March 2015

The Effect of a New Hospital-Based Congestive Heart Failure Care Protocol on Rate of 30-Day Readmission Among CHF Patients

Eric A. Cohen
University of Massachusetts Amherst

THE EFFECT OF A NEW HOSPITAL-BASED CONGESTIVE HEART FAILURE CARE PROTOCOL ON RATE OF 30-DAY READMISSION AMONG CHF PATIENTS

A thesis presented

by

ERIC A. COHEN

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

February 2015

Public Health
Biostatistics and Epidemiology

THE EFFECT OF A NEW HOSPITAL-BASED CONGESTIVE HEART FAILURE CARE PROTOCOL ON RATE OF 30-DAY READMISSION AMONG CHF PATIENTS

A thesis presented

by

ERIC A. COHEN

Approved as to style and content by:

_____

Penelope Pekow, Chair

_____

Nicholas Reich, Member

_____

Brian Whitcomb, Member

_____

Edward Stanek, Department Head
Department of Biostatistics

ABSTRACT

THE EFFECT OF A NEW HOSPITAL-BASED
CONGESTIVE HEART FAILURE CARE PROTOCOL
ON RATE OF 30-DAY READMISSION AMONG CHF PATIENTS

FEBRUARY 2015

ERIC A. COHEN

B.A., SWARTHMORE COLLEGE

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Penelope Pekow

Approximately 20% of congestive heart failure (CHF) patients are readmitted within 30 days of hospital discharge, a rate which may be affected by in-hospital and post-discharge care. Reducing this rate is important to hospitals, both to improve outcomes and to avoid reductions in Medicare reimbursement. Assessing outcomes within a short post-discharge window best measures the impact of the care, planning, and followup of that admission; but most research on the effects of changes in CHF care has measured outcomes over periods longer than 30 days, adding the unpredictable long-term course of CHF to the factors affecting the outcome. As well, almost no studies to date have included the appreciable effects of CHF comorbidities in their analyses.

This study addresses these needs by measuring rates of 30-day all-cause readmission, and by adjusting for comorbidities and demographic factors in our analysis.

We hypothesize that an improved CHF care protocol including both in-hospital and post-discharge components will reduce the risk of readmission, and may alter the rate of change of that risk.

We have analyzed as an interrupted time series data on 2764 discharges of CHF patients from a hospital that implemented such a change to assess the effect of the new protocol on the readmission risk and on the trend in that risk, comparing outcomes in the 22 months preceding introduction of the new protocol to those in the first 31 months of full implementation. Using multiple logistic regression, we have tested for an association between the new protocol and both the unadjusted risk of readmission, and that risk in a model including comorbidities and demographic factors as covariates.

Neither model found a statistically significant association between introduction of the protocol and log-odds of readmission (unadjusted $p = 0.847$, adjusted $p = 0.755$) or between introduction of the protocol and change in risk of readmission over time (unadjusted $p = 0.437$, adjusted $p = 0.313$).

These results, in comparison with other published results, can clarify what changes to care protocols have been shown to be effective. Further, post hoc power analysis of this study can inform study design for further research.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Congestive heart failure (CHF) is a substantial health burden in the United States. In 2012 prevalence in U.S. adults was 2.4% (1), annual incidence in developed countries is estimated at 5 to 10 per 100,000 (2), and a U.S. resident's lifetime risk for CHF is one in five (3). Outcomes are poor as well: one-year mortality ranges from 7% for cases of mild CHF to 28% for cases of severe CHF (4), and among Medicare beneficiaries hospitalized for CHF, half will not survive three years (1). Secondary to the human cost, CHF imposes a substantial economic burden: Healthcare spending on CHF is 1% to 2% of total worldwide healthcare spending (2), and the indirect economic loss — due to mortality and morbidity — has been estimated at over nine billion dollars in the U.S. in 2012 (1).

Readmission is a major feature of both poor patient outcomes and healthcare cost. A patient hospitalized for CHF will often be quickly readmitted for the same or a related complaint — the average 15-day readmission rate for CHF is 12.5% (5). Wide differences in outcome show that a high readmission rate is not inevitable: adjusted for severity, 15% of US hospitals have CHF readmission rates 5% or more higher than expected, and 5% have rates 5% or more lower (5). This suggests that the issue is amenable to public health intervention: certain practices can reduce the risk of CHF readmission. Recent healthcare policy has also affected this concern, as the Affordable Care Act of 2010 provides for escalating penalties for hospitals that do not reduce readmissions categorized as unnecessary readmissions (6). Identifying measures that reduce CHF patient readmission, particularly readmission within 30 days of discharge, is

therefore central to both improving patient outcomes and to allowing hospitals to avoid financial penalties.

Studies have shown that improved care protocols reduce readmissions. Interventions including patient education, telephone followup, medication management, and coordinated transitional care have been shown in various studies to reduce readmissions in Medicare patients with coronary artery disease, congestive heart failure, diabetes, chronic pulmonary disease, or some combination of those conditions (7). Medication reconciliation and scheduling followup appointments prior to discharge have been shown to reduce readmission in CHF patients (8).

However, although many CHF care protocols have been studied, all improved protocols have included primarily post-discharge interventions and few in-hospital interventions. But for a hospital attempting to improve care, or avoid reimbursement penalties, in-hospital care may be the preferred focus of effort. Changes to in-hospital procedures can be implemented more quickly and easily than changes requiring coördination with external agencies or care providers. As well, the costs and reimbursability of in-house care practices are more readily forecast.

As well, despite the many studies, little information is available focusing particularly on 30-day readmission; all previous studies concentrate on longer-term outcomes of interest. Although the long-term course of CHF progression is of real importance to assessing patient care, we believe that the unpredictability of this course, dependent as it is on the complex interaction of many comorbidities, life events, and the choices of patients and caregivers, decreases the utility of longer-term measures in assessing the quality of a given hospital admission.

Finally, although comorbidities are significant predictors of readmission risk in CHF patients (9), little of the available literature adjusts for these covariates.

Addressing these deficits, we have studied the effects on 30-day all-cause readmission of a CHF care protocol introduced by Baystate Medical Center in January 2009. This new protocol was centered on in-hospital changes to care. The available data includes demographic information and information on comorbidities at discharge, allowing us to adjust for these effects when testing the study hypotheses.

Epidemiology of CHF care protocols and risk of readmission

Multiple studies have investigated the association between improved CHF care protocols and risk of readmission in CHF patients.

Restricting ourselves to studies conducted since 2000, and concentrating on studies addressing (at least tangentially) measures similar to our outcome of interest, the studies surveyed have all been randomized controlled trials (RCTs) and have almost uniformly found a reduction in the risk of all-cause readmission for CHF patients receiving care under an improved protocol, relative to those receiving the standard of care (10–17). Of the eight studies surveyed, six found an improvement in this outcome (11,12,14,16–18), one found a statistically insignificant improvement (at 95% confidence) (15), and only one found no association between the new protocol and lower risk of all-cause readmission (10). Meta-analyses are common in this area, and of seven meta-analyses surveyed (all addressed only RCTs), six found an overall tendency towards lowered rates of all-cause readmission in patients treated under an improved care protocol relative to the standard of care (19–25), and only one found no association (26).

