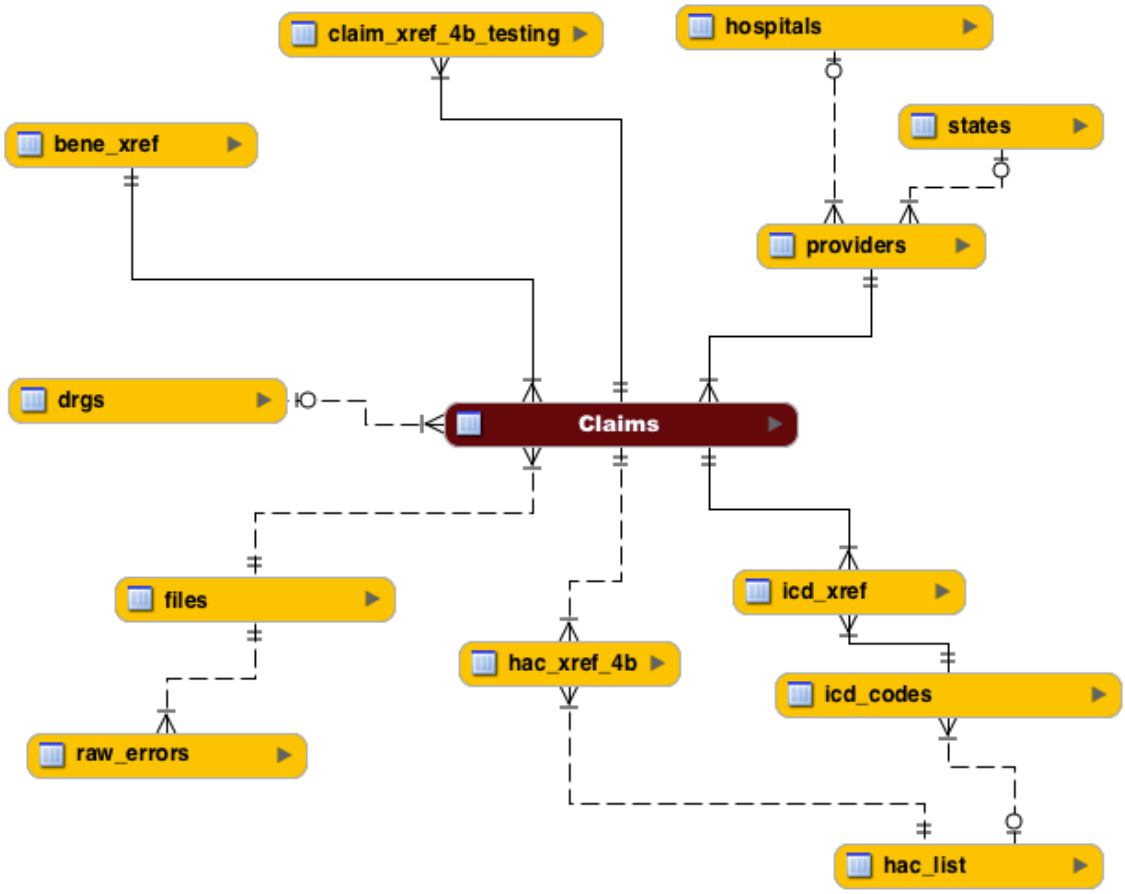



*Figure 7. Research database schema.*

Dimension tables have two types: normalized dimension tables used to uniquely identify a given dimension, and cross-reference tables, used to increase join efficiency between central fact and dimension data. An example of the latter is the "icd_codes" table and the central fact table "claims". In this case, I introduce an xref table to cross link ICD9 codes to rows in the fact table which results in vastly reduced query times at a cost of additional SQL code. Xref tables are also beneficial when, as in this case, there

are multiple sources for the data: ICD-9-CM definitions are available from multiple sources, and the administrative claims data is available from CMS. The integrity of each table can be validated independently from the others because they are only connected at the xref table.

Dimension tables ebb and flow as different questions arise. The schema presented in Figure 7 is an example of a point-in-time snapshot of the research schema. It is a conceptual representation showing tables as the colored boxes and connections between them indicating relationships. The symbols on the lines next to the tables indicate cardinality of the relationship. Provider and Claims, for example, contains two small lines next to Providers meaning "every claim record has one and only 1 provider associated with it." There is a relationship in the reverse direction as well indicated by what has become known as "crow's foot" notation. The reverse relation indicates "every provider is associated with one or more claims records." These concepts are codified in the database itself. Note also "claim_xref_4b_testing" and "hac_xref_4b" These tables are examples of the iterative analysis process (only two were shown for brevity). Table contents and relationships change as knowledge expands and as wider dimensions are combined into the mix. From these tables, extracts are drawn and fed into R [38] for visualizations, first-order statistics, and eventual inference.

The input data or base data used for this research comes from many sources. Within this section, I identify each data source, versions if applicable, relevant fields within the source and any additional notes.

### 3.2.1 Administrative Claims Data

This study used Medicare's administrative inpatient claims data, 5% LDS spanning the calendar the years 2007 through 2011; five years in total. POA indicators and MS-DRGs along with main ICD-9-CM diagnosis codes did not exist before 2007Q4, and so I omitted it from the research database. This data is de-identified. It contains no patient-identifiable data, all locations are limited to state level and 3-digit zip code (if the zip code is present at all), and no physician identifiers are present. It does contain the claim's true THRU_DT values, making it possible to construct an accurate inpatient history of

events. This is different than the usual LDS files where dates are truncated to the year and quarter.

The following fields were used within this study. Their definitions can be found in ResDAC "LDS Inpatient SNF [Skilled Nursing Facility] Claim Record Data Dictionary."[39]

#### *3.2.1.1 Coded Fields*

Administrative claims data contains numerous fields whose values are keys, the actual meaning of which could be simple enough to be defined in the LDS Data Dictionary. For example, GNDR_CD contains the code definitions: "1=Male, 2=Female, or 0=Unknown." When these codes represent data shared across several CMS databases, the definitions are maintained elsewhere; ICD-9-CM definitions are not defined within the data dictionary, but are available from the CMS web site. Still other fields have definitions maintained by ResDAC, "a CMS contractor that provides free assistance to academic, government and non-profit researchers interested in using Medicare and/or Medicaid data for their research." [40] These field definitions or decodes are available per field by using the identified table name in the LDS Data Dictionary for the particular field.

All administrative clams fields used in the research database have definitions taken from the ResDAC web site. I saved these definitions into individual <field>.raw text files, where "field" is represents names such as "bene_gndr", "fac_type", or "ptntstus". I manually scanned the raw text for junk, invalid characters, confusing definition strings, historical encodings, notes, sub-types based on other fields not present in the given definition file, etc. Then I manually coerced these files into <field>.data file; a form capable of being imported directly into a database or into R. I stored all codes and definitions in a table called names_codes along with a category identifying the field to which they belong. I do not show this table in Figure 7 above because it has so many relationship links it obscures the snowflake design.

#### 3.2.1.2 ICD-9-CM Codes and Definitions

The admitting diagnosis field (AD_DGNS), primary diagnosis field (PDGNS_CD) and ten claim diagnoses codes (DGNSCDx) contain ICD-9-CM diagnoses codes. A claim's six procedures codes (PRCDRCDx) contain treatments performed on the beneficiary. These codes may vary slightly year to year as new conditions or treatments become known and as old ones become disused. ICD-9-CM codes are reviewed periodically and updated on CMS's fiscal year, which starts each 01-Oct. I loaded each ICD-9-CM code set for the time span 2007Q4 through 2011Q3 into a separate table. I assigned version numbers to each code for back-checking and code normalization. All ICD-9-CM diagnosis and procedure codes I normalized to a common value in order to allow year-to-year comparisons using code and definition files provided by CMS.[41]

#### 3.2.1.3 Claim POA Flags

The ten POA fields (CLMPOAx) accompany the ten diagnosis fields (DGNSCDx). CMS does not process these fields in any way. These fields contain raw data from the providers. Consequently, as part of the data loading and validation step, I translated 'y' values to 'Y', etc. All unknown values I considered invalid and treated as "no data", "null", or "NA" for statistical purposes.

#### 3.2.1.4 Beneficiary Date of Birth

The beneficiaries in the LDS files used in this study I binned to 5-year values as described in the LDS Data Dictionary.

#### 3.2.1.5 Provider ID

The provider identification is the institutional provider Medicare certifies to provide services to the beneficiary. The provider ID field within the administrative claims data is a non-atomic a 6-character "ID" composed of two parts with up to four meanings:

- Two character positions encode digits that represent the Social Security Administration's (SSA) state ID code. (See 3.2.3 State Code Crosswalk File below.)

- Four character positions uniquely encode a "hospital ID" within the given SSA state.
- The hospital ID is a range of numbers within a series which further classify the hospital's type. (For example, a hospital ID in the range 0001 to 0879 identifies a "short-term hospital.")
- The hospital ID may have optional character overlays creating an additional layer of identification. (For example, codes 'E' and 'F' indicate an Emergency Hospital, and code 'M' indicates a Psychiatric Unit within a Critical Access Hospital.)

I preserved all of these available codes and their meanings within the research database.

#### *3.2.1.6 Claim ID and Beneficiary ID*

The CLAIM_NO field uniquely identifies a single administrative claims record. The BENE_ID field is a de-identified beneficiary ID that has been replaced with a random string to prevent re-identification of the patient. It does, however, allow me to construct a patient's claim history.

#### *3.2.1.7 Claim Through Date and Utilization Day Count*

THRU_DT is the last day of a beneficiary's hospital stay. UTIL_DAY is a number that represents the total number of days the patient remained within the given provider's facility for treatment. It can be a tricky number to use. HACs can increase a patient's treatment at an inpatient facility, a side effect of which may be a longer stay in the hospital: a higher day-count. It is also true that longer stays may increase the likelihood of a HAC occurring.[42] Because HACs act as both a cause of longer stays and an effect of staying in a hospital longer, it becomes difficult to separate them when examining administrative claims data, particularly with a 5% LDS sample. The patient's admission date is a calculated value: THRU_DT – UTIL_DAY. Admission date is used to derive any days between hospital visits for readmission analysis.

#### *3.2.1.8 Financial Fields*

TOT_CHRG represents the provider's "retail" cost for services supplied to the beneficiary. PMT_AMT is the actual amount CMS paid to the provider. BENE_DDCTBL_AMT is the beneficiary's deductible amount given to the provider. Beneficiaries may also have coinsurance with a liability amount: BENE_COIN_LBLTY. The following derived fields I use in the research database:

Patient's Cost = BENE_COIN_LBLTY + BENE_DDCTBL_AMT

Amount Allowed = PMT_AMT + Patient's Cost

### 3.2.2 MS-DRGs

All administrative claims starting with 2007Q4 have an associated MS-DRG code. The MS-DRG codes are available from CMS [43]. This file includes the MS-DRG numbers plus their one of 26 Major Diagnostic Category (MDC) codes, a type of either "surgery" or "medical", and a brief description. Contained within the description string is an optional "with/without" "CC/MCC" status indicator, which has been extracted and saved for DRG complexity comparisons. I have linked all of the MS-DRG codes in the research database to DRG definitions within this file.

### 3.2.3 State Code Crosswalk File

The SSA state codes are a 2-digit number linking a 2-character abbreviation and a text string representing a U.S. state, a territory, or some other geographic region. This numbering does not match the Federal Information Processing Standard (FIPS) state numbering scheme used by the rest of the U.S. government. To link to U.S. census data, I converted the SSA state ID numbers to FIPS state numbers using a crosswalk file. [44] I save both the SSA state information and the normalized FIPS standard within the research database.

Because the hospital ID number within a state is also used to encode a hospital type by using ranges of hospital ID numbers, some states ran out of space in their numbering scheme. To solve this, SSA created "new" state numbers to preserve the existing hospital type encoding at a cost of duplicating state IDs: for example, Texas uses SSA state IDs

45, 67, and 74 while the FIPS system uses only 48. SSA state 2-digit ID number to FIPS state name decodes are available from multiple sources. [45]

### 3.2.4 US Census Population Data

I added population data available from the US. Census Bureau to the state tables. Census data uses the FIPS state and county codes. I used the crosswalk file to link population, demographic, and economic data to the SSA codes for the corresponding territory.

### 3.2.5 Hospital Attributes

Basic hospital information is available from CMS via their hospital cost data files.[46] The data available contains hospital name, location, and attributes such as number of beds, number of Medicare patients served, and total number of patient discharges during the calendar year. This data is self-reported, compiled annually, and may be missing or rounded. The address information may refer to "head office," "administrative building," etc., and not the actual hospital location. Within the hospital information, some bed fields were blank and could not then be used for analysis. They represent eight hospital records attached to seventeen administrative claims, and I deem them acceptable to omit. Then, I merged hospital attributes with the administrative claims data set via the providers dimension table.

### 3.2.6 Data Verification Checks

I programmed all database loading steps to increase reliability and to make iterative design changes possible. The data input process validated each field for unrecognized input as it was read from the raw files. When I discovered such an item, I wrote the file name, input line number, database row number, field, and error message to an error table. Then I reviewed and corrected these issues as a post-build process. Items included detecting punctuation characters instead of digits.

I also created a full loop-back check mechanism. No data was harmed in the making of the research database. Therefore, a given row could be read, reassembled, and

compared to the original line from the given file to determine database import process integrity. This process ensures that I did not accidentally transmogrify any data.

I checked all creation, manipulation, and checking programs, and source data files into GIT,[47] a version control program used to facilitate both change tracking to source code and correct system versioning across system pieces.

## *3.3 Creating Experimental Data Sets*

### 3.3.1 POA data

As a reminder, administrative claims in this study are limited to between 0 and 10 diagnosis codes and corresponding POA indicators. A POA=no indicates a secondary condition that occurred in the hospital after the patient was admitted. CMS does not filter or process these values when assembling LDS files, meaning the POA fields contain "raw" entries directly from the providers. I removed invalid values as described in Section 3.2.1.3.

For the purposes of adjusting reimbursement, CMS equates a POA=u (unknown) indicator with a POA=no. Likewise, they equate a POA=w (cannot make clinical determination) with POA=yes. All of the statistical data sets I created for this study follow this rule.

### 3.3.2 General Techniques

The purpose of this study is to understand and assess the usefulness of POA indicators as a potential hospital readmissions predictor. This section describes several methods used to explore data, relationships, and form a basis from which statistical models may be drawn.

In most cases, the first step in understanding data is running simple queries against the research database. I formed queries using MySQL Workbench, and revealed first cut statistics via simple aggregation. Once I determined an interesting result set for a query, I exported the set into comma-separated value files where the data set can be imported into the next tool, typically Excel. I repeated the process adding additional dimension data

such as hospital attributes, patient demographics, and admission quarter or month. At some point, the data set becomes impractical for Excel and the analysis space turns over to R, enabling statistical modeling along with more complex visualization.

Analysis can also extend beyond what a simple SQL statement could provide. As with the change of tools above, there is a data aggregation tool change. For this research project, I have chosen to use Perl [48] to provide all of the data transmogrification and programming needs. I created and validated small sample sets using a county or state as an upper limit on size. Then I checked the extracted sets via summary statistics, charts and other visualizations. I repeated this process to include additional columns making the data set richer, or to add more rows including more observations. With confidence in the incremental data, I could draw a complete data set and examine internal data relationships in fine detail.

### 3.3.3 Exploring POA Indicators via Excel and R

Aside from the usual spreadsheet functions such as sum(), mean(), quartile(), which are available in SQL or any programming language, Excel has two key features useful for analytics: quick charts and pivot tables. Data visualization is often referred to as presentation: the culmination of the analytic process represented in a graphic form. Backing up to the initial stages of analysis, this ability of letting the eye consume and relate information makes it possible to conduct visual experiments quickly and efficiently. Text and tables of numbers work well for small amounts of data. Subtleties may easily escape the reader. But a picture has a better chance at conveying a message hidden within the data than a large table or block of text. It is this visual efficiency that makes early charting compelling as an analytic tool.

Excel's pivot tables extend charting by easily changing variables of interest. They accomplish this by summing and cross-summing data. For example, converting a list of claims records, one can create a pivot using hospitalizations by state and by then by month. The resulting table is now a new way to consider the original data. It may be valuable to look at as is, or by creating a new chart. Either is appropriate at the initial stage of data analysis: generating initial data sets.

When an initial data set is determined to have more potential than another set, it is useful to have more extensive tools with which to examine the set. For this extended tool set, I have chosen R, a statistical programming language and R Studio, an environment to interact with the data. I can create R scripts to process input data, transforming it for detailed analysis using a variety of inference algorithms. R itself has numerous built-in statistical functions just like all the other statistical packages: T-Tests, ANOVA, Chi-Square, GLM for linear and logistic regression. R is an open-source system allowing users to create and share modules with the R community. Among these modules are newer statistical techniques or visualizations including: ggplot2 and Caret. The former is widely believed to be the standard for R data visualizations and the later contains a best-of-breed implementation of the Random Forest ensemble machine learning technique.

### 3.3.4 Beneficiary Hospitalization Viewer

The next step to understanding administrative claims data is to look beyond the single claim and look into individual patient histories. Several of the studies listed in Section 2 aggregate data at the hospital level. Others aggregate data at the state level. I found a need to understand diseases through patients who had them because the disease conditions interact with each other. I created a patient hospitalization viewing tool to allow me to see patient history rolled up in an aggregated format. Of specific interest is seeing providers phase in POA reporting from year to year. Another point of interest is seeing a patient's hospitalizations, their duration and time until the next (Figure 8). All patient identifiable information has been redacted. HAC conditions are highlighted in yellow. POA=no conditions are highlighted in blue, and days between claims are highlighted in green indicating more than 30 days and red indicating less than 30 days. Note that this particular patient has AMI and PN primary diagnosis conditions marked by 'P', but also has HF as a secondary condition.

Just building this tool provided a glimpse of the administrative data researchers may not ever get. It allows claim examination if an individual patient's disease progresses. It allows seeing downstream complications of various procedures, and allows for assessing

how well the 30-day readmission logic seems to be working. I came back to this tool time and time again to understand what data exists and what my tools had implied.

| Claims Codes | | 1 | 2 | 3 | 4 | Description |
|---|---|---|---|---|---|---|
| 1) 486 | D | P1-Y | - | A | - | Pneumonia, organism unspecified |
| 2) 41401 | D | 2-Y | 8-Y | - | 2-Y | Coronary atherosclerosis of native coronary artery |
| 3) 4941 | D | - | AP1-Y | - | - | Bronchiectasis with acute exacerbation |
| 4) 25000 | D | 3-Y | - | 7-Y | - | Diabetes mellitus without mention of complication, type II or unspecified type, not stated as uncontrolled |
| 5) 2948 | D | 5-Y | - | 8-Y | - | Other persistent mental disorders due to conditions classified elsewhere |
| 6) 4019 | D | 4-Y | 7-Y | - | - | Unspecified essential hypertension |
| 7) 41011 | D | - | - | - | P1-Y | Acute myocardial infarction of other anterior wall, initial episode of care |
| 8) 41090 | D | - | - | - | A | Acute myocardial infarction of unspecified site, episode of care unspecified |
| 9) 48242 | D | - | - | P1-Y | - | Methicillin resistant pneumonia due to Staphylococcus aureus |
| 10) 51881 | D | - | 2-Y | - | 8-N | Acute respiratory failure |
| 11) 5849 | D | - | - | 2-N | 7-N | Acute kidney failure, unspecified |
| 12) 5990 | D | - | 3-Y | - | 4-Y | Urinary tract infection, site not specified |
| 13) 78605 | D | A | - | - | - | Shortness of breath |
| 14) V103 | D | 7-1 | 9-1 | - | - | Personal history of malignant neoplasm of breast |
| 15) 03289 | D | - | 4-Y | - | - | Other specified diphtheria |
| 16) 04112 | D | - | - | - | 5-Y | Methicillin resistant Staphylococcus aureus in conditions classified elsewhere and of unspecified site |
| 17) 1121 | D | - | - | 6-Y | - | Candidiasis of vulva and vagina |
| 18) 40390 | D | - | - | 4-Y | - | Hypertensive chronic kidney disease, unspecified, with chronic kidney disease stage I through stage IV, or unspecified |
| 19) 412 | D | 6-1 | - | - | - | Old myocardial infarction |
| 20) 42731 | D | - | - | - | 9-Y | Atrial fibrillation |
| 21) 4280 | D | - | - | - | 6-Y | Congestive heart failure, unspecified |
| 22) 49121 | D | - | - | 3-Y | - | Obstructive chronic bronchitis with (acute) exacerbation |
| 23) 5163 | D | - | 5-Y | - | - | Idiopathic fibrosing alveolitis |
| 24) 5853 | D | - | - | 5-Y | - | Chronic kidney disease, Stage III (moderate) |
| 25) 82021 | D | - | - | - | 3-Y | Closed fracture of intertrochanteric section of neck of femur, HAC(5.1)=[Fracture] |
| 26) E8844 | D | - | - | - | 10-U | Accidental fall from bed |
| 27) V0254 | D | - | - | 9-1 | - | Carrier or suspected carrier of Methicillin resistant Staphylococcus aureus |
| 28) V1582 | D | - | 6-1 | - | - | Personal history of tobacco use |
| 29) 3722 | P | - | - | - | X | Left heart cardiac catheterization |
| 30) 7935 | P | - | - | - | X | Open reduction of fracture with internal fixation, femur |
| 31) 8856 | P | - | - | - | X | Coronary arteriography using two catheters |
| 32) 966 | P | - | - | - | X | Enteral infusion of concentrated nutritional substances |
| 33) 9672 | P | - | - | - | X | Continuous invasive mechanical ventilation for 96 consecutive hours or more |
| 34) 9904 | P | - | - | - | X | Transfusion of packed cells |
| 1) 177 MED | | - | - | X | - | w MCC Respiratory infections & inflammations, mdc=04 |
| 2) 190 MED | | - | X | - | - | w MCC Chronic obstructive pulmonary disease, mdc=04 |
| 3) 195 MED | | X | - | - | - | w/o CC/MCC Simple pneumonia & pleurisy, mdc=04 |
| 4) 981 SURG | | - | - | - | X | w MCC Extensive O.R. procedure unrelated to principal diagnosis, mdc= |

| Hospital | 1 | 2 | 3 | 4 | Total days, Hospital Info (if available) |
|---|---|---|---|---|---|
| Admission Status | 7 | 7 | 7 | 4 | 4 = Transfer from hospital - The patient was admitted as an inpatient transfer from an acute care facility. 7 = Emergency room - The patient was admitted upon the recommendation of this facility's emergency room physician. |
| Admission Dates | [redacted] | | | | |
| [redacted] | | | | 24 | 24 days. |
| [redacted] | 2 | 3 | 7 | | 12 days. |
| Through Dates | [redacted] | | | | 36 total days |
| Days Between | 156 | 16 | 43 | | |
| Claim Status | | | | | |
| Discharge Status | 06 | 01 | 03 | 20 | 01 = Discharged to home/self care (routine charge). 03 = Discharged/transferred to skilled nursing facility (SNF) with Medicare certification in anticipation of covered skilled care -- (For hospitals with an approved swing bed arrangement, use Code 61 - swing bed. For reporting discharges/ transfers to a non-certified SNF, the hospital must use Code 04 - ICF. 06 = Discharged/transferred to home care of organized home health service organization. 20 = Expired (did not recover - Christian Science patient). |
| Claim Number | [redacted] | | | | |

*Figure 8. Sample AMI patient viewer output.*

After working with claims data for a while, it is easy to make the leap from seeing this data as "clinical" in nature overlooking its "administrative" aspects. Under the former, one can come to expect the data to be in a certain format, nice and orderly, where all data is accurate, complete, without conflicts. Under the latter, there are changes, amendments, refusals; a patient can be seen at multiple hospitals, transferred to long-term care and back. These changes can be seen when looking through individual patient histories where they may not be seen through aggregated results or filtered out as "outliers."

