INFECTIONS DUE TO MEDICAL CARE

IN OREGON HOSPITALS, 2003 – 2005

by

James P. Oliver

A THESIS

Presented to the Department of Public Health and Preventive Medicine

and the Oregon Health & Science University

School of Medicine

in partial fulfillment of

the requirements for the degree of

Master of Public Health

November, 2007

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Master's thesis of

James P. Oliver

has been approved

Thomas M. Becker, MD, PhD

Donseok Choi, PhD

John McConnell, PhD

TABLE OF CONTENTS

List of Tables and Figures (in order of appearance)

List of Abbreviations

AHRQ	Agency for Healthcare Research and Quality
AIC	Akaike's information criterion
AICC	Akaike's information criterion corrected for small samples
APR-DRG	All Patient Refined Diagnosis-Related Group
BIC	Bayesian information criterion
CDC	Centers for Disease Control and Prevention
CGS	Core Grouping Software
CMS	Centers for Medicare and Medicaid Services
DF	degrees of freedom
DRG	diagnosis-related group
ER	emergency room
GI	gastro-intestinal
HAI	healthcare-acquired infection
HCUP	Health Care Utilization Project
HDD	hospital discharge data
ICC	intra-class correlation
ICD-9-CM	International Classification of Diseases, Ninth Edition, Clinical Modification
IOM	Institute of Medicine
IV	intra-venous
LOS	length of stay
MDC	major diagnostic category

n	number
OHPR	Office for Oregon Health Policy and Research
OR	operating room
Pr	probability
PSI	Patient Safety Indicator(s)
pts.	patients
Q-Q	quantile-quantile
ROM	risk of mortality
SE	standard error
SOI	severity of illness
SPSS	SPSS statistics software
SAS	SAS statistics software
UCSF	University of California at San Francisco
VIF	variance inflation factor
vs.	versus

Acknowledgments

There is an extensive list of people to whom I owe my sincerest gratitude:

- My parents, who consistently encouraged me to finish and also provided a very timely dose of financial support
- Thomas Gragnola (my doctor) who, by application of evidence-based medicine and perhaps a touch of parlor magic, was able to tame a major health barrier that nearly prevented me from finishing this thesis
- Tina Edlund (my boss), who hired me, inspired me, and allowed me to take this project farther than even I imagined
- Bruce Goldberg (Director of the Department of Human Services), for being savvy enough to ask the question that spawned this project
- The Oregon Legislature, who during the waning days of the 2007 regular session passed a bill mandating reporting of certain healthcare acquired infections, indirectly making a compelling case for finishing the manuscript
- Thomas M. Becker, who took a personal interest in seeing that I finished up, and this was truly the key reason why this thesis was completed
- Dongseok Choi and John McConnell, who kept a close eye on the biometry and economics
- Katie Riley, who gracefully endured way too many inquiries from me regarding a wide range of topics
- Ian and Colin Cullison, who did their best to not bother me while I was working on the manuscript

- Dinsdale (the 106 pound lap dog), who remained loyal even though he didn’t understand why I could not go outside and play
- Kelley Cullison, who through great personal sacrifice gave me everything else that I needed and more

Abstract

*Background***:** About 2 million cases of Healthcare Acquired Infections (HAI) occur in hospitals each year, causing approximately 90,000 deaths and generating $4.5 billion in excess medical expenses. HAI complicate medical care causing worse clinical outcomes, extended hospital stays, and higher rates of mortality. Identifying adverse events due to medical care and the excess costs associated with them, along with designing interventions to decrease their prevalence, may offer a chance to redirect increasingly scarce healthcare dollars to more productive purposes. *Specific Aims:* Identify cases of HAI in Oregon hospital discharge data; estimate statewide costs associated with cases of HAI; and determine if severity of illness or comorbid conditions are important predictors of estimated costs per case of HAI. *Study Design:* Cross-sectional secondary analysis of hospital discharge data. *Human Subjects:* Adult Oregonians treated in Oregon hospitals who were at risk of developing an HAI during calendar years 2003, 2004, and 2005. *Methods:* The outcomes of interest were diagnosis of an HAI in the discharge record and the estimated costs per discharge. Cost-to-charge ratios for each hospital were estimated from audited financial statements. Costs per discharge were estimated from total charges multiplied by the hospital-specific cost-to-charge ratio and adjusted for inflation to 2005 dollars. Estimated costs per discharge for patients with HAI were compared to estimated costs per discharge for patients without HAI. Regression modeling was used to assess whether severity of illness or comorbidities are important predictors of estimated costs per discharge. *Results:* 1034 prevalent cases of HAI were identified. Costs per discharge averaged over $20,000 higher for patients with HAI compared to patients without HAI. The statewide excess costs were estimated to be at least $21 million, or approximately $7 million per year. Presence of an HAI, severity of illness, presence of a severe comorbidity, and surgical DRG were important predictors of estimated cost per discharge. *Conclusions:* HAI cases in Oregon hospitals are common, the statewide excess costs are substantial, and the excess costs are not fully explained by severity of illness or severe comorbidities.

Background

Literature Review

In 1999 the Institute of Medicine (IOM) published a landmark report, "To Err is Human," describing the morbidity and mortality from adverse clinical events due to medical care. In this report, the IOM estimated that annually 44,000 to 98,000 inpatients died from adverse clinical events due to medical care. The IOM further suggested that between $17 billion and $29 billion in annual excess costs to society were generated by preventable medical errors. About half of the costs associated with preventable medical errors were direct medical expenses (IOM, 1999).

Hospital-acquired infections (HAI), also known as nosocomial infections, are one type of adverse clinical event due to medical care. Although they have been discussed in hundreds of peer-reviewed articles during the past five years, there currently is no consensus about how to define and measure HAI. HAI generally refer to infections contracted by patients in a healthcare setting while they are being treated for other conditions. The Centers for Disease Control and Prevention estimates that 2 million cases of HAI occur in hospitals each year, causing approximately 90,000 deaths and generating $4.5 billion in excess medical expenses (CDC, 2006).

Recent studies suggest that 5-15% of all hospitalized patients suffer from HAI and that these cases are widely under-reported (Weinstein, Siegel, and Brennan, 2006; Smith, et al, 2004; Graves, 2004; and Eggimann and Pettet, 2001). Commonly reported HAI include ventilator-associated pneumonia, infections at surgical or trauma sites, and bacteremia caused by the use of IV devices (such as catheters). HAI complicate medical care causing worse clinical outcomes, extended hospital stays, and higher rates of mortality (Pepin, Valiquette, and Cossette, 2005; Safdar, et al, 2005; and DeRyke, et al, 2005). Furthermore, HAI result in excess direct costs for medical care, even after controlling for common risk factors and severity of illness (Elward, et al, 2005; Roberts, et al, 2003.; Whitehouse, et al, 2002; Rello, et al, 2002, Tambyah, Knasinski, and Maki, 2003; and Zhao, et al, 2002).

Attempts to quantify HAI have generated mixed reviews. Surveillance methods and HAI definitions vary widely. Some have suggested that existing methods cannot accurately and consistently detect HAI cases (Brossette, et al, 2006). This is at least partially because HAI, especially surgical site infections, frequently manifest after discharge (Nan, et al, 2005). Others have reported that surveillance methods produce good agreement with “gold standard” clinical diagnoses (Miller, et al, 2006; Layde, et al, 2005). Some studies have implicated “the process of care” as the primary cause of HAI (Yogari, Elward, and Fraser, 2002; and Grundmann, et al, 2005). In any case, a wide and growing body of literature utilizing a variety of methods has reported dramatic differences in clinical outcomes and costs in hospitalized patients with HAI compared to hospitalized patients without HAI.

There is very little doubt that HAI have serious clinical, financial, and policy implications for patients, hospitals, and the State of Oregon. Currently there are approximately 576,000 Oregonians without health insurance, including about 117,000 children (OHPR, 2007). Policymakers are searching for every possible opportunity to stretch existing healthcare funding. Identifying adverse events due to medical care and the excess costs associated with them, along with designing interventions intended to decrease them, may offer a chance to redirect increasingly scarce healthcare dollars to more productive purposes. Policymakers increasingly demand local data to inform their decisions, and HAI data specific to Oregon are not yet available in the peer-reviewed literature. Accordingly, the specific aims of this project are:

- Identify cases of HAI in Oregon hospital discharge data.
- Estimate statewide costs associated with cases of HAI.
- Determine if severity of illness or comorbid conditions are important predictors of estimated costs per case of HAI.

AHRQ Patient Safety Indicators

The Agency for Healthcare Research and Quality (AHRQ) has developed a series of evidence-based algorithms, known as Patient Safety Indicators (PSIs), to help measure inpatient safety. The PSIs were developed and refined by a panel of clinicians and peer reviewers facilitated by the Evidence-based Practice Center at UCSF-Stanford. The algorithm is intended to identify adverse clinical events and potentially preventable complications in hospital discharge records (AHRQ, 2003). Specifically, the PSIs attempt to determine the proportions of adverse events and potentially preventable complications that are caused by inpatient medical care. Version 3.01a, released in May, 2006, consists of twenty indicators that are measured at the hospital level and seven that are measured at a regional level (state or metropolitan statistical area).

One of these, Selected Infections Due to Medical Care (PSI 07) is intended to detect HAI cases. AHRQ has produced programming code to flag HAI (using secondary diagnoses) in hospital discharge data sets and calculate risk-adjusted proportions to account for case-mix differences across hospitals (AHRQ, 2006). The AHRQ algorithm flags ICD-9-CM codes 999.3 (infection following infusion, injection, transfusion, or vaccination) and 996.62 (infection due to vascular device, implant, and graft) in adult medical or surgical patients hospitalized for at least two days, excluding patients with immuno-compromised conditions. Most commonly flagged are infections due to intravenous infusion lines and vascular catheters. Other types of infections, such as ventilator-associated pneumonia and infections at surgical sites, are not included in the AHRQ algorithm. This definition tends to very conservatively flag HAI cases in order to minimize false positives, but it probably also underestimates the true number of HAI cases.

AHRQ also produces separate programming code to flag severe comorbidities in hospital discharge data sets. The programming creates indicator variables for 29 common comorbidities based on diagnosis-related groups (DRGs) and International Classification of Diseases, Ninth Edition, Clinical Modification (ICD-9-CM) diagnosis codes. The complete list of comorbidities flagged by the AHRQ programming is presented in Appendix A. The programming is updated annually to reflect changes in ICD-9-CM codes and

DRG definitions; version 3.1 includes the October 2005 updates to the ICD-9-CM codes and DRG definitions.

Severity of Illness

3M Health Information Systems first introduced the APR-DRG classification system in 1990. The intent was to offer an alternative to DRGs published by the Centers for Medicare and Medicaid Services (CMS) that would simplify risk-adjustment. Risk-adjustment is important to hospitals since some, particularly referral and trauma centers, may tend to serve a larger proportion of severely ill or traumatically injured patients. The APR-DRG software attempts to compensate for these differences with algorithms that assign a proprietary DRG along with severity of illness and risk of mortality scores (3M Health Information Systems, 2005). This allows more legitimate comparisons of morbidity and mortality between hospitals. The APR-DRG classification system is widely used for both clinical and research purposes, including in the AHRQ Inpatient Quality Indicators.

The APR DRG Severity of Illness (SOI) score is assigned based on diagnoses, procedures, age, gender, and patient status at discharge (3M Health Information Systems, 2005). While the SOI score is actually a categorical classification (minor, moderate, major, extreme), it is denoted by an integer for intuitive simplicity. Thus, patients with higher SOI scores fall into the categories representing more severe illnesses or conditions. In patients who are otherwise identical, patients with HAI can generally be expected to be more severely ill, and have higher SOI scores, than patients without HAI.

I undertook this study to document the prevalence of HAI in Oregon inpatients and to estimate the excess costs associated with HAI cases. By examining the importance of comorbid conditions and severity of illness, the study also anticipates concerns from hospitals that may imply their patients are more acutely ill or present with more complex cases. In addition, Oregon's recently convened Patient Safety Commission currently lacks studies about HAI that focus on Oregon hospitals, a gap in the literature addressed by this work.

Methods

<u>Study Design</u>

This study was a cross-sectional secondary analysis of pooled Oregon Hospital Discharge Data (HDD) from calendar years 2003, 2004, and 2005. The primary outcomes of interest were diagnosis of an HAI in the discharge record and the estimated costs per discharge. Cost-to-charge ratios for each hospital were estimated from audited financial statements. Costs per stay for patients with HAI were compared to patients without HAI. Regression modeling was used to assess whether severity of illness or comorbidities were important predictors of costs per stay. Prior to analysis extensive validations were performed on the HDD, according to the Healthcare Utilization Project (HCUP) quality assurance protocols (HCUP, 2006), and these are fully described in Appendix B.