Of the studies surveyed, the COACH study described by Jaarsma et al. is notable for being by far the largest trial surveyed (1,023 patients) and for finding no association between the intervention and any of its outcomes (10). This 2003–2005 RCT was conducted across 17 hospitals in the Netherlands. Patient inclusion criteria included hospital admission for CHF and age ≥ 18 years (27). Patients were randomized to one of three exposure groups: The control group (n = 339) received standard care. A "Basic Support Group" (n = 340) received in-hospital education, nine additional post-discharge clinic visits with a CHF nurse, and post-discharge education and strategies for adherence to medical advice. An "Intensive Support Group" (n = 344) received the above, but with 18 clinic visits with a CHF nurse, two CHF nurse home visits, two multidisciplinary consultation sessions, and telephone followup from a CHF nurse.

The study followed patients for 18 months. Analysis of time to first event of either CHF hospitalization or death showed a hazard ratio of 0.96 (95% confidence interval (CI) 0.76 – 1.21, p = 0.73) for patients in the basic support group relative to patients receiving standard care, and 0.93 (95% CI 0.73 – 1.17, p = 0.53) for patients in the intensive support group relative to patients receiving standard care — that is, no significant effect of the interventions. Some information on our outcome of interest, all-cause readmission, is available from this study but not over a span of 30 days. Over their 18-month surveillance period, the study found an incidence rate ratio of 1.07 (95% CI 0.83 – 1.37, p = 0.62) for readmissions (counting multiple admissions per patient) for the combined intervention groups versus the control group. Jaarsma et al. have proposed that the null results may be due to the high standard of care in the control group: an average of over five cardiologist followup visits per control-group patient in the 18-month period, a high

standard of GP care in the Netherlands, and high patient compliance even in the control group (28).

The study most similar to our own is that of Anderson et al (18). Results of this study include rate of all-cause 30-day readmission, and the intervention included some in-hospital components, making it particularly relevant to our concerns. (Both of these features are rare in the literature surveyed). The study was also unusual in at least partially addressing the issue of comorbidities: number of patient comorbidities was assessed at study intake, and although the nature of the comorbidities was not considered and the number of comorbidities was not a factor in the analysis, the two groups were not statistically significantly different on this measure. This study was an RCT enrolling patients with primary diagnosis of CHF admitted to Bridgeport Hospital, Bridgeport, CT, between January 1, 1996 and March 31, 1997. Total population was 121 patients (77 control, 44 intervention). The in-patient portions of the intervention included a one-hour consultation with a cardiac nurse specialist; half-hour consultations with each of a physical therapist and a dietician; and discharge planning by a cardiac nurse specialist, coordinated with home health nurses who would be providing post-discharge care. The post-discharge intervention included a followup phone evaluation with a nurse case manager, and 6–20 visits (depending on patient progress) by a home health nurse. Patients were followed for six months.

The study found the intervention to be effective: patients in the intervention group had an 11.4% rate of all-cause six-month readmission compared with a 44.2% readmission rate for patients in the control group ($p = 0.01$, risk ratio and confidence intervals not

reported). For our outcome of interest, the rate of all-cause 30-day readmission was 6.0% in the intervention group versus 22.1% in the control group ($p = 0.01$).

Despite the randomized design, estimates in this study may have been affected by imbalance in some covariates. Of patients in the intervention group, 13% were married and 50% lived alone; by comparison, of those in the control group, 35% were married ($p = 0.09$) and 27% lived alone ($p = 0.01$). The support of a companion at home in activities of daily living will have an appreciable effect on health outcomes in a frail or elderly population such as this. Patients receiving the intervention were thus in more difficult circumstances than those not — this would bias Anderson's results towards the null, and the effects of the intervention may, in fact, be stronger than shown in this study.

Overall, the existing literature finds a positive association between care under an improved CHF care protocol incorporating primarily post-discharge interventions, and reduced risk of readmission for CHF patients.

Motivation

CHF, and readmission of CHF patients within 30 days of discharge, are substantial health burdens in the United States. Reducing the rate of CHF readmission will help to better utilize healthcare resources and improve patient outcomes. Studies generally show that improved care protocols reduce the risk of readmission, but the existing studies do not provide adequate information for U.S. hospitals considering changes to CHF care.

First, changes to the in-hospital portions of care are those most readily implemented by a hospital, but little information is available on interventions based primarily on in-hospital interventions. Of the eight protocols surveyed, six mandated post-discharge home visits

(10,12,14,15,17,18), one included a possible physician home visit (11), and one included nurse home visits if clinic visits were impractical (16). These are protocols that may be more difficult for a hospital to implement, as they may involve non-reimbursed medical or nursing activities, and coordination with external private or governmental organizations. Only two studies of those surveyed (14,18) included any in-hospital interventions. The issue of home visits may be particularly problematic, as two meta-analyses found that in-person communication was effective whereas telephonic communication was ineffective (19) or of borderline significance (22); whereas in-person communication can be harder for a hospital to implement or find reimbursement for. (One meta-analysis (20) found in-person and telephone communication to be equally effective.)

Second, the outcomes of interest in the studies surveyed are very diverse, but none has a 30-day outcome as a primary outcome of interest; followup periods in the surveyed studies range from 60 days to two years. A short followup window, unlike those in the studies surveyed, provides a measure of proximate care quality that is not affected by the unpredictable long-term course of CHF and its frequent comorbidities. The 30-day measure allows more focused information on the index admission: quality of care, medical stability at discharge, and quality of discharge planning. A 30-day measure is also consistent with the other readmission measures tracked by the National Quality Forum, and with data available from the Centers for Medicare and Medicaid Services.

Finally, only one (12) of the eight studies surveyed adjusted the analysis for patient comorbidities. It seems intuitive that a 90-year-old patient with diabetes and chronic kidney disease should be at greater risk of readmission than a 65-year-old with no

comorbidities; and that a model that compares one to the other without allowing for these unequal risks may be inaccurate. In fact, comorbidities are common and do affect risk of readmission: Dahlström (29) found anemia in over 20% of CHF patients; Havranek et al.(30) found that over 30% of CHF patients also had COPD, 40% had diabetes, and more than 50% had each of coronary heart disease and a history of hypertension. These and other comorbidities appreciably affect risk of readmission, as shown by Silverstein et al. (9) and Philbin and DiSalvo (31).