### 3.3.5 Data Exclusion Criteria

CMS uses AMI, HF, and PN within their hospital readmission evaluation process. This study limits itself to these conditions as well. Three sets of study data need to be produced: one for each condition of interest: AMI, HF, and PN. However, the process of establishing readmission requires looking between claims as well as within claims. Each patient's claim history must be established. Index events and subsequent readmissions must be established, all as described in Section 1.2.2. CMS describes the process adopted by NQF and uses the following conditions to remove certain claims records from consideration with identifying potential index claims. CMS contracted with Yale University/Yale-New Haven Hospital-Center for Outcomes Research & Evaluation (Yale-CORE) to define preliminary hospital readmission measure methodologies for AMI, HF, and PN, publish these measures, solicit comments, and incorporate feedback into a final set of measures.[49, 50, 2] These exclusion criteria are true for all three data sets. These conditions are:

- Claim not from U.S. state and D.C.
- Claim not from short-term acute-care hospitals
- Claim not an IPPS claim type
- Patient < 65 years old at time admission
- Patients who die in-hospital
- Patients who are transferred out for escalated care to other facilities

- 30-day exclusion for same diagnosis. See Figure 5, option 1. Claim can be an index or a readmission, but not both.

AMI has two additional exclusion criteria: 1) Claims with zero utilization days (length of stay) are not considered "valid" AMI conditions. 2) Claims with "staged procedures," where a hospital discharges a patient pending a scheduled procedure, are not under a standardized clinical guideline making them difficult to filter in or out. This study uses the first exclusion criteria, but not the second. HF also has an additional exclusion based on a beneficiary's enrollment in Medicare Part A and Part B. This study does not utilize Part B data, so this criteria will be ignored.

### 3.3.6 Data Inclusion Criteria

Once I deemed claims eligible for inclusion in the study, I assigned each claim an index and readmission status across all years using methodology defined in Figure 5, option 2 (a claim can be either index or readmission or both). A second pass reduced this list to option 1 (a claim can be either index or readmission but not both). I extracted all potential index claims having a primary diagnosis code found in Appendix B – Principal Diagnostic Codes for AMI, HF, and PN into one of the three disease data sets. I used these data sets to develop the base statistical hospital readmission model. I kept any claims with incomplete data in the data set and marked the missing information as "null".

## *3.4 Defining Statistical Models*

### 3.4.1 Approach

I used logistic regression to evaluate the usefulness of POA data within the hospital readmission estimation process. With regard to hospital readmission modeling in combination with the data available for this study, it makes sense to identify a model using a minimal set of predictors. The study's purpose is to understand the impact of incorporating POA data into readmission models, not to create a "new and improved" readmission model nor duplicate efforts of Federal agencies. I deemed it sufficient to approximate existing hospital readmission models. Therefore, I developed one statistical

model as a base model representing hospital readmission. I tested the base model against each primary diagnosis condition's data set varying POA data included and varying readmission time period as defined below.

Hospital readmission periods are defined as: 7 days, 15 days, or 30 days from the given claim's through-date field. CMS uses 30 days as its standard, which was adopted from the NQF 30-day hospital readmission. MedPAC refers to 7-, 15-, and 30-day readmission windows.[8] See Figure 9. Seven days was the lowest number I found in the literature. Researchers thought seven days was long enough to capture a reasonable proportion of side effects and complications following surgery. However, seven days is not long enough to capture Hospital Acquired Infections (HAIs) adding credibility to fifteen days. Other studies reference thirty days. I have included all three readmission periods because they are all in-use. The section below describes additional detail about the contents of the model, the independent and dependent variable.

**TABLE 5–2** **Potentially preventable hospital readmission rates**

| | Patients readmitted to hospital within: | | |
|---|---|---|---|
| | **7 days** | **15 days** | **30 days** |
| Rate of potentially preventable readmissions | 5.2% | 8.8% | 13.3% |
| Spending on potentially preventable readmissions (in billions) | $5 | $8 | $12 |

Source: 3M analysis of 2005 Medicare discharge claims.

***Figure 9. Potentially preventable readmissions.***

### 3.4.2 Hypotheses, POA Explanatory Variables, and Final Base Model

The data and methodology describe a base readmission dataset and a repeatable method to identify hospital readmissions. It is upon this work that POA data can be incorporated and analyzed. I now develop the following hypotheses and predictor variables:

$H_1$ *The presence of POA=no indicators on an administrative claim is a predictor of hospital readmission.*

***p1*** is defined as a dichotomous variable set to 1 if *any* POA=no indicators were found on the administrative claim.

$H_2$ *The number of POA=no indicators on an administrative claim is a predictor of hospital readmission.*

***p2*** is defined as the count of the number of POA=no indicators found on the given administrative claim.

$H_3$ *The rate at which a hospital fills POA fields across all of its administrative claims is a predictor of hospital readmission.*

***p3*** was set as an aggregated variable at the hospital level by counting *all* claims with *any* POA indicator set on *any* of its administrative claims. Due to the fluctuation in POA adoption, CMS directions or guidance, and overall adherence within the early years of the HAC/POA program, I divided this variable into rates per hospital per claim year.

The predictors ***p1*** and ***p2*** extend the readmission model by adding data richness, making the data set wider versus making the data set longer by adding more rows of data at the individual claim level. ***p3*** attempts to use POA adoption rates aggregated at the hospital level to add new data to the readmission model as a method to add statistical richness. This rate uses all data available for all of a hospital's claims, not just subsets for individual diagnoses.

I harvested the following variables from claims, aggregated them, or mashed them into the diagnosis condition datasets. These variables did not vary over the course of testing, and provide a reasonable approximation of the CMS data given the limited set used for this study. Combining these variables produces a base hospital readmission model.

***B*** represents a beneficiary matrix containing 5-year binned ages and gender. I included race in the initial estimation model development, but I removed it due to sparseness.

***H*** represents hospital characteristics matrix containing state, number of beds, a rural/urban dichotomous variable, and CMS region. I used state data to compare state-to-state results. I dropped CMS regions from the model as they were always statistically insignificant.

***S*** represents a claim severity matrix containing MS-DRG information. I included this information as a way to account for the severity of conditions or of complications at the claim level. There may be some interaction between POA and the severity, but the relationship is not causal.

***Y*** represents the estimation model's dependent variable of hospital readmission. I set it to 1, indicating the current claim has an associated readmission claim within the defined test interval, and I set it to 0 otherwise. For each patient, I converted the number of days between potential index events into a dichotomous readmission variable. Estimates having a positive sign indicate increased log odds of readmission within the defined test period, and negative estimators indicate a reduced log odds of readmission.

Combining all of the predictors produces the following hospital readmission model:

$$Yn_{bcht} = \beta_0 + \beta_1 B_b + \beta_2 P_{cht} + \beta_3 H_{ht} + \beta_4 S_{ct} + \beta_5 C_c + \varepsilon_c$$

Such that the variables represent:

$Yn_{bcht}$ dependent variable: 1 if administrative claim is a readmission within *n* days of index discharge, where *n*={7, 15, or 30}; or 0 otherwise

$\beta_0$ model intercept

$B$ beneficiary matrix: age, gender, race, urban/rural, state, etc.

$C$ administrative claim: utilization days, cost, MS-DRG, etc.

$P$ POA data (see above)

$H$ hospital attributes

$S$ severity measure for a given year

$\varepsilon$ error term

Where the subscripts represent:

$b$ beneficiary subscript

$c$ claim subscript

$h$ hospital subscript

$t$ time (year) subscript

This model was tested using three datasets, one for each AMI, HF, and PN. Each model tested hospital readmission at three different readmission periods: 7-day, 15-day, and 30-day. Each model tested each hypothesis, and so resulted in

3 conditions
3 readmission periods
x 3 POA variables
27 total models

# 4 Results

*Presents results of base hospital readmission model, POA characterization, and hypotheses tests.*

## *4.1 Descriptive Statistics for Hospital Readmission*

The purpose of this section is to establish the equivalence of this study's hospital readmission model to other readmission studies.

### 4.1.1 Ability to Identify Hospital Readmissions

I filtered each administrative claim for eligibility in readmission processing to best match the CMS criteria implementation given both the limitation of this study's input dataset and the limitation of the POA data. As seen in Table 6 below, I removed about half of available claims from the study for one or more reasons.

*Table 6. Summary of ineligible claims*

| *Exclusion Criteria (these categories overlap)* | *Claim Count* | *Claim Percent* |
|---|---|---|
| Claim from cancer hospital | 5677 | 0.20% |
| Non-payment, zero claims | 241 | 0.01% |
| Not a US state or DC | 9933 | 0.35% |
| Not main hospital (psych or rehab units) | 120962 | 4.25% |
| Not IPPS claim | 166016 | 5.83% |
| Not short-term (acute care) hospital claim | 203146 | 7.14% |
| Patient <65 or age unknown | 778864 | 27.36% |
| Patient died or still a patient (intermediate claim) | 102271 | 3.59% |
| Zero-day claim or delta (claim-to-claim) | 181110 | 6.36% |
| Claim replacement of previous claim | 406310 | 14.27% |

| | | |
|---|---|---|
| Claims excluded | 1417117 | 49.78% |
| Claims included | 1429489 | 50.22% |
| Total claims in source data | 2846606 | 100.00% |

### 4.1.2 Readmission Rates Compared to Published Rates

CMS publishes Risk-Standardized Readmission Rates (RSRRs) using a process which compares an estimated readmission rate per hospital to observed rates per hospital

per year, (Table 7). The "Medicare Hospital Quality Chartbook 2012: Performance Report on Outcome Measures" [51] shows median value by condition per year.

*Table 7. Medicare Hospital Quality Chartbook 2012 readmission rates.*

**Table A.3. Trend in Median Hospital RSRRs**

| | Median (Range) of Hospital's RSRR (%) | | |
|---|---|---|---|
| | 2008 | 2009 | 2010 |
| AMI | 19.9 (16.9, 24.1) | 19.7 (15.7, 26.4) | 19.4 (15.4, 23.8) |
| Heart Failure | 24.9 (20.1, 32.7) | 24.7 (20.3, 31.5) | 24.6 (20.1, 30.8) |
| Pneumonia | 18.2 (14.1, 23.5) | 18.5 (14.4, 24.0) | 18.4 (14.9, 24.7) |

This study used a 5% LDS sample, and there was insufficient data available to determine a median value of hospital-based readmission rates when aggregating at the principal diagnosis code. AMI in this study has fewer claims records per year than the number of hospitals, resulting in median values of 100%, 50%, 33%, and down to 0%. Therefore, this study used sample proportions across all states and hospitals (Table 8). Consequently, there are two measures of central tendency: sample proportion from this study and CMS's median readmission. AMI proportion values appear to be lower than the median numbers, while HF and PN appear to be the same.

*Table 8. Readmission proportions with 95% confidence intervals.*

| *Conditions* | *2008* | *2009* | *2010* | *2011* |
|---|---|---|---|---|
| AMI | 16.1 (14.6, 17.7) n=2158 | 16.8 (15.2, 18.5) n=2032 | 16.5 (14.9, 18.2) n=1911 | 16.7 (15.0, 18.4) n=1927 |
| HF | 24.8 (24.2, 25.4) n=19,457 | 24.9 (24.3, 25.5) n=20,605 | 24.4 (23.8, 25.0) n=18,278 | 24.6 (23.9, 25.2) n=19,305 |
| PN | 17.8 (17.2, 18.4) n=15,502 | 18.2 (17.6, 18.9) n=15,080 | 18.1 (17.5, 18.8) n=13,648 | 18.1 (17.5, 18.7) n=15,360 |

Both sets of numbers fall within a range of published literature for the three conditions of interest. Repeatability is complicated by a number of factors. Predating the national standard, readmission was calculated by a single hospital and extrapolated upwards, or by aggregated state data and then extrapolated upward. The NQF and AHRQ publish guideline measures and associated rates, but those may contain Veterans Affairs (VA) data or data for persons <65 years of age.

Given the overall published range and wide delta from study to study, the numbers calculated and listed above by this study are reasonable.

## 4.2 Descriptive Statistics for POA Indicators

There were few studies within the published literature that contain raw statistics detailing POA data itself. In this section, brief characterizations of different POA aspects are made to provide a better understanding of the data used within the regressions.

### 4.2.1 Raw POA Numbers

In FY2008, a large number of providers regularly included POA flags for all of their administrative claims, not just claims containing a HAC condition. Other providers only supplied POA data when a HAC existed in the secondary diagnosis fields. By 2010 with additional guidance from CMS, all providers regularly use POA indicators for all secondary diagnosis codes. Overall, within the study time period 2007Q4 through 2011Q4, 67% or approximately 1.9M claims contain valid POA flags.

The raw administrative claims data contains ten POA fields defined as one character in length. CMS defined six possible values to be used within the POA fields (Table 5). The raw data contains additional characters. They are to contain the values listed in Table 5. However, after adjusting for uppercase/lowercase settings, these fields contain a wide assortment beyond the expected CMS defined values seen in Table 9 below.

*Table 9. POA indicator frequencies.*

| *POA Indicator Value* | | *Count* | *Percent* |
|---|---|---|---|
| Defined by CMS | Y | 13,946,037 | 80.62% |
| | W | 4,644 | 0.03% |
| | N | 967,712 | 5.59% |
| | U | 221,491 | 1.28% |
| | 1 | 1,518,508 | 8.78% |
| Undefined | 0,2,3,4,5,6,7,8,9,D,E,I,P,S,V,X,' | 640,651 | 3.70% |

It should be noted here that POA=no does not necessarily imply wrongdoing by the provider. Many of these conditions indicated as POA=no are complications of other conditions. Unknown drug allergy, for example, may cause additional complications and

therefore will be recorded as POA=no because the allergy happened within the hospital. The 'U' flag represents a condition for which there isn't enough clinical documentation to determine if the patient had the condition at admission. As noted in Section 2.3.3.2, for the purposes of categorizing "hospital acquired" status, if the medical record is insufficient to show POA=no, it is also insufficient to show conclusively POA=yes, and therefore researchers should consider these cases equivalent to POA=no.

The undefined POA flags represent approximately 3.8% of all the valid flags found within the research database. After verifying these characters were indeed in the raw LDS files, I contacted ResDAC concerning the unexpected POA flag values. Their response stated provider codes "are not cleaned and can have errors. This information would indicate … very high background error rate."[52] In addition to extra characters, there may be missing POA flags, as noted in a CMS working document "intended to communicate format and changes … found within the 2008 LDS Standard Analytical Files (SAFs)."[53]

> The POA fields are included in the 2008 inpatient LDS SAF. However, there may be some null POA values due to an error which occurred in the CMS Common Working File. Claims processed during July through October 2008 did not include POA, therefore, any adjustments processed during this time for claims with 2008 service dates were affected. This primarily impacts claims with service dates of October through December 2008.

So, I encoded any POA not approved by CMS as "null" for statistical processing.

### 4.2.2 Raw POA by Diagnostic Column Position

Charting POA indicator versus which administrative diagnostic column the POA indicator is associated with (Figure 10), shows a number of interesting things. Note column 1 (principal diagnosis) is overwhelmingly POA=yes; most POA=no secondary diagnosis are found in column 2 and taper off; POA=no represent a relatively small proportion of claims; the number of columns available for reporting secondary diagnoses codes expanded from 10-25 in 2011 as noted by the drop-off to zero.

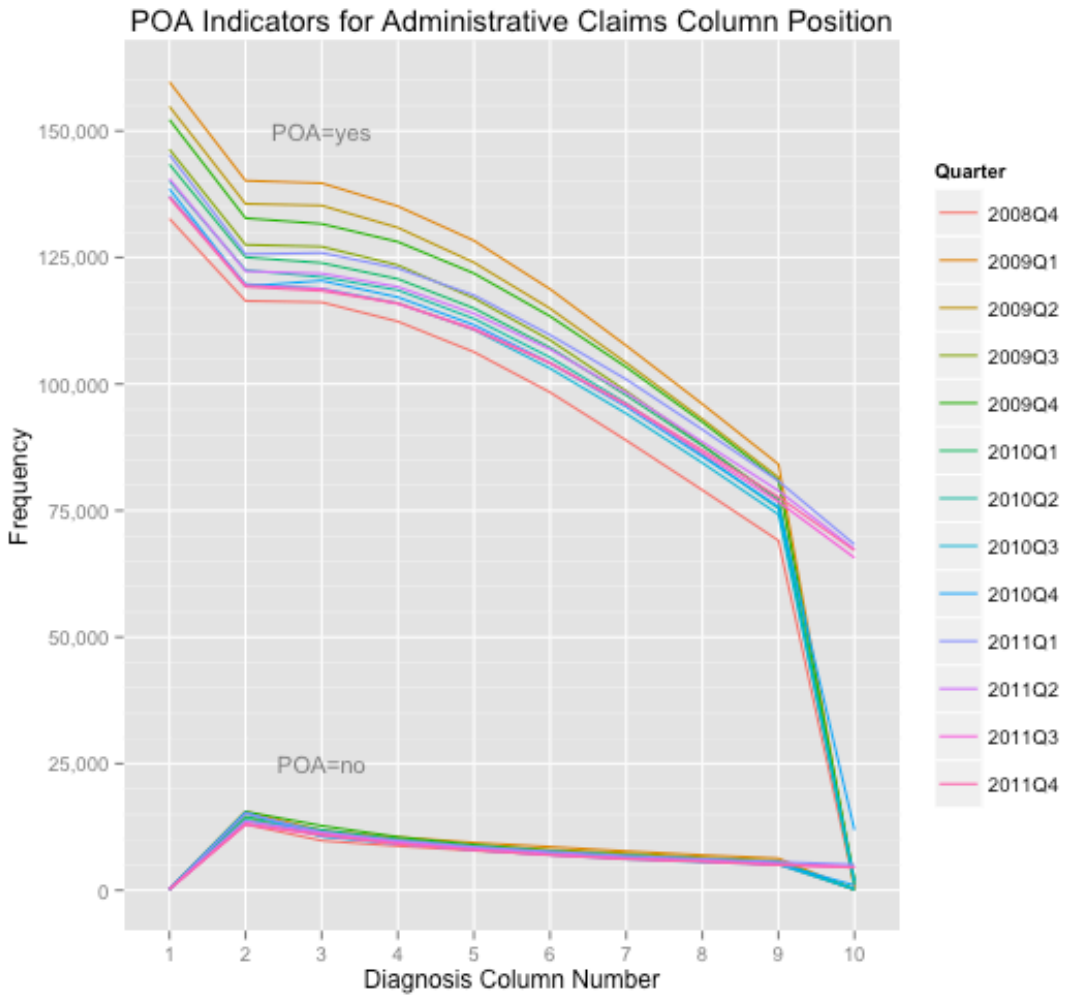



*Figure 10. POA indicators per administrative claims diagnosis column.*

Frequencies are remarkably consistent year-to-year as seen by raw counts, not adjusted values, from a 5% LDS Inpatient file. Not all fields are used to represent a patient's diagnosis as noted by the gradual trailing off of "Y" values with a curious step-down from column 1 to column 2 caused by the principal diagnosis code being not hospital acquired. The odd right-hand tail in column 10 is caused by some providers who were able to submit ten diagnostic values compared to the majority who submitted only nine secondary diagnosis columns. The most frequent place to find an "N" is in column 2. There was speculation raised in the HAC/POA committee minutes after the HAC/POA program inception and before the 2008Q4 effective date, suggesting hospitals would try to game the system by "burying" their POA=no conditions farther down in the secondary diagnosis list in an effort to reduce the appearance of HACs at their institution. This doesn't appear to be the practice, as column 2 contains three times the number of "N" values as column 9. The number of secondary diagnosis fields expanded to 25 in 2011, which did not effect the trajectory of use. (As a reminder, this study used only columns 1 through 10 to be compatible with previous years not having as many fields.) This pattern persisted through providers, states, and CMS regions. POA=no flags are found mostly in the second position decreasing thereafter and almost zero for first diagnosis position.

### 4.2.3 POA Rates for Investigated Conditions

How do POA=no rates compare to frequencies of the ICD-9 codes under which they fall? Appendix D – POA Frequency Data contains a full listing of ICD-9 primary diagnosis codes and the associated POA frequency within that diagnosis. If POA data represents in-hospital complications for patients, one would expect these distributions to be comparable across the different ICD-9 codes. In addition, it would also appear to be reasonable to assume POA rates are comparable across different facets of data. To test this, I created bullet charts to visually compare this idea. The bullet charts used beneficiary gender and age as facets. Each bullet has a larger bar representing the frequency of the corresponding ICD-9 primary diagnosis code and a smaller bar

representing the proportions of any POA=no on the administrative claim. The line on the small bar is its standard error. Each small chart is a facet of the contrasting variable (Figure 11).