The population of interest was adult Oregonians treated in Oregon hospitals who were at risk of developing an HAI. The specific inclusion criteria were:

1. Inpatient at an acute care Oregon Hospital
2. Discharged between January 1, 2003 and December 31, 2005.
3. Patient from Oregon
4. Medical or surgical DRG (as identified by AHRQ PSI application)
5. Age at least 18 or MDC 14 (pregnancy and childbirth)

The exclusion criteria were:

1. Inpatient at an ineligible hospital (VA, psychiatric, or currently closed)
2. Discharge record fails one or more HCUP validations (see Appendix B)

3. Discharge record is unable to be grouped by the APR-DRG software
4. Principal diagnosis is 999.3 or 996.62
5. Length of stay is less than two days
6. Cancer DRG
7. ICD-9-CM code for any immuno-compromised condition

Preliminary Data Preparation

AHRQ provides SPSS and SAS code to identify patient safety indicators (PSIs) within hospital discharge data sets and calculate a variety of proportions. Using the clean HDD, data sets were recoded to meet the requirements of the PSI code (AHRQ, 2004). Variables were also renamed to maintain consistency with the AHRQ SPSS code. Unneeded variables were deleted. Federal fiscal year and a unique random record identifier were added. A dummy variable for race was created since this variable was not included in the HDD sets. For each procedure, the number of days between the admit date and the procedure data was calculated.

Records were limited to patients from Oregon counties (FIPS codes from 41001 to 41071), stays at eligible acute care hospitals in Oregon, and those where the length of stay was at least two days. Records were then further limited to those in which the age in years was at least 18, or the MDC is 14 (pregnancy and childbirth). One data file with new, recoded, and renamed variables was created for each calendar year from 2003 to 2005. A year variable was added so that the data sets could be combined into a single research data file.

A separate fixed-length text data file was created for use with the APR-DRG Core Grouping Software (CGS). This text file consisted of the random record identifier, fiscal year, gender, date of birth, admit date, discharge date, discharge status, diagnosis codes, and procedure codes. The data were mapped by fiscal year to the appropriate version of the ICD-9-CM diagnosis and procedure codes. The CGS assigns the severity of illness and risk of mortality classifications, and these were linked to the research data file by the random record identifier. Records that could not be grouped by the CGS were excluded.

AHRQ PSI Application

The research data file was then run through the AHRQ PSI application in order to identify prevalent cases of HAI and important comorbid conditions. The PSI application verifies inclusion criteria 4 and 5 and exclusion criteria 3 through 6, although verification of inclusion criterion 5 and exclusion criterion 4 occurred in advance to reduce the size of the data file. The PSI application also creates an indicator variable to flag HAI cases. HAI cases were aggregated to the State level, and subsequently stratified by severity of illness or primary payer. A crude proportion of HAI cases per 1,000 discharges was calculated and was directly adjusted based on severity of illness score.

A separate section of AHRQ programming code identifies 29 common comorbid conditions in secondary diagnoses, which are flagged with 29 indicator variables. A separate indicator variable was created to indicate the presence of any comorbidity. The comorbidity variables were aggregated to the State level by HAI status, and subsequently stratified by severity of illness and principal diagnoses with at least 30 observations. The median number of comorbidities for patients with HAI was compared to the median number of comorbidities for patients without HAI.

Estimated Costs

The Oregon HDD includes total inpatient charges, even though essentially nobody pays this amount for inpatient care. For calendar years 2003, 2004, and 2005, summaries of audited financial statements from each Oregon hospital were reviewed. Gross patient revenue and total expenses were extracted from each summarized financial statement. An annual hospital-specific cost-to-charge ratio was estimated by dividing gross patient revenue by total expenses. The cost-to-charge ratio was calculated for each hospital for each calendar year and the hospital-specific cost-to-charge ratio was assigned to each record in the data set.

The costs per stay for each record were then estimated by calculating the product of total charges and the hospital-specific cost-to-charge ratio, and then adjusted for inflation to 2005 dollars. The costs were natural log-transformed in order to produce a normal distribution. Cost outliers, defined as greater than four standard deviations from the log-transformed mean, were excluded from further analysis. Finally, the untransformed estimated costs were aggregated to the State level by year and HAI status, and were subsequently stratified by severity of illness and either expected primary payer or principal diagnoses with at least 30 observations. Estimated costs per discharge for patients with HAI were compared to estimated costs per discharge for patients without HAI. The estimated excess costs were calculated in two steps:

1. Subtracting the estimated costs per discharge for patients without HAI from the estimated costs per discharge for patients with HAI.
2. Multiplying the result by the number of HAI cases.

Regression Modeling

From previous work with the HDD, it is known the log-transformed estimated costs are a normally distributed continuous variable. Simple linear regression was used to assess whether either severity of illness or comorbid conditions are important predictors of estimated total costs. Using log-transformed total costs as the dependent variable, models were tested using age, gender, expected primary payer, length of stay, APR-DRG risk of mortality, discharge status, and other administrative variables as independent predictors. Interactions and squared terms were also tested. Plots of observed vs. predicted values and residuals vs. predicted values, Q-Q plots, and the sums of squares were used to assess model fit. Collinearity of the independent variables was assessed by examining the variance inflation factors, condition indices, and proportions of variation.

Severity of illness was tested several ways: as a categorical variable, recoded as an indicator variable (minor/moderate=0 and major/extreme =1), and as a discrete variable with integer values 1, 2, 3, and 4. Comborbid conditions were similarly tested: as an indicator variable (0=no comorbidity and 1=at least 1 comorbidity), as a count of comorbidities, and indicator variables (0=no and 1 =yes) for the presence of common comorbidities such as hypertension and obesity. In addition, a separate model was tested using log-transformed costs per day (estimated total costs divided by length of stay) as the dependent variable. Since the data were clustered by hospital, a mixed model was then fit to more fully account for the effect of hospitals. All tests of statistical significance used $\alpha = .05$.

<u>Data Security</u>

The research was a secondary analysis of existing de-identified data sets. While the HDD were not individually identified, they did contain protected health information that was potentially identifiable in combination with other data sets. In order to minimize risk, a random record identifier was assigned and existing record identifiers were removed so that the records could be unlinked from the original data sets. Variables that could potentially be identifiable in combination with other data, such as date of birth, diagnosis codes, and procedure codes, were be removed from the research data sets when they were no longer needed to complete the research. The products of this project included statistics, charts, and data tables aggregated to the State level; they did not require disclosing patient-level or provider-level protected health information.

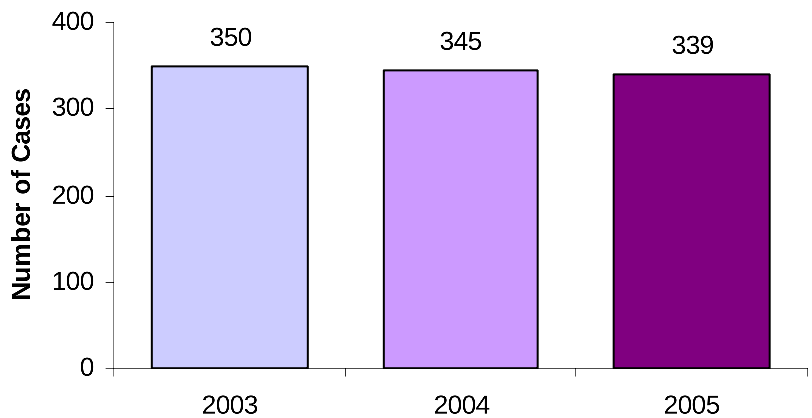
The investigator housed the data files in a secure office environment on a password-protected desktop computer. When not in use, the data files were encrypted in a password-protected archive. After creating and backing up the research data sets, the original data sets were destroyed. Research data files will be maintained for five years, and then will be destroyed. A data use agreement was signed with the Office for Oregon Health Policy and Research, the government agency that maintains the HDD and approves research uses for the data. The research protocol was found to be exempt from review by the Institutional Review Board at Oregon Health and Science University.

Results

<u>Cases of Healthcare Acquired Infections</u>

Overall the number of HAI cases flagged by the AHRQ algorithm remained relatively unchanged from 2003 to 2005 (Figure 1). The crude

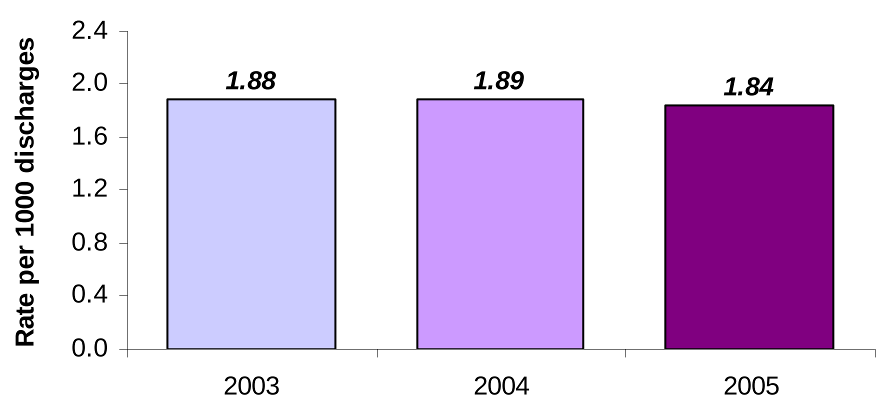
Figure 1: Number of cases of healthcare acquired infections



Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

proportions of HAI cases per 1000 discharges also remained essentially unchanged from 2003 to 2005 (Figure 2). Medicare and commercial health plans

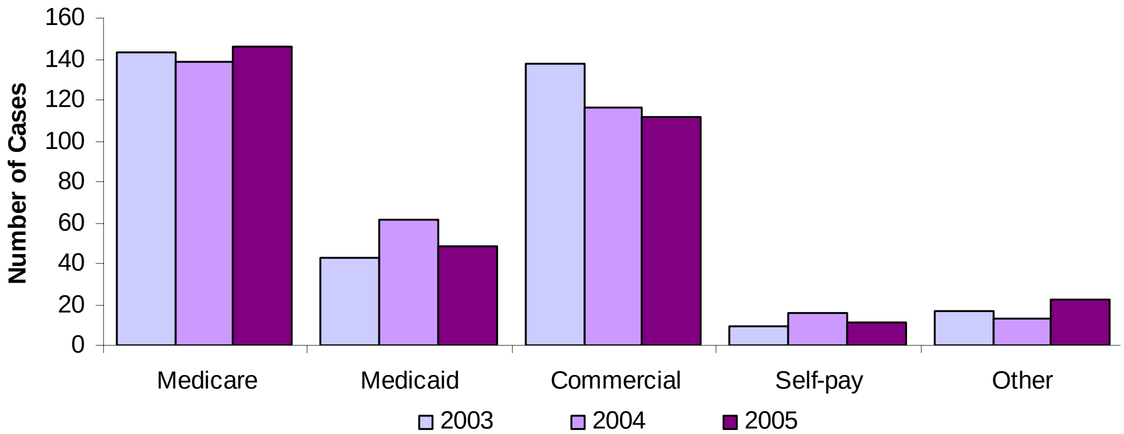
Figure 2: Crude proportion of HAI cases per 1000 discharges



Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

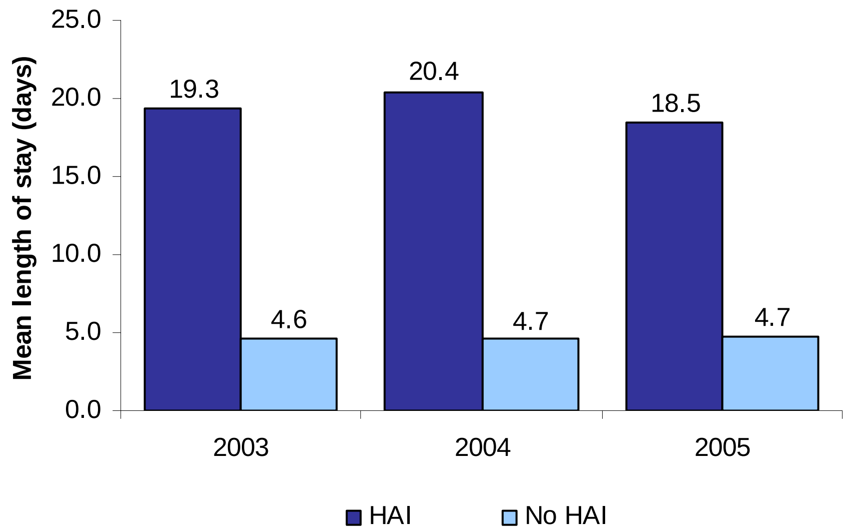
were the primary payers for most HAI cases (Figure 3). Patients with HAI had substantially longer mean length of stay compared to patients without HAI (Figure 4), and this also remained mostly unchanged from 2003 to 2005.