It may be argued that the RCT nature of the existing studies obviates the need for including covariates in their analyses, but there is substantial potential for residual confounding. Particularly for studies with small sample size, stating that two study groups did not significantly differ along some factor is a weaker analysis than controlling or adjusting for that factor. Of the studies surveyed, the median sample size was 173 divided among two or more groups, and excluding Jaarsma et al. (which found no effect and had N = 1,023) the maximum was 239. By contrast, our study covers 2,764 discharges. None of the studies surveyed controlled for covariates by blocking randomization along any dimension other than study site.

Therefore, we have studied the effect of a primarily hospital-based intervention, through the intervention's effects on all-cause 30-day readmission rates, with adjustment for patient demographics and comorbidities present during the admission. We believe this has provided more salient and more accurate information for healthcare planners considering changes to CHF care.

Specific aims and hypotheses

Specific aim: To evaluate the effect of a new CHF care protocol on the risk of all-cause 30-day readmission for patients discharged with a diagnosis of CHF from Baystate Medical Center from September 2007 through July 2012.

Hypothesis 1: The implementation phase of the new protocol will be associated with a lower adjusted risk of all-cause 30-day readmission in CHF patients than would be expected on the basis of the baseline risk and secular trend in risk of readmission during the pre-implementation phase.

Hypothesis 2: The implementation phase of the new protocol will be associated with a trend in the adjusted risk of all-cause 30-day readmission in CHF patients that is not equal to the trend in the adjusted risk during the pre-implementation phase.

CHAPTER 2

METHODS

Study design

This trial assessed the association between care under a new protocol for CHF care, and risk of all-cause hospital re-admission within 30 days of discharge, among CHF patients at Baystate Medical Center, in Springfield, MA.

In July 2009, Baystate Medical Center began implementation of a new protocol for care of CHF patients. This change provided an opportunity for a natural experiment to assess the effect of treatment under this new protocol. We compared the risk of all-cause 30-day readmission in the period following the introduction of the new protocol (the *implementation phase*) to the risk in the period preceding the introduction of the new protocol (the *pre-implementation phase*).

Information on discharge diagnoses and on readmissions was collected for a period of 59 months beginning in September, 2007, and extending through and after the implementation of the new protocol (Figure 1).

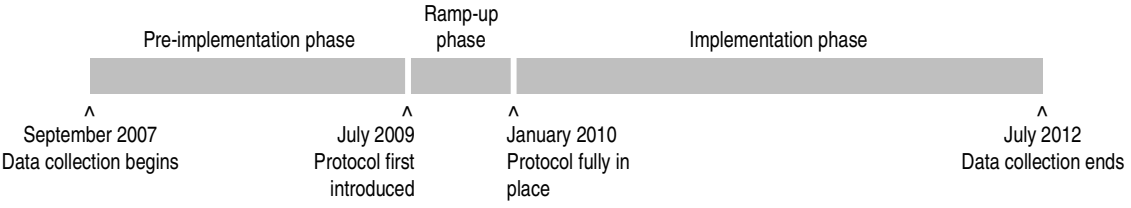


Figure 1: Data-gathering and exposure timeline for implementation of the new CHF care protocol at Baystate Medical Center.

Each datapoint consists of a hospital discharge during this period, including month and year of discharge; medical and demographic information on the patient at the time of

discharge; and whether the patient was readmitted for any cause within 30 days of that discharge.

We have not included data from the ramp-up period in our analysis, a decision made prior to analysis. We believed that any conclusions regarding this relatively brief span (327 of 3091 total discharges, over six months) would have too large a variance to be useful, particularly regarding trend over time; and that amalgamating this data with that from either the preceding or succeeding phases would only bias any results towards the null.

Study population

Baystate Medical Center is a tertiary care hospital serving an ethnically and socioeconomically diverse population in Springfield, MA, a city of 150,000 in western Massachusetts. Participants in this study received care on the dedicated CHF care ward in Baystate Medical Center's main hospital building on its main campus.

The study covered all patients aged 18 or older discharged alive from the dedicated CHF care ward, with a primary discharge diagnosis of CHF, during the study periods of September 2007 – July 2009 (the pre-implementation phase) and January 2010 – July 2012 (the implementation phase), whose discharge did not occur within 30 days of a prior Baystate Medical Center discharge (i.e., whose admission was not a readmission from an admission included in the study). CHF patients admitted to other wards (as, possibly, if the CHF ward was full), and patients discharged from the CHF ward with a primary discharge diagnosis other than CHF, were not included in the source population. Researchers determined patient hospitalization dates and discharge diagnoses (based on

ICD-9 codes) through administrative data provided from Baystate Medical Center's McKesson financial decision support system (see APPENDIX for ICD-9 codes used).

Exposure

The exposure of interest was care under the new CHF care protocol. This protocol included changes from the prior standard of care both in-hospital and post-discharge. In-hospital, under the new protocol, CHF patients are screened on admission to identify those at high risk of readmission. CHF patients are geographically centralized on a single ward, a policy already in place as the preceding standard of care, but the new protocol instituted supervision of CHF patient cases by a dedicated CHF nurse manager, coordinated multi-disciplinary rounds for CHF patients, as well as monthly multi-disciplinary meetings aimed at CHF care. While in hospital, CHF patients are given a detailed medication reconciliation, specialized patient education, and each patient's first followup appointment is scheduled prior to discharge. Following discharge, patients are followed by phone.

Care under the reference condition — during the pre-implementation phase — was Baystate Medical Center's standard of care for CHF patients in place at that time.

Exposure was assessed by comparing a subject's discharge date with the dates defining the pre-implementation and implementation phases. Exposure for each discharge was measured as a single dichotomous variable (Table 1).

Outcome

The outcome of interest was readmission to a hospital, for any reason, within 30 days of the index discharge from Baystate Medical Center. The outcome for each discharge was

measured as a single dichotomous variable (Table 1). The possibility of inaccuracy through admission to a facility other than Baystate is addressed in *Study limitations*, below.

The validity of inclusion, exposure, and outcome assessment depend largely on the accuracy and completeness of hospital administrative records. We believe the hospital administrative records to be highly complete and accurate for our required information — admission and discharge dates, accurate matching on unique patient ID, and admission diagnosis — due to the hospital's need for these records for legal compliance, liability limitation, reimbursement, and census tracking.