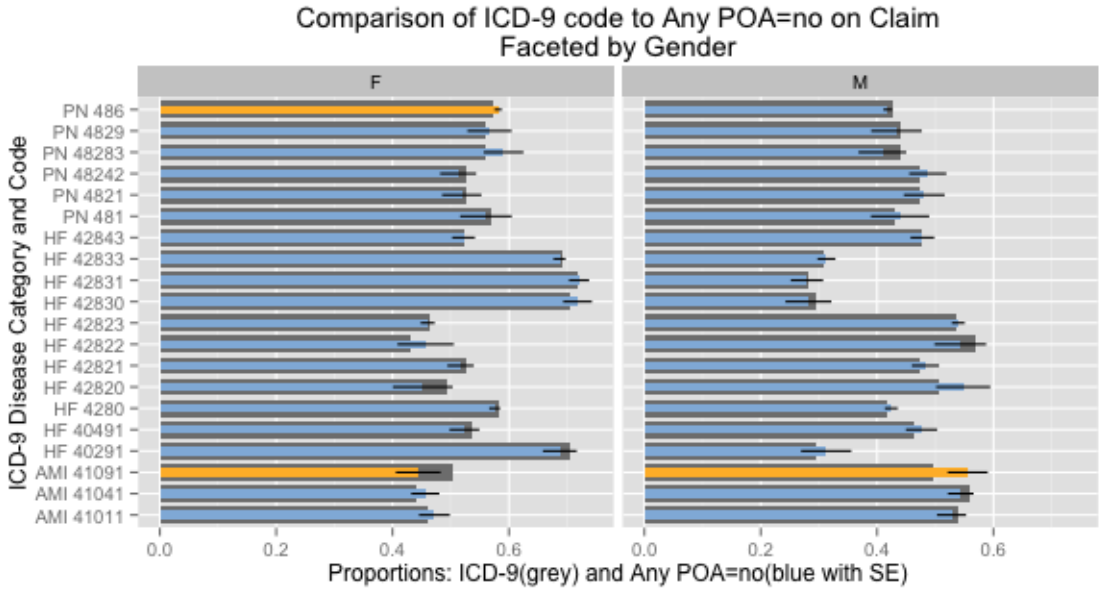



***Figure 11. Proportion bullet chart for gender.***

When faceting by gender, the general assumption appears true: having any POA=no on an administrative claim is proportional to the frequency of the primary diagnosis code. Using HF 42830 as an example, female patients represent about 65% of the total patients, and of those female patients, their "any POA=no on administrative claim?" rate is about 66% with a standard error of about ±1.5%, meaning the two rates are roughly equivalent. There are three exceptions highlighted in Figure 11 for conditions: PN 486 (female) and AMI 41091 (male and female). In the first case, the numbers are significant, but only just. For the second condition, the rates really appear to be different.

I repeated this process, faceting the data by beneficiary age (Figure 12). Across all of the age facets, having any POA=no on an administrative claim is proportional to the frequency of the primary diagnosis code with the exception of those conditions highlighted: PN48242 (80 to 84), HF 42833 (>84), HF 42830 (80 to 84), and HF 42843 (>84). In these cases, the rates are at the point of significance, not really exceeding it.

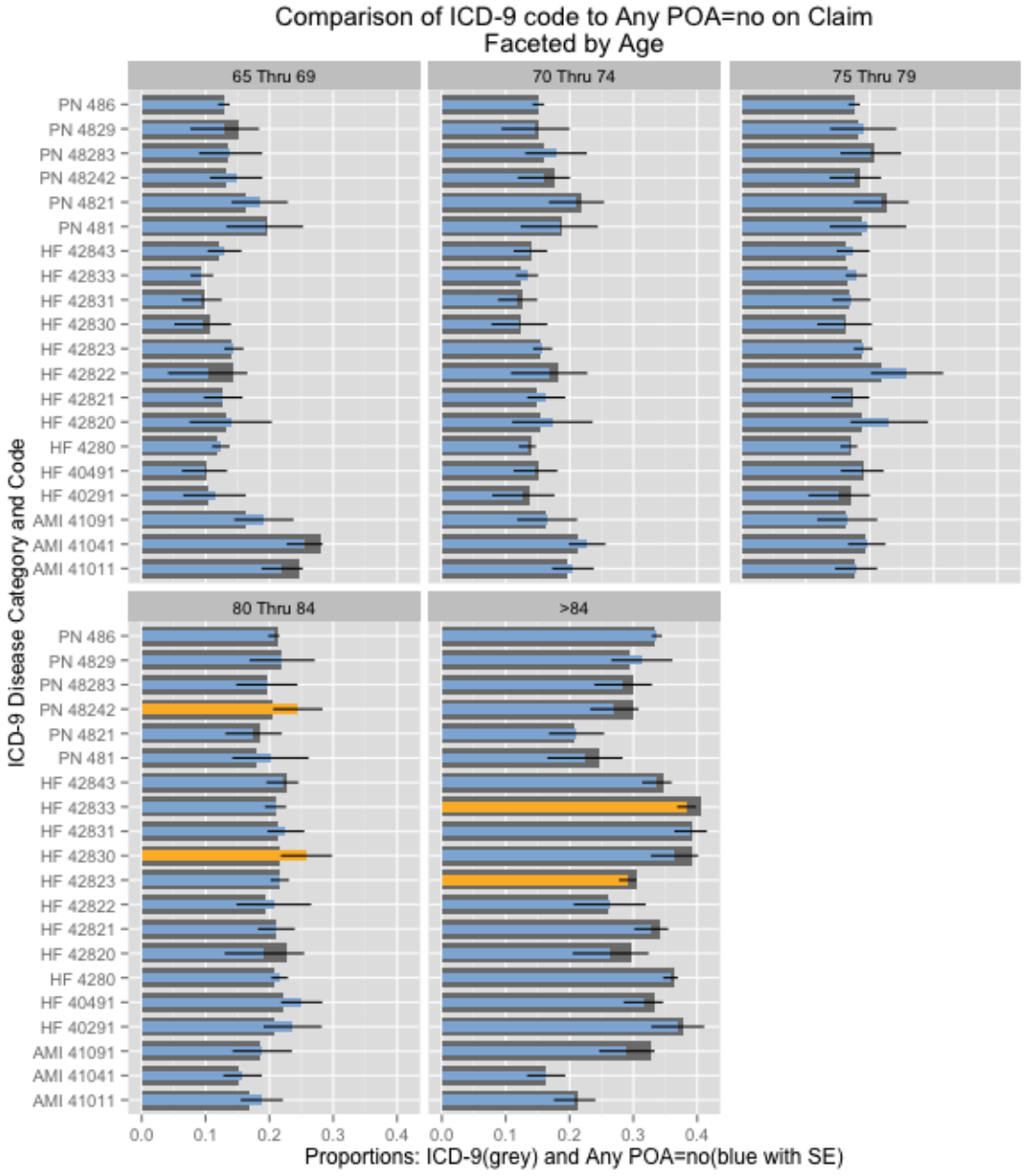



***Figure 12. Proportion bullet chart for age.***

I investigated the highlighted codes further. The distributions do not show any other relationship between ICD-9 code, age or gender, and POA. Whatever statistical differences exist may not be clinical in nature.

### 4.2.4 CMS Regions

Hospital CMS regions were not statistically significant in any of the HF nor PN data sets (Figure 13). Within AMI, the Seattle region was statistically significant compared to its peers.

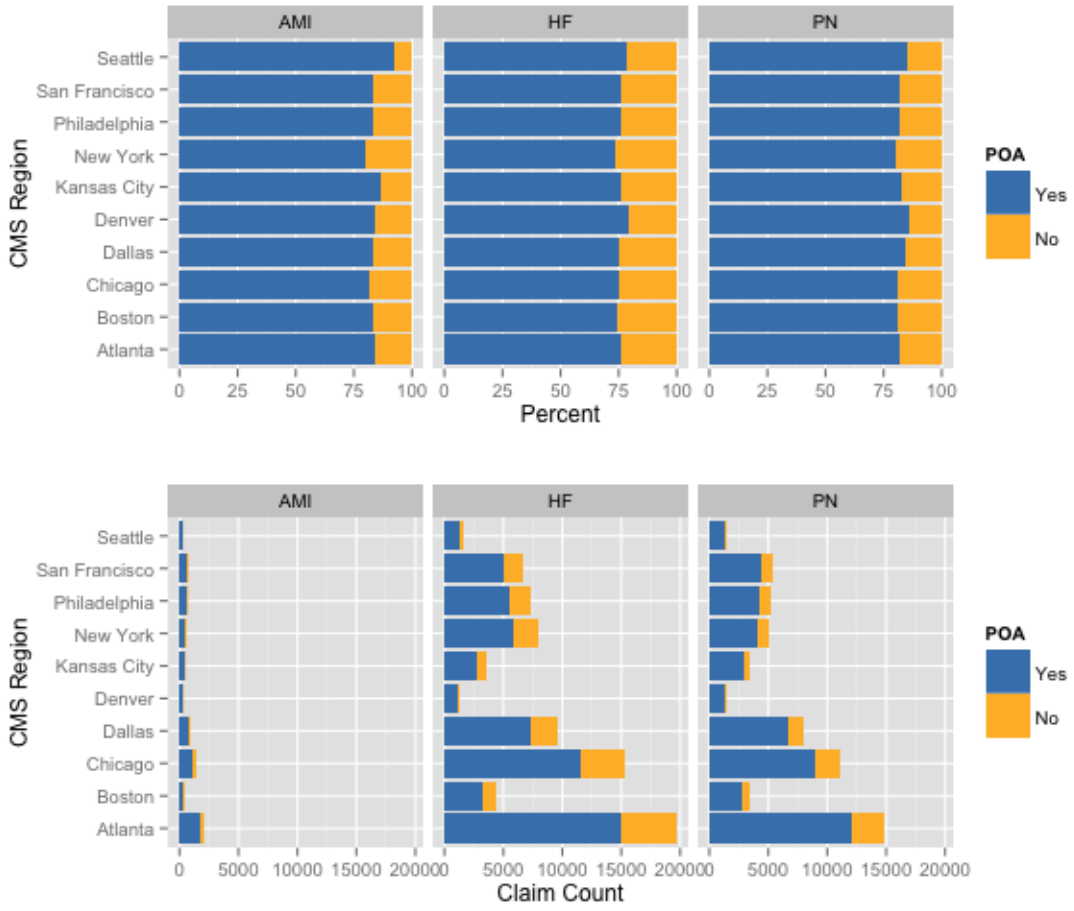



*Figure 13. POA percent and claim counts of 30-day readmissions by CMS regional centers.*

This chart shows the Seattle area's smaller POA=no percent for 30-day readmissions. Examining actual claim counts, however, reveals a probable data anomaly caused by insufficient data. There is no information in the literature suggesting AMI patients have fewer in-hospital secondary conditions compared to any other region. Therefore, I removed CMS regions as a predictor.

### 4.2.5 Q-Q Plots for Hospital-level POA Rates

Hospital-level all-condition POA rates are shown below (Figure 14). The blue line represents the Quantile-Quantile line. Hospitals with no claims in a given year have been removed.

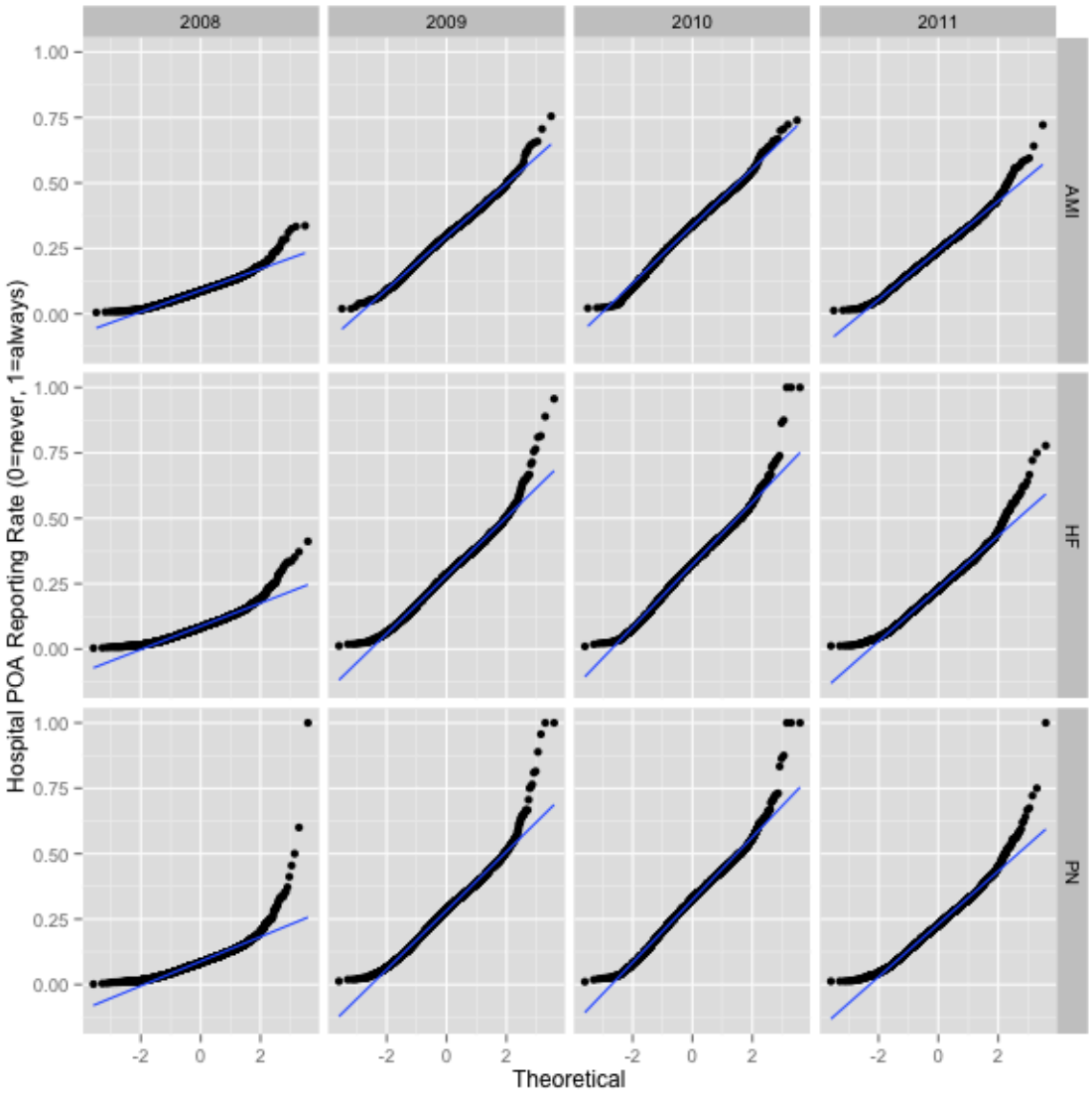



*Figure 14. Q-Q plots of hospital-level POA reporting rates.*

The years 2008 and 2009 were early years for the HAC/POA program. This ramp-up period can bee seen in 2008. In 2009, the situation became more stable. In 2010 and 2011 only slight changes can be seen between HF and PN. POA reporting by providers

was spotty in both coverage and content with some providers not supplying indicators, as seen by the long left-hand tails in 2008. Other hospitals supplied POA data all the time, as a matter of policy, as seen by the rising right-hand tails. Examining claims from the early years by these providers yields POA values of '1' (exempt from reporting), in all fields, but these represent a small percent of total claims. The '1' flag as a POA indicator was phased out mid-2010 and is no longer a source for confounding. Still other hospitals only supply POA data when a HAC condition was present on the administrative claim resulting in an overall low rate for those providers. CMS clarified its POA reporting policy for FY2010 and overall rates climbed. Each chart is slightly different for each condition. In all charts, a vast majority of points coincide with the Q-Q line.

## *4.3 POA Regression Results*

This section presents the result of logistic regression modeling of hospital readmission.

### 4.3.1 Evaluation of POA Predictors Overall

As defined in Section 3.4.2, ***p1*** "binary variable (p_no_flag) indicating claim has a POA=no condition." This predictor was statistically significant within all of the AMI datasets and the 15-day HF dataset, while not statistically significant for the 7-day and 30-day HF and all PN datasets. In each condition and in each readmission period investigated, the odds ratios are greater than 1.0, indicating an increase in readmission, (Figure 15). Green lines are statistically significant and red lines are not.

AMI patients can be treated in a number of ways, some with procedures (81%) and some without (19%). As a follow-up exercise, Chi-Square tests were generated comparing AMI variables for procedure status, POA status, and 30-day readmission status. As seen in Table 10 below, the variable procedure status and readmission status are independent. However, the variables procedure status and POA status are not independent. Likewise, the variables POA status and 30-day readmission status are also not independent. There is some systematic connection between POA status and

procedure status indicating a more complicated patient stay leading to a greater likelihood of hospital readmission.

*Table 10. Chi-Square test of independence between POA, ICD-9-CM procedures, and readmission.*

| *AMI claims* | *Pearson's Chi-squared test, df=1* | *P-value* |
|---|---|---|
| Procedure status<br>30-day readmission | 0.1299469 | 0.7184871 |
| Procedure status<br>POA events | 295.1906 | < 0.000001 |
| POA events<br>30-day readmission | 40.3326 | < 0.000001 |
| Procedure status and POA events<br>30-day readmission | 41.58003 | < 0.000001 |

From this table, I make the following observations: First, AMIs requiring surgical intervention carries a greater chance of complications, which result in both increased POA=no reporting and increased readmissions. Second, HF and PN are medical conditions usually accompanied by comorbidities. The patient is treated for the most serious condition first and then once stabilized, will be treated for other conditions. The timeframe between admission and subsequent treatment could account for the muddling of POA indicators because the subsequent treatment did not happen on the first day of the claim.

The AMI odds ratios have a wider confidence interval due to the smaller sample sizes. 30-day readmission appears to have a greater correlation to readmission than 7-day or 15-day.

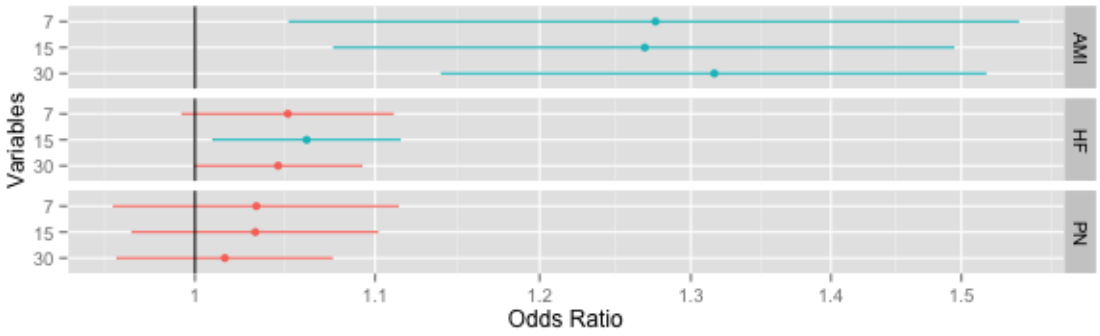



*Figure 15. Odds ratio of POA flag for 7-day, 15-day, and 30-day readmission periods.*

In Section 3.4.2, ***p2*** is defined as "numeric variable (p_no_count) indicating the number from (0 through 10) of POA=no conditions on a claim record." This predictor was approaching statistical significant for each dataset. However, with odds ratios so close to 1.0, it is unlikely these results are anything but noise not reflecting clinical patterns. This variable needs more data to be an effective predictor. It may also be the case that the regression conflates the POA count with utilization days to produce patterns where none exist. For these reasons, I deemed this variable an unreliable indicator and removed it from the final consideration. I conducted no further analysis on ***p2***.

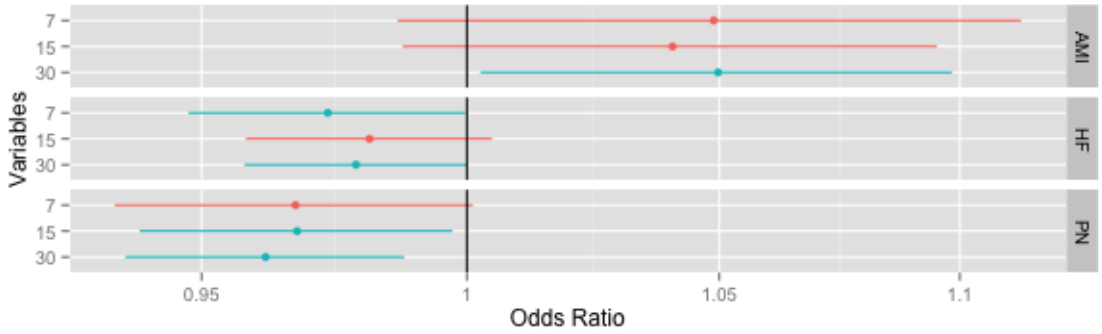



*Figure 16. Odds ratio of POA=no count for 7-day, 15-day, and 30-day readmission periods.*

In Section 3.4.2, ***p3*** is defined as "hospital POA adoption or usage rate." All of the AMI readmission periods across all 4 years were statistically insignificant (Figure 17). Green lines are statistically significant and red lines are not. PN 30-day readmission for 2009 and 2010 were also statistically insignificant, but the remainder were all significant. For AMI, the odds ratios themselves varied above and below 1.0 based on readmission period (all the 7-day odds ratios were > 1.0, while all of the 30-day odds ratios were < 1.0).

For the statistically significant odds ratios, all of the 2008 confidence intervals are wider than other years due to lack of data as the HAC/POA program got under way. The 2008 odd ratios are all askew from their peers as well.

