Regularization and Hierarchical Prior Distributions for Adjustment with Health Care Claims Data: Rethinking Comorbidity Scores

Jacob V Spertus¹, Samrachana Adhikari¹, and Sharon-Lise T Normand¹,²
1: Department of Health Care Policy, Harvard Medical School
2: Department of Biostatistics, Harvard TH Chan School of Public Health

Abstract

Health care claims data refer to information generated from interactions within health systems. They have been used in health services research for decades to assess effectiveness of interventions, determine the quality of medical care, predict disease prognosis, and monitor population health. While claims data are relatively cheap and ubiquitous, they are high-dimensional, sparse, and noisy, typically requiring dimension reduction. In health services research, the most common data reduction strategy involves use of a comorbidity index – a single number summary reflecting overall patient health. We discuss Bayesian regularization strategies and a novel hierarchical prior distribution as better options for dimension reduction in claims data. The specifications are designed to work with a large number of codes while controlling variance by shrinking coefficients towards zero or towards a group-level mean. A comparison of drug-eluting to bare-metal coronary stents illustrates approaches. In our application, regularization and a hierarchical prior improved over comorbidity scores in terms of prediction and causal inference, as evidenced by out-of-sample fit and the ability to meet falsifiability endpoints.

Keywords: Bayesian methods; comorbidity scores; propensity score weighting; regularization; coronary stent.

1 Introduction

Health services researchers focus on assessing the effectiveness of treatments and policy interventions, caracterizing the quality of medical care delivered to specific populations, and developing payment systems, with the goal of improving health outcomes. A fundamental data source used in health services research is administrative billing claims data. Claims data refer to a wide range of administrative databases capturing health care utilization information for reimbursement purposes and are generated by routine interactions between patients, the health system, and insurance providers [12]. This type of data stands in contrast to other formats, like clinical registry data, which are prospectively collected for the explicit purpose of later analysis and are of considerably higher quality [6] [27] [12]. However claims data have many desirable features: they are generated during the normal work flow and thus avoid additional effort and expense; they provide an accounting of denominator populations to ensure rates are well-defined; and they have standard definitions...
to help with consistency of the information collected across the nation and around the world.

In this article, we focus on claims data comprised of alpha-numeric codes for conditions from the International Classification of Diseases (9th Clinical Modification, [ICD-9]) system [9]. The structure of ICD-9 codes is a tree-like hierarchy with each code constructed as a binary variable. The codes contain 3-5 alpha-numeric strings where the first 3 characters define a group of diagnoses (or procedures) and the additional two characters provide further specificity. The ICD-9 codes include approximately 17,000 distinct codes at level 5, many of which represent rare conditions that are unlikely to be observed in any finite population. Coding practices can also vary across health delivery settings such that codes are noisy representations of true conditions. A revision of ICD-9 codes (ICD-10) took effect in October 2015 in the United States, and contains approximately 68,000 codes having a similar hierarchical structure to the ICD-9 system. Thus, while claims data are relatively cheap and ubiquitous, they are high-dimensional, sparse, and noisy. Consequently, some form of dimension reduction for inference is required.

Researchers have proposed a variety of dimension reduction approaches when using claims data. The most common strategies either use clinical expertise to select a small set of codes and include separate binary indicators for each discrete diagnosis or procedure, or implement standard variable selection approaches such as step-wise regression. Because these approaches may omit important variables, inclusion of a summary measure, called a comorbidity index or score, helps with adjusting for other health features [29]. Two of the most frequently used comorbidity indices include the Charlson and the Elixhauser [7][8]. These indices were constructed by selecting and validating ICD-9 codes based on a combination of clinical expertise and statistical analysis. The Elixhauser index has been shown to predict long-term clinical outcomes better than the Charlson index [29]. Both indices may be implemented as a set of dummy variables indicating the presence of groups of conditions (e.g. hypertension, diabetes, etc) if feasible or as a single continuous variable constructed from a weighted sum of the individual condition groups [7][39]. A recent paper by Gilbert et al argued that using a comorbidity index with fixed historical weights is as good as including individual claims coefficients in causal survival analysis [11]. However the assumptions associated with the strategy are rather strong, inferences may be compromised by ignoring uncertainty from the estimation of the weights, and the findings do not translate to other approaches like propensity score analyses or logistic regression for mortality prediction.

Our point of departure from typical approaches is to consider the dimension reduction task as one of regularization. Fundamentally some outside information, typically an assumption that some of the conditions have little or no effect on the outcome of interest, must be incorporated when the number of variables available exceeds the capacity of the data for appropriately precise estimation.

We note that other sophisticated methods have been developed to use high-dimensional claims data for various analysis goals. Syed et al proposed using a support vector machine to enhance risk prediction using current procedural terminology codes [36]. Perotte et al proposed exploiting the hierarchical nature of ICD-9 codes using a support vector machine for diagnosis code assignment [20]. Singh used the group lasso to improve outcome prediction by accounting for the ICD-9 hierarchy in a frequentist setting [30]. In the causal inference domain, Schneeweiss et al developed a frequentist algorithm to select claims confounders based on their prevalence and univariate associations with both treatment and
outcome \[28\].

Our work is motivated by determining the comparative effectiveness of coronary stents once released into the U.S. market. Stents are small scaffolds placed into the heart via a percutaneous coronary intervention (PCI) in order to treat arteries blocked by plaque and keep them clear and open. Drug-eluting stents (DESs) are coated in a drug designed to keep plaque from re-forming, while bare-metal stents (BMSs) have no such coating. Clinical trials have consistently shown that DESs are efficacious as they reduce the need for target vessel revascularization (TVR), i.e., follow-up procedures in the same artery, but have no benefit on survival \[2\][34]. However, DESs necessitate adherence to prolonged dual-antiplatelet therapies to prevent rare but fatal stent thrombosis. Several patient characteristics such as age, comorbidity, and socio-economic factors may impact adherence to dual-antiplatelet therapy, and thus such patients are likely to be implanted with BMSs \[33\]. Because these characteristics also tend to increase the risk of adverse outcomes, observed differences in average outcomes between BMS and DES treated patient groups are heavily confounded. Even after statistical adjustments, various observational analyses have generally found that DES provides a survival benefit in usual care populations \[14\][37]. Although there has historically been speculation that this benefit was real but not observed in randomized trial populations, a recent quasi-randomization analysis and the results of a large pragmatic clinical trial make a strong case that any observed survival benefit is due to residual confounding \[42\][2]. As a result, it remains an open challenge for observational causal inference to accurately estimate the effects of DES versus BMS using observational data. It may be possible to achieve better results by including claims data in some form.

While we utilize a Bayesian propensity score framework, the ideas apply to other techniques such as frequentist propensity scores or Bayesian outcome modeling. A Bayesian approach is particularly well aligned with high-dimensional inference with claims data, as it provides a number of intuitive tools for model building and incorporates all major sources of uncertainty, which may ultimately lead to better inferential performance \[31\]. Our focus is on improving over a typical analysis using comorbidity indices and to this end, we make three interdependent arguments: (1) administrative claims data can help analysts meet the ignorability assumption in observational studies, but standard comorbidity indices may be insufficient; (2) using a considerably larger and more granular set of claims data may lead to better results compared with a comorbidity index; and (3) regularization techniques are very useful for controlling variance when using a high-dimensional set of claims data. Combining these principles, we find that employing high-dimensional claims data with Bayesian regularization techniques leads to significant improvement over a comorbidity index in terms of treatment model fit and causal inference.

In section 2, we briefly remind the reader the necessary assumptions for causal inference with observational data and review propensity score approaches. Our proposed methods are presented in detail in section 3. Section 4 illustrates various strategies for dimensionality reduction by revisiting the DES vs BMS problem. We conclude with a discussion of the advantages of regularization over comorbidity indices for causal inference, as well as potential improvements and modifications for future research.
2 Causal Inference and Propensity Scores

2.1 Causal Inference

Causal inference involves various assumptions, three of which include stable unit treatment value assumption (SUTVA), positivity, and treatment assignment ignorability. We focus on binary treatments where a new treatment is compared to a standard treatment. In this setting, SUTVA implies that treatment does not vary within treatment groups and that the treatment received by one subject does not affect the outcomes of another. Positivity requires that every subject has a chance of receiving the new treatment and we assume that this probability is bounded away from 0 and 1. Positivity can be empirically tested in that blatant violations are apparent by examination or modeling of the data. Propensity scores help facilitate assessing the positivity assumption \( [21] \). Finally, ignorability assumes that potential outcomes are independent of treatment assignment \( [24] \). In observational designs, investigators must adjust for confounders, variables that affect both treatment assignment and outcome, to uncover the causal effect of a particular treatment. This task is among the most challenging in observational causal inference and has inspired many attempts to define methods, either statistical or expertise driven, to select confounders for adjustment \( [3] [28] [31] \). High-dimensional datasets may provide important variables that help investigators meet the ignorability assumption. In particular, health care claims data provide an exciting source of potential confounders, presenting a host of opportunities as well as challenges for researchers.

2.2 Propensity Scores

Propensity scores are the probabilities of receiving the new treatment relative to the standard treatment. They are a commonly used tool for observational causal inference. One particularly important feature involves facilitating an outcome free analysis in that they separate outcome modeling from treatment modeling which can help to maintain objectivity \( [25] \). This feature is often a requirement by regulators, such as the U.S. Food and Drug Administration, who place a premium on objective findings \( [13] \). Propensity scores can be particularly useful in high-dimensional settings when event rates are low or when the number of confounders is large relative to number of events. As a result, many investigators interested in observational causal inference with high-dimensional data have focused on how to build propensity score models. Variable selection is the standard response to dealing with high-dimensional data, and researchers have historically focused on selecting important confounders based on subject matter expertise and rules of thumb \( [3] [35] \). More algorithmic variable selection techniques have also been explored, including the high-dimensional propensity score algorithm of Schneeweiss et al, but this uses outcome information to select variables \( [28] \). Other approaches have proposed more technical variance reduction strategies through machine learning methods, such as elastic net regularization, tree-based algorithms, and learning ensembles \( [32] [15] [38] \). Each of these methods has shown promise in estimating causal effects from large databases. We use a Bayesian framework because it propagates uncertainty from the propensity score model and provides intuitive and sophisticated variance reduction strategies via prior distributions, an advantage that we leverage through a hierarchical prior specification.
2.2.1 Bayesian Propensity Scores

We adopt a Bayesian treatment model and maintain an outcome free analysis, although jointly modeling the propensity score and outcomes has been proposed \cite{16,26}. The particular algorithm we use resembles frequentist propensity score weighting \cite{31}. The algorithm uses a conjugate prior-likelihood pair in the form of a beta-binomial: if \( \tilde{Y} \) represents the total number of subjects having an event out of \( n \) subjects, e.g., \( \tilde{Y} \overset{iid}{\sim} \text{Bin}(n,p) \), then assuming \( p \) arises from \( \text{Beta}(a,b) \) \textit{a-priori}, yields \textit{a-posteriori} \( p \mid \tilde{Y} \sim \text{Beta}(a+Y, b+n-Y) \). We use this fact to integrate over draws from the propensity score model while conditionally drawing from the outcome model, so that uncertainty is incorporated from both the propensity score model and the outcome model.