Figure 3: HAI cases by primary payer



Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Figure 4: Mean length of stay in days



Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Overall the demographic differences between patients with HAI and patients without HAI were modest (Table 1). Patients without HAI were marginally more likely to be female than patients with HAI. Patients with HAI were slightly older than patients without HAI and, because of this, were slightly more likely to be covered by Medicare.

Table 1: Patient characteristics

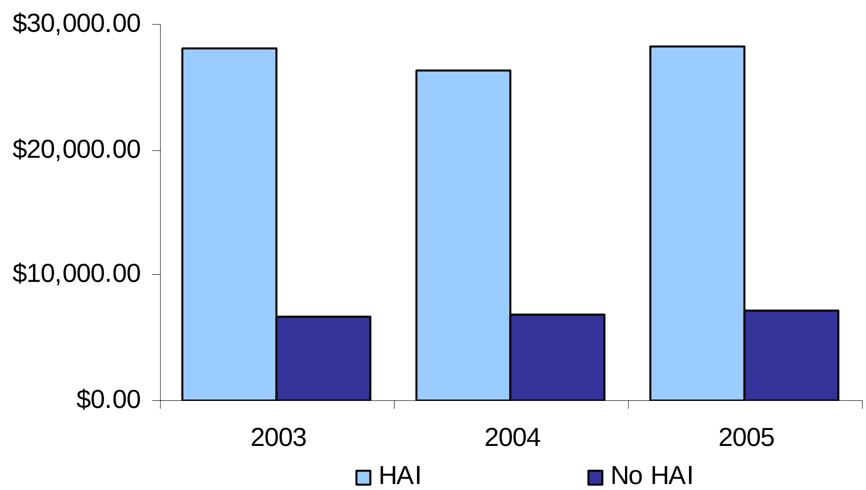
	Medical/surgical pts. without HAI	Medical/surgical pts. with HAI
n	553,487	1034
Mean age	54.6	58.4
% female	65.5	52.9
% urban	60.3	60.1
% Medicare	36.9	41.4
% Medicaid	14.9	14.7

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Statewide Estimated Costs

There was a dramatic difference in mean estimated cost for patients with HAI compared to patients with no HAI (Figure 5). The potential excess costs, or

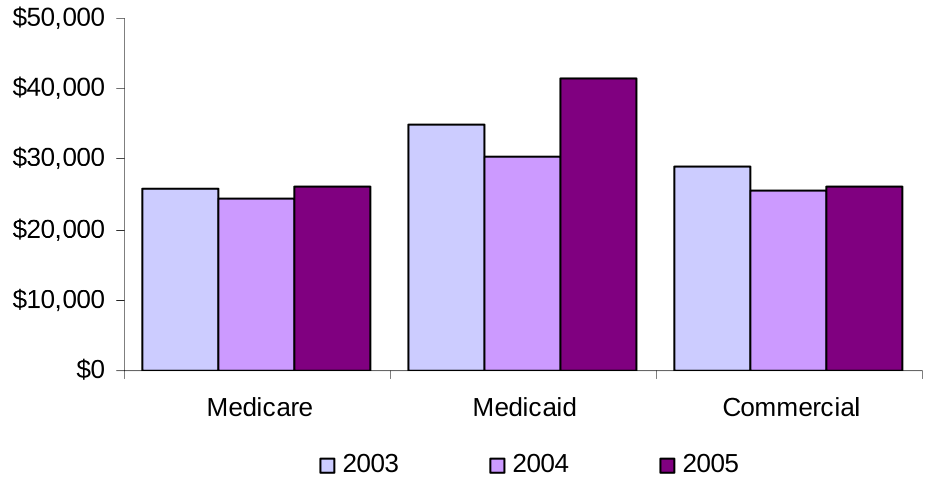
Figure 5: Estimated cost per discharge



Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

the additional costs per case for patients with HAI, averaged over $20,000 per patient from 2003 to 2005. The estimated costs per discharge for patients with HAI were substantially higher for Medicaid compared to Medicare and commercial health plans (Figure 6). Statewide the potential excess cost for

Figure 6: Estimated cost per HAI discharge by payer



Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

patients with HAI was over $21 million from 2003 to 2005. When grouped by DRGs with at least 20 observations, patients with HAI had substantially higher estimated costs per discharge than patients without HAI (Table 2). Similar results

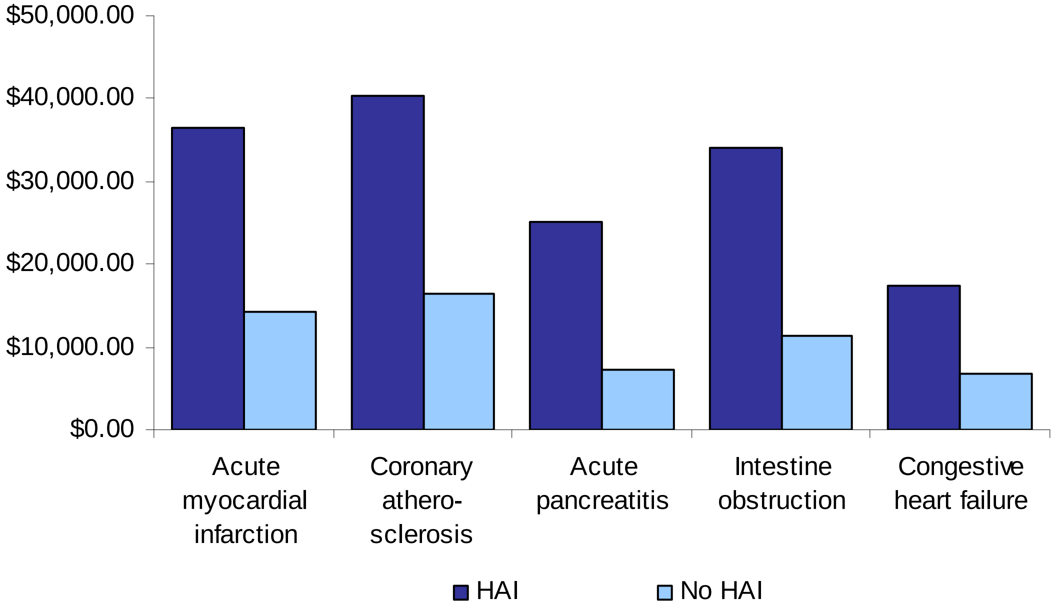
Table 2: Mean estimated cost by DRG (at least 20 observations)

DRG	Medical/surgical pts. without HAI	Medical/surgical pts. with HAI	Definition
110	$27,861	$69,843	Major cardiovascular procedures with complications
148	$19,497	$42,916	Major small/large bowel procedures with complications
154	$18,583	$51,328	Upper GI procedures age>17 with complications
174	$6,575	$13,919	GI hemorrhage with complications
182	$5,303	$9,643	Miscellaneous upper GI disorders with complications
204	$6,260	$18,902	Disorders of the pancreas except malignancy
415	$15,444	$43,739	OR procedure for infectious/parasitic diseases
416	$8,884	$17,413	Septicemia age>17
483	$85,306	$110,415	Tracheotomy with mechanical ventilation >96 hours

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

were obtained when the data were stratified by principal diagnoses with at least 20 observations (Figure 7).

Figure 7: Estimated cost per discharge by principal diagnosis (≥ 20 observations)

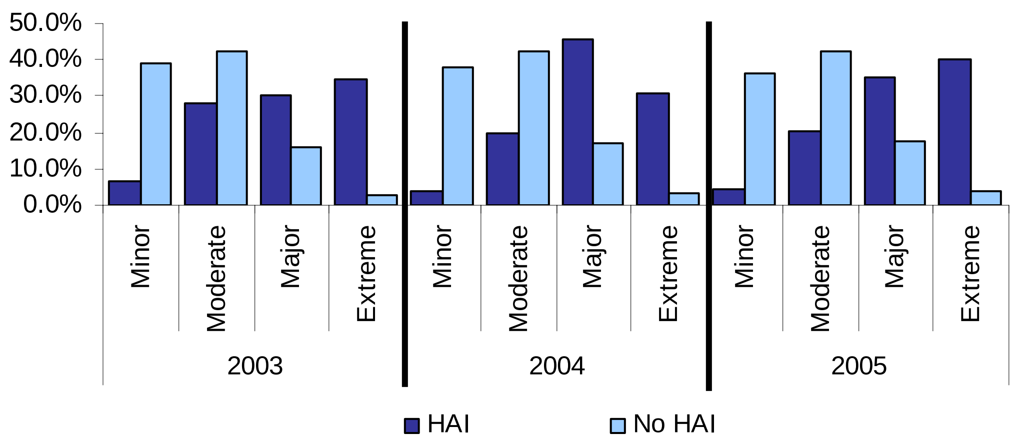


Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

<u>Severity of Illness</u>

Oregon's medical and surgical inpatients with HAI had higher proportions of major and extreme severity of illness than those without HAI (Figure 8).

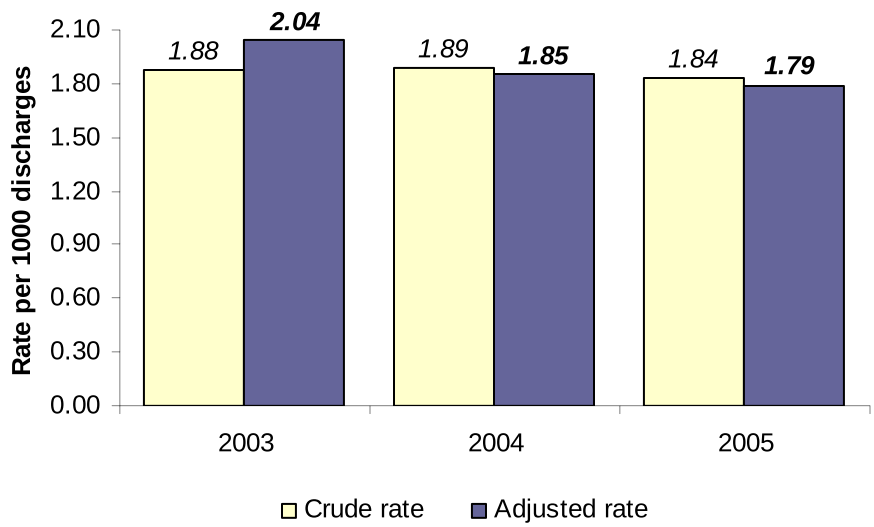
Figure 8: Percentage of discharges by severity of illness category



Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Direct adjustment by severity of illness produced very small differences in the proportions of HAI cases per 1000 discharges during calendar years 2003-2005 (Figure 9). When stratified by SOI category, the dramatic differences in mean estimated cost generally remained (Table 3).

Figure 9: Crude and adjusted proportion of HAI per 1000 discharges



Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Table 3: Estimated costs per discharge

Year	SOI	Medical/surgical pts. without HAI	Medical/surgical pts. with HAI
2003	Minor	$5,350.79	$8,857.32
	Moderate	$6,561.67	$13,615.99
	Major	$9,710.86	$30,091.56
	Extreme	$24,391.74	$63,640.16
2004	Minor	$5,513.75	$6,809.00
	Moderate	$6,747.99	$11,825.34
	Major	$9,798.65	$27,474.13
	Extreme	$25,235.32	$52,785.93
2005	Minor	$5,647.68	$11,327.63
	Moderate	$6,877.43	$14,072.81
	Major	$9,946.74	$23,155.79
	Extreme	$24,052.64	$54,339.14

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Comorbidities

The number of severe comorbidities was not substantially different in patients with HAI compared to patients without HAI (Table 4). The differences

Table 4: Number of severe comorbidities

Year	Mean: Medical/surgical pts. without HAI	Mean: Medical/surgical pts. with HAI	Median: Medical/surgical pts. without HAI	Median: Medical/surgical pts. with HAI
2003	1.4	1.6	1	1
2004	1.4	1.6	1	2
2005	1.5	1.5	1	1

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

were also modest when the data were stratified by severity of illness category (Table 5). The differences were marginally larger when the data were stratified by

Table 5: Number of severe comorbidities by severity of illness category

SOI	Mean: Medical/surgical pts. without HAI	Mean: Medical/surgical pts. with HAI	Median: Medical/surgical pts. without HAI	Median: Medical/surgical pts. with HAI
Minor	0.7	1.2	0	1
Moderate	1.6	1.5	2	1
Major	2.4	1.8	2	2
Extreme	2.2	2.0	2	1

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

principal diagnoses with at least 20 observations, although patients with HAI generally had fewer severe comorbidities (Table 6). When the data were stratified by DRGs with at least 20 observations, patients with HAI again generally had fewer severe comorbidities, although these differences at most were very modest (Table 7).