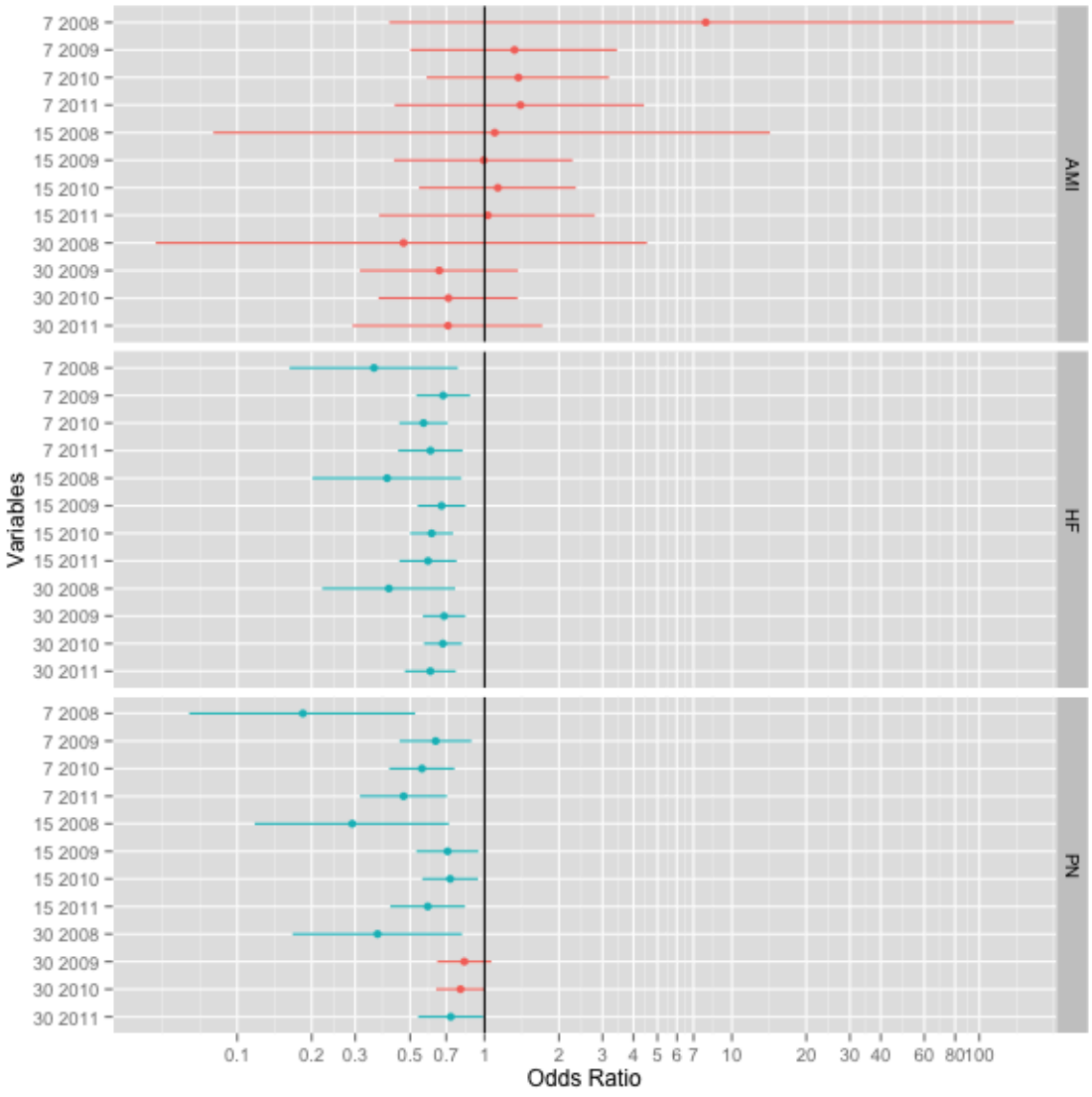



*Figure 17. Odds ratios for hospital POA rate for 7-day, 15-day, and 30-day readmission periods.*

### 4.3.2 ROC Analysis

To assist in model comparison, I created area under receiver operating characteristic ROC curves across each condition for the base model and base model plus POA data, and compared to 30-day readmission measures identified by Yale-CORE readmission measures created under contract for CMS.[49,50, 2] See Table 11. This study's rates are

lower than the published rates, but are above the age and gender models referenced. The addition of POA data improves upon each model slightly.

*Table 11. Comparison of area under ROC curves.*

| *Condition* | *Base Model* | *Base Model plus POA* | *CMS Methodology* |
|---|---|---|---|
| AMI | 0.6149 | 0.6157 | 0.63 |
| HF | 0.5558 | 0.5578 | 0.60 |
| PN | 0.5894 | 0.5896 | 0.63 |

### 4.3.3 Evaluation for AMI

Hypothesis $H_1$ indicates a connection between AMI administrative claims having POA=no and patient readmission. Comparing selected odds ratios for 30-day readmission shows the effect of the POA=no flag is greater than the effect for either gender or length of stay, but less than age 80 to 84 or age > 84 (Figure 18).

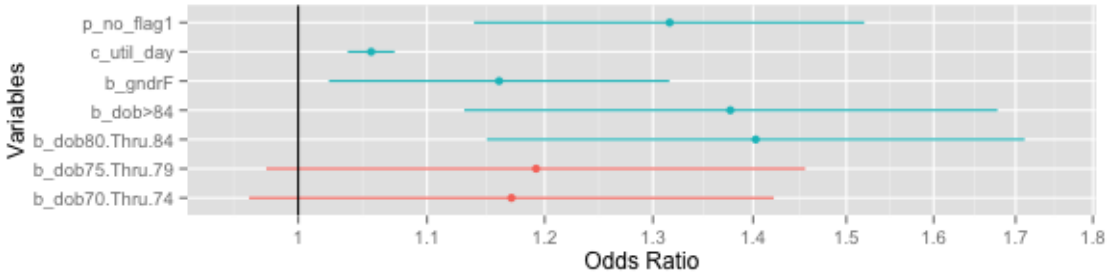



*Figure 18. Odds ratios for AMI 30-day readmission for base model with POA.*

I ran a backward Generalized Linear Models (GLM) variable elimination process on the AMI 30-day readmission model. These results are presented in Figure 19 below. The overall odds ratios became smaller, but the relative significance stayed the same.

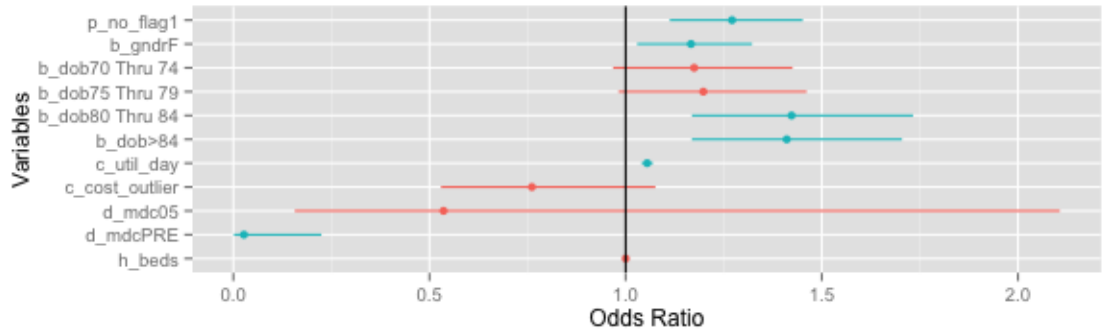



*Figure 19. Odds ratios for AMI 30-day readmission after GLM backward elimination procedure.*

To further understand the effect, I plotted AMI POA=no trends per quarter from 2009Q1 through 2011Q4. Several AMI ICD codes did not have enough data to make reasonable trends across the time period. So I discarded them from further analysis. Charting was limited to the 3 most prevalent codes: 41011, 41041, and 41091. See Figure 20 below.

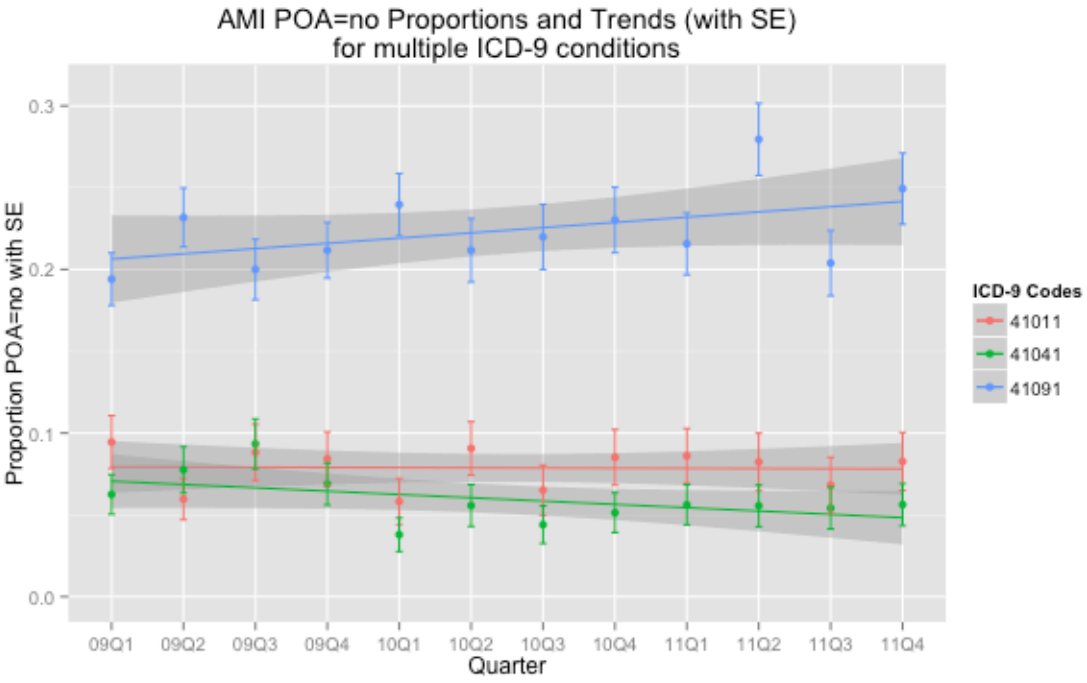



*Figure 20. POA=no proportions of AMI diagnosis per quarter.*

Codes 41011 and 41041 have stable or slightly declining rates respectively. 41091 has an increasing trend over time. This data does not appear to have any time-dependent trend. I generated a second set of trends containing POA=no, POA=yes, and total claims for ICD code 41091 (Figure 21). The overall incident rate falls over time (blue line), yet the POA=no rate remains steady over time (red line) resulting in an overall rising trend seen in the previous chart. I generated similar charts for 41011 and 41041, and they appear similarly: overall rates decrease slightly and the POA=no rate is constant quarter over quarter, resulting in a steady trend (41011) and a slightly decreasing trend (41041) over time. The difference between 41091 and the others is the magnitude of POA=no: it is 1.5 through 2.5 times the others with a wider variance. These rates appear to be in line with the American Journal of Cardiology report "Recent Trends in Hospitalization for Acute Myocardial Infarction" [54], in which the authors cite decreased mortality for AMI with increased complications. It is, perhaps, these complications detected through the POA=no data.

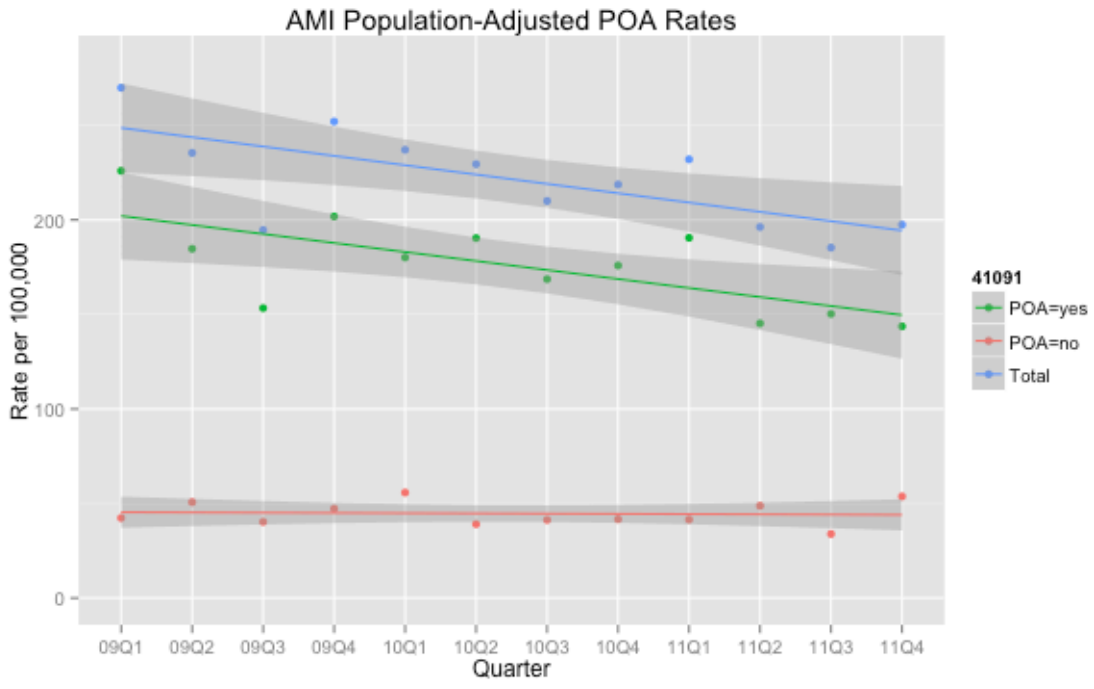



*Figure 21. Population-adjusted POA values by quarter for principal diagnostic code 41091.*

### 4.3.4 Evaluation for HF

I plotted HF POA=no trends per quarter from 2009Q1 through 2011Q4 (Figure 22). Several HF ICD codes did not have enough data to make reasonable trends and were discarded from further analysis. I grouped other HF ICD codes using the first 4 ICD-9 digits as follows:

```
4282x = {42820 42821 42822 42823}
4283x = {42830 42831 42833}
```

HF POA=no rates are with few exceptions smaller than AMI rates.

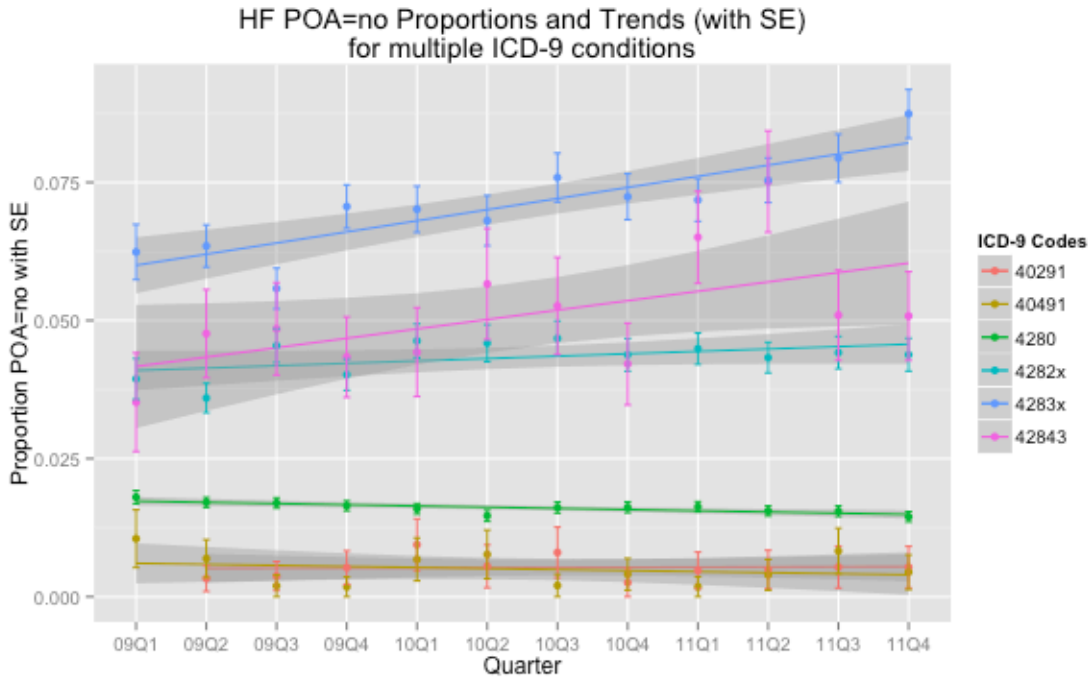



*Figure 22. POA=no proportions of HF diagnosis per quarter.*

I created and compared independent plots for each individual ICD-9 code group. All of these charts indicate a small POA=no proportion. I chose ICD-9 40491 as representative and charted to visualize the POA=yes/POA=no proportions (Figure 23). Notice how low the POA=no values per quarter are compared to the total. With such a consistently small set of data with which to work, it is surprising to see the "any POA=no

on administrative claim" predictor variable on the edge of statistical significance, in this case indicating increased readmission odds.

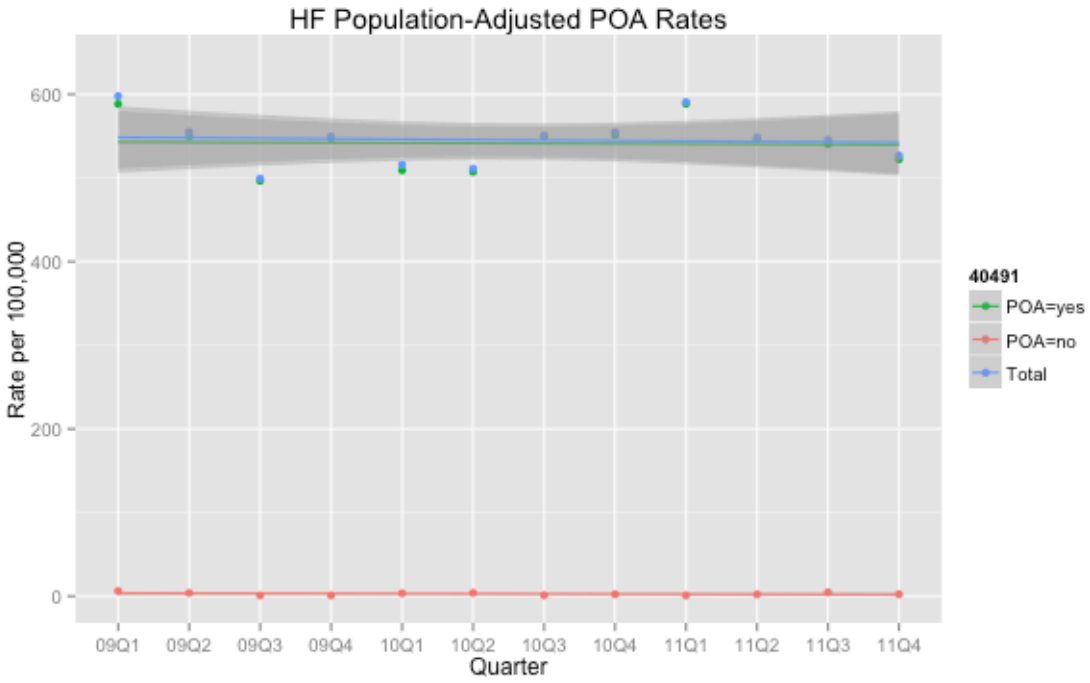



*Figure 23. Population-adjusted POA values by quarter for principal diagnostic code 40491.*

Another measurement from which to assess POA data is CMS's excess readmission rations as calculated for the 2013 hospital readmission reduction program. Comparisons between HF excess readmissions (CMS data) and POA rate (this study) are both per-hospital measures (Figure 24). Excess readmissions ratios less than 1.0 indicate hospitals achieving better than expected readmission results. Excess hospital readmissions are calculated to have similar distributions regardless of hospital size. Note hospitals with rates lower than 1.0 have higher POA usage rates. As hospital size gets larger, POA rate increases. This can be seen in the distribution ribbon at the x-axis of each chart. The POA utilization rate distributions do vary with hospital size as seen by the distribution ribbon along the y-axis of each chart. Hospitals in the first quartile size have a lower POA utilization rate. Those in the forth quartile hospital size have high POA utilization rates. The slope direction across all quartiles is negative indicating some correlation between lower expected readmission rates and higher POA utilization, and between

higher readmission and lower POA utilization. Once can see a consistent change from quartile to quartile of an increasing POA rate, however the difference between groups is not significant.

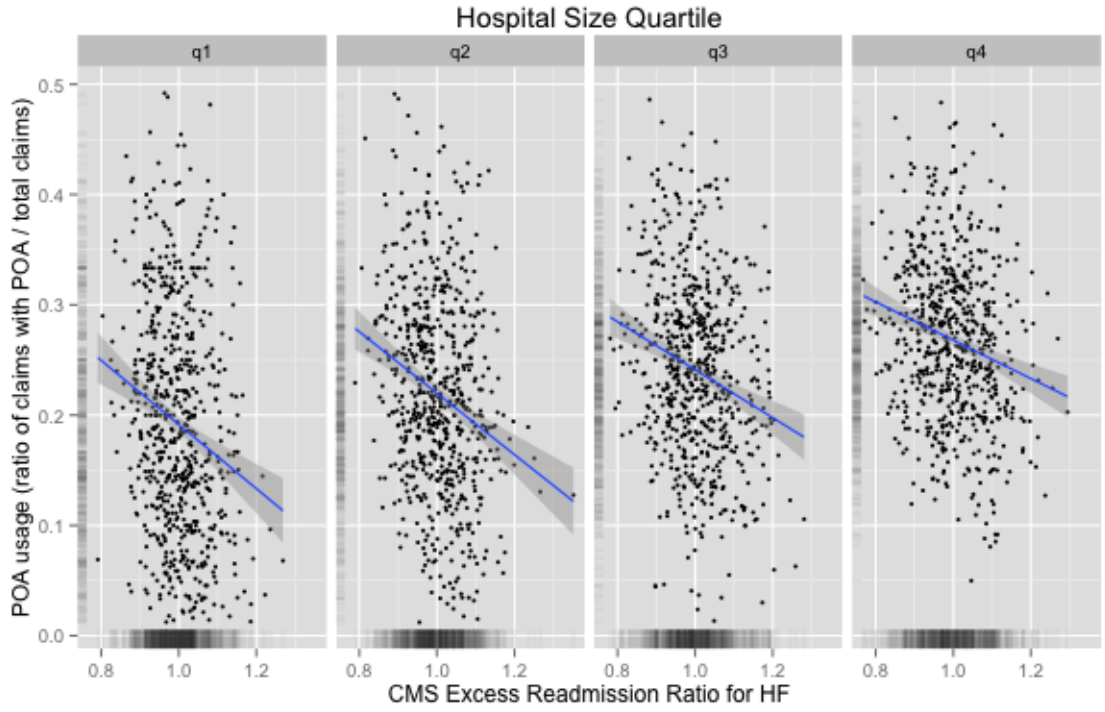



*Figure 24. Scatter plot of excess HF readmission compared to hospital's POA usage.*

### 4.3.5 Evaluation for PN

I plotted PN POA=no trends per quarter from 2009Q1 through 2011Q4 (Figure 25). Several PN ICD codes did not have enough data to make reasonable trends and were discarded from further analysis. PN POA=no rates lay between HF rates and AMI rates. I created and compared independent plots for each individual ICD-9 code group.