Briefly, let \( Y_i \) denote binary outcome for subject \( i \), \( X_i \) binary treatment, and \( \pi_i \) the propensity score, \( \mathbb{P}(X_i = 1) \). The parameter \( \pi \) is the vector of propensity scores for all subjects, \( Y \) the vector of outcomes, and \( D = \{X, Y\} \) their concatenation into an \( n \times 2 \) matrix. Together, \( \pi \) and \( D \) are sufficient to estimate an unbiased treatment effect under an ignorable treatment assignment mechanism. Let \( p_1 \) and \( p_0 \) be the marginal probability of the event in treatment group 1 and 0 respectively. These can be estimated in an unbiased fashion by weighting the original populations according to the inverse of their propensity scores.

We let \( \tilde{Y}(1) \) denote the total number of potential events in treatment group 1 (all subjects for which \( X_i = 1 \)), and \( \tilde{Y}(0) \) the number of events in treatment group 0. By assuming a binomial likelihood for \( \tilde{Y}(1) \) and \( \tilde{Y}(0) \), and selecting flat priors \( p_0 \) and \( p_1 \sim \text{Beta}(1,1) \), we can take advantage of conjugacy to get closed form solutions for the posterior distributions for \( p_0 \) and \( p_1 \). The priors are augmented by weighted counts of successes and failures in each treatment group to generate beta posteriors, \( p_g \mid D, \pi \sim \text{Beta}(a^*_g, b^*_g) \) for \( g = 0,1 \) where

\[
\begin{align*}
a^*_g &= 1 + \gamma_g \left( g \sum_{i=1}^{n} \frac{X_iY_i}{\pi_i} + (1 - g) \sum_{i=1}^{n} \frac{(1 - X_i)Y_i}{1 - \pi_i} \right), \\
b^*_g &= 1 + \gamma_g \left( g \sum_{i=1}^{n} \frac{X_i(1 - Y_i)}{\pi_i} + (1 - g) \sum_{i=1}^{n} \frac{(1 - X_i)(1 - Y_i)}{1 - \pi_i} \right)
\end{align*}
\]

and \( a^*_g \) (\( b^*_g \)) describes the weighted number of events (non-events) in treatment group \( g \). The quantity \( \gamma_g \) is used to normalize the weighted population to its original size and is defined as \( \gamma_g = g \sum_{i=1}^{n} \frac{(1 - X_i)(1 - Y_i)}{1 - \pi_i} + (1 - g) \sum_{i=1}^{n} \frac{(1 - X_i)(1 - Y_i)}{1 - \pi_i} \). The causal parameter, such as the average treatment effect, can be calculated as simple functions of draws from the respective posteriors of \( p_0 \) and \( p_1 \) - for example the average treatment effect defined by the posterior risk difference, \( \Delta = p_1 - p_0 \). Estimates can be obtained using Markov Chain Monte Carlo simulation and incorporate the uncertainty from both the propensity score and outcome models (see in Spertus et al \cite{31}).

3 Proposed Risk Adjustment Approaches

We now consider estimating propensity scores \( \pi_i = \mathbb{P}(X_i \mid B_i, C_i) \), where \( B_i \) is a vector of baseline confounders, for example demographic or clinical covariates from a clinical registry,
and \( C_i \) is a vector of 4-digit ICD-9 diagnosis codes. We use 4-digit rather than 5-digit codes for simplicity and without loss of generality. We define the propensity score model as

\[
\pi_i = \logit^{-1} \{ \beta_0 + \beta_B B_i + \beta_C C_i \}.
\]  

(1)

The intercept \( \beta_0 \) and coefficients associated with the baseline confounders \( \beta_B \) are standard logistic regression parameters. We set a non-informative prior for the intercept as \( \beta_0 \sim \mathcal{N}(0, 10) \). The 95% interval for \( \beta_0 \) is about \([-20, 20]\), which on the logit scale encompasses the vast majority of possible values for the marginal treatment rate. We assume that each component of \( \beta_B \) is distributed \( \text{a priori as } t_5(0, 2.5) \) and components are mutually independent. This prior is referred to as “weakly informative” in the Bayesian data analysis literature \[10\]. It places the majority of prior mass within \([-5, 5]\) though heavy tails allow larger estimates if the data warrants. Weakly informative priors help produce a well-defined posterior but otherwise do little shrinkage compared to a maximum likelihood estimate. Having specified priors for the intercept and baseline variables, we turn to various approaches to handling the claims variables \( C_i \) using regularization.

### 3.1 Regularization via Prior Distributions

#### 3.1.1 Weakly Informative Prior Distributions

A natural choice for the prior distribution of \( \beta_C \) is weakly informative such that \( \beta_C \sim t_5(0, 2.5) \). These priors could be applied to groups of codes at the 3-digit level, or to more granular 4-digit codes. The latter is likely to have higher variance but less bias, and we explore both possibilities in our application. However, alternatives may be required because it is likely that the majority of the ICD-9 codes are rare and have no relationship with the treatment assignment, so they will introduce considerable variance.

#### 3.1.2 Sparsity-Inducing Prior Distributions

An alternative \textit{sparsity-inducing} prior is the horseshoe prior where for the \( j \)th component of \( \beta_C \). We assume the associated coefficient is

\[ \beta_{Cj} \sim \mathcal{N}(0, \lambda_j^2 \tau^2); \text{ where } \lambda_j \sim \text{Cauchy}^+(0, 1) \text{ and } \tau \sim \text{Cauchy}^+(0, 1). \]  

(2)

The notation \( \text{Cauchy}^+(0, 1) \) denotes a half-Cauchy distribution defined over the positive real line. The horseshoe performs global shrinkage through \( \tau \) and local shrinkage through \( \lambda_j \), allowing signals to remain near their maximum likelihood estimates, while shrinking noise variables aggressively towards zero. It has shown impressive predictive properties in simulations and applied problems \[5\][19]. For easier sampling, we replace the priors for \( \lambda_j \) with \( t_{3.5}^+(0, 1) \) priors which users find have little effect on coefficient estimates in practice \[22\].

Under this specification, 50% of the prior mass on \( \beta_{Cj} \) is placed between -0.4 and 0.4, which is quite close to zero. However the 95% prior interval is about \([-11, 11]\), reflecting the fact that the horseshoe places a lot of prior weight near zero but also has “fat tails” in order to avoid shrinking signals.
3.1.3 Hierarchical Prior Distributions

As an alternative to standard weakly informative or sparsity-inducing priors, we propose a hierarchical prior on $\beta_C$ that takes advantage of the natural tree-like hierarchy in ICD-9 codes. Assuming $p$ elements for $C_i$, each element falls into one of $L$ 3-digit categories, where category $l$ is associated with $p_l \geq 1$ codes. The codes in category $l$ are denoted by $C_{il}$. The vector of coefficients for $C_i$ has a special structure. It is partitioned into $\beta_C = \{\beta_{C_1}, \beta_{C_2}, \ldots, \beta_{C_L}\}$ where $\beta_{C_l}$ is the vector of coefficients for the grouped variables $C_{il}$. To emphasize the grouped structure in the regression, we rewrite equation 1 as:

$$\pi_i = \logit^{-1}\{\beta_0 + \beta_B B_i + \sum_{l=1}^L \sum_{k=1}^{p_l} \beta_{C_{lk}} C_{ilk}\}. \quad (3)$$

Capitalizing on the hierarchical structure, we define a hierarchical prior distribution as

$$\beta_{C_l} \sim \text{Laplace}(\mu_l, .5) \text{ where } \mu_l \sim t_5(0, 2.5). \quad (4)$$

The $t_5(0, 2.5)$ priors on $\mu_l$ are again weakly informative. The 4-digit ICD-9 code coefficients are shrunk to the 3-digit (group) mean $\mu_l$ according to a Laplace distribution with a relatively tight scale. For example, a priori each 4-digit code has a probability of 0.86 of being within 1 of the group mean coefficient $\mu_l$. The choice of the Laplace distribution over the Normal distribution forces noisier 4-digit codes to be shrunk more aggressively to the 3-digit group mean, while fatter tails permits less shrinking of signals [10]. If there is only one 4-digit code in a 3-digit category, then we just assume $\beta_{C_l} = \mu_l$ and specify the weakly informative prior $\beta_{C_l} \sim t_5(0, 2.5)$.

To fix ideas, imagine the group of ICD-9 codes corresponds to the category "hypertensive heart disease" and that it has an overall effect of $\mu_l = 2$ (Figure 1). The 4-digit codes for this category include benign, unspecified, and malignant hypertensive heart disease. We might expect the benign condition to in fact have a small effect on stent selection such that $\beta_{C_{l1}} = 0$; the malignant form of the condition may have a larger effect than the group mean, so $\beta_{C_{l3}} = 3$; and the unspecified condition is more ambiguous and receives an estimate equal to the group mean, $\beta_{C_{l2}} = 2$. The fundamental advantage is that some information is shared among similar conditions, while different forms and severities are able to receive larger or smaller effect estimates as the data warrant.

3.2 Regularization via Comorbidity Scoring

Comorbidity indices are the results of a dimension reduction strategy that have the related advantages of low variance and parsimonious interpretation [11]. In typical usage, comorbidity indices compress hundreds or thousands of binary-valued ICD-9 codes into a smaller set of indicator variables or into a scalar summary, either of which can then be used for prediction or risk adjustment. Either way, the resulting claims data set is much smaller, with many of the original diagnosis data discarded or aggregated. The Elixhauser comorbidity index, for example, reduces all ICD-9 codes into 30 indicators, which can then be combined further into a scalar score using fixed severity weights [8][39]. The severity weights are obtained from a regression model and are available in programs distributed by the developer of the index. In our setting of causal inference, the association of the
Elixhauser comorbidities are nuisance parameters – their use is in meeting the ignorability of the treatment assignment assumption.