Covariates

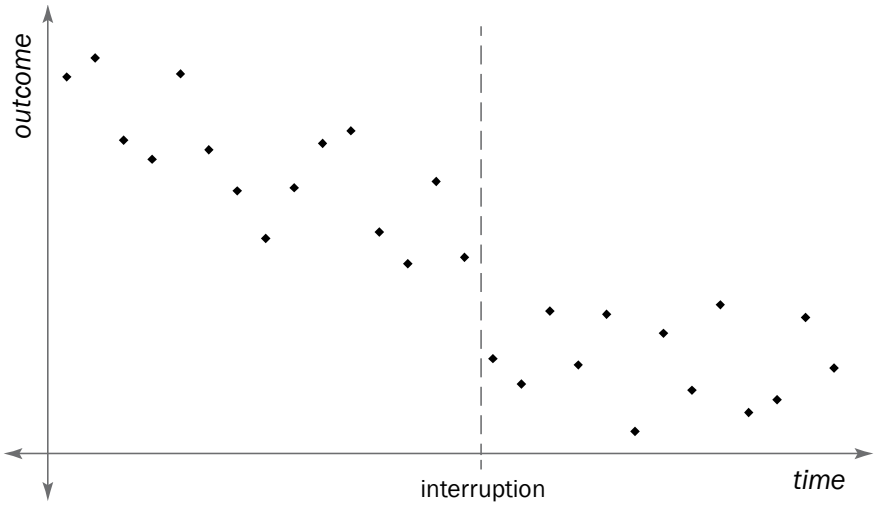
Potential predictors of readmission risk were chosen from those in the hospital administrative data based on covariates used in previous studies of CHF readmission (10–12,14,15,17,18). We chose age, gender, race/ethnicity, source of payment for medical costs, primary diagnosis, secondary diagnoses, length of hospital stay, and discharge disposition (such as home with self care, home with visiting nurse assistance, skilled nursing facility, etc.) as potential predictors of readmission risk in multivariable model-building.

Information on the covariates was gathered from the administrative records mentioned above. Covariates are coded as described in Table 1.

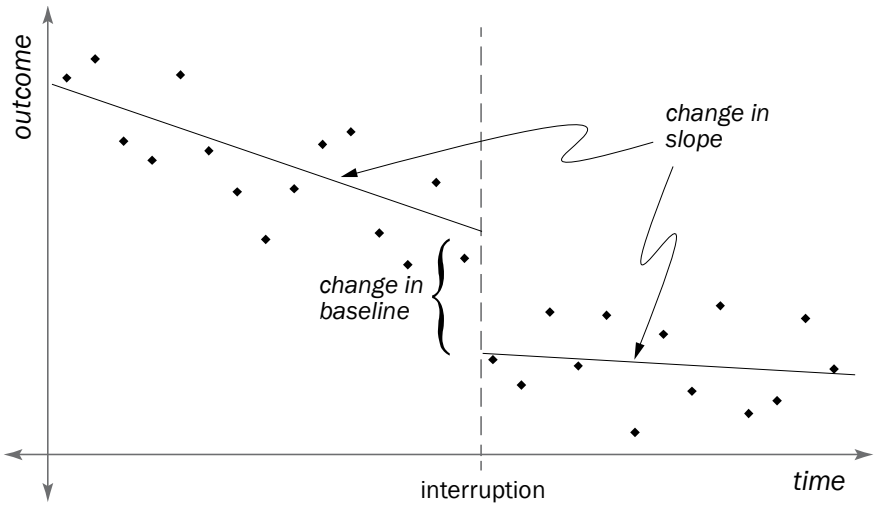
Interrupted time-series analysis

Interrupted time-series analysis is a quasi-experimental protocol that allows comparison of a series of outcome values before and after some change in conditions — the

"interruption". For instance, a study could examine the daily number of emergency-department visits for asthma, before and after a ban on smoking in public places.



Rather than simply comparing mean values before and after the interruption, this type of analysis can account for a secular trend in the outcome over time. (If rates of asthma admission were declining independent of the smoking ban, then a lower rate of admission after the ban should not, alone, be used to indicate an effect of the ban.) As well, it can detect a change in trend associated with the interruption. (Perhaps, with lower asthma admission rates after the smoking ban, the secular trend towards lower rates was reduced given the smaller room for improvement.)



As a regression model, this is expressed as

$$outcome = \beta_0 + \tau * time + \gamma * i + \rho * time * i + other\ predictors$$

Equation 1

where $\tau * time$ is the secular trend over time; i is an indicator variable indicating before/after the interruption, so that $\gamma * i$ is the change in baseline after the interruption; and ρ is the change in trend after the interruption.

As the outcome in this study is a binary outcome, we tested regression models of the form

$$logit(p(readmission)) = \beta_0 + \tau * time + \gamma * i + \rho * time * i + other\ predictors$$

Equation 2

The questions of interest were whether $\gamma \neq 0$ (whether there was a change in baseline log-odds of readmission), and whether $\rho \neq 0$ (whether there was a change in the trend of log-odds of readmission); in particular, we hypothesized that $\gamma < 0$.

Protocol implementation had, in fact, two interruptions — start of the ramp-up period for implementation of the protocol, and full implementation of the protocol (Figure 1). As we did not incorporate data from the ramp-up period, this was appropriately modelled with a single interruption. Omitting this data also means that the date for defining before/after the interruption, that is, the time after which the indicator $i = 1$, can be defined as any time within the ramp-up period without affecting the model.

We tested both a raw model, incorporating only the predictors *time, i*, and *time* $*$ *i*, and the best logistic regression model that we arrived at incorporating those as well as medical and demographic covariates.

Time-series analysis

The regression analysis described requires that the outcomes be independent, whereas time-based measurements often do not to meet this assumption. Time-based outcomes can be autocorrelated: the expected value for the outcome at one time can depend on the observed values at previous times. (The expected closing price of the Dow Jones index for one day is not independent of the observed closing price the previous day.) When autocorrelation is present, analysis via an ARIMA model rather than simple regression is required. Time-based outcomes often also exhibit cyclical behavior: hospital admissions for influenza exhibit a regular seasonal pattern. We examined correlograms of the probabilities of readmission, amalgamated both monthly and quarterly, to detect autocorrelation and seasonality.

Logistic regression model

Covariates

The candidate covariates available for logistic regression models of risk of readmission were:

- age
- sex
- race/ethnicity (white, black, Hispanic, other/unknown);

- medical cost payor (commercial medical insurance/HMO, Medicaid, Medicare, other)
- discharge disposition (against medical advice; home to self-care; home with visiting nurse assistance; hospice; rehabilitation or other skilled nursing facility; transfer)
- length of hospital stay
- primary diagnosis
- secondary diagnoses

Primary diagnosis was expressed as an ICD-9 code. To address cell sparsity, we condensed these diagnoses from the ICD-9 four-digit subcategories and five-digit subdivisions into their three-digit ICD-9 categories (so, for instance, "428.21 Acute systolic heart failure" and "428.9 Heart failure, unspecified" both mapped to "428 Heart failure"). The resulting categories were coded as indicator variables and used for regression.

We hypothesized that length of stay might have a nonmonotonic relation with outcome, so we binned length of stay into short, medium, and long, and built separate regression models treating length of stay as either continuous or categorical. The bin cutpoints were determined upon examining the distribution of length of stay in our data, but prior to any data analysis.

Gagne score

Patient comorbidities are important risk factor for readmissions, and information on these is available for this study from the set of secondary diagnoses for the patient at discharge. These are coded as ICD-9 diagnosis codes. In our data, each discharge is associated with

from 0 to 48 distinct secondary diagnoses, and 1930 secondary diagnoses are represented. Cell sparsity forbids treating these as distinct predictors, so some binning or summary measure is called for.