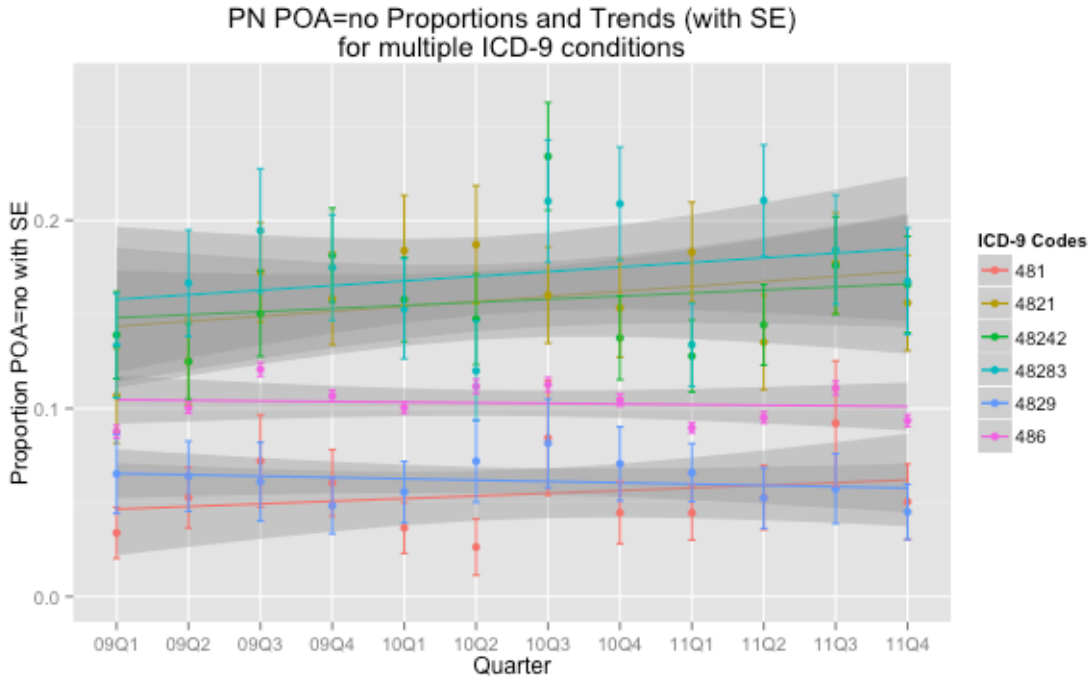



*Figure 25. POA=no proportions of PN diagnosis per quarter.*

Trends for the single largest principal diagnostic ICD code 486 (Figure 26) show a larger POA=no count overall compared to HF and not quite as much as for AMI. There appears to be some seasonality as seen by the total and POA=yes points: consistent peaks of Q1 and valleys of Q3. The POA=no points indicate a remarkably consistent quarter-to-quarter rate.

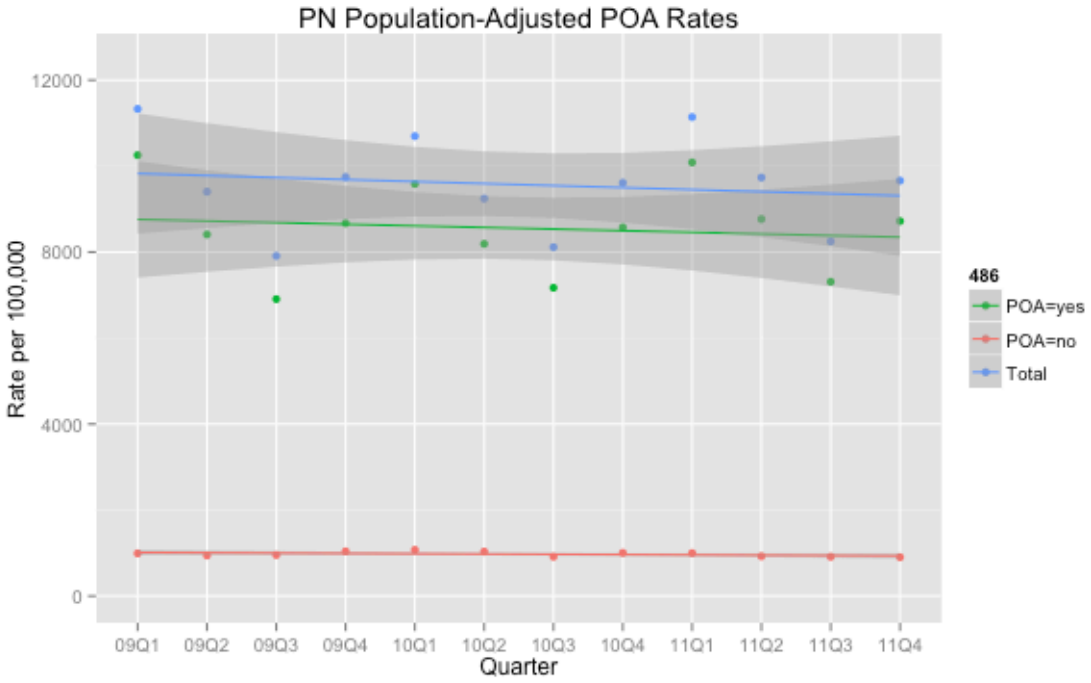



*Figure 26. Population-adjusted POA values by quarter for principal diagnostic code 486.*

## 4.4 Model Summary

There is a difference between predictors for AMI and for HF and PN. These appear to be related to either the patient (disease progression, comorbidities, etc.) or specific procedure as demonstrated by the non-overlapping predictors and significance results. It is possible that changing from a 5% sample to a 20% sample would increase the confidence surrounding the odds ratios without moving the estimation. It is also possible more data could reduce the effect to noise and insignificance.

# 5 Conclusion

*The purpose of this study is to understand and assess the usefulness of POA indicators as a predictor of hospital readmission.*

## *5.1 Findings*

"Would POA indicators be a useful set of data to include when modeling hospital readmissions?" To test this question, based on the literature search results of Section 2.4, three hypotheses were proposed in Section 3.4.2, and detailed results presented in Section 4.3. For convenience, the hypotheses and POA variables are restarted here.

$H_1$ *The presence of POA=no indicators on an administrative claim is a predictor of hospital readmission.*

***p1*** is defined as a dichotomous variable set to 1 if *any* POA=no indicators were found on the administrative claim.

$H_2$ *The number of POA=no indicators on an administrative claim is a predictor of hospital readmission.*

***p2*** is defined as the count of the number of POA=no indicators found on the given administrative claim.

$H_3$ *The rate at which a hospital fills POA fields across all of its administrative claims is a predictor of hospital readmission.*

***p3*** was set as an aggregated variable at the hospital level by counting *all* claims with *any* POA indicator set on *any* of its administrative claims. Due to the fluctuation in POA adoption, CMS directions or guidance, and overall adherence within the early years of the HAC/POA program, I divided this variable into rates per hospital per claim year.

An idea from the literature search that POA data can cluster around certain diseases, codes, or procedures led to the possibility of linking diseases groups together. This led to the development of hypotheses $H_1$ and $H_2$ paired with variables:

p1: Were there any POA=no conditions on the given administrative claim?

p2: How many POA=no conditions exist on the given administrative claim?

$H_1$ proved to be statistically significant for AMI at all readmission periods, and for HF at the 15-day readmission period. All had an odds ratio value associated with an increase in likelihood of hospital readmission. $H_2$ was not statistically significant for any disease nor any readmission period. I speculate that using a 5% sample did not leave enough claims in the resulting pool.

Another idea from the literature search, the idea that a provider's POA rates may vary from its peers, led to creation of a hospital-based POA metric that could be tracked over time. Hypothesis $H_3$ led to development of variable p3 defined as any POA=no occurring on any administrative claim for a given hospital. Rates were statistically significant for HF and PN, reflecting a reduced likelihood of a readmission event. I speculate, based upon the literature, that the reduced likelihood reflects documentation practices of the hospital (staff training, technology, ensuring the correct staff document POA assessments, etc.).

As a conclusion, I restate an observation made in Section 2.4: There is a nuance to POA data. Defining POA metrics and applying them over a wide array of conditions may not produce consistent results.

## 5.2 Future Directions

A logical follow-on to this study is to predict hospital readmissions using POA data. A new set of predictive algorithms have been developed recently, adapted and evolved from a number of sources. These are generally non-parametric techniques built on the computational power of repeated calculations called ensembles. For example, one older technique builds a decision tree by repeatedly partitioning the input dataset into smaller portions until a clear yes/no prediction can be made for any input to the tree. This

technique learns datasets very well, and can over-fit the data resulting in poor performance when applied to new data. The power of the newer predictive algorithms lies in their ability to create thousands of these decision trees (a forest of decision trees) in a relatively short period of time. With computational power comes yet another approach: combine a small random change into each decision tree; a change such as omitting a predictor variable, (a random forest of decision trees). The random aspect works by choosing to not use one of the predictor variables during the construction of the internal decision points. The final prediction is the sum of all predictions over all of the decision trees in the forest (the ensemble). These random permutations of the datasets into ensembles outperform traditional statistical techniques.

Shams, Ajorlou, and Yang built a predictive model using several of the newer techniques to help the VA understand which patients were likely to be readmitted before 30 days.[55] The following diagram (Figure 27) summarizes their findings. PHSF is a variant of random forest. There are newer neural network algorithms that use the random permutation and ensemble ideas that are on-par with random forests.

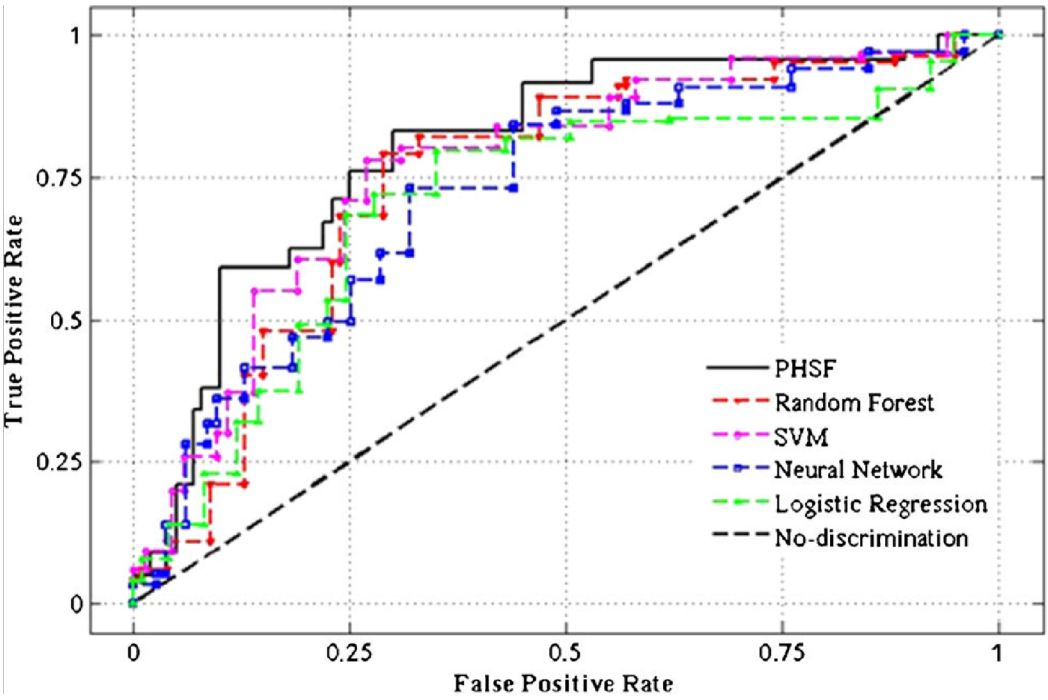



*Figure 27. Comparison of newer predictive techniques compared to logistic regression.*

Most of these new ensemble techniques have no limit to the width of the input dataset, meaning POA data could be added without penalty and with benefit. Adding raw POA data may have some benefit. More than likely, development of new POA variables would have a much better effect. I would also like to pursue the study of using graph theory to connect POA data to primary diagnosis, secondary diagnoses, procedures, and to DRGs or their attributes. Rather than hunt and peck for possible relationships, find them automatically via machine learning.

All of the information presented is inpatient data. None of the patient's other health care is represented here. Another avenue to follow up is combining POA data with outpatient records, Part D drug benefit information, etc., building up an entire patient-centric view of disease and its management for the purposes of tracking care and outcomes.

## *5.3 Summary*

Within this study, I have assembled a data mining research database the purpose of which was to learn about POA data in general and POA data in conjunction with hospital readmission modeling in particular. I believe the approach used within this study captures the essence of Biomedical Health Informatics because it demonstrates precisely a technical sophistication defined by pulling health data from disparate sources and integrating them into a working whole, in combination with statistical approaches to needed pose hypotheses and evaluate them using statistical models. In addition, this effort required an acquired in-depth understanding of U.S. health care administrative claims data. The study shows an ability to draw inferences and test them in an area where little work has been done to date using contemporary statistical methods.

# References

Note: All URLs checked as of publication date.


1 "Health Policy Brief: Medicare Hospital Readmissions Reduction Program," *Health Affairs*, November 12, 2013.

2 Harlan Krumholz, M.D., S.M. et al. "Hospital 30-Day Pneumonia Readmission Measure Methodology" Submitted By Yale University/Yale-New Haven Hospital Center for Outcomes Research & Evaluation (Yale-CORE) to Centers for Medicare & Medicaid Services (CMS), 09-Jun-2008.

3 Benbassat, Jochanan, MD & Mark Taragin, MD, MPH. "Hospital Readmission as a Measure of Quality of Care – Advantages and Limitations." *Archives of Internal Medicine*, 2002, 160: 1074-1081.

4 "Medicare Prescription Drug, Improvement, And Modernization Act of 2003", 108th Congress, Public Law 108 – 173. STAT. 2066, 08-Dec-2003, Page 117.

5 "Monitoring and Improving Inpatient Hospital Care," under History tab. http://www.iqrsupport.org/iqr-program;jsessionid=9abf966b6d01af19adf6d784e2c2 Talligen.com is a CMS contractor supporting IRQ program data collection and processing.

6 "Deficit Reduction Act of 2005", 109th Congress, Public Law 109-171, 120 STAT. 28, 08-Feb-2006. Title V, Subtitle A, Section 5001, Paragraph (a). This law authorized CMS to expand the IQR program and make data available to the public. http://www.gpo.gov/fdsys/pkg/STATUTE-120/pdf/STATUTE-120-Pg4.pdf

7 Jencks SF, Williams MV, Coleman EA. "Rehospitalizations among patients in the Medicare fee-for-service program." *New England Journal of Medicine*. 2009; 360(14):1418-1428.

8 "Payment Policy for Inpatient Readmissions," *MedPAC Jun-2007 Report*, Chapter 5. http://www.medpac.gov/chapters/jun07_ch05.pdf

9 "Patient Protection and Affordable Care Act of 2010," 111th Congress, Public Law 111-148, STAT 119, page 124, 23-Mar-2010. SEC. 3025 "Hospital Readmission Reduction Program." http://www.gpo.gov/fdsys/granule/PLAW-111publ148/PLAW-111publ148/content-detail.html

10 "Readmission Reduction Program", CMS website. http://www.cms.gov/Medicare/Medicare-Fee-for-Service-'Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html

11 "Working Paper: Simple Methods of Measuring Hospital Readmission Rates," America's Health Insurance Plans, Center for Policy and Research, Feb-2012.

12 Richardson, William C. (Committee Chair), et al. *To Err Is Human: Building a Safer Health System*. Institute of Medicine Committee on the Quality of Health Care in America, Nov-1999. Chapter 4 recommends establishing a federal Center for Patient Safety. Chapter 7 identifies safety stakeholders and ways to create appropriate incentives for health care systems.

13 "Establishment of the QuIC Task Force" Presidential Directive, William J. Clinton, 13-Mar-1998, http://archive.ahrq.gov/quic/about/clintonestablish.htm

14 "Quality Interagency Coordination Task Force QuIC Fact Sheet," Agency for Healthcare Research and Quality web site archives. http://www.ahrq.gov/research/findings/factsheets/quality/quic/

15 “Funding,” National Quality Forum. http://www.qualityforum.org/About_NQF/Funding.aspx

16 “Serious Reportable Adverse Events in Health Care”, Advances in Patient Safety, Vol 4, pp 339-352. http://www.ahrq.gov/professionals/quality-patient-safety/patient-safety-resources/resources/advances-in-patient-safety/vol4/Kizer2.pdf

17 Rosenthal, Jill. “Advancing Patient Safety through State Reporting Systems”, Agency for Healthcare Research and Quality, Jun-2007.

18 “Adverse health Events in Minnesota Hospitals – First Annual Public Report”, Jan-2005, Minnesota Department of Health. http://www.health.state.mn.us/patientsafety/ae/aereport0105.pdf

19 "Deficit Reduction Act of 2005", 109th Congress, Public Law 109-171, 120 STAT. 28, 08-Feb-2006. Title V, Subtitle A, Section 5001, Paragraph (b)(2)(D)(c) “Quality Adjustment in DRG Payments for Certain Hospital Acquired Infections” The Federal Register, V73 N161, 19-Aug-2008, at or about page 48444. Law as passed by Congress end of 2005, signed 2006; gives CMS authority to alter DRG payments to limit reimbursements for conditions that could reasonably have been prevented; allows revising condition definitions. CMS has extracted the relevant text and made it available: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Downloads/DeficitReductionAct2005.pdf

20 “Not Paying for Medical Errors”, *New York Times*. 21-Aug-2007, Editorial in Opinion section. http://www.nytimes.com/2007/08/21/opinion/21tue1.html The AHRQ Patient Safety Network web site links to this article . http://www.psnet.ahrq.gov/resource.aspx?resourceID=5818. This is an opinion article and not a peer reviewed journal article and is intended to show the sentiment towards ill-care. It is also interesting to note AHRQ cites it for the same reason.

21 Heidi L. Wald, MD, MSPH; Andrew M. Kramer, MD. “Nonpayment for Harms Resulting From Medical Care: Catheter Associated Urinary Tract Infections”, *Journal of the American Medical Association*. 2007;298(23):2782-2784. DOI:10.1001/jama.298.23.2782.

22 “Hospital-Acquired Conditions (HAC) in Acute Inpatient Prospective Payment System (IPPS) Hospitals Fact Sheet”, CMS http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Downloads/HACFactsheet.pdf

23 73 Federal Register, 19-Aug-2008, pages 48471-48491. See “General Background” for description of HACs, There are sever table containing cost data and ICD-9-CM codes definition for HACs. Number Index for Volume 73.

24 “CMS Medicare Severity Diagnosis Related Groups (MS-DRG) Grouper”, Software product from CMS, available from http://www.ntis.gov/products/grouper.aspx

25 MM7024 Revised, Medicare Learning Network Publication, 13-Aug-2010. http://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNMattersArticles/downloads/MM7024.pdf

26 75 Federal Register, No. 157, 16-Aug-2010, Page 50082. Rules and Regulations. “All Discharges” refers to short-term acute care hospitals participating in the inpatient prospective payment system (IPPS).

27 DRA Section 5001(c). Explained to providers in “Hospital Acquired Conditions and Present on Admission Indicator Reporting Listening Session”, 17-Dec-2007. http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/downloads/HAC-POA-Listening12-17-2007.pdf

28 http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/

29 http://www.rti.org/reports/cms/ HAC-POA annual monitoring reports and data prepared by RTI for CMS.

30 Heidi Wald, Angela Richard, Victoria V Dickson and Elizabeth Capezuti. "Chief nursing officers' perspectives on Medicare's hospital-acquired conditions non-payment policy: implications for policy design and implementation", Implementation Science 2012, 7:78 doi:10.1186/1748-5908-7-78. 28-Aug-2012. http://www.implementationscience.com/content/7/1/78

31 Rogers, Catherine. "Improving Processes to Capture Present-on-Admission Pressure Ulcers." *Advances in Skin & Wound Care.* Issue: Volume 26(12), December 2013, p 566–572. DOI: 10.1097/01. In the conclusion of this study, the author finds this method was not statistically significant in capturing POA information due in part to alternate means of detection; pressure ulcer rates were decreasing, but not because of this method of intervention.

32 Armstrong, David. "New opportunities to improve pressure ulcer prevention and treatment: implications of the CMS inpatient hospital care Present on Admission (POA) indicators/hospital-acquired conditions (HAC) policy. A consensus paper from the International Expert Wound Care Advisory Panel." *Journal of Wound, Ostomy, and Continence Nursing.* [1071-5754] 2008, vol35, issue 5, pg 485-492.

33 Weiss AJ, Elixhauser A, Bae J, Encinosa W. "Origin of Adverse Drug Events in U.S. Hospitals," 2011. HCUP Statistical Brief #158. Jul-2013. Agency for Healthcare Research and Quality, Rockville, MD. http://www.hcup-us.ahrq.gov/reports/statbriefs/sb158.pdf

34 Fokkema, Margriet, et al. "The impact of the present on admission indicator on the accuracy of administrative data for carotid endarterectomy and stenting." *Journal of Vascular Surgery*, 27-Aug-2013, ISSN 0741-5214, http://dx.doi.org/10.1016/j.jvs.2013.07.006. or http://www.sciencedirect.com/science/article/pii/S0741521413012949

35 Hughes, John S., Jon Eisenhandler, Norbert Goldfield, Patti G. Weinberg and Richard Averill. "Postadmission Sepsis as a Screen for Quality Problems: A Case−Control Study." *American Journal of Medical Quality*. 13-Nov-2013. DOI: 10.1177/1062860613509002 http://ajm.sagepub.com/content/early/2013/11/12/1062860613509002

36 Bankowitz, Richard A., Barbara Doyle, Michael Duan, Eugene Kroch and John Martin. "Identifying Hospital-Wide Harm: A Set of ICD-9-CM-Coded Conditions Associated With Increased Cost, Length of Stay, and Risk of Mortality". *American Journal of Medical Quality*. 30-Sep-2013 DOI: 10.1177/1062860613503896. http://ajm.sagepub.com/content/early/2013/09/27/1062860613503896

37 http://www.MySQL.org. MySql Database Engine, Version 5.5.8. MySql Workbench, Version 5.2.31.

38 http://www.r-project.org Statistical engine: R, Verison 3. Tools: R Studio, Version 0.97.449.

39 "Research Data Distribution Center LDS Inpatient SNF Claim Record Data Dictionary", CMS.gov web site. http://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/downloads/SAFldsSNFNov2009.pdf

40 http://www.resdac.org/about-resdac. Quote taken from the "About" tab. Data field definitions are located on ResDAC web site, but the site structure has changes several times through out this study. The most reliable way to find the definitions is via Google.

41 "ICD-9-CM Addenda, Conversion Table, and Guidelines", CMS. http://www.cdc.gov/nchs/icd/icd9cm_addenda_guidelines.htm http://www.cdc.gov/nchs/icd/icd9cm_maintenance.htm. These page provides guidance on a year-to-year changes to the base code set. ICD-9_CM committee minutes, recommendations, and deadlines are documented in "ICD-9-CM Coordination and Maintenance Committee."

42 Healy, Deborah, and Jerry Cromwell. "Hospital-Acquired Conditions–Present on Admission: Examination of Spillover Effects and Unintended Consequences", Final Report, Sep-2012. RTI International, Inc., under contract with CMS. Page 30.

43 "List of Diagnostic Related Groups (DRGS), FY2008", CMS, http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MedicareFeeforSvcPartsAB/downloads/DRGdesc08.pdf.