Table 6: Number of severe comorbidities by principal diagnosis (≥ 20 observations)

Principal diagnosis	Mean		Median	
	Medical/surgical pts. without HAI	Medical/surgical pts. with HAI	Medical/surgical pts. without HAI	Medical/surgical pts. with HAI
Acute myocardial infarction	1.6	1.3	1	1
Coronary athero-sclerosis	1.7	1.1	2	1
Acute pancreatitis	2.0	2.0	2	2
Intestine obstruction	1.7	1.2	2	1
Congestive heart failure	2.1	1.3	2	1

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Table 7: Number of severe comorbidities by DRG (≥ 20 observations)

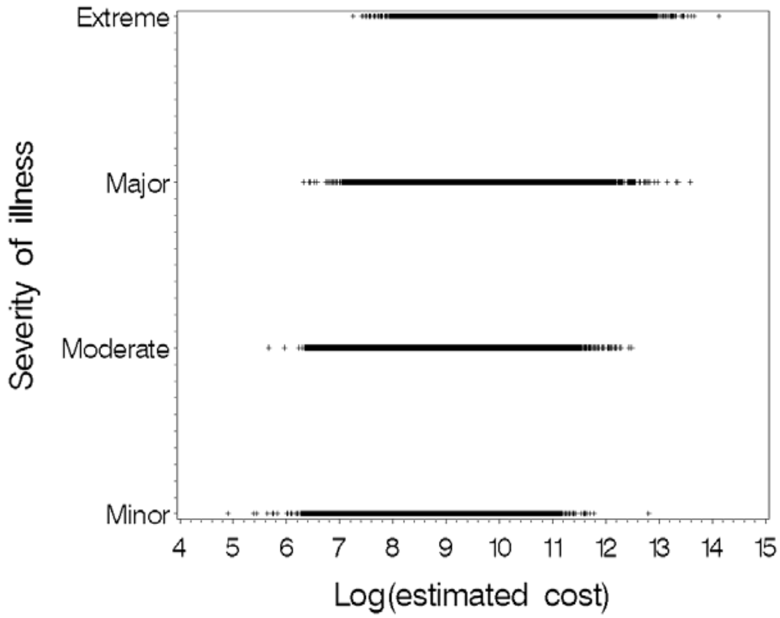
DRG	Mean		Median	
	Medical/surgical pts. without HAI	Medical/surgical pts. with HAI	Medical/surgical pts. without HAI	Medical/surgical pts. with HAI
110	1.2	1.2	2	1
148	1.5	1.4	1	1
154	1.7	1.1	2	1
174	2.6	2.7	3	3
182	2.4	1.9	2	2
204	2.1	1.9	2	2
415	1.6	1.7	1	1
416	2.4	2.1	2	2
483	1.6	1.5	2	1

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Predictors of Estimated Costs

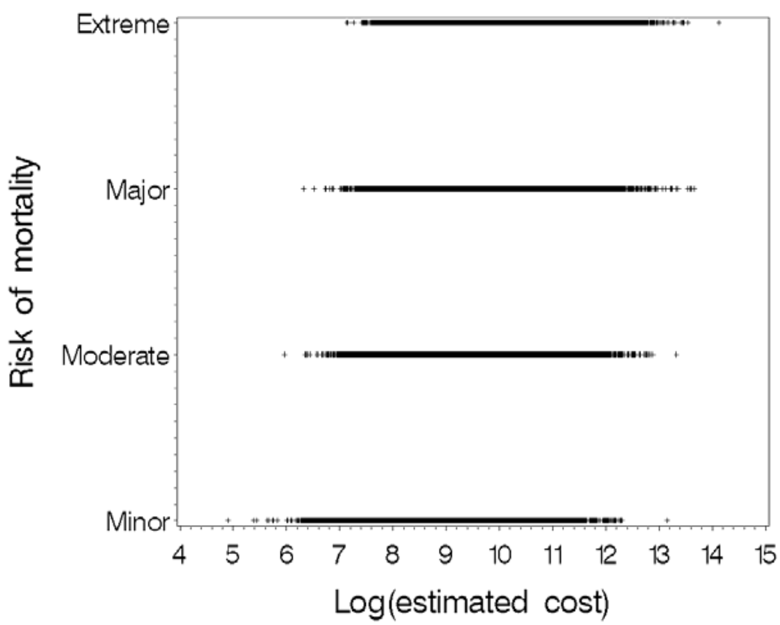
Initial plots of the dependent variable (log of estimated cost) vs. independent variables revealed only modest associations, if any. Three ordinal variables, severity of illness, risk of mortality, and number of procedures, are plotted against the log of estimated cost in Figures 10, 11, and 12. All three plots show slightly positive correlations between the dependent variable and the independent variable, given that the independent variables are ordered from lowest to highest.

Figure 10: Plot of dependent variable (log of estimated cost) vs. independent variable (severity of illness)



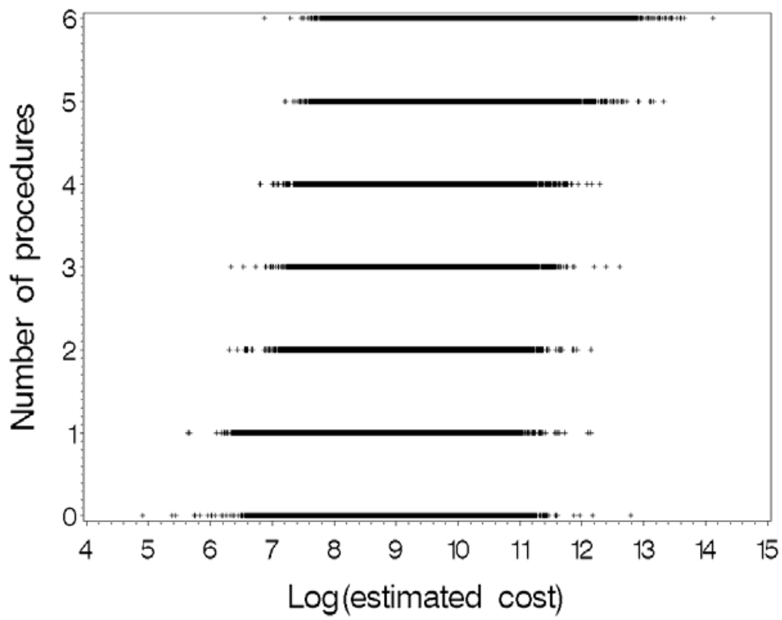
Infections Due to Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Figure 11: Plot of dependent variable (log of estimated cost) vs. independent variable (risk of mortality)



Infections Due to Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Figure 12: Plot of dependent variable (log of estimated cost vs. independent variable (number of procedures)



Infections Due to Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

The linear regression modeling revealed that age, gender, length of stay, number of procedures, severity of illness, risk of mortality, primary payer, source of admission, and number of severe comorbidities were significant predictors of the log-transformed estimated total costs (Table 8). In this model, the reference patient was female, was covered by a commercial health plan, was routinely admitted by her physician, had mild severity of illness and risk of mortality, was billed based on a medical DRG, did not have an HAI during the hospital stay, and was routinely discharged from inpatient care. Both SOI and comorbidities were better predictors as categorical variables rather than as indicator variables or discrete random variables, and for this reason were included in the model as categorical variables. The SE of the parameter estimates are all small enough to warrant no further examination of potential collinearity problems with the model,

Table 8: Regression parameter estimates (dependent variable: log of estimated cost)

Variable	Definition	DF	Estimate	SE	t value	Pr>|t|	VIF
Intercept	--	1	7.715	0.003	2545.93	<.0001	
hai	HAI indicator	1	0.567	0.017	33.75	<.0001	1.010
soi2	Moderate severity of illness	1	0.159	0.002	85.26	<.0001	1.600
soi3	Major severity of illness	1	0.393	0.003	135.98	<.0001	2.227
soi4	Extreme severity of illness	1	0.872	0.006	138.19	<.0001	2.440
comorb	Indicator of any comorbidity	1	0.125	0.002	61.55	<.0001	1.729
npr	Number of procedures	1	0.129	0.001	248.78	<.0001	1.460
ndx	Number of diagnoses	1	0.003	0.000	6.36	<.0001	2.190
medsurg	Surgical DRG indicator	1	0.745	0.002	408.51	<.0001	1.479
age	Age in years	1	0.003	0.000	50.28	<.0001	2.795
gender	Male gender indicator	1	0.143	0.002	89.41	<.0001	1.097
disch	Non-routine discharge indicator	1	0.210	0.002	108.64	<.0001	1.317
mcare	Medicare indicator	1	0.008	0.002	3.99	<.0001	1.905
mcaid	Medicaid indicator	1	-0.027	0.002	-11.71	<.0001	1.302
self	Self-pay indicator	1	0.081	0.004	21.34	<.0001	1.106
other	Other insurance indicator	1	0.051	0.003	14.50	<.0001	1.080
ER	Admitted from ER indicator	1	0.213	0.002	119.65	<.0001	1.471
ltc	Admitted from long-term care	1	0.224	0.011	19.58	<.0001	1.011
trans	Transfer from another hospital	1	0.299	0.004	80.21	<.0001	1.084
rom2	Moderate risk of mortality	1	0.012	0.002	5.13	<.0001	1.929
rom3	Major risk of mortality	1	0.119	0.004	31.71	<.0001	1.869
rom4	Extreme risk of mortality	1	0.181	0.007	25.68	<.0001	2.165

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

and this is confirmed by the small variance inflation factors. The r-square value was .566 and the model produced a highly significant F value (Table 9).

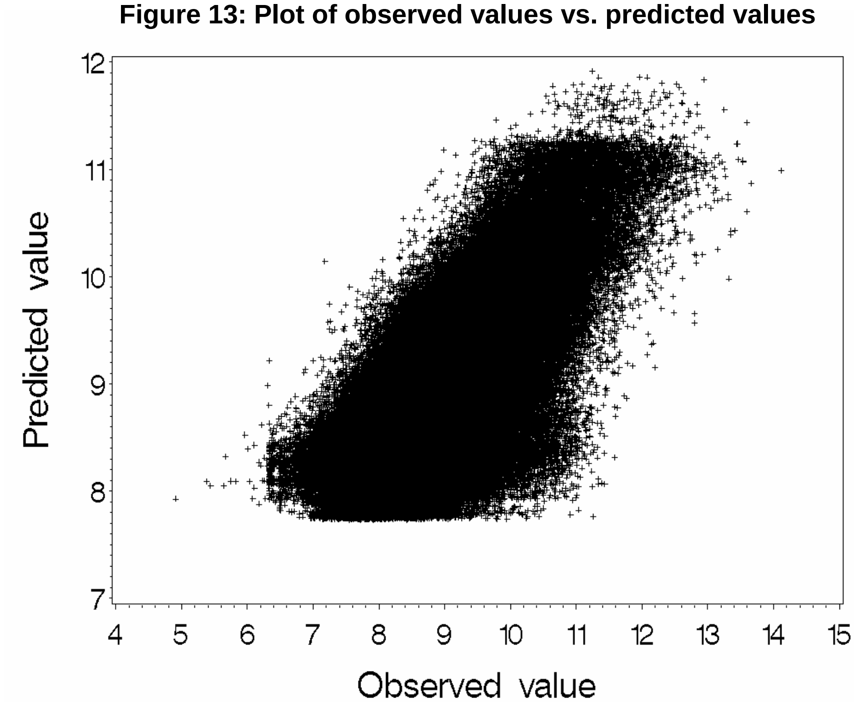
Table 9: Analysis of Variance (dependent variable: log of estimated cost)