We elected to summarize the information from the secondary diagnoses using the Gagne comorbidity score (33). Developed to predict 30-day mortality in an elderly population, this has been shown to have better predictive value (33) in this population than summary scores based on the Elixhauser comorbidity classification system (34) or the Charlson Index (35), and was favorably reviewed in a comparison of scoring systems for predicting mortality in the elderly (36). We did not elect to use the risk score developed by Philbin et al. specifically for prediction of hospital readmission for heart failure (31) as this measure incorporates several measures that are included as separate covariates in our model (race/ethnicity, medical charge payor, discharge disposition) and requires patient medical information not available in our data. Although our interest is morbidity rather than mortality, the study population is elderly (Table 4), as is that in the data used to create the Gagne score, and we believe the Gagne score to be an appropriate candidate covariate for adjusting risk of readmission in our logistic regression model.

Model-building

For models that did not include any covariates, we determined to test three unadjusted logistic regression models:

1. predictors *time, i*, and $time * i$;

2. *i* only (test for change in baseline);

3. *time* only (test for secular trend).

Time was included as month of study, centered on the beginning of the intervention phase (*time = month of study – 28*). We centered the time variable to reduce the artificial collinearity between *time* and *time∗ i*.

For a regression model including covariates, we first built a model including:

i. predictors *time, i*, and *time∗ i;*

ii. predictors *age* and *sex,* which we had determined to include in all models;

iii. candidate predictors *primary diagnosis* (condensed to 3-digit ICD-9 categories), *race/ethnicity, medical charge payor, discharge disposition, Gagne score,* and *length of hospital stay*.

We then built a new model including the predictors from *i)* and *ii)* above, and only the predictors from *iii)* that showed p values ≤ 0.25 by the Wald significance test. We examined the remaining predictors from *iii)* for collinearity with the predictors of interest, *time* and *i*. If collinearity was satisfactory and all remaining predictors had Wald significance $p \leq 0.25$, we chose this as the regression model to test. We did not continue eliminating predictors to arrive at the most parsimonious model, as our goal was to adjust for confounding factors rather than to arrive at a predictive model or investigate which variables might be most strongly associated with the outcome.

We repeated the above procedure including length of stay, categorized as described above, in place of length of stay as a continuous variable, in the model-building.

We repeated the above procedure replacing the predictors in *i)* above with *i* only (test for change in baseline) and *time* only (test for secular trend).

Model evaluation

Model significance

We used a likelihood-ratio test, comparing each model to the naïve (intercept-only) model, to test the evidence for rejecting the null hypothesis that the given model was no better than the naïve model.

Goodness of fit

We evaluated the goodness of model fit, for regression models of interest, via the Hosmer-Lemeshow test. Due to the sensitivity of this test to the choice of cutpoints (37) (38), we evaluated the test for several numbers of groups, and supplemented those results with the result from the le Cessie-van Houwelingen-Copas-Hosmer unweighted sum of squares test statistic ($\hat{S}$ in (37), *USS* in (38)).

We also evaluated the unadjusted regression model — the model including only predictors *time, i*, and $time * i$ — via its Pearson residuals. The discharge dates provided in the source data are quantized by month (year and month of discharge are reported, but not date). For the model including only predictors *time, i*, and $time * i$, then, it is meaningful to group the model results by covariate patterns, which results in one distinct pattern for each month of the study, and compare the observed number of readmissions in each month with that predicted by the model. (The relativity small number of covariate patterns — 53 study months — and high number of events — several dozen discharges per month — avoid the issue of cell sparsity in this analysis.) We also evaluated the normality of the Pearson residuals in this model.

Finally, we investigated the appropriateness of the unadjusted regression model by examining a plot of the outcome (risk of readmission, aggregated by month) as a function of the only continuous predictor, time, for any patterned deviation from linearity which would suggest that a transform of the data might be in order prior to analysis.

Statistical power

We conducted a post hoc power analysis of the study to test what effect size could be detected, in a population similar to this in a study of this type. We tested this via simulation, for an array of effect sizes, to determine the study's power at each effect size.

The procedure, for a single effect size, was:

1. Referring to the best regression model arrived at above,

$$logit(p) = \beta_0 + \tau * time + \gamma * i + \rho * time * i + other\ predictors$$

   choose a value for γ (the change in baseline) and a value for ρ (the change in trend). This is the effect size that will be simulated.

2. Simulate a set of data points that is of the same size as the actual study dataset, and as similar as possible to the study data, calculating the outcome for each data point via the regression model:

   - The coefficients γ and ρ have been chosen above;
   - β_0, τ, and the coefficients for any other predictors can be set at their best-fit estimates from the regression model being used;

- from those values, and from the covariate values for that data point, calculate a value of p for that datapoint. Set the outcome for that datapoint to true (readmission) with probability p.

Applied to all datapoints in the simulated data set, this produces a set of simulated data whose outcomes are those that would be expected if there truly were a change in baseline of size γ and a change in trend of size ρ in the process underlying the outcomes for population being sampled.

3. Fit a logistic regression model to this simulated dataset. Test: does this model (correctly) reject the null hypothesis that $\gamma = 0$ with statistical significance of 95%? Does this model (correctly) reject the null hypothesis that $\rho = 0$ with statistical significance of 95%?

4. Repeat steps 2 and 3 for 500 repetitions, to calculate the proportion of times that this procedure correctly rejects each null hypothesis for the given effect size — that is, for each hypothesis calculate the power of the study.

We repeated this procedure for an array of values of γ and ρ to determine the power of the study at a variety of effect sizes.

Step 2, simulating a set of datapoints, involved some decisions. The true dataset, in addition to summary properties — 50% male, median age 76, and so on — has a complex covariance structure: for instance, age is not independent of sex. This structure should be preserved in simulated data that is intended to represent samples from a similar population. Therefore, we decided to simulate data via bootstrapping: each simulated

datapoint is an element drawn from the set of actual datapoints, with replacement. (The outcome for each datapoint is calculated as described above).

When calculating the datapoint outcomes for a given effect size, we also adjusted β_0, the intercept, to maintain an overall risk of readmission over the entire sample close to that in the actual dataset. Increasing or decreasing the value of γ or ρ from the best-fit value derived from the actual dataset, and using that value to calculate risk of readmission for each simulated datapoint, will inflate or deflate the overall risk of readmission in the simulated dataset relative to that in the true data. To adjust for this we arrived empirically at a formula for adjusting β_0 as a piecewise linear function of γ and ρ to keep the overall simulated readmission risk within 10% of the overall readmission risk in the original data.