44 "CMS's SSA to FIPS State and County Crosswalk," The National bureau of Economic Research. http://www.nber.org/data/ssa-fips-state-county-crosswalk.html This file is available from multiple sources.

45 Social Security 2-digit ID to State Name Decode Table. Available from multiple sources: CMS, SSA, and ResDAC. http://www.resdac.org/sites/resdac.org/files/STATE_TB.txt

46 http://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/CostReports/Hospital-1996-form.html, "Hospital Form 2552-96". http://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/CostReports/REPORTS/HOSPITAL1996-REPORTS.ZIP

47 https://github.com GIT, Version 1.7.5.4.

48 http://www.Perl.org Perl, Version 5.12.3.

49 Harlan M. Krumholz, MD, SM, et al. "Hospital 30-Day Acute Myocardial Infarction Readmission Measure – Methodology," Submitted By Yale University/Yale-New Haven Hospital-Center for Outcomes Research & Evaluation (Yale-CORE) to Centers for Medicare & Medicaid Services (CMS), 09-Jun-2008.

50 Harlan Krumholz, M.D., S.M. et al. "Hospital 30-Day Heart Failure Readmission Measure Methodology" Submitted By Yale University/Yale-New Haven Hospital Center for Outcomes Research & Evaluation (YNHH-CORE) to Centers for Medicare & Medicaid Services (CMS), 23-Apr-2008.

51 "Medicare Hospital Quality Chartbook 2012: Performance Report on Outcome Measures", Prepared by Yale New Haven Health Services Corporation Center for Outcomes Research and Evaluation, Sep-2012. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/MedicareHospitalQualityChartbook2012.pdf

52 Unpublished. This information was provided by ResDAC via email in response to a request for clarification [#DRD-106-25696].

53 "Limited Data Set 2008 Standard Analytical Files November 2009", Paragraph titled, "Present On Admission (POA)." http://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/Downloads/LDSSAF2008Changes.pdf,

54 "Recent Trends in Hospitalization for Acute Myocardial Infarction," *American Journal of Cardiology*, Volume 109, Issue 11, Pages 1589-1593, 01-Jun-2012. http://www.ajconline.org/article/S0002-9149(12)00597-8/abstract

55 Issac Shams, Saeede Ajorlou, Kai Yang. "A predictive analytics approach to reducing 30-day avoidable readmissions among patients with heart failure, acute myocardial infarction, pneumonia, or COPD." *Health Care Management Science* [1386-9620], 03-May-2014.

# Appendix A – Glossary

Understanding anything in the U.S. health care system requires an understanding of the alphabet soup of agencies, laws, programs, committees, advisory panels, etc., driving health care change. This table is a short list created to clarify HAC/POA and 30 day readmission. The following table is a glossary of terms used in this thesis.

| ***Acronym / Organization*** | ***Definition*** |
|---|---|
| ACA / PPACA | *Patient Protection and Affordable Care Act*<br>Public Law 111-148, 2010. |
| AHRQ | *Agency for Healthcare Research and Quality*<br>1998, Part of HHS. http://www.AHRQ.Gov<br>Misson: "produce evidence to make health care safer, higher quality, more accessible, equitable, and affordable, and to work with the U.S. Department of Health and Human Services (HHS) and other partners to make sure that the evidence is understood and used." |
| CMS | *Center for Medicare and Medicaid Services*<br>Established: 1965 |
| HCAHPS | *Hospital Consumer Assessment of Healthcare Provider Systems*<br>Established: 2002. Purpose: CMS partners with AHRQ to develop national patient survey. 2005, NQF endorsed methods and approach, CMS implemented survey in 2006. |
| HCUP | *Healthcare Cost and Utilization Project*<br>1988, under AHRQ<br>Purpose: tracks treatments and outcomes |
| HHS | *Department of Health and Human Services*<br>Part of the U.S. Federal government which contains CMS |
| IPAB | *Independent Payment Advisory Board*<br>2010, Agency created under ACA. Charged with identifying and implementing savings in Medicare without effecting coverage or quality. |
| IQR | *Hospital Inpatient Quality Reporting System*<br>Section 501(b) Medicare Prescription Drug Improvement and Modernization Act of 2003 0.4% reduction of payments to hospitals that do not report quality measures. Increased to 2.0% in Deficit Reduction Act of 2005. A portion of quality data collected is posted on http://www.HospitalCompare.HHS.Gov website. |
| Joint Commission | An independent, not-for-profit organization that accredits and certifies health care organizations in the United States |
| MedPAC | *Medicare Payment Advisory Committee*<br>Established: 2007. Purpose: Advise Congress about Medicare payments, quality of care delivered, and access to care. |
| MMA | *Medicare Prescription Drug, Improvement and Modernization Act*<br>Authorized Medicare Part D coverage for prescription drugs, health savings accounts, and national reporting program |
| NIH | *National Institutes of Health*<br>Established: 1931. Purpose: Preventing and curing disease |
| NQF | *National Quality Forum*<br>Mission: "1) Building consensus on national priorities, 2) Endorsing national consensus standards, and 3) Promoting goals through education and outreach programs. |
| QuIC | *Quality Interagency Coordinating Committee*<br>Mission: To ensure that all Federal agencies involved in purchasing, providing, studying, or regulating health care services are working in a coordinated way toward the common goal of improving quality of care. |
| VA | *Veterans Affairs* |

| ***Acronym*** | ***Definition*** |
|---|---|
| ACA | *Patient Protection and Affordable Care Act of 2010* |
| ADE | *Adverse Drug Events* |
| AMI | *Acute Myocardial Infarction* |
| CABG | *Coronary Arterial Bypass Graft* |
| CAS | *Carotid Angioplasty and Stenting* |
| CAUTI | *Catheter-Associated Urinary Tract Infection* |
| CC | *Complications or Comorbidities* |
| CEA | *Carotid Endarterectomy* |
| CDS | *Clinical Decision Support* |
| DRA | *Deficit Reduction Act* |
| DRG | *Diagnostic Related Group* |
| EMR | *Electronic Medical Record* |
| FIPS | *Federal Information Processing Standard* |
| GLM | *Generalized Linear Models* |
| HAC | *Hospital Acquired Condition* |
| HF | *Heart Failure* |
| HIMSS | *Health Information Management Systems Society* |
| HIT/IT | *Health Information Technology / Information Technology* |
| ICD-9-CM | *International Classification of Disease, 9th Revision, with Clinical Modification* |
| IPPS | *Inpatient Prospective Payment System* |
| LDS | *Limited Data Set* |
| LOS | *length of stay* |
| MCC | *Major Complications or Comorbidities* |
| MILI | *University of Minnesota Carlson School, Medical Industry Leadership Institute* |
| MMA | *Medicare Prescription Drug, Improvement and Modernization Act* |
| MS-DRG | *Medicare Severity Diagnostic Related Group* |
| PIC | *Potential Inpatient Conditions* |
| POA | *Present on Admission* |
| PN | *Pneumonia* |
| ResDAC | *Research Data Assistance Center* |
| RSRR | *Risk-Standardized Readmission Rates* |
| SAF | *Standard Analytical Files* |
| SNF | *Skilled Nursing Facility* |

## Appendix B – Principal Diagnostic Codes for AMI, HF, and PN

CMS currently assesses hospital readmission penalties for acute myocardial infarction AMI, HF, PN. This study limits conditions to this set as defined by the administrative claim principal diagnosis field using the following table.

| *Condition* | *ICD-9-CM Diagnosis Code* | *Description* |
|---|---|---|
| AMI | 41000 | Acute myocardial infarction of anterolateral wall, episode of care unspecified |
| AMI | 41001 | Acute myocardial infarction of anterolateral wall, initial episode of care |
| AMI | 41010 | Acute myocardial infarction of other anterior wall, episode of care unspecified |
| AMI | 41011 | Acute myocardial infarction of other anterior wall, initial episode of care |
| AMI | 41020 | Acute myocardial infarction of inferolateral wall, episode of care unspecified |
| AMI | 41021 | Acute myocardial infarction of inferolateral wall, initial episode of care |
| AMI | 41030 | Acute myocardial infarction of inferoposterior wall, episode of care unspecified |
| AMI | 41031 | Acute myocardial infarction of inferoposterior wall, initial episode of care |
| AMI | 41040 | Acute myocardial infarction of other inferior wall, episode of care unspecified |
| AMI | 41041 | Acute myocardial infarction of other inferior wall, initial episode of care |
| AMI | 41050 | Acute myocardial infarction of other lateral wall, episode of care unspecified |
| AMI | 41051 | Acute myocardial infarction of other lateral wall, initial episode of care |
| AMI | 41080 | Acute myocardial infarction of other specified sites, episode of care unspecified |
| AMI | 41081 | Acute myocardial infarction of other specified sites, initial episode of care |
| AMI | 41090 | Acute myocardial infarction of unspecified site, episode of care unspecified |
| AMI | 41091 | Acute myocardial infarction of unspecified site, initial episode of care |
| HF | 40201 | Malignant hypertensive heart disease with heart failure |
| HF | 40211 | Benign hypertensive heart disease with heart failure. Unspecified hypertensive heart disease with heart failure |
| HF | 40291 | Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified |
| HF | 40401 | Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage V or end-stage renal disease |
| HF | 40403 | Hypertensive heart and chronic kidney disease, benign, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified |
| HF | 40411 | Hypertensive heart and chronic kidney disease, benign, with heart failure and chronic kidney disease stage V or end-stage renal disease |
| HF | 40413 | Hypertensive heart and chronic kidney disease, unspecified, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified |
| HF | 40491 | Hypertensive heart and chronic kidney disease, unspecified, with heart failure and chronic kidney disease stage V or end stage renal disease |
| HF | 40493 | Congestive heart failure, unspecified. Left heart failure |
| HF | 4280 | Systolic heart failure, unspecified. Acute systolic heart failure |
| HF | 4281 | Chronic systolic heart failure. Acute on chronic systolic heart failure |
| HF | 42820 | Diastolic heart failure, unspecified |
| HF | 42821 | Acute diastolic heart failure |
| HF | 42822 | Chronic diastolic heart failure |
| HF | 42823 | Acute or chronic diastolic heart failure |
| HF | 42830 | Combined systolic and diastolic heart failure, unspecified |
| HF | 42831 | Acute combined systolic and diastolic heart failure |
| HF | 42832 | Chronic combined systolic and diastolic heart failure |

| *Condition* | *ICD-9-CM Diagnosis Code* | *Description* |
|---|---|---|
| HF | 42833 | Acute on chronic combined systolic and diastolic heart failure |
| HF | 42840 | Heart failure, unspecified |
| HF | 42841 | Malignant hypertensive heart disease with heart failure |
| HF | 42842 | Benign hypertensive heart disease with heart failure. Unspecified hypertensive heart disease with heart failure |
| HF | 42843 | Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified |
| HF | 4289 | Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage V or end-stage renal disease |
| PN | 4800 | Pneumonia due to adenovirus |
| PN | 4801 | Pneumonia due to respiratory syncytial virus |
| PN | 4802 | Pneumonia due to parainfluenza virus |
| PN | 4803 | Pneumonia due to SARS-associated coronavirus |
| PN | 4808 | Pneumonia due to other virus not elsewhere classified |
| PN | 4809 | Viral pneumonia, unspecified |
| PN | 481 | Pneumococcal pneumonia [Streptococcus pneumoniae pneumonia] |
| PN | 4820 | Pneumonia due to Klebsiella pneumoniae |
| PN | 4821 | Pneumonia due to Pseudomonas |
| PN | 4822 | Pneumonia due to Hemophilus influenzae [H. influenzae] |
| PN | 48230 | Pneumonia due to Streptococcus, unspecified |
| PN | 48231 | Pneumonia due to Streptococcus, group A |
| PN | 48232 | Pneumonia due to Streptococcus, group B |
| PN | 48239 | Pneumonia due to other Streptococcus |
| PN | 48240 | Pneumonia due to Staphylococcus, unspecified |
| PN | 48241 | Methicillin-susceptible pneumonia due to Staphylococcus aureus |
| PN | 48242 | Methicillin-resistant pneumonia due to Staphylococcus aureus |
| PN | 48249 | Other Staphylococcus pneumonia |
| PN | 48281 | Pneumonia due to anaerobes |
| PN | 48282 | Pneumonia due to escherichia coli [E. coli] |
| PN | 48283 | Pneumonia due to other gram-negative bacteria |
| PN | 48284 | Pneumonia due to Legionnaires' disease |
| PN | 48289 | Pneumonia due to other specified bacteria |
| PN | 4829 | Bacterial pneumonia, unspecified |
| PN | 4830 | Pneumonia due to mycoplasma pneumoniae |
| PN | 4831 | Pneumonia due to chlamydia |
| PN | 4838 | Pneumonia due to other specified organism |
| PN | 4841 | Pneumonia in cytomegalic inclusion disease |
| PN | 4843 | Pneumonia in whooping cough |
| PN | 4845 | Pneumonia in anthrax |
| PN | 4846 | Pneumonia in aspergillosis |
| PN | 4847 | Pneumonia in other systemic mycoses |
| PN | 4848 | Pneumonia in other infectious diseases classified elsewhere |
| PN | 485 | Bronchopneumonia, organism unspecified |
| PN | 486 | Pneumonia, organism unspecified |
| PN | 4870 | Influenza with pneumonia |

## Appendix C – Detailed Regression Output

The following pages contain all of the regression output from model 3.1, shortened to save space. Each regression contains all of the variables described in Section 3.4.2. In particular, the following text remains unchanged for each regression. POA variables are highlighted.

```
Call:
glm(formula = yyy ~ p_no_flag + I(t_2008 * h_2008_poa_rate) +
    I(t_2009 * h_2009_poa_rate) + I(t_2010 * h_2010_poa_rate) +
    I(t_2011 * h_2011_poa_rate) + b_gndr + b_dob + c_util_day +
    c_cost + c_cost_outlier + d_type + d_mdc + h_beds + h_urban,
    family = binomial(logit), data = df)

Significance codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Dispersion parameter for binomial family taken to be 1


The dataset 'df' was replaced with one of the following:
       AMI     Acute myocardial infarction dataset
       HF      Heart Failure dataset
       PN      Pneumonia dataset

The dependent variable, 'yyy' was replaced with one of the following in each dataset:
       y_7     7-day readmission event
       y_15    15-day readmission event
       y_30    30-day readmission event

This results in 3x3 combination of logistic regressions, the output of which is listed
below.
```

**AMI 7-day hospital readmission**

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2189  -0.4673  -0.4116  -0.3554   2.5226

Coefficients:
                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                -1.790e+00  7.384e-01  -2.424 0.015346 *
p_no_flag1                  2.437e-01  9.857e-02   2.472 0.013439 *
I(t_2008 * h_2008_poa_rate) 2.059e+00  1.483e+00   1.389 0.164906
I(t_2009 * h_2009_poa_rate) 2.784e-01  4.906e-01   0.568 0.570364
I(t_2010 * h_2010_poa_rate) 3.151e-01  4.330e-01   0.728 0.466753
I(t_2011 * h_2011_poa_rate) 3.349e-01  5.922e-01   0.565 0.571765
b_gndrF                     3.018e-01  8.633e-02   3.496 0.000472 ***
b_dob70 Thru 74             1.664e-01  1.353e-01   1.230 0.218638
b_dob75 Thru 79             2.812e-01  1.351e-01   2.082 0.037340 *
b_dob80 Thru 84             2.678e-01  1.388e-01   1.930 0.053664 .
b_dob>84                    2.538e-01  1.363e-01   1.862 0.062636 .
c_util_day                  5.274e-02  1.131e-02   4.663 3.12e-06 ***
c_cost                      6.278e-07  4.453e-06   0.141 0.887902
c_cost_outlier             -2.432e-01  2.447e-01  -0.994 0.320296
d_typeSURG                 -1.752e-01  1.029e-01  -1.703 0.088597 .
d_mdc05                    -1.188e+00  7.051e-01  -1.684 0.092111 .
d_mdcPRE                   -1.551e+01  2.466e+02  -0.063 0.949832
h_beds                      1.068e-04  1.698e-04   0.629 0.529660
h_urbanUrban               -7.114e-02  1.292e-01  -0.551 0.581951
---
    Null deviance: 4411.4  on 7197  degrees of freedom
Residual deviance: 4307.1  on 7179  degrees of freedom
  (1559 observations deleted due to missingness)
AIC: 4345.1

Number of Fisher Scoring iterations: 14

                                        OR        2.5 %       97.5 %
(Intercept)                 1.669502e-01 3.338094e-02 6.636678e-01
p_no_flag1                  1.275909e+00 1.050739e+00 1.546560e+00
I(t_2008 * h_2008_poa_rate) 7.837393e+00 4.119999e-01 1.379607e+02
I(t_2009 * h_2009_poa_rate) 1.321012e+00 5.016160e-01 3.433016e+00
I(t_2010 * h_2010_poa_rate) 1.370402e+00 5.832431e-01 3.185098e+00
I(t_2011 * h_2011_poa_rate) 1.397763e+00 4.329396e-01 4.415455e+00
b_gndrF                     1.352347e+00 1.142170e+00 1.602329e+00
b_dob70 Thru 74             1.181064e+00 9.057468e-01 1.540146e+00
b_dob75 Thru 79             1.324744e+00 1.016543e+00 1.727069e+00
b_dob80 Thru 84             1.307102e+00 9.955436e-01 1.716245e+00
b_dob>84                    1.288952e+00 9.875531e-01 1.685906e+00
c_util_day                  1.054155e+00 1.030810e+00 1.077595e+00
c_cost                      1.000001e+00 9.999917e-01 1.000009e+00
c_cost_outlier              7.841223e-01 4.756823e-01 1.244738e+00
d_typeSURG                  8.393278e-01 6.864319e-01 1.027461e+00
d_mdc05                     3.049581e-01 8.217347e-02 1.447582e+00
d_mdcPRE                    1.830461e-07 4.654191e-80 9.882497e-79
h_beds                      1.000107e+00 9.997655e-01 1.000432e+00
h_urbanUrban                9.313355e-01 7.260521e-01 1.205468e+00
```

**AMI 15-day hospital readmission**

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3790  -0.5464  -0.4829  -0.4198   2.8672

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                 -1.602e+00  6.749e-01  -2.373  0.01764 *
p_no_flag1                   2.380e-01  8.385e-02   2.838  0.00453 **
I(t_2008 * h_2008_poa_rate)  9.465e-02  1.322e+00   0.072  0.94292
I(t_2009 * h_2009_poa_rate) -7.824e-03  4.242e-01  -0.018  0.98528
I(t_2010 * h_2010_poa_rate)  1.237e-01  3.723e-01   0.332  0.73967
I(t_2011 * h_2011_poa_rate)  2.975e-02  5.121e-01   0.058  0.95367
b_gndrF                      2.146e-01  7.328e-02   2.929  0.00340 **
b_dob70 Thru 74              1.268e-01  1.144e-01   1.108  0.26770
b_dob75 Thru 79              2.136e-01  1.155e-01   1.849  0.06441 .
b_dob80 Thru 84              3.363e-01  1.157e-01   2.906  0.00366 **
b_dob>84                     2.695e-01  1.152e-01   2.339  0.01932 *
c_util_day                   5.442e-02  9.787e-03   5.560  2.7e-08 ***
c_cost                       1.126e-06  3.817e-06   0.295  0.76797
c_cost_outlier              -3.217e-01  2.096e-01  -1.534  0.12494
d_typeSURG                  -1.356e-01  8.813e-02  -1.539  0.12388
d_mdc05                     -9.751e-01  6.479e-01  -1.505  0.13233
d_mdcPRE                    -3.711e+00  1.227e+00  -3.025  0.00249 **
h_beds                       1.976e-04  1.420e-04   1.392  0.16406
h_urbanUrban                -9.856e-03  1.124e-01  -0.088  0.93013
---
    Null deviance: 5696.8  on 7487  degrees of freedom
Residual deviance: 5564.1  on 7469  degrees of freedom
  (1269 observations deleted due to missingness)
AIC: 5602.1

Number of Fisher Scoring iterations: 5

                                       OR       2.5 %     97.5 %
(Intercept)                 0.20156404 0.048674067  0.7331511
p_no_flag1                  1.26869369 1.075765856  1.4944982
I(t_2008 * h_2008_poa_rate) 1.09927642 0.080028879 14.2618241
I(t_2009 * h_2009_poa_rate) 0.99220612 0.430036889  2.2687152
I(t_2010 * h_2010_poa_rate) 1.13168424 0.543613526  2.3398245
I(t_2011 * h_2011_poa_rate) 1.03020002 0.374831414  2.7913097
b_gndrF                     1.23939075 1.073670667  1.4310633
b_dob70 Thru 74             1.13519672 0.906901365  1.4205809
b_dob75 Thru 79             1.23807101 0.986998370  1.5525426
b_dob80 Thru 84             1.39973330 1.115656280  1.7564844
b_dob>84                    1.30925404 1.045162802  1.6420297
c_util_day                  1.05592330 1.035749772  1.0762930
c_cost                      1.00000113 0.999993539  1.0000085
c_cost_outlier              0.72494864 0.474327398  1.0805484
d_typeSURG                  0.87318538 0.734936649  1.0383025
d_mdc05                     0.37714611 0.109387317  1.4930412
d_mdcPRE                    0.02444599 0.001103334  0.2119443
h_beds                      1.00019761 0.999914237  1.0004714
h_urbanUrban                0.99019278 0.796894559  1.2384916
```

**AMI 30-day hospital readmission**

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4397  -0.6235  -0.5526  -0.4822   2.8573

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                 -1.552e+00  6.671e-01  -2.326 0.020001 *
p_no_flag1                   2.748e-01  7.364e-02   3.731 0.000191 ***
I(t_2008 * h_2008_poa_rate) -7.544e-01  1.166e+00  -0.647 0.517763
I(t_2009 * h_2009_poa_rate) -4.215e-01  3.753e-01  -1.123 0.261405
I(t_2010 * h_2010_poa_rate) -3.367e-01  3.302e-01  -1.019 0.307974
I(t_2011 * h_2011_poa_rate) -3.407e-01  4.507e-01  -0.756 0.449681
b_gndrF                      1.486e-01  6.423e-02   2.314 0.020678 *
b_dob70 Thru 74              1.578e-01  9.895e-02   1.594 0.110836
b_dob75 Thru 79              1.760e-01  1.016e-01   1.731 0.083369 .
b_dob80 Thru 84              3.384e-01  1.014e-01   3.336 0.000850 ***
b_dob>84                     3.197e-01  1.006e-01   3.178 0.001483 **
c_util_day                   5.401e-02  8.816e-03   6.127 8.97e-10 ***
c_cost                      -7.282e-07  3.445e-06  -0.211 0.832615
c_cost_outlier              -2.574e-01  1.833e-01  -1.404 0.160335
d_typeSURG                  -4.898e-02  7.790e-02  -0.629 0.529560
d_mdc05                     -6.467e-01  6.461e-01  -1.001 0.316805
d_mdcPRE                    -3.591e+00  1.221e+00  -2.940 0.003280 **
h_beds                       2.471e-04  1.238e-04   1.997 0.045849 *
h_urbanUrban                 1.080e-02  9.911e-02   0.109 0.913204
---
    Null deviance: 7024.6  on 7832  degrees of freedom
Residual deviance: 6867.8  on 7814  degrees of freedom
  (924 observations deleted due to missingness)
AIC: 6905.8