Source	DF	Sum of Squares	Mean Square	F value	Pr > F
Model	21	193852	9218.2	32912.4	<.0001
Error	529818	148393	0.280		
Total	529839	341975			

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

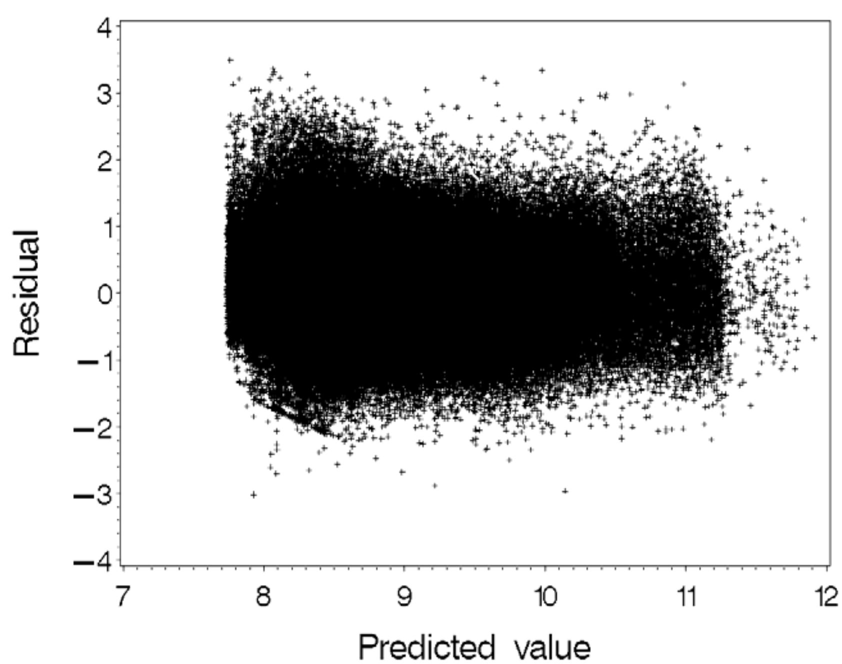
The plot of predicted values vs. observed values demonstrated a good linear relationship (Figure 13). The plot of residuals vs. predicted values indicated no egregious variance problems with the model (Figure 14), although this plot did reveal perhaps a slight narrowing of the range of residuals for log-transformed

Figure 13: Plot of observed values vs. predicted values



Infections Due to Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

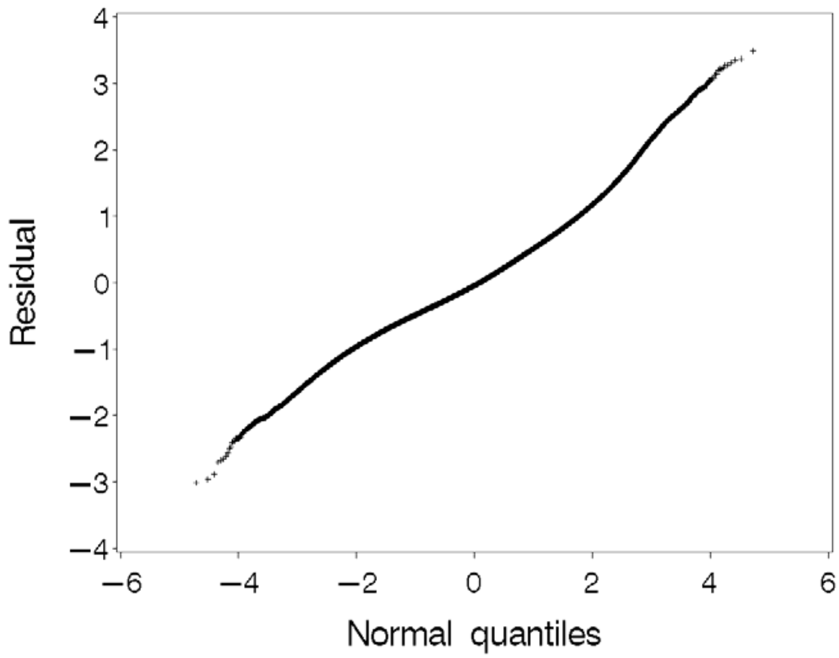
Figure 14: Plot of predicted values vs. residuals



Infections Due to Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

estimated costs greater than approximately 9. The Normal Q-Q plot of the residuals showed a reasonably normal distribution (Figure 15).

Figure 15: Normal Q-Q plot



Infections Due to Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Overall the mean predicted values were substantially higher for patients with HAI compared to patients without HAI (Table 10). When stratified by severity

Table 10: Mean predicted values (dependent variable: log of estimated cost)			
Stratifier	**Predicted values (In dollars per discharge)**		
None	**Medical/surgical pts without HAI**	**Medical/surgical pts with HAI**	**Pr > |t|**
	8.846	10.371	<.0001
Severity of illness	**Medical/surgical pts without HAI**	**Medical/surgical pts with HAI**	**Pr > |t|**
Minor	8.619	9.239	<.0001
Moderate	8.814	9.640	<.0001
Major	9.196	10.210	<.0001
Extreme	10.177	11.131	<.0001
Severe co-mobidities	**Medical/surgical pts without HAI**	**Medical/surgical pts with HAI**	**Pr > |t|**
No	8.602	10.410	<.0001
Yes	8.969	10.363	<.0001

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

of illness or presence of any severe comorbidities, the mean predicted values were again substantially higher for patients with HAI compared to patients without HAI.

Linear regression modeling of the log-transformed estimated cost per day revealed similar findings. Statistically significant predictors were the same as the previous model, although the parameter estimates were substantially smaller since estimated cost per day is a smaller quantity than estimated cost (Table 11).

Table 11: Regression parameter estimates (dependent: log of estimated cost per day)

Variable	Definition	DF	Estimate	SE	t value	Pr>|t|	VIF
Intercept	--	1	6.873	0.002	2771.32	<.0001	
hai	HAI indicator	1	-0.100	0.014	-7.24	<.0001	1.010
soi2	Moderate severity of illness	1	0.020	0.002	12.90	<.0001	1.600
soi3	Major severity of illness	1	0.030	0.002	12.87	<.0001	2.227
soi4	Extreme severity of illness	1	0.084	0.005	16.17	<.0001	2.440
comorb	Indicator of any comorbidity	1	0.073	0.002	43.84	<.0001	1.729
npr	Number of procedures	1	0.099	0.000	232.82	<.0001	1.460
ndx	Number of diagnoses	1	-0.013	0.000	-38.92	<.0001	2.190
medsurg	Surgical DRG indicator	1	0.582	0.001	390.18	<.0001	1.479
age	Age in years	1	0.004	0.000	99.94	<.0001	2.795
gender	Male gender indicator	1	0.084	0.001	64.22	<.0001	1.097
disch	Non-routine discharge indicator	1	-0.065	0.002	-41.26	<.0001	1.317
mcare	Medicare indicator	1	-0.030	0.002	-17.80	<.0001	1.905
mcaid	Medicaid indicator	1	-0.009	0.002	-4.84	<.0001	1.302
self	Self-pay indicator	1	0.071	0.003	23.00	<.0001	1.106
other	Other insurance indicator	1	-0.011	0.003	-3.87	<.0001	1.080
ER	Admitted from ER indicator	1	0.120	0.001	82.40	<.0001	1.471
ltc	Admitted from long-term care	1	-0.165	0.009	-17.60	<.0001	1.011
trans	Transfer from another hospital	1	-0.077	0.003	-25.28	<.0001	1.084
rom2	Moderate risk of mortality	1	0.005	0.002	2.43	<.0001	1.929
rom3	Major risk of mortality	1	0.063	0.003	20.56	<.0001	1.869
rom4	Extreme risk of mortality	1	0.189	0.006	32.80	<.0001	2.165

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

It is noteworthy that, in addition to relatively large changes in magnitude compared to most other predictors, the parameter estimates for the HAI indicator variable and the non-routine discharge indicator variable both have a negative sign in the cost per day model. As with the previous model, there are no problems with collinearity based on small standard errors of the parameter estimates and small variance inflation factors. The anova table shows that this model also produced a highly significant F-value, but the r-square value (.453) was marginally smaller than the previous model (Table 12).

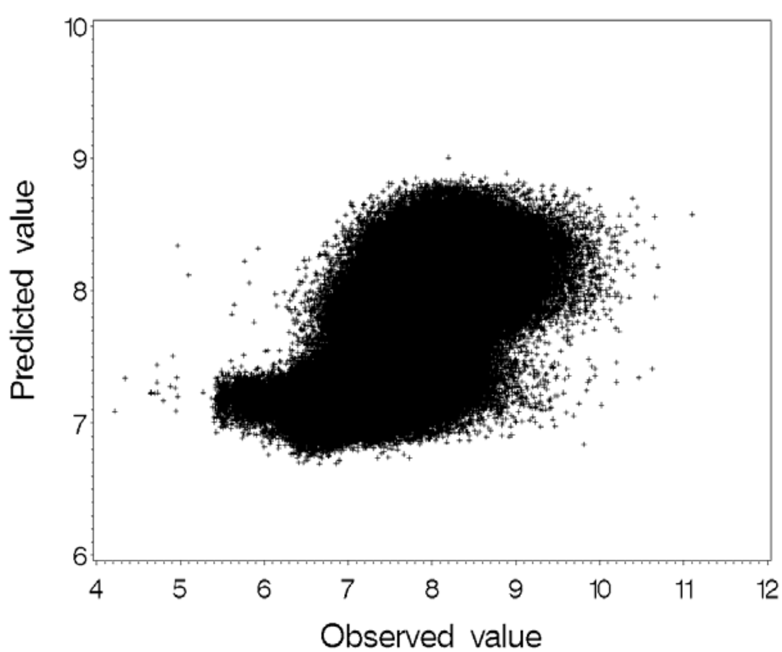
Table 12: Analysis of Variance (dependent variable: log of estimated cost per day)

Source	DF	Sum of Squares	Mean Square	F value	Pr > F
Model	21	82315	3919.8	32912.4	<.0001
Error	529818	99400	.188		
Total	529839	181715			

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

The plot of predicted values vs. observed values again demonstrated a good linear relationship (Figure 16). The plot of predicted values vs. residuals showed no overt pattern, indicating that the variance remains fairly constant

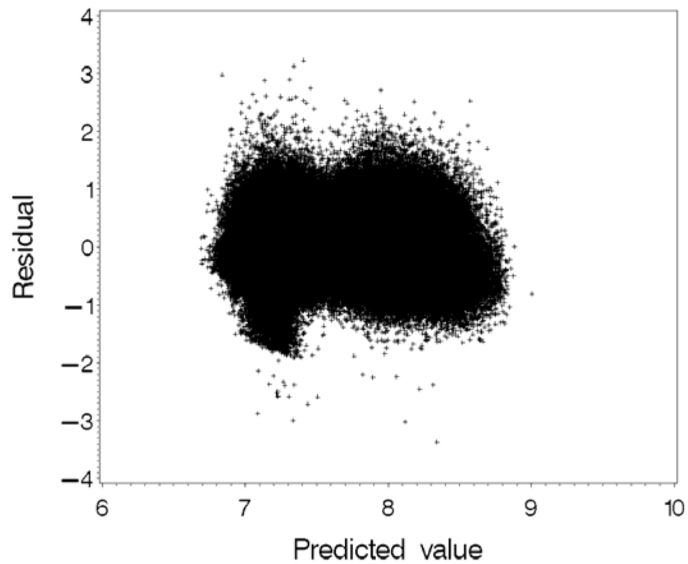
Figure 16: Plot of observed values predicted values



Infections Due to Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

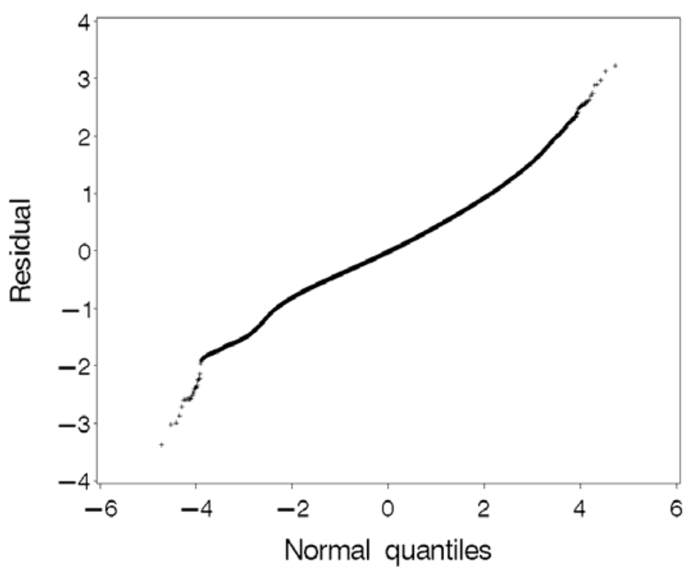
across the entire range of values (Figure 17). The normal Q-Q plot again reveals a reasonably normal distribution (figure 18). The mean predicted values are

Figure 17: Plot of predicted values vs. residuals



Infections Due to Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Figure 18: Normal Q-Q plot



Infections Due to Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

smaller and the differences are smaller between patients with an HAI and patients without an HAI (Table 13). For patients with minor severity of illness, the mean predicted values are higher for patients without an HAI. For patients with moderate severity of illness, the predicted values are not statistically different.