4 Application

We implement a variety of methods to model propensity scores for DES vs BMS using the Bayesian propensity score approach to compare inferences using a variety of regularization strategies. We utilize ICD-9 diagnosis codes that are present-on-admission for a cohort of adults undergoing coronary stenting in hospitals.

4.1 Claims Data and the Mass-DAC Registry

We used claims data from the Massachusetts Center for Health Information and Analysis (CHIA). The CHIA data included 15 free-response fields with Present On Admission (POA) ICD-9 codes. Because POA codes describe patients at hospital arrival, they are not impacted by treatment decisions. To set a lower bound on sparsity we considered all 4-digit POA ICD-9 for which 10 or more patients were coded as having the condition.

The claims data were supplemented with baseline demographic and clinical variables from the Mass-DAC registry, a state mandated database harvesting clinical information for all percutaneous coronary interventions (PCIs) performed in adults (age ≥ 18 years) in all
non-Federal Massachusetts’ hospitals annually. We identified 131 potential confounders in the Mass-DAC registry, including variables capturing demographic characteristics (e.g. age, race), pre-existing conditions (e.g. heart disease, diabetes), or procedure information (e.g. treated vessel, heart attack severity, hospital). The registry data are prospectively collected by hospital personnel who use the National Cardiovascular Data Registry instrument from the American College of Cardiology. A very small number of cells were missing data, and we imputed a single dataset for analysis.

We compared a number of outcomes between subjects implanted with a drug eluting stent (DES) and subjects implanted with bare metal stents (BMS) using Bayesian propensity score methods: 2-day mortality, 30-day mortality, 1-year mortality, and target vessel revascularization (TVR). Two-day and 30-day mortality are designed to serve as “falsifiability endpoints.” Because any difference in the stent performance would not manifest after such a short period, if we observe a difference on one of these outcomes it is almost certain it is due to residual confounding. We also generally expect no difference, or a very small difference, on 1-year mortality based on previous findings from RCTs and instrumental variable analyses [2][42]. We do expect a significant benefit to DES on TVR rates.

4.2 Applied Methods

In all methods, the Mass-DAC variables were included and their coefficients, $\beta_B$, were specified with weakly informative $t_{2.5}$ priors. Therefore only the modeling of claims data varied.

As baseline analyses, we fit two different Bayesian propensity score models to the log-odds of DES based on the Elixhauser index. We first grouped variables using the mapping of Quan et al from ICD-9 codes to the 30 original Elixhauser categories [8][23]. This mapping was accomplished in R using the package medicalrisk [17]. We then implemented a model with indicator variables for each category and used a weakly informative $t_{2.5}$ prior. We refer to this model as “Elixhauser indicator.” Second, we replaced the 30 binary indicators with a comorbidity score created by summing the product of each indicator with its fixed severity weight.[39] We refer to this model as the “Elixhauser continuous.”

We used the full claims set (defined as all 4-digit POA ICD-9 for which 10 or more patients were coded as having the condition) with weakly informative, sparsity-inducing, and hierarchical priors to estimate the propensity score. As standard approaches a logistic regression using 3-digit codes and $t_{2.5}$ priors and a logistic regression using 4-digit ICD-9 codes and $t_{2.5}$ priors were fit [18]. Using 3-digit rather than 4-digit codes is a typical dimension reduction technique. We have found $t_{2.5}$ priors to behave similarly to non-informative priors (analogous to maximum likelihood estimation) in this setting, but have slightly more posterior stability and easier sampling [10]. As an alternative, we used horseshoe priors for strong regularization on the full 4-digit ICD-9 code set. Finally, we place a hierarchical prior on the coefficients so that coefficients associated with 4-digit codes are shrunk together towards the coefficients of the 3-digit codes.

The propensity score models were fit using Stan for Hamiltonian Monte Carlo sampling from each posterior [4]. The convergence of the samplers was assessed by examining trace plots and by the Gelman-Rubin “r-hat” statistic, where a value less than 1.1 indicated sufficient convergence [10]. To assess the out-of-sample fit of different models, we calculated the leave-one-out information criterion (LOO-IC). LOO-IC is an approximation to leave-one-out cross validation that can be computed from a point-wise log-likelihood matrix.
using Pareto-smoothed importance sampling, a procedure implemented in R through the \texttt{loo} package \cite{King2019}. A lower LOO-IC indicates a model with better predictive accuracy.

To assess adjustment for measured confounders, we calculated how well the propensity scores balanced the observed covariates by calculating standardized differences, defined as the difference in means between the weighted BMS and DES populations over their pooled standard deviation. Because we have a different weighted population for each propensity score draw, this quantity is calculated once for each draw, giving a standardized difference distribution for each variable \cite{Petersen2019}. We assess the overall balance by determining the percentage of covariates with posterior mean standardized difference falling within (-10,10) \cite{Ivey2019}. Ideally, the 95% intervals for standardized differences will also be contained within (-10,10) for most variables.

We assessed how well each method met our falsifiability endpoints and their causal effect estimates for 2-day, 30-day, and 1-year TVR. We were particularly interested in the coverage and width of the 95% posterior intervals and whether they included zero.

### 4.3 Results

After eliminating diagnoses with fewer than 10 observed occurrences, we were left with a total of 334 4-digit ICD-9 confounders. There were fewer than 10 patients with individual codes in the “psychoses” Elixhauser category, so the Elixhauser indicator sets included 29 diagnoses rather than 30.

Grouping by the first 3-digits of the ICD-9 codes resulted in 69 groups containing at least 2 4-digit codes and 117 single 4-digit codes with no higher level group. The condition with the largest number of 4-digit codes in our sample was acute myocardial infarctions, i.e., heart attacks, and contained 10 4-digit codes. Thirty-seven of the 69 groups had only 2 4-digit codes. The complete list of all outcomes and confounders along with their prevalences (or means and standard deviations) appears in our appendix below.

Table \ref{table:propensity_models} displays characteristics of the propensity score models including number of claims variables used, LOO-IC, and summaries of covariate balance. In terms of fit, Horsehoe regularization achieved the lowest LOO-IC score, indicating it had the best predictive accuracy. The Elixhauser indicator also had good LOO-IC, achieved through a considerable reduction in the dimension of the regression, though further reduction into a score degraded the quality of the fit considerably as indicated by a higher LOO-IC for Elixhauser continuous. The hierarchical prior provided lower LOO-IC than either the 4-digit or 3-digit $t_5(0, 2.5)$ prior approaches.

In terms of balance, the Elixhauser methods and 3-digit codes with weakly informative priors failed to bring posterior mean standardized difference for all variables to within (-10,10). Regressions using the full claims set with weakly informative or hierarchical priors had many confounders for which the upper (or lower) 95% posterior interval limit is beyond 10% (or smaller than -10%) after weighting. In contrast, the confounder balance when using a horseshoe prior results in only 6 variables with standardized difference posterior intervals outside the 10% window.

Figure \ref{fig:regularization_behavior} illustrates the behavior of the different approaches to regularization when using the full claims set for two groups (e.g., two 3-digit diagnosis codes): cardiac dysrhythmias and drug/alcohol abuse. For cardiac dysrhythmias, the group mean for the hierarchical prior coefficients is very close to zero because the associations of the individual codes have opposite signs on the logit of the probability of receiving a DES. The hierarchical prior coef-
Table 1: **Characteristics of propensity score models.** Number of claims derived variables used in model, leave-one-out information criterion, and summaries of covariate balance for unadjusted analysis and each propensity score analysis. Column “Means $\notin (-10,10)$” tabulates the number of variables with posterior mean standardized differences outside the interval -10 to 10, which is typically considered good balance. Column “CIs $\notin (-10,10)$” tabulates number of variables where some part of the 95% posterior interval lies outside (-10,10). LOO-IC = leave-one-out information criterion; SE = standard error. *=: This is a model with an intercept only for the purpose of calculating LOO-IC and standardized differences reflect covariates in unweighted population.

| Approach                        | # Claims Variables | LOO-IC (SE) | Means $\notin (-10,10)$ | CIs $\notin (-10,10)$ |
|--------------------------------|--------------------|-------------|-------------------------|----------------------|
| Unadjusted                     |                    |             |                         |                      |
| Elixhauser Methods             |                    |             |                         |                      |
| $t_5(0,2.5)$ Prior Indicator Variables | 29              | 9537 (94)  | 3                       | 11                   |
| $t_5(0,2.5)$ Prior Continuous Score | 1               | 9640 (92)  | 4                       | 12                   |
| Full Claims Set                |                    |             |                         |                      |
| $t_5(0,2.5)$ Prior 3-Digits     | 186               | 9662 (100) | 1                       | 59                   |
| $t_5(0,2.5)$ Prior 4-Digits     | 334               | 9741 (107) | 0                       | 150                  |
| Horseshoe Prior 4-Digits        | 334               | 9468 (88)  | 0                       | 6                    |
| Hierarchical Prior 3 and 4-Digits | 334             | 9639 (104) | 0                       | 74                   |

The Elixhauser indicator, Elixhauser continuous, and horseshoe approaches had the lowest uncertainty with posterior interval widths of 1.9, 1.8, and 2.0 respectively. The weakly informative 3-digit approach had an average width of 2.3, while the weakly informative 4-digit approach had an average width of 3.0. The hierarchical prior approach had an average credible interval width of 2.5, about 20% tighter than the 4-digit approach. The weakly informative and hierarchical specifications thus incur more variance than the smaller or more regularized Elixhauser and horseshoe approaches, which is ultimately reflected in

coefficients are slightly pooled towards their group mean, while coefficients under the horseshoe are pulled towards zero, in some cases quite sharply. The weakly informative coefficients are universally larger in magnitude than the others. A similar pattern emerges for drug/alcohol abuse coefficients. The group overall is associated with a lower probability of receiving a DES, with some granular codes exhibiting much smaller or larger effects. For both coefficient groups, a weakly informative prior using a single indicator for the 3-digit code category fails to capture the nuances apparent at lower levels.