Number of Fisher Scoring iterations: 5

                                      OR       2.5 %    97.5 %
(Intercept)                 0.2118320 0.051804144 0.7583994
p_no_flag1                  1.3162281 1.138868008 1.5200679
I(t_2008 * h_2008_poa_rate) 0.4702947 0.046841548 4.5364865
I(t_2009 * h_2009_poa_rate) 0.6560667 0.313321907 1.3646005
I(t_2010 * h_2010_poa_rate) 0.7141350 0.372788144 1.3607424
I(t_2011 * h_2011_poa_rate) 0.7112773 0.292524513 1.7122556
b_gndrF                     1.1602227 1.022996558 1.3159354
b_dob70 Thru 74             1.1708965 0.964341249 1.4215516
b_dob75 Thru 79             1.1923811 0.976733755 1.4549850
b_dob80 Thru 84             1.4026982 1.149721371 1.7114039
b_dob>84                    1.3766519 1.130730533 1.6774501
c_util_day                  1.0555004 1.037365355 1.0738679
c_cost                      0.9999993 0.999992442 1.0000060
c_cost_outlier              0.7730890 0.534841366 1.0983647
d_typeSURG                  0.9522031 0.817632024 1.1097084
d_mdc05                     0.5237544 0.152566418 2.0681798
d_mdcPRE                    0.0275586 0.001251847 0.2360677
h_beds                      1.0002471 1.000001361 1.0004870
h_urbanUrban                1.0108611 0.834275414 1.2306134
```

**HF 7-day hospital readmission**

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8696  -0.5933  -0.5591  -0.5222   2.5700

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                 -9.605e+00  7.246e+01  -0.133 0.894547
p_no_flag1                   4.913e-02  2.867e-02   1.714 0.086527 .
I(t_2008 * h_2008_poa_rate) -1.029e+00  3.998e-01  -2.574 0.010041 *
I(t_2009 * h_2009_poa_rate) -3.847e-01  1.273e-01  -3.021 0.002516 **
I(t_2010 * h_2010_poa_rate) -5.678e-01  1.152e-01  -4.930 8.21e-07 ***
I(t_2011 * h_2011_poa_rate) -5.033e-01  1.528e-01  -3.293 0.000992 ***
b_gndrF                      2.831e-02  2.221e-02   1.275 0.202426
b_dob70 Thru 74              5.154e-02  4.128e-02   1.249 0.211796
b_dob75 Thru 79             -4.176e-02  3.988e-02  -1.047 0.295138
b_dob80 Thru 84             -1.052e-01  3.867e-02  -2.720 0.006529 **
b_dob>84                    -2.032e-01  3.637e-02  -5.586 2.33e-08 ***
c_util_day                   4.238e-02  2.843e-03  14.906  < 2e-16 ***
c_cost                      -1.782e-06  2.115e-06  -0.843 0.399260
c_cost_outlier              -2.983e-01  1.103e-01  -2.703 0.006870 **
d_typeMED                    7.973e+00  7.246e+01   0.110 0.912387
d_typeSURG                   7.713e+00  7.246e+01   0.106 0.915229
d_mdc05                     -1.001e-01  1.699e-01  -0.589 0.555594
d_mdcPRE                    -1.695e+00  4.975e-01  -3.408 0.000654 ***
h_beds                       1.113e-04  4.286e-05   2.596 0.009427 **
h_urbanUrban                 2.687e-03  3.134e-02   0.086 0.931658
---
    Null deviance: 57392  on 66719  degrees of freedom
Residual deviance: 56995  on 66700  degrees of freedom
  (18265 observations deleted due to missingness)
AIC: 57035

Number of Fisher Scoring iterations: 8

                                       OR        2.5 %      97.5 %
(Intercept)                 6.738131e-05           NA 674.2185903
p_no_flag1                  1.050361e+00 0.9928193339   1.1108972
I(t_2008 * h_2008_poa_rate) 3.572520e-01 0.1626929264   0.7799475
I(t_2009 * h_2009_poa_rate) 6.806702e-01 0.5301324519   0.8732434
I(t_2010 * h_2010_poa_rate) 5.667659e-01 0.4520634022   0.7100185
I(t_2011 * h_2011_poa_rate) 6.045468e-01 0.4477571023   0.8151641
b_gndrF                     1.028712e+00 0.9849295272   1.0745168
b_dob70 Thru 74             1.052890e+00 0.9711375754   1.1417095
b_dob75 Thru 79             9.591040e-01 0.8870888651   1.0372251
b_dob80 Thru 84             9.001662e-01 0.8345979151   0.9712075
b_dob>84                    8.161428e-01 0.7601672322   0.8766672
c_util_day                  1.043286e+00 1.0374861984   1.0491125
c_cost                      9.999982e-01 0.9999940551   1.0000024
c_cost_outlier              7.421145e-01 0.5956514667   0.9182336
d_typeMED                   2.901369e+03 0.0002892812          NA
d_typeSURG                  2.237969e+03 0.0002231548          NA
d_mdc05                     9.047337e-01 0.6543352868   1.2752107
d_mdcPRE                    1.835134e-01 0.0651264772   0.4626698
h_beds                      1.000111e+00 1.0000267678   1.0001948
h_urbanUrban                1.002691e+00 0.9431659906   1.0664445
```

**HF 15-day hospital readmission**

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8428  -0.6729  -0.6378  -0.5934   2.4357

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                 -8.640e+00  4.395e+01  -0.197 0.844167
p_no_flag1                   5.908e-02  2.545e-02   2.322 0.020247 *
I(t_2008 * h_2008_poa_rate) -9.074e-01  3.538e-01  -2.565 0.010330 *
I(t_2009 * h_2009_poa_rate) -4.001e-01  1.132e-01  -3.533 0.000411 ***
I(t_2010 * h_2010_poa_rate) -4.924e-01  1.018e-01  -4.837 1.32e-06 ***
I(t_2011 * h_2011_poa_rate) -5.263e-01  1.359e-01  -3.873 0.000107 ***
b_gndrF                      1.588e-02  1.969e-02   0.807 0.419789
b_dob70 Thru 74              4.776e-02  3.703e-02   1.290 0.197189
b_dob75 Thru 79             -1.627e-02  3.560e-02  -0.457 0.647547
b_dob80 Thru 84             -7.818e-02  3.451e-02  -2.266 0.023467 *
b_dob>84                    -1.599e-01  3.243e-02  -4.931 8.18e-07 ***
c_util_day                   3.959e-02  2.572e-03  15.396  < 2e-16 ***
c_cost                      -1.765e-06  1.912e-06  -0.923 0.355902
c_cost_outlier              -2.598e-01  9.828e-02  -2.644 0.008199 **
d_typeMED                    7.264e+00  4.395e+01   0.165 0.868746
d_typeSURG                   7.018e+00  4.395e+01   0.160 0.873141
d_mdc05                     -7.795e-02  1.526e-01  -0.511 0.609476
d_mdcPRE                    -1.614e+00  4.426e-01  -3.646 0.000266 ***
h_beds                       8.425e-05  3.825e-05   2.202 0.027639 *
h_urbanUrban                 1.578e-02  2.782e-02   0.567 0.570592
---
    Null deviance: 69489  on 70191  degrees of freedom
Residual deviance: 69074  on 70172  degrees of freedom
  (14793 observations deleted due to missingness)
AIC: 69114

Number of Fisher Scoring iterations: 7

                                       OR       2.5 %     97.5 %
(Intercept)                 1.769088e-04          NA 34.0960777
p_no_flag1                  1.060863e+00 1.009138783  1.1150039
I(t_2008 * h_2008_poa_rate) 4.035663e-01 0.201299105  0.8058355
I(t_2009 * h_2009_poa_rate) 6.702774e-01 0.536710502  0.8365970
I(t_2010 * h_2010_poa_rate) 6.111304e-01 0.500440581  0.7459027
I(t_2011 * h_2011_poa_rate) 5.907705e-01 0.452418247  0.7707034
b_gndrF                     1.016012e+00 0.977571602  1.0560135
b_dob70 Thru 74             1.048915e+00 0.975528482  1.1279498
b_dob75 Thru 79             9.838582e-01 0.917635717  1.0550557
b_dob80 Thru 84             9.247951e-01 0.864416386  0.9896303
b_dob>84                    8.522345e-01 0.799891961  0.9083216
c_util_day                  1.040385e+00 1.035153675  1.0456416
c_cost                      9.999982e-01 0.999994475  1.0000020
c_cost_outlier              7.711813e-01 0.634490228  0.9328838
d_typeMED                   1.427258e+03 0.007381521         NA
d_typeSURG                  1.116768e+03 0.005776270         NA
d_mdc05                     9.250099e-01 0.690259256  1.2567010
d_mdcPRE                    1.990949e-01 0.079659979  0.4554552
h_beds                      1.000084e+00 1.000008922  1.0001589
h_urbanUrban                1.015903e+00 0.962141261  1.0730009
```

**HF 30-day hospital readmission**

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8062  -0.7619  -0.7247  -0.6118   2.2535

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                 -8.600e+00  4.395e+01  -0.196 0.844882
p_no_flag1                   4.403e-02  2.278e-02   1.933 0.053275 .
I(t_2008 * h_2008_poa_rate) -8.912e-01  3.156e-01  -2.824 0.004748 **
I(t_2009 * h_2009_poa_rate) -3.760e-01  1.010e-01  -3.724 0.000196 ***
I(t_2010 * h_2010_poa_rate) -3.875e-01  9.036e-02  -4.289 1.79e-05 ***
I(t_2011 * h_2011_poa_rate) -5.045e-01  1.211e-01  -4.165 3.11e-05 ***
b_gndrF                     -4.505e-03  1.754e-02  -0.257 0.797294
b_dob70 Thru 74              4.383e-02  3.307e-02   1.325 0.185030
b_dob75 Thru 79             -3.300e-02  3.183e-02  -1.037 0.299862
b_dob80 Thru 84             -8.199e-02  3.078e-02  -2.663 0.007735 **
b_dob>84                    -1.493e-01  2.887e-02  -5.169 2.35e-07 ***
c_util_day                   3.725e-02  2.336e-03  15.946  < 2e-16 ***
c_cost                      -2.909e-06  1.707e-06  -1.704 0.088318 .
c_cost_outlier              -2.612e-01  8.905e-02  -2.934 0.003350 **
d_typeMED                    7.545e+00  4.395e+01   0.172 0.863715
d_typeSURG                   7.338e+00  4.395e+01   0.167 0.867421
d_mdc05                     -8.886e-02  1.358e-01  -0.654 0.512860
d_mdcPRE                    -1.169e+00  3.624e-01  -3.227 0.001250 **
h_beds                       7.548e-05  3.424e-05   2.204 0.027499 *
h_urbanUrban                 2.770e-02  2.486e-02   1.114 0.265237
---
    Null deviance: 83631  on 74880  degrees of freedom
Residual deviance: 83209  on 74861  degrees of freedom
  (10104 observations deleted due to missingness)
AIC: 83249

Number of Fisher Scoring iterations: 7

                                       OR       2.5 %     97.5 %
(Intercept)                 1.841532e-04          NA 35.4874324
p_no_flag1                  1.045010e+00 0.999303866  1.0926487
I(t_2008 * h_2008_poa_rate) 4.101726e-01 0.220648469  0.7603947
I(t_2009 * h_2009_poa_rate) 6.865848e-01 0.563176971  0.8366954
I(t_2010 * h_2010_poa_rate) 6.787194e-01 0.568453847  0.8100843
I(t_2011 * h_2011_poa_rate) 6.037869e-01 0.476017330  0.7653470
b_gndrF                     9.955053e-01 0.961877855  1.0303386
b_dob70 Thru 74             1.044800e+00 0.979267694  1.1147985
b_dob75 Thru 79             9.675419e-01 0.909077651  1.0298856
b_dob80 Thru 84             9.212843e-01 0.867407741  0.9786595
b_dob>84                    8.613522e-01 0.814049269  0.9116137
c_util_day                  1.037949e+00 1.033209566  1.0427142
c_cost                      9.999971e-01 0.999993736  1.0000004
c_cost_outlier              7.700944e-01 0.645640441  0.9154954
d_typeMED                   1.890527e+03 0.009785442         NA
d_typeSURG                  1.536866e+03 0.007955492         NA
d_mdc05                     9.149756e-01 0.704013343  1.1995965
d_mdcPRE                    3.105415e-01 0.148322924  0.6170213
h_beds                      1.000075e+00 1.000008138  1.0001424
h_urbanUrban                1.028082e+00 0.979291288  1.0795304
```

**PN 7-day hospital readmission**

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.5952  -0.4754  -0.4420  -0.4138   2.8430

Coefficients:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                   -2.287e+00  1.985e-01 -11.522  < 2e-16 ***
p_no_flag1                     3.251e-02  3.860e-02   0.842 0.399677
I(t_2008 * h_2008_poa_rate)   -1.691e+00  5.366e-01  -3.152 0.001621 **
I(t_2009 * h_2009_poa_rate)   -4.553e-01  1.710e-01  -2.663 0.007747 **
I(t_2010 * h_2010_poa_rate)   -5.827e-01  1.547e-01  -3.766 0.000166 ***
I(t_2011 * h_2011_poa_rate)   -7.536e-01  2.067e-01  -3.647 0.000266 ***
b_gndrF                       -1.282e-01  2.951e-02  -4.343 1.41e-05 ***
b_dob70 Thru 74               -6.917e-03  5.436e-02  -0.127 0.898746
b_dob75 Thru 79               -4.230e-03  5.232e-02  -0.081 0.935567
b_dob80 Thru 84               -1.799e-02  5.097e-02  -0.353 0.724083
b_dob>84                      -4.404e-02  4.766e-02  -0.924 0.355434
c_util_day                     5.129e-02  3.817e-03  13.437  < 2e-16 ***
c_cost                         1.982e-05  3.316e-06   5.976 2.28e-09 ***
c_cost_outlier                -2.788e-01  1.334e-01  -2.091 0.036541 *
d_typeSURG                    -3.472e-01  1.237e-01  -2.807 0.005003 **
d_mdc04                       -1.418e-01  1.858e-01  -0.763 0.445304
d_mdc25                       -1.198e+01  9.549e+01  -0.125 0.900135
d_mdcPRE                      -2.407e+00  4.164e-01  -5.780 7.46e-09 ***
h_beds                         1.508e-04  6.205e-05   2.430 0.015085 *
h_urbanUrban                   3.296e-02  3.895e-02   0.846 0.397417
---
    Null deviance: 34406  on 51856  degrees of freedom
Residual deviance: 33885  on 51837  degrees of freedom
  (13726 observations deleted due to missingness)
AIC: 33925

Number of Fisher Scoring iterations: 12

                                        OR        2.5 %       97.5 %
(Intercept)                   1.016017e-01 6.863456e-02 0.1495621445
p_no_flag1                    1.033044e+00 9.574730e-01 1.1139018819
I(t_2008 * h_2008_poa_rate)   1.842500e-01 6.397300e-02 0.5242673364
I(t_2009 * h_2009_poa_rate)   6.342516e-01 4.532291e-01 0.8859454920
I(t_2010 * h_2010_poa_rate)   5.583674e-01 4.119500e-01 0.7556053668
I(t_2011 * h_2011_poa_rate)   4.706594e-01 3.134639e-01 0.7047051302
b_gndrF                       8.797172e-01 8.303095e-01 0.9321354302
b_dob70 Thru 74               9.931065e-01 8.928050e-01 1.1048905305
b_dob75 Thru 79               9.957794e-01 8.989010e-01 1.1035554117
b_dob80 Thru 84               9.821694e-01 8.890330e-01 1.0856680202
b_dob>84                      9.569146e-01 8.719555e-01 1.0510904911
c_util_day                    1.052623e+00 1.044766e+00 1.0605165545
c_cost                        1.000020e+00 1.000013e+00 1.0000263242
c_cost_outlier                7.566753e-01 5.799099e-01 0.9784711347
d_typeSURG                    7.066762e-01 5.513216e-01 0.8957405595
d_mdc04                       8.678123e-01 6.044250e-01 1.2533145937
d_mdc25                       6.250311e-06 6.483132e-18 0.0007485875
d_mdcPRE                      9.010667e-02 3.842014e-02 0.1977271363
h_beds                        1.000151e+00 1.000028e+00 1.0002712804
h_urbanUrban                  1.033511e+00 9.578360e-01 1.1158513759
```

**PN 15-day hospital readmission**

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.5665  -0.5535  -0.5163  -0.4824   2.6565

Coefficients:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                   -1.938e+00  1.695e-01 -11.435  < 2e-16 ***
p_no_flag1                     3.192e-02  3.333e-02   0.958  0.33818
I(t_2008 * h_2008_poa_rate)   -1.232e+00  4.612e-01  -2.670  0.00758 **
I(t_2009 * h_2009_poa_rate)   -3.443e-01  1.478e-01  -2.329  0.01984 *
I(t_2010 * h_2010_poa_rate)   -3.202e-01  1.324e-01  -2.418  0.01561 *
I(t_2011 * h_2011_poa_rate)   -5.288e-01  1.778e-01  -2.974  0.00294 **
b_gndrF                       -1.382e-01  2.547e-02  -5.426 5.75e-08 ***
b_dob70 Thru 74                4.682e-02  4.714e-02   0.993  0.32063
b_dob75 Thru 79                2.418e-02  4.562e-02   0.530  0.59608
b_dob80 Thru 84                3.229e-02  4.430e-02   0.729  0.46618
b_dob>84                      -6.933e-03  4.158e-02  -0.167  0.86757
c_util_day                     5.131e-02  3.382e-03  15.170  < 2e-16 ***
c_cost                         1.718e-05  3.007e-06   5.715 1.10e-08 ***
c_cost_outlier                -3.586e-01  1.213e-01  -2.957  0.00311 **
d_typeSURG                    -2.627e-01  1.064e-01  -2.470  0.01351 *
d_mdc04                       -2.135e-01  1.578e-01  -1.353  0.17602
d_mdc25                       -1.427e+00  7.638e-01  -1.868  0.06175 .
d_mdcPRE                      -2.390e+00  3.702e-01  -6.457 1.06e-10 ***
h_beds                         1.264e-04  5.402e-05   2.340  0.01931 *
h_urbanUrban                   3.491e-02  3.363e-02   1.038  0.29925
---
    Null deviance: 43387  on 53978  degrees of freedom
Residual deviance: 42781  on 53959  degrees of freedom
  (11604 observations deleted due to missingness)
AIC: 42821

Number of Fisher Scoring iterations: 4

                                      OR      2.5 %    97.5 %
(Intercept)                   0.14404023 0.10316278 0.2005425
p_no_flag1                    1.03243534 0.96694544 1.1019019
I(t_2008 * h_2008_poa_rate)   0.29181112 0.11766714 0.7176032
I(t_2009 * h_2009_poa_rate)   0.70873003 0.53013954 0.9462616
I(t_2010 * h_2010_poa_rate)   0.72597931 0.55969459 0.9406741
I(t_2011 * h_2011_poa_rate)   0.58932699 0.41551162 0.8342353
b_gndrF                       0.87091822 0.82852768 0.9155203
b_dob70 Thru 74               1.04793567 0.95552315 1.1495005
b_dob75 Thru 79               1.02447785 0.93698060 1.1204973
b_dob80 Thru 84               1.03281221 0.94710384 1.1267571
b_dob>84                      0.99309131 0.91566590 1.0777600
c_util_day                    1.05265283 1.04569208 1.0596504
c_cost                        1.00001718 1.00001129 1.0000231
c_cost_outlier                0.69867364 0.54889070 0.8832385
d_typeSURG                    0.76895424 0.62189717 0.9438654
d_mdc04                       0.80774745 0.59355629 1.1023757
d_mdc25                       0.24007855 0.03702283 0.8684018
d_mdcPRE                      0.09160096 0.04305309 0.1846164
h_beds                        1.00012638 1.00001970 1.0002315
h_urbanUrban                  1.03552463 0.96967238 1.1063143
```

**PN 30-day hospital readmission**

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4733  -0.6391  -0.5989  -0.5586   2.5760

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                 -1.646e+00  1.491e-01 -11.042  < 2e-16 ***
p_no_flag1                   1.581e-02  2.923e-02   0.541 0.588603
I(t_2008 * h_2008_poa_rate) -9.961e-01  4.025e-01  -2.475 0.013319 *
I(t_2009 * h_2009_poa_rate) -1.880e-01  1.286e-01  -1.462 0.143870
I(t_2010 * h_2010_poa_rate) -2.238e-01  1.156e-01  -1.936 0.052853 .
I(t_2011 * h_2011_poa_rate) -3.147e-01  1.545e-01  -2.038 0.041586 *
b_gndrF                     -1.216e-01  2.225e-02  -5.466 4.60e-08 ***
b_dob70 Thru 74              7.051e-02  4.119e-02   1.712 0.086916 .
b_dob75 Thru 79              2.490e-02  4.001e-02   0.622 0.533762
b_dob80 Thru 84              6.060e-02  3.870e-02   1.566 0.117330
b_dob>84                    -1.090e-03  3.642e-02  -0.030 0.976128
c_util_day                   4.798e-02  3.037e-03  15.799  < 2e-16 ***
c_cost                       1.700e-05  2.707e-06   6.279 3.41e-10 ***
c_cost_outlier              -4.361e-01  1.122e-01  -3.888 0.000101 ***
d_typeSURG                  -1.876e-01  9.254e-02  -2.027 0.042635 *
d_mdc04                     -1.934e-01  1.387e-01  -1.395 0.163102
d_mdc25                     -1.675e+00  7.562e-01  -2.216 0.026725 *
d_mdcPRE                    -2.566e+00  3.487e-01  -7.360 1.84e-13 ***
h_beds                       8.944e-05  4.781e-05   1.871 0.061382 .
h_urbanUrban                 2.133e-02  2.923e-02   0.730 0.465518
---
    Null deviance: 53789  on 56813  degrees of freedom
Residual deviance: 53141  on 56794  degrees of freedom
  (8769 observations deleted due to missingness)
AIC: 53181