Table 13: Mean predicted values (dependent variable: log of estimated cost per day)			
Stratifier	**Predicted values (ln dollars per day)**		
None	**Medical/surgical pts without HAI**	**Medical/surgical pts with HAI**	**Pr > |t|**
	7.546	7.805	<.0001
Severity of illness	**Medical/surgical pts without HAI**	**Medical/surgical pts with HAI**	**Pr > |t|**
Minor	7.537	7.419	<.0297
Moderate	7.519	7.534	<.5475
Major	7.560	7.732	<.0001
Extreme	7.945	8.094	<.0001
Severe co-mobidities	**Medical/surgical pts without HAI**	**Medical/surgical pts with HAI**	**Pr > |t|**
No	7.492	7.878	<.0001
Yes	7.574	7.784	<.0001

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

The SAS mixed procedure creates all necessary indicator variables, so categorical variables for severity of illness, risk of mortality, primary payer, and source of admission were used rather than the corresponding indicator variables (Table 14). The mixed model was fit initially with the same dependent and independent variables as the previous models, except that hospital (identified by HospCode) was added as a random effect and during testing major diagnostic category (MDC) proved to be a statistically significant fixed effect. This model (Mixed model 1) showed a significant effect of hospital (Table 15), although the calculated intraclass correlation (ICC) was quite modest (.078). The model fit statistics are given in Table 16.

ICC = .02277/(.02277+.2659) = .078

Table 14: Indicator to categorical variable crosswalk

Indicator Variable	Definition	Categorical variable
reference	Mild severity of illness	soi
soi2	Moderate severity of illness	soi
soi3	Major severity of illness	soi
soi4	Extreme severity of illness	soi
reference	Commercial insurance	pay1
mcare	Medicare indicator	pay1
mcaid	Medicaid indicator	pay1
self	Self-pay indicator	pay1
other	Other insurance indicator	pay1
reference	Routinely admitted	asource
ER	Admitted from ER indicator	asource
ltc	Admitted from long-term care	asource
trans	Transfer from another hospital	asource
reference	Mild risk of mortality	rom
rom2	Moderate risk of mortality	rom
rom3	Major risk of mortality	rom
rom4	Extreme risk of mortality	rom

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Table 15: Mixed model 1 covariance parameter estimates

Parameter	Subject	Estimate	SE	Z value	Pr > Z
Intercept	HospCode	.02277	.00442	5.15	<.0001
Residual		.2659	.00052	514.29	<.0001

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Table 16: Mixed model 1 fit statistics

-2 Res Log Likelihood	801199.2
AIC (smaller is better)	801203.2
AICC (smaller is better)	801203.2
BIC (smaller is better)	801207.2

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

The mixed model was also tested and fit with patient-level interaction terms with the hospital random effect (Mixed model 2). Primary payer (pay1), severity of illness (soi), major diagnostic category (MDC), surgical DRG indicator (medsurg), and source of admission (asource) all had statistically significant interactions with hospital (Table 17). The model fit statistics are given in Table 18 and all show a substantial decrease, indicating improved model fit. The calculated intraclass correlation (.283) is substantially larger than the ICC from the Mixed model 1.

ICC=(.0225+.0015+.0024+.0176+.0177+.0205)/(.0225+.0015+.0024+.0176+.0177+.0205+.2081) = .283

Table 17: Mixed model 2 covariance parameter estimates

Parameter	Subject	Estimate	SE	Z value	Pr > Z
Intercept	HospCode	.02253	.00630	3.58	<.0001
pay1	HospCode	.00152	.00023	6.65	<.0001
soi	HospCode	.00240	.00038	6.30	<.0001
MDC	HospCode	.01755	.00096	18.30	<.0001
medsurg	HospCode	.01767	.00365	4.84	<.0001
asource	HospCode	.02053	.00041	6.07	<.0001
Residual		.2081	.00052	513.50	<.0001

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Table 18: Mixed model 2 fit statistics

-2 Res Log Likelihood	675522.7
AIC (smaller is better)	675536.7
AICC (smaller is better)	675536.7
BIC (smaller is better)	675550.9

Infections Due To Medical Care in Oregon Hospitals, 2003-2005
Source: Oregon Hospital Discharge Data

Discussion

Key Findings

There were several key findings from this study. First, the crude proportion of HAI cases is much smaller than reported in recent literature, most likely a consequence of the narrow case definition and under-reporting of HAI cases. Second, estimated costs per discharge were much higher for patients with an HAI compared to patients without an HAI, a result that is consistent with recent literature. Third, regression modeling demonstrated that HAI was an important predictor of costs per discharge (log dollars) and that the predicted estimated costs (log dollars) were significantly higher for patients with an HAI compared to those without an HAI, even after controlling for numerous dependent variables including severity of illness and presence of any severe comorbidity. Finally, a linear mixed model accounting for the data being clustered by hospital demonstrated significant interactions with the hospital term, which implies that there is additional hospital-level variability not accounted for in the two simple linear models.

Although recent studies speculate that 5-15% of all inpatients suffer from an HAI, the crude proportions in Oregon hospitals during calendar years 2003 to 2005 were far lower. The crude proportion of HAI cases in 2005 (1.84 per 1000 medical and surgical admissions) is over 27 times smaller than the lower end of the speculated range. This could be due to several factors. First, the AHRQ algorithm has a very narrow definition of HAI, most likely underestimating the true proportion of HAI cases. HAI cases are probably under-reported since this has not been a priority in the past and many may manifest after discharge, again causing the proportion of HAI cases to be underestimated. In addition, the proportion of admissions for common acute conditions, such as pneumonia and appendicitis, is traditionally lower on the west coast compared with the rest of the nation. This trend could possibly extend to HAI cases as well. Finally, the speculated range reported in recent literature may simply overestimate the true proportion of HAI cases.

The adjusted proportions of HAI cases (directly adjusted by severity of illness category) were just slightly different than the crude proportions. During calendar years 2004 and 2005 it is quite plausible that these differences are due to random chance alone. During 2003 random chance may be part of the explanation, but it is much less plausible that the difference is due to random chance alone. This difference could be due to differential misclassification of HAI case, which is consistent with the AHRQ algorithm underestimating the true number of cases. This difference could also be due to differential misclassification of the severity of illness, specifically that lower weighted cases (major and extreme) were misclassified as higher weighted cases (minor and moderate). However, this would require differentially misclassifying thousands of patients in order to produce a very small decrease in the risk-adjusted proportion. This seems implausible unless it is an artifact of major coding or grouping changes. Otherwise, there is no reason to expect that this bias would be substantially larger during calendar year 2003.

The estimated cost per discharge was substantially higher for patients with HAI compared to patients without HAI, and this proved true even after stratification by severity of illness. These results are consistent with recent studies that found excess costs for patients with HAI even after controlling for severity of illness. In the regression models presence of an HAI, severity of illness, and the presence of any severe comorbidity were all statistically significant predictors of estimated cost per discharge. The number of severe comorbidities did not differ substantially in patients with an HAI compared to patients without an HAI regardless of stratifying by severity of illness, principal diagnosis, or DRG.

In the first simple linear model the predicted costs per discharge in log dollars remained substantially higher for patients with an HAI compared to patients without an HAI, even after controlling for numerous dependent variables. This result is also consistent with recent HAI literature. The predictors with the largest magnitudes were extreme severity of illness, surgical DRG, presence of an HAI, major severity of illness. While severity of illness is clearly an important predictor in the model, it does not fully explain the huge cost differences for patients with an HAI compared to patients without an HAI. In fact, when the predicted values were stratified by severity of illness, significant cost differences (log dollars) remained across all severity of illness categories. In addition, the predicted costs per discharge for patients with an HAI are nearly equal when stratified by presence of a severe comorbidity, implying that costs per discharge for patients with an HAI are not different regardless of the presence of a severe comorbidity. When stratified by the presence of a severe comorbidity, the predicted costs per discharge were significantly higher for patients with an HAI compared to patients without an HAI. This result indicates that, although the presence of a severe comorbidity proved to be a significant predictor of costs per discharge (log dollars), this does not fully explain the cost higher inpatient costs for patients with an HAI compared to patients without an HAI.

In the second simple linear model, the predicted costs per day in log dollars were also generally higher for patients with an HAI compared to patients without an HAI, but not for patients with mild and moderate severity of illness. This is important since at least 80% of medical and surgical inpatients had mild or moderate severity of illness in each calendar year from 2003 to 2005. This result implies that, for the vast majority of medical and surgical patients in Oregon, the costs per day for patients with an HAI are not higher than the costs per day for patients without an HAI. This finding is not surprising since the increased costs for patients with an HAI are probably due to longer length of stay compared to patients without an HAI. However, the policy importance is uncertain, as the statewide policy will likely be driven by total costs rather than average cost per day.

In the second model, the regression coefficients for the HAI indicator and the non-routine discharge indicator had a negative slope, which is consistent with a direct relationship to length of stay (inverse relationship to 1/length of stay). This is an intuitive result since one expected outcome of an HAI is increased length of stay and non-routine discharges, such as to long-term care or rehabilitation, also imply extended length of stay. It is interesting to note that the regression coefficient for the surgical DRG indicator was substantially larger than the other regression coefficients in the second model, giving it a disproportionately larger impact on the log of estimated costs per day. This implies that inpatient surgery is the major driving force influencing costs per day for Oregon medical and surgical inpatients. This is also an intuitive result since, if other variables are held constant, a surgical inpatient will generally incur more hospital charges than a medical inpatient.

The linear mixed models identified several hospital-level interactions that produced additional variability by hospital. These included hospital interactions with primary payer, severity of illness, major diagnostic category, presence of a surgical DRG, and the source of admission. The interactions are all intuitive, particularly for primary payer since it is widely known that the mix of payers can vary substantially by hospital. It is also clear that hospitals serve different patient populations, who may present with different diagnoses and different severity of illness. The services offered by the hospital may influence the proportion of patients admitted for surgery, the mix of surgeries performed, and the route of admission (a trauma center may admit a larger proportion of patients through the emergency department). However, the second linear mixed model, while a substantial improvement over the model without interactions, explained only about 28% of the variability. Regardless, there is significant variability between and with hospitals that is not explained by patient-level variables. The variability not accounted for in these models could be due to variability by geography, variability by physician, another factor that causes additional variability by hospital, or some other factor, such as genetics, that causes additional variability by patient. Future work in this area could focus on identifying and modeling the sources of additional variability.

Strengths and Limitations

This study has several key strengths. One of these is the large magnitude of the difference in estimated costs per discharge for patients with HAI compared to patients without HAI. It is extremely unlikely that this result is due to random chance. The AHRQ algorithm does cause limited selection bias by age and gender, but this effect is modest and cannot explain the large difference. A large number of additional cases, perhaps due to widespread under-reporting or misclassification, would reduce this difference but also substantially increase the total statewide estimated excess costs due to HAI cases.

A second key strength is sample space, which originally included three calendar years of discharge records from Oregon inpatients treated in acute care hospitals located in Oregon. Ultimately over 550,000 inpatient discharge records met the inclusion criteria. In addition to alleviating concerns about selection bias, the large sample size provides sufficient statistical power to detect relatively small differences in the log-transformed estimated costs and offers ample degrees of freedom for tests of statistical significance. A third key strength is the extensive use of “out of the box” applications, published protocols, and data sets that are routinely available, which allows the study’s methods to be feasibly replicated in other settings. A final strength of this study is the strong association between the observed values and the predicted values in the first simple linear model. The association is visually obvious and fitting the model did not cause egregious problems with variance of the residuals or with collinearity.