Tools

Data analysis was conducted using the R language and environment for statistical computing version 3.0.3 (39) and SAS software version 9.4 (40). The R packages foreign (41), ggplot2 (42), Hmisc (43), lmtest (44), MKmisc (45), and rms (46) were used as well. The SAS code to calculate Gagne scores was that from the Brigham & Women's Hospital Division of Pharmacoepidemiology and Pharmacoeconomics (47).

CHAPTER 3

RESULTS

Data cleaning

We found no missing or out-of-range values in the dataset. The primary key (admission ID) was unique in the dataset.

Baseline and summary values

A total of 2,764 CHF discharges were included in the study, 26% (716) in the pre-intervention phase and 74% (2048) in the intervention phase. Of the total, 21.2% (586) were readmitted; of those in the pre-intervention phase 21.8% (156) and of those in the intervention phase 21.0% (430) were readmitted (Table 2).

Risk of readmission, aggregated by calendar month, varied from 7.1% to 46.4%, with an interquartile range of 16.7% – 22.9%, and with no evident trend or pattern (Table 3, Figure 2, Figure 3, Figure 4).

A total of 2,114 distinct patients are represented through the three study phases. This patient pool is large enough that we may regard observations as independent: the admission of several dozen patients in a month (the maximum observed was 83, median 56) does not appreciably affect the size or characteristics of the population still available for admission.

Patient characteristics at discharge for discharges in the pre- and post-intervention populations showed statistically significant differences (at $\alpha = 0.05$) in many demographic and medical measures (Table 4,Table 5). We tested differences between the

pre-intervention and post-intervention groups along continuous variables using the Mann-Whitney-Wilcoxon test, and differences along categorical variables using Fisher's exact test.

Time-series analysis

A correlogram of monthly proportions of readmission revealed no autocorrelation, with only small and random deviations from zero for autocorrelation at lags up to 24 months (Figure 5). The Durbin-Watson test indicated no autocorrelation for lag up to 4 months (Table 6). The lack of autocorrelation at lag = 12 indicated no seasonal effects on proportion of readmission, as did the lack of autocorrelation at lag = 4 for values aggregated by quarter (September – November, December – February, March – May, June – August) (Figure 6). We therefore did not investigate ARIMA analysis, and did not adjust for seasonality in risk of readmission.

Non-adjusted regression model

Modeling risk of readmission as a function of time, the pre/post division, and their interaction, we found no statistically significant evidence (at $\alpha = 0.05$) that any of the three predictors had an odds ratio (OR) $\neq 1$ (Table 7). The significance of the model as a whole gave a likelihood-ratio test value of $p = 0.846$, indicating that we cannot reject, at $\alpha = 0.05$, the null hypothesis that the model incorporating these predictors is no better than the naïve model.

Overall, we concluded that there was no strong evidence that linearity of log-odds in the predictors was an inappropriate form for modeling the response. The Hosmer-Lemeshow $\hat{C}$ statistic is far from statistically significant at $\alpha = 0.05$ for any group size tested (Table

8). The le Cessie-van Houwelingen-Copas-Hosmer global goodness of fit p value is 0.030 (Table 8), but in an unpredictive model, such as this one, this test statistic can be greatly inflated (46). The Pearson residuals for this model aggregated by month give $p = 0.210$ (Table 8) and the Pearson residuals are normally distributed (Figure 7), as we would expect for an appropriate model.

Examining risk of readmission, aggregated by month, as a linear function of the only continuous predictor, time, we did not see any pattern in the data indicating that a transform of the outcome would be more appropriate as the outcome in a logistic regression model (Figure 8, Figure 9, Figure 10).

Combining these, we find no evidence that a logistic regression model with risk of readmission as the outcome is an inappropriate model. However, within this framework, the unadjusted model finds no association between the change in protocol and the risk of readmission, or between the change in protocol and a change in the trend in risk of readmission.

We also tested separately for a change in baseline risk following the change in protocol, and for presence of a secular trend in risk of readmission. Testing simply for an association between pre- versus post-implementation and risk of readmission, we again could not reject the null hypothesis of OR = 1 ($p = 0.656$) (Table 9). As well, we could not find evidence for any secular trend in risk of readmission ($p = 0.750$) (Table 10). (Had any of these been significant, the issue of adjustment for multiple comparisons would have been relevant.)

Complete regression model

Building our best regression model using the procedure described above, we arrived at a model incorporating age, sex, race/ethnicity, medical cost payor, and Gagne score as predictors of risk of readmission, as well as the predictors of interest — *time, i,* and *time*i* (Table 11).

All predictors were highly statistically significant in this model, with the exception of the predictors of interest. We found no statistically significant evidence at α =0.05 that any of *time, i,* or *time*i* had an OR $\neq 1$. The significance of the model as a whole gave a likelihood-ratio test value of $p < 0.0001$, indicating that we can reject, at $\alpha = 0.05$, the null hypothesis that the model including these predictors is no better than the naïve model.

The statistical non-significance of the predictors of interest is not due to inflated variance in the estimate for any of the coefficients of interest, caused by collinearity with any of the other predictors in the model. None of the categorical covariates in this model is associated with *i* (Table 5); age and Gagne score are associated with *i* (Table 4), but eliminating either or both of age and Gagne score as predictors did not materially affect the significance of the predictors of interest in the resulting models. None of the covariates in this model is associated with time at $\alpha = 0.05$ (Table 12, Table 13); only race/ethnicity is at all associated with time ($p = 0.070$) but a model omitting race/ethnicity still showed no statistical significance for the predictors of interest.

The model including these covariates shows good fit by all measures we employed. The $\hat{C}$ statistic is far from statistically significant at $\alpha = 0.05$ for any group size tested, and the le Cessie-van Houwelingen-Copas-Hosmer global goodness of fit p value is 0.894 (Table

14). Overall, this is a well-fit logistic regression model, but one which shows no statistical significance for the predictors of interest.

As with the unadjusted regression model, we also tested separately for a change in baseline risk following the change in protocol, and for presence of a secular trend in risk of readmission. For each question we built a regression model as described above, and in each case we arrived at a model including the predictors age, sex, race/ethnicity, medical cost payor, and Gagne score in addition to the predictor of interest (Table 15, Table 16). As with the unadjusted regression models, we could not reject the null hypothesis of no change in baseline risk following the change in protocol ($p = 0.392$) or the null hypothesis of no secular trend in risk of readmission ($p = 0.433$).

Statistical power

As expected, simulated study power to detect a change in baseline risk following the change in protocol ($OR_\gamma \neq 1$) increased with simulated γ further from zero, and power to detect a change in the trend of risk following the change in protocol ($OR_\rho \neq 1$) increased with simulated ρ further from zero (Table 17, Table 18, Table 19). (The two effects were not entirely orthogonal, but the interaction does not have any clear interpretation, or may be meaningless, due to the dependence of β_0 on both γ and ρ in the regression equation used to simulate the data.)