Table 2 presents posterior summaries of causal effects in the Mass-DAC data under each method. The unadjusted difference in means between the two stent groups exhibits considerable confounding of both 30-day and 1-year mortality with large risk reductions for DES of 3.4% and 6.9% respectively. The Elixhauser approaches were the only models unable to adjust away the difference in 2-day mortality, though margins were slim. For 30-day mortality, only the full claims sets with $t_5(0,2.5)$ or hierarchical priors estimated no significant effect. In contrast, aggressively regularizing using horseshoe priors yielded a 30-day mortality benefit to DES that is indicative of residual confounding. On one-year mortality, the model utilizing the hierarchical prior distribution for the ICD-9 coefficients achieved a considerable reduction from the large unadjusted difference, and improved over the comparably large effect estimates generated from the Elixhauser and horseshoe prior approaches. However no method included zero in its 95% credible interval, conflicting with findings from randomized trials. As expected, all methods produced a larger benefit to TVR than indicated by the unadjusted difference.

The Elixhauser indicator, Elixhauser continuous, and horseshoe approaches had the lowest uncertainty with posterior interval widths of 1.9, 1.8, and 2.0 respectively. The weakly informative 3-digit approach had an average width of 2.3, while the weakly informative 4-digit approach had an average width of 3.0. The hierarchical prior approach had an average credible interval width of 2.5, about 20% tighter than the 4-digit approach. The weakly informative and hierarchical specifications thus incur more variance than the smaller or more regularized Elixhauser and horseshoe approaches, which is ultimately reflected in
Figure 2: **Selected posterior mean coefficients.** Values on log-odds scale for 4-digit ICD-9 coefficients with weakly informative priors (orange circle), horseshoe (red triangle), and hierarchical (blue diamond) specifications. Group mean defined by 3-digit code for hierarchical (blue dashed line) and 3-digit weakly informative coefficient estimate (green dashed line) are also shown. Left panel shows coefficients for code group 427, cardiac dysrhythmias; right panel shows coefficients for code group 305, drug or alcohol abuse.

The estimated distributions of causal parameters.

| Approach                      | Mortality                          |
|-------------------------------|------------------------------------|
|                               | 2-Day | 30-Day | 1-Year | 1-Year TVR |
| Unadjusted                    | -1.0 (1.4, -0.6)                  | -3.4 (-4.1, -2.7) | -6.9 (-8.0, -5.7) | -2.4 (-3.7, -1.3) |
| Elixhauser Methods            |       |        |        |            |
| $t_5(0, 2.5)$ Prior Indicator Variables | -0.4 (-0.7, -0.1)                  | -1.2 (-2.0, -0.5) | -3.1 (-4.3, -1.8) | -4.3 (-5.7, -2.9) |
| $t_5(0, 2.5)$ Prior Continuous Score | -0.4 (-0.7, -0.1)                  | -1.3 (-2.0, -0.6) | -3.3 (-4.5, -2.1) | -4.5 (-5.9, -3.0) |
| Full Claims Set               |       |        |        |            |
| $t_5(0, 2.5)$ Prior 3-Digits   | -0.3 (-0.7, 0.1)                   | -1.0 (-2.0, 0.2) | -2.7 (-4.1, -1.1) | -4.4 (-6.1, -2.8) |
| $t_5(0, 2.5)$ Prior 4-Digits   | -0.3 (-0.7, 0.0)                   | -0.7 (-2.0, 1.2) | -2.5 (-4.5, -0.3) | -3.6 (-5.5, -1.8) |
| Horseshoe Prior 4-Digits      | -0.3 (-0.7, 0.0)                   | -1.1 (-1.9, -0.3) | -3.0 (-4.3, -1.7) | -4.1 (-5.6, -2.6) |
| Hierarchical Prior 3 and 4-Digits | -0.3 (-0.7, 0.1)                  | -0.9 (-2.0, 0.5) | -2.6 (-4.2, -0.8) | -3.8 (-5.5, -2.0) |

Table 2: **Casual effect estimates for selected outcomes.** Risk differences for various methods applied to the coronary stent study. Risk differences (95% Credible Intervals) estimated from the Mass-DAC data show benefit of DES compared to BMS in terms of reduced percent chance of given outcome ($\Delta = p_{DES} - p_{BMS}$). TVR = target vessel revascularization.

5 Discussion

In this paper, we proposed the use of regularizing prior distributions rather than comorbidity indices for data reduction strategies in high-dimensional causal inference problems in health care. We found that Bayesian priors improved inferences over comorbidity indices both in terms of out-of-sample fit or the ability to meet falsifiability endpoints. These prior distributions are able to work with a large number of claims variables while controlling variance.
To compare multiple methods with and without hierarchical modeling of ICD-9 codes, we estimated causal effects of DES compared to BMS on a number of outcomes. Specifically, we were looking for our methods to adjust away any difference in 30-day mortality and, ideally, 1-year mortality as well. RCTs have consistently shown no survival benefit to DES, but this has proved to be very difficult to replicate in observational analyses [2][12][34][14]. By using claims data and either a hierarchical ICD-9 or flat weakly informative propensity score model, we succeeded in removing the specious benefit shown to 2-day and 30-day mortality, which were very likely to be caused by residual confounding. In contrast, models using Elixhauser comorbidity indicators, a continuous Elixhauser score, or sparsity-inducing horseshoe priors on the full claims set did not manage to adjust away confounding on 30-day mortality. The fact that these dimension reducing approaches generated the best out-of-sample fit in terms of LOO-IC indicates the importance of using additional checks like falsifiability endpoints in causal inference problems, rather than relying solely on scores of predictive ability.

Unfortunately no method succeeded in removing all confounding of 1-year mortality, but the hierarchical and weakly informative prior distributions had comparably more success, estimating small effect sizes and 95% credible intervals close to 0. Residual confounding has plagued past observational studies of DES and BMS and it appears that, while claims data are able to account for important confounders not found in registries, we are still lacking the necessary variables to completely remove confounding [14]. Likely unmeasured confounders include socio-economic variables, which past studies have shown affect adherence to dual-antiplatelet therapy and ultimately a doctor’s decision to use a DES versus a BMS in usual care settings [33].

While generating similar posterior means, the hierarchical prior approach had considerably smaller variance than the unregularized approach, with about 20% tighter credible intervals on average. All methods showed, as expected, a considerable benefit to DES on the TVR outcome.

We restricted our attention to Bayesian propensity scores throughout, only changing the variable selection or regularization strategy for comparability. However, hierarchical ICD-9 code modeling is not limited to Bayesian propensity scores and could easily be adapted to Bayesian outcome modeling or frequentist techniques. Whatever the specifics, our work demonstrates that investigators performing causal inference in non-randomized settings could benefit by ignoring comorbidity indices and using more claims variables in their analyses. Comorbidity indices will continue to have utility for risk stratification and parsimonious assessments in clinical practice, but as we have shown, there are better alternatives for confounder adjustment.

6 Acknowledgments

Mr Spertus and Dr Normand were supported by a grant from the US Food and Drug Administration (U01FD004493). Mr Spertus, Dr Normand, and Dr Adhikari were supported by a grant from the National Institute of General Medical Sciences (1R01GM111339).

We are indebted to the Massachusetts Department of Public Health (MDPH) and the Massachusetts Center for Health Information and Analysis Case Mix Databases (CHIA) for the use of their data.
References

[1] Peter C Austin and Elizabeth A Stuart. Moving towards best practice when using inverse probability of treatment weighting (iptw) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, 34:3661–3679, 2015.

[2] Kaare H. Bonaa, Jan Mannsverk, Rune Wiseth, Lars Aaberge, Yngvar Myreng, Ottar Nygard, Dennis W. Nilsen, Nils-Einar Klow, Michael Uchto, Thor Trovik, Bjorn Bendz, Sindre Stavnes, Reidar Bjornerheim, Alf-Inge Larsen, Morten Slette, Terje Steigen, Ole J. Jakobsen, Oyvind Bleie, Eigr Fossum, Tove A. Hanssen, Oystein Dahl-Eriksen, Inger Njolstad, Knut Rasmussen, Tom Wilsgaard, and Jan E. Nordrehaug. Drug-eluting or bare-metal stents for coronary artery disease. *The New England Journal of Medicine*, 2016.

[3] M Alan Brookhart, Sebastian Schneeweiss, Kenneth J Rothman, Robert J Glynn, Jerry Avorn, and Til Sturmer. Variable selection for propensity score models. *Practice of Epidemiology*, 163:1149–1156, 2006.

[4] Bob Carpenter, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: a probabilistic programming language. *Journal of Statistical Software*, In Press, 2016.

[5] Carlos M Carvahlo, Nicholas G Polson, and James G Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97:465–480, 2010.

[6] Massachusetts Data Analysis Center. Cardiac study annual reports. 2017.

[7] Mary E Charlson, Peter Pompei, Kathy L Ales, and C Ronald MacKenzie. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases*, 40:373–383, 1987.

[8] Anne Elixhauser, Claudia Steiner, D Robert Harris, and Rosanna M Coffey. Comorbidity measures for use with administrative data. *Medical Care*, 36:8–27, 1998.

[9] Centers for Disease Control and Prevention. International classification of diseases, ninth revision, clinical modification (icd-9-cm). [https://www.cdc.gov/nchs/icd/icd9cm.htm](https://www.cdc.gov/nchs/icd/icd9cm.htm).

[10] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian Data Analysis: Third Edition*. CRC Presss, 2014.

[11] Elizabeth A Gilbert, Robert T Krafty, Richard J Bleicher, and Brian L Egleston. On the use of summary comorbidity measures for prognosis and survival treatment effect estimation. *Health Services Outcomes Research Methods*, https://doi.org/10.1007/s10742-017-0171-2, 2017.

[12] Lisa Iezzoni. Assessing quality using administrative data. *Annals of Internal Medicine*, 127:666–674, 1997.
[13] Mark S Levenson and Lilly Q Yue. Regulatory issues of propensity score methodology application to drug and device safety studies. *Journal of Biopharmaceutical Statistics*, 23:110–121, 2012.

[14] Laura Mauri, Treacy S Silbaugh, Robert E Wolf, Katya Zelevinsky, Ann Lovett, Zheng Zhou, Frederic S Resnic, and Sharon-Lise T Normand. Long-term clinical outcomes after drug-eluting and bare-metal stenting in Massachusetts. *Circulation*, 118:1817–1827, 2008.

[15] Daniel F McCaffrey, Greg Ridgeway, and Andrew R Morral. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods*, 9:403–425, 2004.

[16] Lawrence C McCandless, Paul Gustafson, and Peter C Austin. Bayesian propensity score analysis for observational data. *Statistics in Medicine*, 28:94–112, 2009.

[17] Patrick McCormick and Thomas Joseph. Package ‘medicalrisk’. 2016.