Several limitations should also be noted. The most important of these is that the study relies on indirect identification of cases. By using de-identified data it was not possible to verify the case finding with medical record reviews or evaluate the sensitivity and specificity. It was also not possible to determine if infection diagnoses were present on admission or occurred during the hospital stay. This, in turn, makes it impossible to determine if it is appropriate to exclude patients with HAI as a principal diagnosis. While this exclusion prevents double-counting patients readmitted with the same HAI, it also prevents identifying HAI cases that manifested after discharge and resulted in readmission. However, the magnitude of any bias caused by false positive cases is most likely very small compared to bias caused by under-reporting HAI cases. Again, the anticipated effect of underestimating HAI cases is to reduce the difference in estimated costs per discharge for patients with HAI compared to patients without HAI, but also to increase the total statewide excess costs due to HAI.

Another important limitation is that, even despite the extensive literature devoted to HAI, there currently are no consensus benchmarks for comparison. Some might argue that one case is too many. In reality hospitals reporting zero cases most likely are simply not reporting their HAI cases. So at least in the short run zero cases is not a realistic benchmark. Some might suggest that the benchmark is hospital-specific or physician-specific. This is probably a much more realistic approach since both hospitals and physicians vary and, more importantly, the reporting by both hospitals and physicians probably varies widely. The urban hospital with hand sanitation screen-savers on every computer and hand sanitizer dispensers at every bedside may have a higher proportion of HAI cases because they have the best reporting program (and perhaps the best infection control program), not because it is unsanitary or less safe than other hospitals.

The third limitation concerns calculating the hospital-specific and year-specific cost-to-charge ratio. Since this ratio is calculated from the gross patient revenue and total expenses reported in the audited financial statement, this calculation may be somewhat imprecise for several reasons. First, the gross patient revenue and total expenses also include outpatient revenue and expenses, which may occur at a different ratio than the inpatient revenue and expenses. There are numerous factors that affect inpatient revenue and expenses and this study does not attempt to fully consider or control for these factors. The audited financial statements also reflect the fiscal year, and these do not necessarily coincide with the calendar year. Lastly, local government entities run some hospitals, and these may not be allowed to generate more patient revenue than expenses.

In addition to identifying other sources of variability, future work in this area could focus on validating the PSI 07 algorithm with medical record reviews. Doing this would help quantify the magnitude of bias and also help clarify why the crude proportions reported in this study are far lower than 5%, the lower end of the speculated range reported in recent literature. Although AHRQ is currently working on “gold standard” validation of nationwide data, validation with Oregon data would be much more compelling to Oregon policymakers. After an algorithm is validated and HAI cases can be identified with known sensitivity and specificity, discussions can then shift to quantifying the magnitude and causes of under-reporting. Further work in this area could also include establishing benchmarks and prospectively testing practical interventions intended to reduce the proportion of HAI cases in Oregon hospitals.

Conclusion

Hospital-acquired infections are a common problem in Oregon even despite using AHRQ’s highly conservative estimates of infections due to medical care. The statewide total estimated excess cost due to 1034 HAI cases reported during calendar years 2003 to 2005 is at least $21 million, or approximately $7 million per year on average. Due to the very conservative estimates and the high likelihood of under-reporting, the true number of cases and the true costs are probably much higher. If 5 -15% of all hospitalized patients develop HAI, then at least 17,000 Oregonians suffered from an HAI each year from 2003-2005. In addition, the excess costs cannot be explained away by differences in age, gender, severity of illness, or severe comorbidities.

At a minimum these costs estimates represent an opportunity to redirect precious health care resources towards more productive purposes, perhaps providing health insurance to some of the 576,000 Oregonians who currently are uninsured. Since there is now a “business case” for reducing infections due to medical care, the State should diligently work with hospitals to improve reporting and develop practical evidence-based interventions to eliminate most or all preventable HAI cases. This will reduce excess health care expenditures due to HAI cases, HAI morbidity, and the indirect costs of HAI to Oregonians.

References

3M Health Information Systems. “3M Grouper Products Concepts Manual.” Murray, UT: 3M Health Information Systems, 2005.

Agency for Healthcare Research and Quality. “Guide to Patient Safety Indicators.” Rockville, MD: AHRQ, 2003.

Agency for Healthcare Research and Quality. “Patient Safety Indicators: Software Documentation, Version 3.0.” Rockville, MD: AHRQ, 2006.

Brossette SE, Hacek DM, Gavin PJ, Kamdar MA, Gadbois KD, Fisher AG, and Peterson LR. “A laboratory-based, hospital-wide, electronic marker for nosocomial infection: the future of infection control surveillance?” *Am J Clin Pathol.* 2006: 125 (1), 34-39.

DeRyke CA, Lodise TP, Rybak MJ, and McKinnon PS. “Epidemiology, treatment, and outcomes of nosocomial bacteremic Staphylococcus aureus pneumonia.” *Chest.* 2005: 129 (3), 1414-1422.

Eggimann P and Pittet D. “Infection control in the ICU.” *Chest.* 2001: 120 (6), 2059-2093.

Elward AM, Hollenbeak CS, Warren DK, and Fraser VJ. “Attributable cost of nosocomial primary bloodstream infection in pediatric intensive care patients.” *Pediatrics.* 2005: 115 (4), 868-872.

Grundmann, et al. “How many infections are caused by patient-to-patient transmission in intensive care units?” *Crit Care Med.* 2005: 33 (5), 946-951.

Miller PR, Johnson JC, Karchmer T, Hoth JJ, Meredith JW, and Chang MC. “National nosocomial infection surveillance system: from benchmark to bedside in trauma patients.” *J Trauma.* 2006: 60 (1), 98-103.

Nan DN, Fernandez-Ayala M, Farinas-Alvarez C, Mons R, Ortega FJ, Gonzales-Macias J, and Farinas MC. “Nosocomial infection after lung surgery: incidence and risk factors.” *Chest.* 2005: 128 (4), 2647-2652.

Office for Oregon Health Policy and Research. “Profile of Oregon’s Uninsured, 2006.” Salem, OR: Department of Human Services and Office for Oregon Health Policy and Research, 2007.

Pennsylvania Health Care Cost Containment Council. “Hospital-Acquired Infection Data Collection and Methodology for ICD.9.CM Infection Screening Diagnosis Codes.” Harrisburg, PA: PHC4, 2006.

Pepin J, Valiquette L, and Cossette B. “Mortality attributable to nosocomial Clostridium difficile-associated disease during an epidemic caused by a hypervirulent strain in Quebec.” *CMAJ.* 2005: 173 (9), 1037-1042.

Rello J, Ollendorf DA, Oster G, Vera-Llonch M, Bellm L, Redman R, and Kollef MH. “Epidemiology and outcomes of ventilator-associated pneumonia in a large US database.” *Chest.* 2002: 122 (6), 2115-2121.

Roberts RR, Scott RD, Cordell R, Solomon SL, Steele L, Kampe LM, Trick WE, and Weinstein RA. “The use of economic modeling to determine the hospital costs associated with nosocomial infections.” *Clin Infect Dis.* 2003: 36 (1), 1424-1432.

Safdar N, Dezfulian C., Collard HR, and Saint S. “Clinical and economic consequences of ventilator-associated pneumonia: a systematic review.” *Crit Care Med.* 2005: 33 (10), 2184-2193.

Warren DK, Shukla SJ, Olsen MA, Kollef MH, Hollenbeak CS, Cox MJ, Cohen MM, and Fraser VJ. “Outcome and attributable cost of ventilator-associated pneumonia among intensive care unit patients in a suburban medical center.” *Crit Care Med.* 2003: 31 (5), 1312-1317.

Smith RL, Bohl JK McElearney ST, Friel CM, Barclay MM, Sawyer RG, and Foley EF. “Wound infection after elective colorectal surger.” *Ann Surg.* 2004: 239 (5), 599-605.

Tambyah PA, Knasinski V, and Maki DG. “The direct costs of nosocomial catheter-associated urinary tract infection in the era of managed care.” *J. Urology.* 2003: 170 (1): 339.

Weinstein RA, Siegel JD, and Brennan PJ. “Infection Control Report Cards – Securing Patient Safety. *NEJM.* 2005: 353 (3), 225-227.

Whitehouse JD, Friedman ND, Kirkland KB, Richardson WJ, and Sexton DJ. “The impact of surgical-site infections following orthopedic surgery at a community hospital and a university hospital: adverse quality of life, excess length of stay, and extra cost.” *Infect Control Hosp Epidemiol.* 2002: 23 (4), 174-176.

Yogari JS, Elward AM, and Fraser VJ. “Rate, risk factors, and outcomes of nosocomial primary bloodstream infection in pediatric intensive care unit patients.” *Pediatrics.* 2002: 110 (3), 481-485.

Zhao SZ, Dodge WE, Spalding W, Barr CE, and Li JZ. “Length of hospital stay and cost of Staphylococcus and Streptococcus infections among hospitalized patients.” *Clin Ther.* 2002: 24 (5), 818-834.

Appendix A: Comorbidities Flagged by AHRQ Algorithm

AIDS

Alcohol abuse

Congestive heart failure

Chronic blood loss anemia

Chronic pulmonary disease

Coagulopathy

Deficiency anemia

Depression

Diabetes

Diabetes with chronic complications

Drug abuse

Fluid and electrolyte disorders

Hypothyroidism

Hypertension

Liver disease

Lymphoma

Metastatic cancer

Obesity

Other neurological disorders

Paralysis

Peptic ulcer disease with bleeding

Peripheral vascular disease

Pulmonary circulation disease

Psychoses

Renal failure

Rheumatoid arthritis

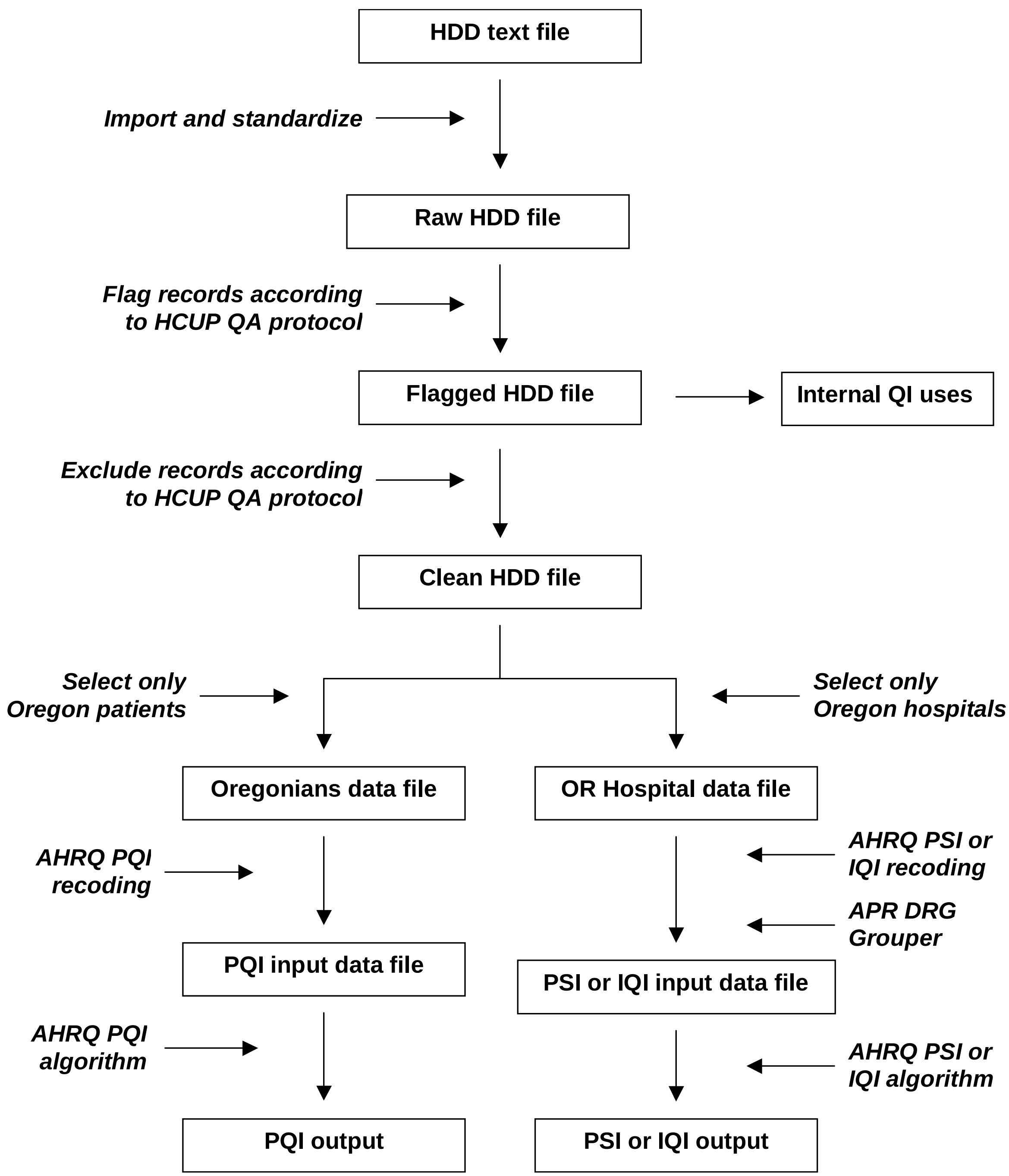
Solid tumor without metastasis

Valvular disease

Weight loss

Appendix B: Preliminary Data Preparation

Before conducting any type of analysis on the Oregon hospital discharge data (HDD), an extensive standardization and quality assurance protocol is followed. This is illustrated in the flow chart below:



<u>Standardization</u>

Hospital discharge data text files for calendar years 2003 to 2005 were imported into SPSS version 14.0.0. The data layout, variable names, variable formats, and coding of some categorical variables have evolved over time, although they have not changed since 2001. The 2001 HDD was used as the standard for formats and variable names.