Power to detect $OR_\gamma \neq 1$ was 86% for an OR_γ of 0.5 (at $OR_\rho = 1$; power values covered the range from 73% to 88% for the range of OR_ρ values 0.94 – 1.06). Power was 75% to detect an OR_γ of 2.0 (again, at $OR_\rho = 1$; power was 50% at $OR_\rho = 1.06$, and 85% at $OR_\rho = 0.94$). However, this power fell rapidly for values of OR_γ closer to 1: power was 58% to

detect $OR_\gamma = 0.6$ and only 33% for $OR_\gamma = 1.5$ (both at $OR_\rho = 1$). For the strongest effects, power was 98% to detect $OR_\gamma = 0.4$, but still only 90% for $OR_\gamma = 2.5$ (both at $OR_\rho = 1$).

As the effect of a covariate on the risk of the outcome depends upon the values of the other covariates in a logistic regression model, there is no single mapping from these odds ratios to effects on risk of readmission. To get an indication of the magnitude of the effects of a given OR, however, we can examine a predicted risk as calculated over the population studied: to do this, we apply our regression model (that in Table 11) with the given counterfactual γ and all other coefficients unchanged, to predict risks of readmission for all actual discharges (datapoints) observed.

Doing so, an OR_γ of 0.5 results in an overall predicted risk of readmission of 0.11 during the implementation phase, in contrast to the observed risk of 0.21 during that phase (and risk of 0.22 observed during the pre-implementation phase). An OR_γ of 0.6 results in an overall predicted risk of readmission of 0.13 during the implementation phase; and at the other extreme, OR_γ values of 2.0 and 1.5 result in overall predicted risks of readmission of 0.33 and 0.27 during the implementation phase.

The power to detect $OR_\rho \neq 1$ was 81% for an OR_ρ of 0.96 (at $OR_\gamma = 1$; power values covered the range from 74% to 85% for the range of OR_γ values 0. 4 – 2.5). Power was also 81% for an OR_ρ of 1.05 (at $OR_\gamma = 1$; power varied widely, covering the range from 65% for $OR_\gamma = 2.5$ to 93% for $OR_\gamma = 0.4$). As with power to detect γ, power to detect ρ fell off quickly: power was 56% to detect $OR_\rho = 0.97$ and 66% to detect $OR_\rho = 1.04$. At the most extreme values simulated, power was 99% to detect $OR_\rho = 0.94$ and 92% to detect $OR_\rho = 1.06$ (at $OR_\gamma = 1$). (For scale, recall that the coefficient ρ is for trend in log-

odds of readmission by month, so $OR_\rho = 0.96$ indicates an odds ratio of 0.61 for a year's difference in discharge dates.)

As with γ, we can estimate what the effect of a counterfactual ρ would be on the population studied. As ρ affects the effect of time on risk of readmission, we cannot express the effect of a given ρ as an overall risk of readmission. Rather, we can examine the effect of a counterfactual ρ on the risk of readmission in each month of the intervention phase, as against the risk, in that month, as predicted by the best regression model. (Recall that the best regression model gave a point estimate of $\rho = 0.015$, but with a p value of 0.312, so we did not conclude that ρ was significantly different nonzero at $\alpha = 0.05$.) In particular, we can compare the predicted risk of readmission in the first month of the intervention phase with that in the last month of the intervention phase to see the most extreme effects of a counterfactual ρ.

Also note that to examine the effects of a counterfactual ρ, we will apply the predictive model for that ρ, and the month of interest, to the entire sample, not to the sample in that month. This will eliminate the effects of random variation in the covariates among months.

The expected risk of readmission by month of study (as predicted by the best regression model, that in Table 11, applied to all discharges in the pre- and post-intervention phases) decreases slightly over time in the intervention phase, more than offset by a slightly increasing risk of readmission over time in the intervention phase (with neither effect significantly nonzero at $\alpha = 0.05$). In particular, the expected risk of readmission is 0.204 in the first month of the intervention phase (study month 29) and 0.213 in the last month

of the intervention phase (study month 59). These values are the basis for comparing predicted risks of readmission by month resulting from various simulated values of ρ.

Projecting in this way, $OR_\rho = 0.94$ results in an overall predicted risk of readmission of 0.19 in month 29, and a minuscule 0.02 in month 59. At an OR_ρ just below one, $OR_\rho = 0.99$ results in an overall predicted risk of readmission of 0.20 in month 29, and a still-small 0.11 in month 59. At the other extreme, $OR_\rho = 1.06$ results in an overall predicted risk of readmission of 0.21 in month 29, and a huge 0.50 in month 59. At an OR_ρ just above one, $OR_\rho = 1.01$ results in an overall predicted risk of readmission of 0.20 in month 29, and 0.19 in month 59 — at values of ρ close to this, the secular trend in risk and the change in trend due to the interruption roughly cancel each other.

Summarizing these many projections, it seems our retrospective power is small. A change in baseline risk of readmission, in a population of like this one, from 0.22 in the pre-intervention phase to 0.13 in the intervention phase, would be a very substantial real-world effect; but our simulated power to detect this never rises as high even as 65% (Table 18). Power is similarly low to detect what would be substantial effects of a change in trend.

CHAPTER 4

DISCUSSION

Our results indicate no statistically significant relation between introduction of the modified CHF care protocol and either the rate of all-cause 30-day readmission or the trend in that rate. This result does not support most previous findings, which show an association between an improved care protocol and reduction in readmission rate. We believe that limitations of the study may have biased results towards the null, but that the total effect of any such effects was small.

Study limitations

We find essentially no possibility for misclassification of exposure. We also believe that inaccurate assessment of CHF status was unlikely. Quan et al. evaluated the use of ICD-9 codes in administrative data, and found ICD-9 codes to have sensitivity of 71.6%, and specificity of 99.3% in detecting CHF when compared to a gold standard of diagnosis from chart review (32). The high specificity indicates that we did not include any appreciable proportion of non-CHF patients in our study. The low sensitivity does not apply to our situation, as Quan et al. were assessing detection of any of the conditions present in a patient — they found a 28% chance that a patient with CHF as one of their health conditions would not have that recorded as an ICD-9 code. For inclusion in our source population, however, a patient would need to have CHF severe enough at admission that it should be the primary diagnosis; we believe that almost no CHF this severe would not be coded. The final possibility for misclassification of exposure lies in the possibility of a patient with several health conditions, including CHF, being coded as

a CHF admission, and placed on the CHF ward, when in fact that admission was due to another health condition — pneumonia, for instance. This situation might bias results in either direction, in the unlikely case that the intervention were either more or less effective in these patients than in the intended population, but there is no reason to believe that these unlikely miscategorizations occurred more frequently in either phase of the study.

Selection bias may have occurred due to our inclusion in the source population of only patients discharged from the CHF ward. If that ward tended to be fuller during one of the study phases, and if — as seems likely — more severe patients tended to get the beds available on that ward, then readmission could have been more likely during the phase with higher utilization of the CHF ward. As the population of the catchment area did not change appreciably over the course of the study, we believe this potential bias, which could have biased results either towards or away from the null, was small.