[18] Sharon-Lise T Normand, Carl N Morris, Karen S Fung, Barbara J McNeil, and Arnold M Epstein. Development and validation of a claims based index for adjusting for risk of mortality: the case of acute myocardial infarction. *Journal of Clinical Epidemiology*, 48:229–243, 1995.

[19] Tomi Peltola, Aki S Havulinna, Veikko Salomaa, and Aki Vehtari. Hierarchical bayesian survival analysis and projective covariate selection in cardiovascular event risk prediction. *Proceedings of the 11th Annual UAI Conference on Bayesian Modeling Applications Workshop*, 1218:79–88.

[20] Adler Perotte, Rimma Pivovarov, Karthik Natarajan, Nicole Weiskopf, Frank Wood, and Noemie Elhadad. Diagnosis code assignment: models and evaluation metrics. *Journal of the American Medical Informatics Association*, 21:231–237, 2014.

[21] Maya L Petersen, Kristin E Porter, Susan Gruber, Yue Wang, and Mark J van der Laan. Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research*, 21:31–54, 2010.

[22] Juho Piironen and Aki Vehtari. Projection predictive variable selection using stan + r. 2015.

[23] Hude Quan, Vijaya Sundarajan, Patricia Halfon, Andrew Fong, Bernard Burnand, Jean-Christophe Luthi, L Duncan Saunders, Cynthia A Beck, Thomas E Feasby, and William A Ghali. Coding algorithms for defining comorbidities in icd-9 and icd-10 administrative data. *Medical Care*, 43:1130–1139, 2005.

[24] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55, 1983.

[25] Donald B Rubin. For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, 2:808–840, 2008.

[26] Olli Saarela. On bayesian estimation of marginal structural models. *Biometrics*, 71:279–301, 2015.
[27] Mary Vaughan Sarrazin and Gary E Rosenthal. Finding pure and simple truths with administrative data. *Journal of the American Medical Association*, 307:1433–1435, 2012.

[28] Sebastian Schneeweiss, Jeremy A Rassen, Robert J Glynn, Jerry Avorn, Helen Mogun, and M Alan Brookhart. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*, 20:512–522, 2009.

[29] Mansour TA Sharabiani, Paul Aylin, and Alex Bottle. Systematic review of comorbidity indices for administrative data. *Medical Care*, 50:1109–1118, 2012.

[30] Anima Singh. Exploiting hierarchical and temporal information in building predictive models from ehr data. *PhD Dissertation for Massachusetts Institute of Technology*, 2015.

[31] Jacob V Spertus and Sharon-Lise T Normand. Bayesian propensity scores for high-dimensional causal inference: A comparison of drug-eluting to bare-metal coronary stents. *Biometrics*, page In submission, 2018.

[32] Jacob V Spertus, Sharon-Lise T Normand, Robert Wolf, Matt Cioffi, Ann Lovett, and Sherri Rose. Assessing hospital performance following percutaneous coronary intervention using big data. *Circulation: Cardiovascular Quality and Outcomes*, In submission, 2016.

[33] John A Spertus, Richard Kettelkamp, Clifton Vance, Carole Decker, Philip G Jones, John S Rumsfeld, John C Messenger, Sanjaya Khanal, Eric D Peterson, Richard G Bach, Harlan M Krumholz, and David J Cohen. Prevalence, predictors, and outcomes of premature discontinuation of thienopyridine therapy after drug-eluting stent placement: Results from the premier registry. *Circulation*, 113:2803–2809.

[34] Christoph Stettler, Simon Wandel, Sabin Allemann, Adnan Kastrati, Marie Claude Morice, Albert Schomig, Matthias E Pfisterer, Gregg W Stone, Martin B Leon, Jose Suarez de Leon, Jean-Jacques Goy, Seung-Jung Park, Manel Sabate, Maarten J Suttrop, Henning Kelbaek, Christian Spaulding, Maurizio Menichelli, Paul Vermeersch, Maurits T Dirkens, Pavel Cervinka, Anna Sonia Petronio, Alain J Nordmann, Peter Diem, Bernhard Meier, Marcel Zwahlen, Stephan Reichenbach, Sven Trelle, Stephan Windecker, and Peter Juni. Outcomes associated with drug-eluting and bare-metal stents: a collaborative network meta-analysis. *The Lancet*, 370:937–948, 2007.

[35] Elizabeth A Stuart, Eva DuGoff, Michael Abrams, David Salkever, and Donald Steinwachs. Estimating causal effects in observational studies using electronic health data: Challenges and (some) solutions. *EGEMS*, 1, 2013.

[36] Zeeshan Syed, Ilan Rubinfeld, Joe H Patton, Jennifer Ritz, Jack Jordan, Andrea Doud, and Vic Velanovich. Using procedural codes to supplement risk adjustment: a nonparametric learning approach. *Journal of the American College of Surgeons*, 212:1086–1093, 2011.

[37] Jack V Tu, James Bowen, Maria Chiu, Dennis T Ko, Peter C Austin, Yaolhua He, Robert Hopkins, Jean-Eric Tarride, Gord Blackhouse, Charles Lazzam, Eric A Cohen,
and Ron Goeree. Effectiveness and safety of drug-eluting stents in Ontario. *The New England Journal of Medicine*, 357:1393–1402, 2007.

[38] Mark J. van der Laan and Sherri Rose. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer, 2011.

[39] Carl van Walraven, Peter C Austin, Alison Jennings, Hude Quan, and Alan J Forster. A modification of the elixhauser comorbidity measures into a point system for hospital death using administrative data. *Medical Care*, 47:626–633, 2009.

[40] Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, 27:1413–1432, 2017.

[41] Aki Vehtari, Andrew Gelman, Jonah Gabry, Juho Piironen, and Ben Goodrich. Package ‘loo’. 2017.

[42] Lakshmi Venkitachalam, Yang Lei, Elizabeth A Magnuson, Paul S Chan, Joshua M Stolker, Kevin F Kennedy, Neal S Kleiman, and David J Cohen. Survival benefit with drug-eluting stents in observational studies: Fact or artifact? *Circulation*, 4:587–594, 2011.

### A Covariate Tables

#### A.1 Registry Covariates

Table 3: Prevalence or mean (SD) of outcomes in the Mass-DAC registry. “Cases per Site” gives min, mean (SD), and max of numbers treated at the 25 hospitals, overall and in each treatment group. The 31 coronary artery segment indicators are not shown. HMO = health maintenance organization; CAD = coronary artery disease; CVD = cardiovascular disease; PAD = peripheral artery disease; CABG = coronary artery bypass graft; STEMI = ST-elevated myocardial infarction; LVSD = left ventricular systolic dysfunction; NYHA = New York heart association; LAD = left anterior descending (artery); RCA = right coronary artery.

|                      | Overall | BMS   | DES  |
|----------------------|---------|-------|------|
| Number Treated (n)   | 8718    | 3081  | 5637 |
| **Registry Outcomes**|         |       |      |
| % 2 Day Mortality    | 0.50    | 1.14  | 0.16 |
| % 30 Day Mortality   | 2.01    | 4.19  | 0.82 |
| % 1 Year TV Revascularization | 7.38 | 8.96  | 6.51 |
| % 1 Year Mortality   | 5.75    | 10.19 | 3.32 |
| **Registry Confounders**|       |       |      |
| % Male               | 69.39   | 68.16 | 70.05|
| Mean (SD) Age in Years | 64.66 (12.5) | 66.37 (11.7) | 63.72 (13.5) |
| Mean (SD) Height in cm | 170.67 (10.5) | 170.41 (10.6) | 170.81 (10.4) |
| Mean (SD) Weight in kg | 86.48 (20.3) | 84.89 (19.9) | 87.34 (20.8) |
| **Race**             |         |       |      |


| % Caucasian | 91.26 | 90.72 | 91.56 |
| % Black     | 3.79  | 4.87  | 3.19  |
| % Asian     | 2.31  | 2.43  | 2.24  |
| % Native American | < 0.1 | < 0.3 | < 0.2 |
| % Native Pacific | < 0.1 | < 0.3 | < 0.2 |
| % Hispanic or Latino | 4.55 | 4.8 | 4.42 |

**Payor**

| % Government | 51.57 | 59.17 | 47.42 |
| % None       | 2.56  | 4.12  | 1.7   |
| % Non-US     | < 0.1 | < 0.3 | < 0.2 |
| % Private Commercial or HMO | 45.82 | 36.68 | 50.82 |

| % Smoker     | 74.68 | 70.85 | 76.78 |
| % Hypertension | 78.3 | 76.6  | 79.23 |
| % Dyslipidemia | 81.82 | 77.96 | 83.93 |
| % Diabetes   | 32.16 | 30.38 | 33.14 |
| % Family History CAD | 28.96 | 24.08 | 31.63 |
| % Chronic Lung Disease | 14.21 | 16.42 | 13   |
| % Current Dialysis | 1.82 | 2.21  | 1.61  |
| % Prior CVD  | 10.23 | 11.33 | 9.63  |
| % Prior PAD  | 11.96 | 12.76 | 11.53 |
| % Prior Myocardial Infarction | 26.67 | 25.84 | 27.12 |
| % Prior Heart Failure | 10.61 | 12.46 | 9.6   |
| % Prior Valve Surgery | 1.56 | 2.34  | 1.14  |
| % Prior PCI  | 27.56 | 19.77 | 31.83 |
| % Prior CABG | 12.3  | 11.46 | 12.76 |
| % Prior Cardiogenic Shock | 1.9 | 3.89 | 0.82 |
| % Prior Cardiac Arrest | 2.29 | 4.15  | 1.28  |

**CAD Presentation**

| % No Angina | 5.32 | 5.78 | 5.07 |
| % Symptom Unlikely to be Ischemic | 1.00 | 1.17 | 0.90 |
| % Stable Angina | 11.88 | 5.81 | 15.2 |
| % Unstable Angina | 30.19 | 23.4 | 33.9 |
| % Non-STEMI | 27.23 | 28.17 | 26.72 |
| % STEMI   | 24.37 | 35.67 | 18.2 |
| % Thrombolytic Therapy | 0.8 | 1.23 | 0.57 |
| % Cardiomyopathy or LVSD | 9.36 | 11.1 | 8.41 |

**Anginal Canadian Classification**

| % 0      | 9.64 | 11.65 | 8.53 |
| % I      | 2.4  | 1.59  | 2.84 |
| % II     | 12.2 | 7.66  | 14.69 |
| % III    | 28.83 | 24.15 | 31.38 |
| % IV     | 46.94 | 54.95 | 42.56 |