Number of Fisher Scoring iterations: 4

                                      OR      2.5 %    97.5 %
(Intercept)                 0.19281095 0.14381271 0.2580375
p_no_flag1                  1.01593730 0.95922144 1.0756947
I(t_2008 * h_2008_poa_rate) 0.36931038 0.16731951 0.8104148
I(t_2009 * h_2009_poa_rate) 0.82863235 0.64371362 1.0657538
I(t_2010 * h_2010_poa_rate) 0.79946182 0.63714680 1.0024049
I(t_2011 * h_2011_poa_rate) 0.72999342 0.53896848 0.9874436
b_gndrF                     0.88547125 0.84769454 0.9249622
b_dob70 Thru 74             1.07306042 0.98989804 1.1633749
b_dob75 Thru 79             1.02521140 0.94797902 1.1089743
b_dob80 Thru 84             1.06247403 0.98501848 1.1463686
b_dob>84                    0.99891080 0.93029526 1.0730636
c_util_day                  1.04915008 1.04292068 1.0554113
c_cost                      1.00001700 1.00001171 1.0000223
c_cost_outlier              0.64657704 0.51754106 0.8035335
d_typeSURG                  0.82894718 0.68975658 0.9915368
d_mdc04                     0.82413363 0.62843235 1.0826608
d_mdc25                     0.18724450 0.02919180 0.6650142
d_mdcPRE                    0.07681597 0.03771027 0.1485825
h_beds                      1.00008944 0.99999519 1.0001826
h_urbanUrban                1.02156400 0.96480683 1.0819582
```

## Appendix D – POA Frequency Data

The following pages contain cross tabulation output from R for each ICD code used in this study showing POA counts and proportions. This data represents "potential index" claims from the data sets extracted for statistical modeling. POA in this table is defined in Section 3.4.2. It does not contain tabulations for all available claims.

The format of each cell is defines as:

| | POA | | |
|---|---|---|---|
| AMI ICD Codes | 0 (yes) | 1 (no) | Row Total |
| 41000 | 1 | 1 | 2 |
| | 0.500 | 0.500 | 0.000 |
| | 0.000 | 0.000 | |
| 41001 | 348 | 226 | 574 |
| | 0.606 | 0.394 | 0.066 |
| | 0.040 | 0.026 | |
| 41010 | 7 | 2 | 9 |
| | 0.778 | 0.222 | 0.001 |
| | 0.001 | 0.000 | |
| 41011 | 1474 | 750 | 2224 |
| | 0.663 | 0.337 | 0.254 |
| | 0.169 | 0.086 | |
| 41020 | 1 | 0 | 1 |
| | 1.000 | 0.000 | 0.000 |
| | 0.000 | 0.000 | |
| 41021 | 336 | 170 | 506 |
| | 0.664 | 0.336 | 0.058 |
| | 0.038 | 0.019 | |
| 41031 | 205 | 112 | 317 |
| | 0.647 | 0.353 | 0.036 |
| | 0.023 | 0.013 | |
| 41040 | 6 | 3 | 9 |
| | 0.667 | 0.333 | 0.001 |
| | 0.001 | 0.000 | |
| 41041 | 1900 | 936 | 2836 |
| | 0.670 | 0.330 | 0.324 |
| | 0.217 | 0.107 | |
| 41050 | 1 | 1 | 2 |
| | 0.500 | 0.500 | 0.000 |
| | 0.000 | 0.000 | |
| 41051 | 197 | 113 | 310 |

| | | | |
|---|---|---|---|
| | 0.635 | 0.365 | 0.035 |
| | 0.023 | 0.013 | |
| 41081 | 177 | 45 | 222 |
| | 0.797 | 0.203 | 0.025 |
| | 0.020 | 0.005 | |
| 41090 | 33 | 4 | 37 |
| | 0.892 | 0.108 | 0.004 |
| | 0.004 | 0.000 | |
| 41091 | 1322 | 376 | 1698 |
| | 0.779 | 0.221 | 0.194 |
| | 0.151 | 0.043 | |
| **Column Total** | **6008** | **2739** | **8747** |

| | POA | | |
|---|---|---|---|
| **HF ICD Codes** | **0 (yes)** | **1 (no)** | **Row Total** |
| 4280 | 25124 | 4705 | 29829 |
| | 0.842 | 0.158 | 0.351 |
| | 0.296 | 0.055 | |
| 4281 | 155 | 28 | 183 |
| | 0.847 | 0.153 | 0.002 |
| | 0.002 | 0.000 | |
| 4289 | 20 | 3 | 23 |
| | 0.870 | 0.130 | 0.000 |
| | 0.000 | 0.000 | |
| 40201 | 176 | 28 | 204 |
| | 0.863 | 0.137 | 0.002 |
| | 0.002 | 0.000 | |
| 40211 | 35 | 13 | 48 |
| | 0.729 | 0.271 | 0.001 |
| | 0.000 | 0.000 | |
| 40291 | 1892 | 368 | 2260 |
| | 0.837 | 0.163 | 0.027 |
| | 0.022 | 0.004 | |
| 40401 | 163 | 71 | 234 |
| | 0.697 | 0.303 | 0.003 |
| | 0.002 | 0.001 | |
| 40403 | 66 | 10 | 76 |
| | 0.868 | 0.132 | 0.001 |
| | 0.001 | 0.000 | |
| 40411 | 51 | 16 | 67 |
| | 0.761 | 0.239 | 0.001 |
| | 0.001 | 0.000 | |
| 40413 | 6 | 3 | 9 |
| | 0.667 | 0.333 | 0.000 |
| | 0.000 | 0.000 | |
| 40491 | 1822 | 722 | 2544 |
| | 0.716 | 0.284 | 0.030 |
| | 0.021 | 0.009 | |
| 40493 | 416 | 146 | 562 |

| | | | |
|---|---|---|---|
| | 0.740 | 0.260 | 0.007 |
| | 0.005 | 0.002 | |
| 42820 | 1124 | 208 | 1332 |
| | 0.844 | 0.156 | 0.016 |
| | 0.013 | 0.002 | |
| 42821 | 3699 | 948 | 4647 |
| | 0.796 | 0.204 | 0.055 |
| | 0.044 | 0.011 | |
| 42822 | 1094 | 232 | 1326 |
| | 0.825 | 0.175 | 0.016 |
| | 0.013 | 0.003 | |
| 42823 | 11279 | 3738 | 15017 |
| | 0.751 | 0.249 | 0.177 |
| | 0.133 | 0.044 | |
| 42830 | 2429 | 461 | 2890 |
| | 0.840 | 0.160 | 0.034 |
| | 0.029 | 0.005 | |
| 42831 | 3669 | 934 | 4603 |
| | 0.797 | 0.203 | 0.054 |
| | 0.043 | 0.011 | |
| 42832 | 701 | 154 | 855 |
| | 0.820 | 0.180 | 0.010 |
| | 0.008 | 0.002 | |
| 42833 | 8920 | 2885 | 11805 |
| | 0.756 | 0.244 | 0.139 |
| | 0.105 | 0.034 | |
| 42840 | 312 | 61 | 373 |
| | 0.836 | 0.164 | 0.004 |
| | 0.004 | 0.001 | |
| 42841 | 726 | 214 | 940 |
| | 0.772 | 0.228 | 0.011 |
| | 0.009 | 0.003 | |
| 42842 | 206 | 51 | 257 |
| | 0.802 | 0.198 | 0.003 |
| | 0.002 | 0.001 | |
| 42843 | 3548 | 1247 | 4795 |
| | 0.740 | 0.260 | 0.056 |
| | 0.042 | 0.015 | |
| **Column Total** | **67633** | **17246** | **84879** |

| | POA | | |
|---|---|---|---|
| **PN ICD Codes** | **0 (yes)** | **1 (no)** | **Row Total** |
| 481 | 841 | 223 | 1064 |
| | 0.790 | 0.210 | 0.016 |
| | 0.013 | 0.003 | |
| 485 | 600 | 121 | 721 |
| | 0.832 | 0.168 | 0.011 |
| | 0.009 | 0.002 | |
| 486 | 44662 | 10510 | 55172 |

| | | | |
|---|---|---|---|
| | 0.810 | 0.190 | 0.843 |
| | 0.682 | 0.161 | |
| 4800 | 3 | 3 | 6 |
| | 0.500 | 0.500 | 0.000 |
| | 0.000 | 0.000 | |
| 4801 | 23 | 6 | 29 |
| | 0.793 | 0.207 | 0.000 |
| | 0.000 | 0.000 | |
| 4802 | 4 | 2 | 6 |
| | 0.667 | 0.333 | 0.000 |
| | 0.000 | 0.000 | |
| 4808 | 10 | 4 | 14 |
| | 0.714 | 0.286 | 0.000 |
| | 0.000 | 0.000 | |
| 4809 | 205 | 55 | 260 |
| | 0.788 | 0.212 | 0.004 |
| | 0.003 | 0.001 | |
| 4820 | 278 | 104 | 382 |
| | 0.728 | 0.272 | 0.006 |
| | 0.004 | 0.002 | |
| 4821 | 999 | 422 | 1421 |
| | 0.703 | 0.297 | 0.022 |
| | 0.015 | 0.006 | |
| 4822 | 183 | 65 | 248 |
| | 0.738 | 0.262 | 0.004 |
| | 0.003 | 0.001 | |
| 4829 | 939 | 300 | 1239 |
| | 0.758 | 0.242 | 0.019 |
| | 0.014 | 0.005 | |
| 4830 | 76 | 22 | 98 |
| | 0.776 | 0.224 | 0.001 |
| | 0.001 | 0.000 | |
| 4831 | 4 | 3 | 7 |
| | 0.571 | 0.429 | 0.000 |
| | 0.000 | 0.000 | |
| 4838 | 36 | 10 | 46 |
| | 0.783 | 0.217 | 0.001 |
| | 0.001 | 0.000 | |
| 4870 | 438 | 99 | 537 |
| | 0.816 | 0.184 | 0.008 |
| | 0.007 | 0.002 | |
| 48230 | 109 | 33 | 142 |
| | 0.768 | 0.232 | 0.002 |
| | 0.002 | 0.001 | |
| 48231 | 4 | 6 | 10 |
| | 0.400 | 0.600 | 0.000 |
| | 0.000 | 0.000 | |
| 48232 | 10 | 2 | 12 |
| | 0.833 | 0.167 | 0.000 |
| | 0.000 | 0.000 | |

| | | | |
|---|---|---|---|
| 48239 | 62 | 19 | 81 |
| | 0.765 | 0.235 | 0.001 |
| | 0.001 | 0.000 | |
| 48240 | 42 | 13 | 55 |
| | 0.764 | 0.236 | 0.001 |
| | 0.001 | 0.000 | |
| 48241 | 690 | 126 | 816 |
| | 0.846 | 0.154 | 0.012 |
| | 0.011 | 0.002 | |
| 48242 | 831 | 507 | 1338 |
| | 0.621 | 0.379 | 0.020 |
| | 0.013 | 0.008 | |
| 48249 | 39 | 14 | 53 |
| | 0.736 | 0.264 | 0.001 |
| | 0.001 | 0.000 | |
| 48281 | 12 | 5 | 17 |
| | 0.706 | 0.294 | 0.000 |
| | 0.000 | 0.000 | |
| 48282 | 154 | 83 | 237 |
| | 0.650 | 0.350 | 0.004 |
| | 0.002 | 0.001 | |
| 48283 | 977 | 352 | 1329 |
| | 0.735 | 0.265 | 0.020 |
| | 0.015 | 0.005 | |
| 48284 | 62 | 30 | 92 |
| | 0.674 | 0.326 | 0.001 |
| | 0.001 | 0.000 | |
| 48289 | 31 | 11 | 42 |
| | 0.738 | 0.262 | 0.001 |
| | 0.000 | 0.000 | |
| **Column Total** | **52324** | **13150** | **65474** |

```
> CrossTable(AMI$c_surgery, AMI$y30b, chisq=TRUE);


   Cell Contents
|-------------------------|
|                       N |
| Chi-square contribution |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  8028


              | AMI$y30b
AMI$c_surgery |         0 |         1 | Row Total |
--------------|-----------|-----------|-----------|
            0 |      1223 |       236 |      1459 |
              |     0.018 |     0.089 |           |
              |     0.838 |     0.162 |     0.182 |
              |     0.182 |     0.178 |           |
              |     0.152 |     0.029 |           |
--------------|-----------|-----------|-----------|
            1 |      5481 |      1088 |      6569 |
              |     0.004 |     0.020 |           |
              |     0.834 |     0.166 |     0.818 |
              |     0.818 |     0.822 |           |
              |     0.683 |     0.136 |           |
--------------|-----------|-----------|-----------|
 Column Total |      6704 |      1324 |      8028 |
              |     0.835 |     0.165 |           |
--------------|-----------|-----------|-----------|


Statistics for All Table Factors


Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  0.1299469     d.f. =  1     p =  0.7184871

Pearson's Chi-squared test with Yates' continuity correction
------------------------------------------------------------
Chi^2 =  0.1033545     d.f. =  1     p =  0.7478407
```

```
> CrossTable(AMI$c_surgery, AMI$p_no_flag, chisq=TRUE);


   Cell Contents
|-------------------------|
|                       N |
| Chi-square contribution |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  8747


              | AMI$p_no_flag
AMI$c_surgery |         0 |         1 | Row Total |
--------------|-----------|-----------|-----------|
            0 |      1378 |       210 |      1588 |
              |    75.653 |   165.946 |           |
              |     0.868 |     0.132 |     0.182 |
              |     0.229 |     0.077 |           |
              |     0.158 |     0.024 |           |
--------------|-----------|-----------|-----------|
            1 |      4630 |      2529 |      7159 |
              |    16.781 |    36.810 |           |
              |     0.647 |     0.353 |     0.818 |
              |     0.771 |     0.923 |           |
              |     0.529 |     0.289 |           |
--------------|-----------|-----------|-----------|
 Column Total |      6008 |      2739 |      8747 |
              |     0.687 |     0.313 |           |
--------------|-----------|-----------|-----------|


Statistics for All Table Factors


Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  295.1906     d.f. =  1     p =  3.678035e-66

Pearson's Chi-squared test with Yates' continuity correction
------------------------------------------------------------
Chi^2 =  294.1638     d.f. =  1     p =  6.156239e-66
```

```
> CrossTable(AMI$y30b, AMI$p_no_flag, chisq=TRUE);


   Cell Contents
|-------------------------|
|                       N |
| Chi-square contribution |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  8020


             | AMI$p_no_flag
    AMI$y30b |         0 |         1 | Row Total |
-------------|-----------|-----------|-----------|
           0 |      4739 |      1958 |      6697 |
             |     2.064 |     4.658 |           |
             |     0.708 |     0.292 |     0.835 |
             |     0.853 |     0.795 |           |
             |     0.591 |     0.244 |           |
-------------|-----------|-----------|-----------|
           1 |       819 |       504 |      1323 |
             |    10.445 |    23.581 |           |
             |     0.619 |     0.381 |     0.165 |
             |     0.147 |     0.205 |           |
             |     0.102 |     0.063 |           |
-------------|-----------|-----------|-----------|
Column Total |      5558 |      2462 |      8020 |
             |     0.693 |     0.307 |           |
-------------|-----------|-----------|-----------|


Statistics for All Table Factors


Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  40.74791     d.f. =  1     p =  1.731882e-10

Pearson's Chi-squared test with Yates' continuity correction
------------------------------------------------------------
Chi^2 =  40.3326     d.f. =  1     p =  2.14205e-10
```

```
> CrossTable(AMI$y30b, AMI$test_poa_surgery, chisq=TRUE);


   Cell Contents
|-------------------------|
|                       N |
| Chi-square contribution |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  8028


             | AMI$test_poa_surgery
    AMI$y30b |         0 |         1 | Row Total |
-------------|-----------|-----------|-----------|
           0 |      4900 |      1804 |      6704 |
             |     1.944 |     4.913 |           |
             |     0.731 |     0.269 |     0.835 |
             |     0.852 |     0.793 |           |
             |     0.610 |     0.225 |           |
-------------|-----------|-----------|-----------|
           1 |       852 |       472 |      1324 |
             |     9.844 |    24.878 |           |
             |     0.644 |     0.356 |     0.165 |
             |     0.148 |     0.207 |           |
             |     0.106 |     0.059 |           |
-------------|-----------|-----------|-----------|
Column Total |      5752 |      2276 |      8028 |
             |     0.716 |     0.284 |           |
-------------|-----------|-----------|-----------|


Statistics for All Table Factors


Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  41.58003     d.f. =  1     p =  1.131429e-10

Pearson's Chi-squared test with Yates' continuity correction
------------------------------------------------------------
Chi^2 =  41.15087     d.f. =  1     p =  1.409205e-10
```

# Appendix E

It is tempting to arrive at the first answer and stop without being mindful of alternatives or without completing the rigor necessary to demonstrate even to yourself that *your* answer really is *the* answer. This, however, is not consistent with the scientific method, which strives for considered experimentation, reproducibility, and communication of results. In this section, I would like to address the middle item: reproducibility.

After assessing the state of the raw input data files, the need for care became obvious when several "common" fields did not contain data encoded in a similar way. Ignoring this difference in coding could have led to a significant waste of time or possibly even erroneous results. Each LDS file is accompanied by a data dictionary defining the individual data fields, layout, data type, and contents. Extending these definitions to secondary data such as ResDAC field decode definitions, U.S. census data, CMS ICD-9 definitions, etc., means managing dozens of data sources and their associated documentation. Research is an iterative process of considering options to choose a direction, extending an effort in that direction, and building upon what has come before. Working through a few iterations is manageable, but when numerous iterations flow in quick succession, it can be difficult to judge when an error slipped into the mix. Managing the iterations means both process (what are you doing and how are you doing it) and product (a database, queries, graphs, extracts, models, regression output).

To manage and ensure accountability for all items in the research effort, I found the following steps beneficial, as they gave me confidence in my work its results by making the process repeatable.

## Secure the Input Data

Create a data management directory containing the following subdirectories: "raw", "docs", "load". Save all data used into the "raw" directory or subdirectory. Toggle file permissions to "read-only" mode to prevent accidents. Run an MD5 checksum on the file. Add a comment to a Sources document located in "docs" containing: the name of

your source file, its checksum, where the file originated, a link to it, and its data dictionary, help page, or manual. Save the data dictionary into "docs" as well. Make this file writeable and add comments, highlighting, and references. In my case, I created a database to normalize my data. Script everything that builds the ultimate dataset. Manual processes are subject to error even if those steps are "documented." All scripts, tools, and applications in the afferent flow are saved in "load" directory. Within all of my code, I left enough landmarks via print statements to isolate an error when it would arise. Which application was running? What was it doing? All of this information is saved in log files located in "docs" for future reference.

## Automate Your Work

In cases such as with input data for this study, the original source data was not directly usable. It had to be altered to a normalized form. A small set of changes to input data can be done confidently. A mountain of changes cannot. Likewise, manually checking a few hundred rows of data for "equivalence" is not a bit task. Doing so repeatedly is tedious. Doing so for 2.8M rows is impossible. In this study, I built a loop-back data comparison tool to read data from the database and compare it to the original data file. Automating the build and check functions makes it possible to rebuild the database overnight. Build in some fault tolerance. Don't stop on the first error. Some data fields do contain garbage. Note it and continue. Then review the output in the log file. One-off errors are probably not a worry, but 190,000 of the same error may need looking into.

Input source data is *their* data. The resulting dataset it *your* data. You get to define relationships and meaning. Once the input process is secured, build the intermediate data, which will become the base for the statistical extract. Automate this step as well. Save these scripts into their own directory apart from others. These scripts tend to change as understanding of data and identification of ultimate target change over time. Break these apart into distinct pieces. For this study, patient history represents one distinct set of data and will need a separate set of activities to produce them. These were distinct from integrating hospital information into the mix, and therefore had its own

directory for source. Likewise for efferent flow. Create separate directories for each type of extract. Document what you are doing for yourself.

## Use Version Control

Version control helps manage the daily source change by tracking it for you. Once daily effort has led to a successful build, save a copy of the project. Save all source files and documentation. Source data doesn't change, and doesn't need to be part of the versioning process. Versioning captures the state of all the pieces used. Some pieces vary more than others. Once those are at a stable point, version the project again, and then continue to modify. For a small group project, one doesn't need the spectrum of features available for large projects. I chose Git[47]. Read the documentation and incorporate it into daily workflow. When things do go awry, the version tool can show changes between what is in the source files now and what was there when it last worked.

Version control shows what changes. It can't show why it changes. Keep a running changes document in the "docs" directory. This is for you to keep track of yourself and your actions. Such a document is useful when planning your next set of changes.

Snap a version when you show someone your results, even if at an intermediate state, particularly statistical results, graphs, charts, tables. Use temporary directories when making moment runs. Once stable, save visualizations and the code used to make the visualization in a directory in the project directory. Snap a version with a comment.

## Iterate

The daily grind is not iteration; it's work. Iteration is the longer view. What happened last month? What do I need to accomplish this week? How does that fit into my ultimate goal? This mindset preserves the work and products in such a way as to make future work more efficient. Making changes to a database or dataset as understanding of data and relationships changes is easy once the process is automated. It is repeatable. The repeatability fosters both confidence in the result and an ability to experiment. Capture what is going on in a version file in the "docs" directory. Avoid the temptation to just dive into the problem. Think about the immediate changes within the

context of the larger iteration. Finish one change before starting another. Snap a version. Update your own documentation so you don't forget something, or loose track of what you were doing, or why.

### Backup Your Work

Version control is a way of "backing up" a logical state of your source code. It is not a physical backup. Disk space is cheap. Buy an external drive and use it regularly. Backup everything: original source files, source code, datasets, database, output directories, and documentation, anything you'd be "sad" to loose. Keep the backup safe. Accidents happen. Lightning happens. Theft happens. Stupidity happens. Be an adult. Learn about backup strategies and use them.

### Define a Baseline and Stick to It

Building up a research database requires utilizing dozens of applications, libraries, tools, modules, etc., and writing custom scripts and applications on top of those. In my case, these reside on a laptop. All of these have their own version. Stick to the major version throughout the life of your project. Collect the names, versions, any notes on installation, configuration, setup, etc. into a "baseline" document in your "docs" directory. This baseline is crucial to progress. Applying something such as an operating system update may damage the baseline, rendering it unusable. Correcting a baseline change may not be trivial. Downloading a new version of an application may render additional libraries as "out of date" and could require purchase of new upgrades. This ricochet is time consuming and frustrating at the end of the project. Resist the temptation. Plan for it. Snap a version of all code. Make a backup of the project. Find out what will need changing before applying any change. Security patches are generally not as catastrophic as operating system upgrades.