To standardize the data files, the data layout, variable names, and variable formats were revised to match the 2001 HDD as closely as possible. Categorical variables were not recoded. Variables for calendar year and OHPR hospital ID were added. A unique identifier was also added to each record and, if included in that year's HDD file, the original unique identifier was removed. After completing these steps, a new SPSS data file was saved. This was considered the "raw" HDD file.

<u>HCUP Quality Assurance</u>

The Health Care Utilization Project (HCUP), located within AHRQ, publishes widely accepted quality assurance procedures for hospital discharge data (HCUP, 2006). These were used to clean the raw HDD file. The procedures can be viewed as distinct validations, often requiring multiple distinct steps:

1. Validating the length of stay (LOS)
    a. *Admit date is not missing*
    b. *Discharge date is not missing*
    c. *Admission date is before discharge date*
    d. *LOS is not over 365 days*

2. Validating the age

    a. *Date of birth (DOB) is not missing*

    b. *DOB is not after admission date*

    c. *DOB is not before January 1, 125 years prior to the calendar year of the discharge data*

    d. *Age in years is not greater than 124*

3. Validating the gender

    a. *Gender is not missing*

4. Validating the diagnoses

    a. All diagnosis (dx) codes are valid ICD9-CM diagnosis codes as of the discharge date (invalid diagnoses recoded as missing)

    b. *Principal dx is not missing*

    c. *Dx is not inconsistent with gender*

    d. *Dx is not inconsistent with age*

    e. *Major Diagnostic Category (MDC) is not missing*

    f. *Diagnosis-Related Group (DRG) is not missing*

    g. *DRG is not 469*

5. Validating the procedures

    a. All procedure (px) codes are valid ICD9-CM procedure codes as of the discharge date (invalid procedures recoded as missing)

    b. *Px is not inconsistent with gender*

    c. *Px is not inconsistent with age*

d. Procedure date is within acceptable range (invalid dates and associated procedures recoded as missing)

6. Validating the total charges

    a. *Total charges are over $25*

If a discharge record violated any of the individual steps listed above, a value was written to the appropriate flag variable to identify the violation. A data file of all discharge records and flags was saved in order to aggregate and inspect the violations. If a discharge record violated any of the italicized steps, then that record was excluded.

In order to validate diagnosis and procedure codes as of the discharge dates, data sets of valid ICD9-CM codes for each federal fiscal year were acquired from The Centers for Medicare and Medicaid Services (CMS). The revised codes are implemented each October 1st and the revisions are, with rare exception, published annually. Diagnosis and procedure codes from each calendar year of HDD were compared to the ICD-9-CM data file corresponding to the same federal fiscal year. Discharges after September 30th were also validated against new codes added to (or deleted from) the subsequent year's ICD-9-CM data file.

If a diagnosis code in the HDD file was contained in that year's ICD9-CM data set (or in the subsequent year's new diagnosis codes for discharges after September 30th), then it was considered valid as of the discharge date. Similarly, if a procedure code in the HDD file was contained in that year's ICD9-CM (or in the subsequent year's new procedure codes for discharges after September

30th), it was considered valid as of the discharge date. Invalid diagnosis and procedure codes caused values to be written to the appropriate flag variable.

HCUP also publishes ICD-9-CM code lists for flagging maternal, neonatal, female, and male diagnoses and procedures. The flags are used to validate certain diagnoses and procedures against age and/or gender (HCUP, 2006). The code lists are updated as needed to maintain consistency with revisions to the ICD-9-CM or to correct errata. The list appropriate for each year of HDD was used for this evaluation (see Appendix B.1); some lists covered multiple years. Errata in published code lists were corrected prior to performing the validation. The following comparisons were used to determine inconsistency with age and/or gender:

1. Age in years greater than zero and presence of a neonatal diagnosis
2. Age in years under 10 and presence of a maternal diagnosis or procedure
3. Age in years over 55 and presence of a maternal diagnosis or procedure
4. Female gender and presence of a male diagnosis or procedure
5. Male gender and presence of a female diagnosis or procedure

Over time some existing hospitals closed and several new hospitals opened. This required validating the hospital identification variable. OHPR maintains a hospital database for reporting purposes; this database was queried to create lists of hospitals that were open during each calendar year. The HDD for each calendar year was then limited to records from hospitals contained in that year's list.

The remaining data set, records meeting all the HCUP quality assurance standards and containing a valid hospital identification variable, was saved as a new SPSS file. This was considered the “clean” HDD file for that calendar year. One raw, flagged, and clean data set was created for each year from 2003 to 2005. In addition to the annual ICD-9-CM revisions and dynamic hospital lists, the layout and formatting of HDD text file also evolved. Essentially it was necessary to write unique programming code for each calendar year.

Appendix B.1: HCUP ICD9-CM Diagnosis and Procedure Codes

<u>Maternal diagnosis codes</u>

Beginning with calendar year 2002 data: 630 to 677; 796.5; V22.0 to V24.2; V27.0 to V27.9

Beginning with calendar year 2003 data: 630 to 677; 796.5; V22.0 to V24.2; V27.0 to V27.9, V65.11

<u>Maternal procedure codes</u>

Beginning with calendar year 2003 data: 720 to 7537, 754 to 7599

<u>Neonatal diagnosis codes</u>

Beginning with calendar year 2001 data: 277.01, 762.0 to 770.6, 770.8 to 778.5, 778.7 to 779.9, V29.0 to V29.9, and V30.00 to V39.2

Beginning with calendar year 2005 data: 277.01, 762.0 to 770.6, 770.8 to 778.5, 778.7 to 779.9, 796.6, V29.0 to V29.9, V30.00 to V39.2

<u>Female diagnosis codes</u>

Beginning with calendar year 2003 data: 016.60 to 016.76, 054.11, 054.12, 098.15 to 098.17, 098.35 to 098.37, 112.1, 131.01, 174.0 to 174.9, 179 to184.9, 198.6, 218.0 to 221.9, 233.1 to 233.3, 236.0 to 236.3, 256.0 to 256.9, 302.73, 302.76, 306.51, 306.52, 456.6, 611.5, 611.6, 614.0 to 677, 716.30 to 716.39, 752.0 to 752.49, 792.3, 795.00 to 795.09, 796.5, 867.4, 867.5, 878.4 to 878.7, 902.55, 902.56, 902.81, 902.82, 939.1, 939.2, 947.4, 996.32, V07.4, V10.40 to V10.44, V13.1 to V13.29, V22.0 to V25.01, V25.1, V25.3, V25.41 to V25.43, V25.5, V26.1, V26.51, V27.0 to V28.9, V45.51, V45.52, V49.81, V50.42, V52.4, V67.01, V72.3 , V72.4, V76.2, V76.46, V76.47

Beginning with calendar year 2004 data: 016.60 to 016.76, 054.11, 054.12, 098.15 to 098.17, 098.35 to 098.37, 112.1, 131.01, 174.0 to 174.9, 179 to184.9, 198.6, 218.0 to 221.9, 233.1 to 233.3, 236.0 to 236.3, 256.0 to 256.9, 302.73, 302.76, 306.51, 306.52, 456.6, 611.5, 611.6, 614.0 to 677, 716.30 to 716.39, 752.0 to 752.49, 752.81, 792.3, 795.00 to 795.09, 796.5, 867.4, 867.5, 878.4 to 878.7, 902.55, 902.56, 902.81, 902.82, 939.1, 939.2, 947.4, 996.32, V07.4, V10.40 to V10.44, V13.1 to V13.29, V22.0 to V25.01, V25.1, V25.3, V25.41 to V25.43, V25.5, V26.1, V26.51, V27.0 to V28.9, V45.51, V45.52, V49.81, V50.42, V52.4, V65.11, V67.01, V72.3 , V72.4, V76.2, V76.46, V76.47

Beginning with calendar year 2005 data: 016.60 to 016.76, 054.11, 054.12, 098.15 to 098.17, 098.35 to 098.37, 112.1, 131.01, 174.0 to 174.9, 179 to184.9, 198.6, 218.0 to 221.9, 233.1 to 233.3, 236.0 to 236.3, 256.0 to 256.9, 302.73, 302.76, 306.51, 306.52, 456.6, 611.5, 611.6, 614.0 to 677, 716.30 to 716.39, 752.0 to 752.49, 752.81, 792.3, 795.00 to 795.09, 796.5, 867.4, 867.5, 878.4 to 878.7, 902.55, 902.56, 902.81, 902.82, 939.1, 939.2, 947.4, 996.32, V07.4, V10.40 to V10.44, V13.1 to V13.29, V22.0 to V25.01, V25.1, V25.3, V25.41 to V25.43, V25.5, V26.1, V26.51, V27.0 to V28.9, V45.51, V45.52, V49.81, V50.42, V52.4, V65.11, V67.01, V72.3 to V72.41, V76.2, V76.46, V76.47, V84.02, V84.04

Female procedure codes

Beginning with calendar year 1996 data: 650 to 7599, 8781 to 8789, 8846, 8878, 8926, 9141 to 9149, 9217, 9614 to 9618, 9644, 9724, 9726, 9771 to 9775, 9816 to 9817, 9823, 9998

Male diagnosis codes

Beginning with calendar year 2001 data: 0164.0 to 0165.6, 054.13, 072.0, 098.12 to 098.14, 098.32 to 098.34, 131.03, 175.0 to 175.9, 185 to 187.9, 214.4, 222.0 to 222.9, 233.4 to 233.6, 236.4 to 236.6, 257.0 to 257.9, 302.74, 302.75, 456.4, 600 to 608.9, 752.51, 752.52; 752.63 to 752.69, 758.7, 788.32, 790.93, 792.2, 878.0 to 878.3, 939.3, V10.45 to V10.49, V13.61, V26.52, V50.2, V76.44, V76.45

Beginning with calendar year 2004 data: 0164.0 to 0165.6, 054.13, 072.0, 098.12 to 098.14, 098.32 to 098.34, 131.03, 175.0 to 175.9, 185 to 187.9, 214.4, 222.0 to 222.9, 233.4 to 233.6, 236.4 to 236.6, 257.0 to 257.9, 302.74, 302.75, 456.4, 600 to 608.9, 752.51, 752.52; 752.63 to 752.69, 758.7, 788.32, 790.93, 792.2, 878.0 to 878.3, 939.3, 959.13, V10.45 to V10.49, V13.61, V26.52, V50.2, V76.44, V76.45

Beginning with calendar year 2005 data: 0164.0 to 0165.6, 054.13, 072.0, 098.12 to 098.14, 098.32 to 098.34, 131.03, 175.0 to 175.9, 185 to

187.9, 214.4, 222.0 to 222.9, 233.4 to 233.6, 236.4 to 236.6, 257.0 to 257.9, 302.74, 302.75, 456.4, 600 to 608.9, 752.51, 752.52; 752.63 to 752.69, 758.7, 788.32, 790.93, 792.2, 878.0 to 878.3, 939.3, 959.13, V10.45 to V10.49, V13.61, V26.52, V50.2, V76.44, V76.45, V84.03

Male procedure codes

Beginning with calendar year 1988 data: 600 to 649.9, 879.1 to 879.9, 982.4, 999.4 to 999.6