**Anti-Anginal Medications**

| % Beta Blockers | 57.79 | 54.56 | 59.55 |
| % Calcium Channel Blockers | 11.44 | 11.39 | 11.46 |
| % Long Acting Nitrates | 11.49 | 9.41 | 12.63 |
| NYHA Class          | % 0  | % I  | % II | % III | % IV | % Compassionate Use | % Cardiogenic Shock | % Mechanical Ventricular Support | % Ejection Fraction < 30% |
|---------------------|------|------|------|-------|------|---------------------|---------------------|-------------------------------|--------------------------|
|                     | 87.57 | 0.91 | 3.06 | 4.16  | 4.3  | 0.91                | 1.85                | 0.58                          | 2.88                     |
|                     | 83.51 | 0.78 | 3.67 | 5.39  | 6.65 | 2.01                | 3.8                 | 1.14                          | 3.7                      |
|                     | 89.78 | 0.98 | 2.73 | 3.49  | 3.02 | 0.3                 | 0.78                | 0.28                          | 2.43                     |

| Coronary Anatomy    | % Left Dominant | % Right Dominant | % Left Dominant | % Left Main Disease | Mean Left Main Stenosis (SD) | Mean Proximal LAD Stenosis (SD) | Mean Mid-Distal LAD Stenosis (SD) | Mean Circumflex Stenosis (SD) | Mean RCA Stenosis (SD) |
|---------------------|-----------------|------------------|-----------------|---------------------|-----------------------------|---------------------------------|-------------------------|-------------------------|------------------------|
|                     | 8.30            | 86.19            | 5.51            | 5.93                | 7.98 (19.8)                 | 36 (39.7)                      | 48.2 (39.7)             | 47.95 (40.8)            | 60.66 (39.9)           |
|                     | 8.08            | 85.88            | 6.04            | 6.52                | 8.74 (19.3)                 | 35.1 (39.7)                    | 47.64 (39.7)            | 47.94 (40.7)            | 65.25 (40)             |
|                     | 8.42            | 86.36            | 5.22            | 5.61                | 7.57 (20.7)                 | 36.5 (39.5)                    | 48.51 (39.6)            | 47.96 (40.8)            | 58.15 (39.2)           |

| Status              | % Urgent        | % Emergent       | % Other         | Mean (SD) Thrombectomies | Mean (SD) Total Stents Used | Mean (SD) Lesions | Mean (SD) Lesion Length | % In-stent Restenosis | % Chronic Total Occlusion |
|---------------------|-----------------|------------------|-----------------|--------------------------|---------------------------|-------------------|------------------------|-------------------------|--------------------------|
|                     | 52.49           | 26.69            | 20.82           | 0.12 (0.3)                | 1.46 (0.8)                | 1.3 (0.6)         | 18.08 (10)             | 6.96                    | 1.73                    |
|                     | 48.78           | 38.30            | 25.21           | 0.16 (0.3)                | 1.41 (0.8)                | 1.24 (0.6)        | 17 (10.4)              | 2.34                    | 1.3                     |
|                     | 54.51           | 20.35            | 25.14           | 0.09 (0.4)                | 1.5 (0.8)                 | 1.33 (0.5)        | 18.68 (9.3)            | 9.49                    | 1.97                    |

| PCI Indication      | % Immediate PCI for STEMI | % STEMI (Unstable, > 12 hours) | % STEMI (Stable, > 12 hours) | % STEMI (Stable, thrombolitics) | % STEMI (Rescue, failed thrombolitics) | % High risk Non-STEMI | % Staged | % Other | Mean (SD) Thrombectomies | Mean (SD) Total Stents Used | Mean (SD) Lesions | Mean (SD) Lesion Length | % In-stent Restenosis | % Chronic Total Occlusion |
|---------------------|---------------------------|-------------------------------|------------------------------|---------------------------------|----------------------------------------|----------------------|----------|---------|--------------------------|--------------------------|-------------------|----------------------|-------------------------|--------------------------|
|                     | 21.80                     | 1.67                          | 0.57                         | 0.44                            | 0.46                                   | 48.99                | 0.86     | 25.21   | 0.12 (0.3)                | 1.46 (0.8)                | 1.3 (0.6)         | 18.08 (10)           | 6.96                    | 1.73                    |
|                     | 31.39                     | 2.66                          | 0.94                         | 0.58                            | 0.75                                   | 45.86                | 0.45     | 20.82   | 0.16 (0.3)                | 1.41 (0.8)                | 1.24 (0.6)        | 17 (10.4)            | 2.34                    | 1.3                     |
|                     | 16.55                     | 1.14                          | 0.37                         | 0.35                            | 0.3                                    | 50.7                 | 1.08     | 29.5    | 0.09 (0.4)                | 1.5 (0.8)                 | 1.33 (0.5)        | 18.68 (9.3)          | 9.49                    | 1.97                    |

| Cases per Site (25 Total) | Min # Treated | Mean # Treated (SD) | Max # Treated |
|---------------------------|--------------|--------------------|--------------|
|                           | 47           | 349 (269)          | 896          |
|                           | 12           | 123 (99)           | 352          |
|                           | 14           | 225 (187)          | 605          |
Table 4: Prevalence of present on admission ICD-9 diagnosis codes used in analysis. As with the analysis using a hierarchical prior, codes are grouped together if there were two or more 4-digit codes sharing a 3-digit grouping. w/o = without.

| Diagnosis Code | Description (Present on Admission) | Prevalence |
|----------------|------------------------------------|------------|
| 041            | Bacterial infection of unspecified site | 0.5        |
| 0411           | Staphylococcus infection, unspecified | 0.2        |
| 0414           | Shiga toxin-producing E. coli (STEC) | 0.3        |
| 070            | Viral hepatitis                      | 1.1        |
| 0703           | Viral hepatitis B w/o mention of hepatic coma | 0.2        |
| 0705           | Acute hepatitis C w/o mention of hepatic coma | 0.4        |
| 0707           | Unspecified viral hepatitis C w/o hepatic coma | 0.5        |
| 242            | Thyrotoxicosis with or w/o goiter    | 0.3        |
| 2420           | Toxic diffuse goiter w/o mention of thyrotoxic crisis or storm | 0.1        |
| 2429           | Thyrotoxicosis w/o mention of goiter, thyrotoxic crisis, or storm | 0.2        |
| 244            | Acquired hypothyroidism              | 7.4        |
| 2440           | Postsurgical hypothyroidism          | 0.2        |
| 2448           | Other specified acquired hypothyroidism | 0.2        |
| 2449           | Unspecified acquired hypothyroidism  | 7          |
| 250            | Diabetes mellitus                    | 30.8       |
| 2500           | Diabetes mellitus w/o mention of complication | 27         |
| 2501           | Diabetes with ketoacidosis           | 0.2        |
| 2504           | Diabetes with renal manifestations   | 1.5        |
| 2505           | Diabetes with ophthalmic manifestations | 1.1        |
| 2506           | Diabetes with neurological manifestations | 1.9        |
| 2507           | Diabetes with peripheral circulatory disorders | 0.2        |
| 2508           | Diabetes with other specified manifestations | 0.3        |
| 272            | Disorders of lipoid metabolism       | 69.9       |
| 2720           | Pure hypercholesterolemia            | 13.7       |
| 2721           | Pure hyperglyceridemia               | 1          |
| 2722           | Mixed hyperlipidemia                 | 0.1        |
| 2724           | Other and unspecified hyperlipidemia | 59.4       |
| 2729           | Unspecified disorder of lipoid metabolism | 0.1        |
| 274            | Gout                                | 3.5        |
| 2740           | Gouty arthropathy, unspecified       | 0.3        |
| 2749           | Gout, unspecified                    | 3.2        |
| 275            | Disorders of mineral metabolism      | 0.9        |
| 2750           | Hereditary hemochromatosis           | 0.1        |
| 2752           | Disorders of magnesium metabolism    | 0.3        |
| 2753           | Disorders of phosphorus metabolism   | 0.2        |
| 2754           | Unspecified disorder of calcium metabolism | 0.3        |
| 276            | Disorders of fluid, electrolyte, and acid-base balance | 4.5        |
| 2760           | Hyposmolality and/or hyponatremia    | 0.1        |
| 2761           | Hyposmolality and/or hyponatremia    | 1.2        |
| 2762           | Acidosis                            | 0.7        |
| 2764           | Mixed acid-base balance disorder     | 0.1        |
| 2765           | Volume depletion, unspecified        | 0.7        |
2766 Transfusion associated circulatory overload 0.1
2767 Hyperpotassemia 0.7
2768 Hypopotassemia 1.1
280 Iron deficiency anemias 1.4
2800 Iron deficiency anemia secondary to blood loss (chronic) 0.2
2809 Iron deficiency anemia, unspecified 1.2
281 Other deficiency anemias 0.3
2810 Pernicious anemia 0.1
2819 Unspecified deficiency anemia 0.1
285 Other and unspecified anemias 5.5
2851 Acute posthemorrhagic anemia 0.1
2852 Anemia in chronic kidney disease 1.5
2858 Other specified anemias 0.1
2859 Anemia, unspecified 3.8
287 Purpura and other hemorrhagic conditions 1.3
2874 Posttransfusion purpura 0.1
2875 Thrombocytopenia, unspecified 1.2
294 Other organic psychotic conditions (chronic) 1
2941 Dementia w/o behavioral disturbance 0.2
2948 Other persistent mental disorders 0.8
296 Affective psychoses 1.4
2962 Major depressive affective disorder, single episode 0.2
2963 Major depressive affective disorder, recurrent episode 0.1
2968 Bipolar disorder, unspecified 0.9
2969 Unspecified episodic mood disorder 0.1
300 Neurotic disorders 6.8
3000 Anxiety state, unspecified 4.6
3004 Dysthymic disorder 2.2
305 Nondependent abuse of drugs 21.9
3050 Alcohol abuse, unspecified 2.2
3051 Tobacco use disorder 20.4
3052 Cannabis abuse, unspecified 0.9
3055 Opioid use disorder, unspecified 0.2
3056 Cocaine abuse, unspecified 0.7
3059 Other, mixed, or unspecified drug abuse, unspecified 0.2
333 Extrapyramidal disease and abnormal movement disorders 0.6
3331 Essential and other specified forms of tremor 0.2
3339 Unspecified extrapyramidal disease and abnormal movement 0.4
338 Pain, not elsewhere classified 2
3382 Chronic pain due to trauma 1.8
3384 Chronic pain syndrome 0.2
348 Other conditions of brain 0.5
3481 Anoxic brain damage 0.4
3483 Encephalopathy, unspecified 0.1
355 Mononeuritis of lower limb 0.4
3558 Mononeuritis of lower limb, unspecified 0.1
| Code  | Description                                              | Count |
|-------|----------------------------------------------------------|-------|
| 3559  | Mononeuritis of unspecified site                         | 0.3   |
| 362   | Other retinal disorders                                  | 1.4   |
| 3620  | Background diabetic retinopathy                          | 1     |
| 3625  | Macular degeneration (senile), unspecified              | 0.4   |
| 396   | Diseases of mitral and aortic valves                     | 0.7   |
| 3962  | Mitral valve insufficiency and aortic valve stenosis     | 0.3   |
| 3963  | Mitral valve insufficiency and aortic valve insufficiency| 0.3   |
| 401   | Essential hypertension                                   | 62.4  |
| 4010  | Malignant essential hypertension                         | 0.1   |
| 4011  | Benign essential hypertension                            | 0.7   |
| 4019  | Unspecified essential hypertension                       | 61.6  |
| 410   | Acute myocardial infarction                              | 54.3  |
| 4100  | AMI of anterolateral wall                                | 2.7   |
| 4101  | AMI of other anterior wall                               | 6.4   |
| 4102  | AMI of inferolateral wall                                | 2.2   |
| 4103  | AMI of inferoposterior wall                              | 2.4   |
| 4104  | AMI of other inferior wall                               | 9.5   |
| 4105  | AMI of other lateral wall                                | 0.9   |
| 4106  | True posterior wall infarction                            | 0.2   |
| 4107  | Subendocardial infarction                                | 28.7  |
| 4108  | AMI of other specified sites                             | 0.3   |
| 4109  | AMI of unspecified site                                  | 1.1   |
| 411   | Other acute and subacute form of ischemic heart disease  | 19.8  |
| 4111  | Intermediate coronary syndrome                           | 19.3  |
| 4118  | Acute coronary occlusion w/o myocardial infarction       | 0.5   |
| 414   | Other forms of chronic ischemic heart disease            | 86.7  |
| 4140  | Coronary atherosclerosis of unspecified vessel           | 85.6  |
| 4141  | Aneurysm of heart (wall)                                 | 0.6   |
| 4142  | Chronic total occlusion of coronary artery               | 9.5   |
| 4148  | Other specified forms of chronic ischemic heart disease  | 4.9   |
| 4149  | Chronic ischemic heart disease, unspecified              | 0.2   |
| 424   | Other diseases of endocardium                            | 6.3   |
| 4240  | Mitral valve disorders                                   | 3.5   |
| 4241  | Aortic valve disorders                                   | 2.8   |
| 425   | Cardiomyopathy                                           |       |
| 4254  | Other primary cardiomyopathies                           | 2.5   |
| 4255  | Alcoholic cardiomyopathy                                 | 0.1   |
| 426   | Conduction disorders                                     | 5.2   |
| 4260  | Atrioventricular block, complete                         | 0.8   |
| 4261  | Atrioventricular block, unspecified                      | 1.3   |
| 4262  | Left bundle branch hemiblock                             | 0.1   |
| 4263  | Other left bundle branch block                           | 1.4   |
| 4264  | Right bundle branch block                                | 1.5   |
| 4265  | Bundle branch block, unspecified                         | 0.4   |
| 427   | Cardiac dysrhythmias                                     | 17.3  |
| 4270  | Paroxysmal supraventricular tachycardia                  | 0.2   |
4271 Paroxysmal ventricular tachycardia 2.4
4273 Atrial fibrillation 10
4274 Ventricular fibrillation 1.6
4275 Cardiac arrest 1.3
4276 Premature beats, unspecified 0.7
4278 Sinoatrial node dysfunction 4.3
428 Heart failure 15.5
4280 Congestive heart failure, unspecified 14.5
4282 Systolic heart failure, unspecified 5.6
4283 Diastolic heart failure, unspecified 3
4284 Combined systolic and diastolic heart failure, unspecified 1.6
429 Ill-defined descriptions and complications of heart disease 1.8
4293 Cardiomegaly 0.7
4298 Other disorders of papillary muscle 0.2
4299 Heart disease, unspecified 0.9
433 Occlusion and stenosis of precerebral arteries 1.8
4331 Occlusion/stenosis of carotid artery 1.8
4333 Occlusion/stenosis of multiple precerebral arteries 0.5
440 Atherosclerosis 2.6
4400 Atherosclerosis of aorta 0.3
4401 Atherosclerosis of renal artery 0.4
4402 Atherosclerosis of native arteries of the extremities 1.9
4408 Atherosclerosis of other specified arteries 0.2
441 Aortic aneurysm and dissection 1.2
4412 Thoracic aneurysm w/o mention of rupture 0.1
4414 Abdominal aneurysm w/o mention of rupture 1.1
443 Other peripheral vascular disease 5.6
4430 Raynaud’s syndrome 0.3
4438 Peripheral angiopathy in diseases classified elsewhere 0.3
4439 Peripheral vascular disease, unspecified 5
444 Arterial embolism and thrombosis 0.2
4442 Arterial embolism and thrombosis of upper extremity 0.1
4448 Embolism and thrombosis of iliac artery 0.1
447 Other disorders of arteries and arterioles 0.5
4471 Stricture of artery 0.4
4477 Aortic ectasia, unspecified site 0.1
458 Hypotension 1.7
4580 Orthostatic hypotension 0.2
4582 Hypotension of hemodialysis 0.1
4588 Other specified hypotension 0.6
4589 Hypotension, unspecified 0.7
493 Asthma 4.9
4932 Chronic obstructive asthma, unspecified 1.4
4939 Asthma, unspecified type, unspecified 3.6
518 Other diseases of lung 3
5180 Pulmonary collapse 0.4
| Code   | Description                                                                 | Rate |
|--------|-----------------------------------------------------------------------------|------|
| 5184   | Acute edema of lung, unspecified                                            | 0.1  |
| 5188   | Acute respiratory failure                                                  | 2.5  |
| 530    | Diseases of esophagus                                                      | 17.3 |
| 5301   | Esophagitis, unspecified                                                  | 0.3  |
| 5308   | Esophageal reflux                                                          | 17.1 |
| 536    | Disorders of function of stomach                                           | 0.3  |
| 5363   | Gastroparesis                                                              | 0.1  |
| 5368   | Dyspepsia and other specified disorders of function of stomach             | 0.2  |
| 564    | Functional digestive disorders, not elsewhere classified                    | 1.3  |
| 5640   | Constipation, unspecified                                                  | 0.8  |
| 5641   | Irritable bowel syndrome                                                   | 0.6  |
| 571    | Chronic liver disease and cirrhosis                                        | 0.6  |
| 5715   | Cirrhosis of liver w/o mention of alcohol                                  | 0.2  |
| 5718   | Other chronic nonalcoholic liver disease                                   | 0.4  |
| 578    | Gastrointestinal hemorrhage                                                | 0.4  |
| 5781   | Blood in stool                                                             | 0.1  |
| 5789   | Hemorrhage of gastrointestinal tract, unspecified                          | 0.2  |
| 584    | Acute renal failure                                                        | 2.8  |
| 5845   | Acute kidney failure with lesion of tubular necrosis                       | 0.3  |
| 5849   | Acute kidney failure, unspecified                                          | 2.5  |
| 585    | Chronic kidney disease                                                     | 9.9  |
| 5852   | Chronic kidney disease, Stage II (mild)                                    | 0.4  |
| 5853   | Chronic kidney disease, Stage III (moderate)                               | 1.9  |
| 5854   | Chronic kidney disease, Stage IV (severe)                                  | 0.6  |
| 5855   | Chronic kidney disease, Stage V                                            | 0.2  |
| 5856   | End stage renal disease                                                    | 1.8  |
| 5859   | Chronic kidney disease, unspecified                                        | 5.1  |
| 599    | Other disorders of urethra and urinary tract                               | 1.9  |
| 5990   | Urinary tract infection, site not specified                                | 1.5  |
| 5996   | Urinary obstruction, unspecified                                           | 0.1  |
| 5997   | Hematuria, unspecified                                                     | 0.3  |
| 696    | Psoriasis and similar disorders                                            | 0.8  |
| 6960   | Psoriatic arthropathy                                                      | 0.2  |
| 6961   | Other psoriasis                                                           | 0.6  |
| 707    | Chronic ulcer of skin                                                      | 0.8  |
| 7070   | Pressure ulcer, unspecified site                                           | 0.3  |
| 7071   | Ulcer of lower limb, unspecified                                           | 0.5  |
| 7072   | Pressure ulcer, unspecified stage                                          | 0.3  |
| 710    | Diffuse diseases of connective tissue                                     | 0.4  |
| 7100   | Systemic lupus erythematosus                                               | 0.2  |
| 7101   | Systemic sclerosis                                                         | 0.2  |
| 715    | Osteoarthrosis and allied disorders                                        | 4    |
| 7153   | Osteoarthrosis, localized                                                  | 0.8  |
| 7159   | Osteoarthrosis, unspecified                                                | 3.2  |
| 721    | Spondylosis and allied disorders                                           | 0.7  |
| 7210   | Cervical spondylosis w/o myelopathy                                       | 0.2  |
7213  Lumbosacral spondylosis w/o myelopathy 0.3
7219  Spondylosis of unspecified site, w/o mention of myelopathy 0.2
722  Intervertebral disc disorders 0.2
7221  Displacement of lumbar intervertebral disc w/o myelopathy 0.1
7225  Degeneration of thoracic or thoracolumbar intervertebral disc 0.1
723  Other disorders of cervical region 0.3
7231  Cervicalgia 0.2
7234  Brachial neuritis or radiculitis NOS 0.1
724  Other and unspecified disorders of back 3.4
7240  Spinal stenosis, unspecified region 0.9
7242  Lumbago 1.1
7243  Sciatica 0.2
7245  Backache, unspecified 1.3
729  Other disorders of soft tissues 1.3
7291  Myalgia and myositis, unspecified 0.8
7295  Pain in limb 0.1
7298  Swelling of limb 0.4
733  Other disorders of bone and cartilage 2.1
7330  Osteoporosis, unspecified 1.7
7339  Disorder of bone and cartilage, unspecified 0.4
780  General symptoms 3.5
7800  Coma 0.2
7802  Syncope and collapse 0.7
7803  Febrile convulsions (simple), unspecified 0.2
7804  Dizziness and giddiness 0.4
7805  Sleep disturbance, unspecified 1.4
7806  Fever, unspecified 0.3
7807  Chronic fatigue syndrome 0.2
7809  Fussy infant (baby) 0.3
785  Symptoms involving cardiovascular system 2.4
7850  Tachycardia, unspecified 0.2
7852  Undiagnosed cardiac murmurs 0.3
7855  Shock, unspecified 2
786  Symptoms involving chest or respiratory system 1.5
7860  Respiratory abnormality, unspecified 0.6
7862  Cough 0.1
7863  Hemoptysis 0.1
7865  Chest pain, unspecified 0.7
787  Symptoms involving digestive system 0.8
7870  Nausea with vomiting 0.2
7872  Dysphagia, unspecified 0.3
7879  Diarrhea 0.3
788  Symptoms involving urinary system 0.7
7882  Retention of urine, unspecified 0.3
7883  Urinary incontinence, unspecified 0.3
7884  Urinary frequency 0.1
| Code  | Description                                                                 | Weight |
|-------|------------------------------------------------------------------------------|--------|
| 790   | Nonspecific findings on examination of blood                                  | 2.4    |
| 7902  | Impaired fasting glucose                                                      | 1.8    |
| 7904  | Elevation of levels of transaminase or LDH                                    | 0.2    |
| 7906  | Other abnormal blood chemistry                                                | 0.1    |
| 7909  | Abnormal arterial blood gases                                                 | 0.3    |
| 794   | Nonspecific abnormal results of function studies                             | 1.4    |
| 7943  | Abnormal cardiovascular function study, unspecified                          | 1.2    |
| 7948  | Nonspecific abnormal results of function study of liver                       | 0.2    |
| 996   | Complications peculiar to certain specified procedures                        | 5      |
| 9967  | Other complications due to unspecified device, implant, and graft             | 4.8    |
| 9968  | Complications of transplanted organ, unspecified                              | 0.2    |
| E87   |                                                                              | 0.5    |
| E878  | Transplant of whole organ causing abnormal reaction or complication           | 0.4    |
| E879  | Cardiac catheterization as the cause of reaction or complication              | 0.1    |
| 2780  | Obesity, unspecified                                                          | 12.3   |
| 4039  | Hypertensive chronic kidney disease                                           | 9.1    |
| 4139  | Other and unspecified angina pectoris                                         | 7.9    |
| 496   | Chronic airway obstruction, unspecified                                       | 7.3    |
| 3272  | Organic sleep apnea, unspecified                                              | 4.9    |
| 6000  | Hypertrophy of prostate without urinary obstruction                           | 4.9    |
| 311   | Depressive disorder, unspecified                                              | 4.8    |
| 4168  | Other chronic pulmonary heart diseases                                        | 1.9    |
| 3572  | Polyneuropathy in diabetes                                                    | 1.8    |
| 7140  | Rheumatoid arthritis                                                          | 1.5    |
| 7169  | Arthropathy, unspecified, site unspecified                                    | 1.5    |
| 3659  | Unspecified glaucoma                                                          | 1.4    |
| 486   | Pneumonia, organism unspecified                                               | 1.3    |
| 5621  | Diverticulosis of colon                                                       | 1.2    |
| 3970  | Diseases of tricuspid valve                                                    | 1.2    |
| 4912  | Obstructive chronic bronchitis without exacerbiation                          | 1.1    |
| 2886  | Leukocytosis, unspecified                                                     | 1.1    |
| 3459  | Epilepsy                                                                     | 1.1    |
| 5533  | Diaphragmatic hernia                                                          | 0.9    |
| 3569  | Unspecified peripheral neuropathy                                             | 0.9    |
| 3039  | Alcohol dependence                                                           | 0.9    |
| V498  | Asymptomatic postmenopausal status                                            | 0.8    |
| 2689  | Vitamin D deficiency                                                         | 0.8    |
| 3469  | Migraine                                                                     | 0.8    |
| 5838  | Nephritis and nephropathy                                                     | 0.8    |
| 4598  | Venous insufficiency                                                          | 0.7    |
| 5939  | Unspecified disorder of kidney and ureter                                     | 0.7    |
| 4928  | Other emphysema                                                               | 0.7    |
| 7194  | Pain in joint                                                                 | 0.6    |
| 515   | Postinflammatory pulmonary fibrosis                                           | 0.6    |
| 3899  | Unspecified hearing loss                                                      | 0.6    |
| 185   | Malignant neoplasm of prostate                                                | 0.5    |
| Code  | Description                                                                                     | Weight |
|-------|-----------------------------------------------------------------------------------------------|--------|
| 2777  | Dysmetabolic syndrome                                                                          | 0.5    |
| 725   | Polymyalgia rheumatica                                                                          | 0.5    |
| 3098  | Posttraumatic stress disorder                                                                  | 0.5    |
| 6078  | Balanitis xerotica obliterans                                                                   | 0.4    |
| 5119  | Unspecified pleural effusion                                                                   | 0.4    |
| 3320  | Paralysis agitans                                                                               | 0.4    |
| 2662  | Other B-complex deficiencies                                                                    | 0.4    |
| 2898  | Primary hypercoagulable state                                                                   | 0.4    |
| 3694  | Legal blindness, as defined in U.S.A.                                                           | 0.4    |
| 7990  | Asphyxia                                                                                       | 0.4    |
| 9981  | Hemorrhage complicating a procedure                                                             | 0.4    |
| 4029  | Hypertensive heart disease without heart failure                                                | 0.4    |
| 5569  | Ulcerative colitis                                                                              | 0.4    |
| 412   | Old myocardial infarction                                                                       | 0.4    |
| 5070  | Pneumonitis due to inhalation of food or vomitus                                                 | 0.4    |
| 7840  | Headache                                                                                        | 0.3    |
| 7823  | Edema                                                                                          | 0.3    |
| 5559  | Regional enteritis of unspecified site                                                          | 0.3    |
| V08   | Asymptomatic human immunodeficiency virus                                                       | 0.3    |
| 3669  | Unspecified cataract                                                                            | 0.3    |
| 9971  | Cardiac complications, not elsewhere classified                                                  | 0.3    |
| 5355  | Gastritis and gastroduodenitis                                                                   | 0.3    |
| 5888  | Secondary hyperparathyroidism                                                                    | 0.3    |
| V458  | Aortocoronary bypass status                                                                     | 0.3    |
| 2041  | Chronic lymphoid leukemia                                                                       | 0.3    |
| 4779  | Allergic rhinitis                                                                               | 0.3    |
| E849  | Home accidents                                                                                  | 0.3    |
| 2028  | Other malignant lymphomas                                                                        | 0.2    |
| 2520  | Hyperparathyroidism                                                                             | 0.2    |
| 5965  | Hypertonicity of bladder                                                                        | 0.2    |
| 2387  | Essential thrombocytethemia                                                                      | 0.2    |
| 3310  | Alzheimer’s disease                                                                              | 0.2    |
| 4239  | Disease of pericardium                                                                           | 0.2    |
| 3140  | Attention deficit disorder without mention of hyperactivity                                      | 0.2    |
| 5339  | Peptic ulcer of unspecified site                                                                | 0.2    |
| 6929  | Contact dermatitis and other eczema                                                              | 0.2    |
| 5920  | Calculus of kidney                                                                              | 0.2    |
| 6826  | Cellulitis and abscess of leg, except foot                                                        | 0.2    |
| 7890  | Abdominal pain                                                                                  | 0.2    |
| 4049  | Hypertensive heart and chronic kidney disease, unspecified                                       | 0.2    |
| 4465  | Giant cell arteritis                                                                            | 0.2    |

27
| Code  | Diagnosis                                      | Weight |
|-------|-----------------------------------------------|--------|
| 7288  | Interstitial myositis                         | 0.2    |
| 4824  | Pneumonia due to Staphylococcus, unspecified  | 0.2    |
| 2410  | Nontoxic uniodular goiter                     | 0.2    |
| 0389  | Unspecified septicemia                        | 0.1    |
| 2841  | Antineoplastic chemotherapy induced pancytopenia | 0.1  |
| 2959  | Unspecified schizophrenia, unspecified       | 0.1    |
| 3040  | Opioid type dependence, unspecified          | 0.1    |
| 3540  | Carpal tunnel syndrome                        | 0.1    |
| 570   | Acute necrosis of liver                       | 0.1    |
| 0084  | Intestinal infection due to staphylococcus    | 0.1    |
| 1629  | Malignant neoplasm of bronchus and lung       | 0.1    |
| 1985  | Secondary malignant neoplasm of bone and marrow | 0.1  |
| 2030  | Multiple myeloma                              | 0.1    |
| 4739  | Unspecified sinusitis (chronic)               | 0.1    |
| 501   | Asbestosis                                    | 0.1    |
| 5589  | Noninfectious gastroenteritis and colitis     | 0.1    |
| 5771  | Chronic pancreatitis                          | 0.1    |
| 5790  | Celiac disease                                | 0.1    |
| 7832  | Loss of weight                                | 0.1    |
| 1889  | Malignant neoplasm of bladder                 | 0.1    |
| 2019  | Hodgkin’s disease                             | 0.1    |
| 2536  | Other disorders of neurohypophysis            | 0.1    |
| 2554  | Gluocorticoid deficiency                      | 0.1    |
| 2713  | Intestinal disaccharidase deficiencies and malabsorption | 0.1  |
| 2904  | Vascular dementia, uncomplicated              | 0.1    |
| 3860  | Meniere’s disease                             | 0.1    |
| 4059  | Unspecified renovascular hypertension         | 0.1    |
| 4151  | Iatrogenic pulmonary embolism and infarction  | 0.1    |
| 490   | Bronchitis, acute/chronic unspecified         | 0.1    |
| 6953  | Rosacea                                       | 0.1    |
| 2113  | Benign neoplasm of colon                      | 0.1    |
| 3429  | Hemiplegia affecting unspecified side         | 0.1    |
| 4370  | Cerebral atherosclerosis                      | 0.1    |
| 5742  | Calculus of gallbladder                        | 0.1    |
| 591   | Hydronephrosis                                | 0.1    |
| 7367  | Deformity of ankle and foot, acquired         | 0.1    |
| 7812  | Abnormality of gait                           | 0.1    |
| V125  | History of unspecified circulatory disease    | 0.1    |
| V158  | History of failed moderate sedation           | 0.1    |