Bias towards the null could have resulted from misclassification of the care provided (essentially, ours was an intention-to-treat analysis). First, even prior to the ramp-up phase, some practitioners could have provided aspects of care under the new protocol. We believe the scope for bias here is minimal as most aspects of the new protocol resulted from changes in hospital policies that an individual care provider could not have implemented. More appreciably, provider adherence to the new protocol could have been lax even during the implementation phase (incomplete attendance at multidisciplinary meetings or rounds suggest themselves). We believe that any change to care will be subject to organizational inertia, and that any change requiring additional provider time or effort will necessarily be difficult; therefore, although incomplete adherence would

have been limited by enforcement of hospital protocols, this effect could have biased our results towards the null.

The only plausible inaccurate ascertainment of readmission would occur if a patient were readmitted to a hospital other than Baystate Medical Center. However, Baystate Medical Center was the only tertiary care facility in the region throughout the study, and we estimate that few CHF patients would choose to change care providers for a serious health condition. As well, the brief 30-day window for readmissions reduces the possibility that patients would move out of the catchment area in that time. Thus, this effect would have been extremely small.

The major weakness of an interrupted time-series analysis is in comparing exposed and unexposed groups that were recruited and observed during different time spans. Any change over time affecting the population included in the study and associated with the outcome can be a source of confounding and bias results either towards or away from the null. We minimized such effects via both study design and analysis. The study design kept the exposed and unexposed groups as similar as possible, as both the exposed and unexposed were drawn from a population that did not change appreciably between the unexposed and exposed phases: During the study period, the population of Springfield, MA, did not change appreciably in SES factors or age profile; no area medical centers opened or closed to change the catchment area for Baystate Medical Center; and the inclusion criteria (ICD-9 codes) were constant as well. Interrupted time series analysis adjusted for the effect of any constantly-varying trends across time, and the logistic regression model for changes in any of the covariates included in the model. Any residual confounding could have biased the results in either direction, and we believe the

mitigating effects of both our study design and analysis will have resulted in minimal levels of any such effect.

Thus, we cannot conclude that study limitations biased our results appreciably towards the null; and therefore the results of this study should be regarded as indicating a true lack of association under these study conditions between the modified care protocol and rate of readmission.

Care protocol

We believe there is a substantial possibility that the disparity between the results of this study and those of most previous studies is due to the difference between the care protocols implemented. Of the eight protocols surveyed, two included the option of home visits and six included mandatory home visits as part of the improved protocol. Home visits greatly reduce the rate of noncompliance with followup care, allow a professional to evaluate any negative trends in patient health before they become severe enough to require patient-initiated care-seeking, and allow that professional to evaluate the adequacy of the patient's in-home care. These effects may be major factors in reducing rates of readmission, but home visits are not part of the protocol implemented in this study.

Effect size and direction

The small effect sizes found in this study are also worth noting. Even were the measured effects statistically significant, they indicate an OR of 1.08 for risk of readmission in the post-implementation versus the pre-implementation phase, and a similarly small OR for the change in trend (Table 11). The OR of 1.08 results in an overall predicted risk ratio

(risk ratio as applied to the entire population studied) of 1.065 for risk of readmission in the post-implementation versus the pre-implementation phase. This ratio also indicates, were we to consider it either statistically significant or clinically meaningful, an increased risk under the new protocol.

Post hoc statistical power

The statistical power arrived at by our simulation is very low, even for high simulated values of γ (the coefficient for a post-implementation change in baseline), and for high simulated values of ρ (the coefficient for a post-implementation change in trend) (Table 17 –Table 19). It is tempting, then, to regard our null results as only very poor evidence against the alternative hypothesis of an effect associated with the change in protocol. If the *observed power* — the power of an experiment of this type, calculated post hoc from the observed results — were high, we could interpret this as a high probability of rejecting the null hypothesis if it were in fact false; and thus might regard a statistically non-significant result as providing evidence for the null hypothesis. Conversely, in our case, we might regard our results as adding very little support to the null hypothesis. However, as Hoenig and Heisey show (48), observed power is a monotonic function of the p value, so all the information that can be drawn from the estimated parameters is already present in their p values. Worse, Hoenig and Heisey show that for two otherwise-identical studies with the same outcome, the study with the higher observed power will also be that with the higher p value: if higher observed power is to be interpreted as evidence in favor of the null hypothesis, then higher p values must be so interpreted as well. In this study, then, we cannot interpret our low observed power (classically arrived at by calculation, here by simulation) as indicating that we can dismiss our null results.

Having rejected such an inappropriate interpretation, the post hoc power simulation still provides useful results for further research: in populations similar to this one (in distribution and correlation of covariates), we know what power can be expected to detect what possible values for γ and ρ, and a slight modification of our procedure can estimate the power for other sample sizes as well.

Other results

We find it interesting that Gagne score is such a statistically significant predictor of outcome. As well, the effects of the Gagne score seem to be clinically significant: With an odds ratio of 1.09 for each one point increase in Gagne score, the odds ratio for an increase over the interquartile range — 4 points, from 3 to 7 — is 1.42. The OR of 1.42 for a 4-point increase results in an overall predicted risk ratio (risk ratio as applied to the entire population studied) of 1.31, a change from 18% to 24% in risk of readmission.

This study was not intended to test the validity of an association between Gagne score and risk of readmission, and no significance values have been adjusted for multiple comparisons. But as Gagne score is easily computed, it could be a useful tool in assessing which patients most need extra in-hospital or followup care, and we suggest that this association merits further study.

Generalizability

We would not hesitate to generalize these results to the populations served by most other U.S. major medical centers. Baystate Medical Center serves an ethnically and socioeconomically diverse population, and we have no reason to believe that populations composed of other racial and ethnic groups, more rural populations, or populations from

the central or western U.S. would not react to CHF care as the study population did. We also believe, due to similarities in the distribution and covariance structure of predictors, that the results of our power simulation would apply to studies in these populations.

We would not want to generalize to settings in the less-developed world where the standard of care is very different from that in this study. The unexposed condition in such a setting would not resemble the unexposed condition in this study, and it is not clear what the exposed condition of care under the new protocol would mean when starting from such a different baseline. We would not even want to generalize to settings in Canada or Western Europe, where standards for followup and continuity of care differ appreciably from those in the U.S.

This leads in to the greatest limitation to generalizability, the very specificity of the interventions implemented. The literature on changes to CHF care protocols describes combinations of interventions as diverse as: cross-disciplinary caregiver consultation; geographic centralization of CHF patients; aggressive medication use; medication reconciliation; post-discharge telephone followup; dietary consultation; home RN visits; home dietician visits; 24-hour post-discharge help line; home teaching sessions; and advanced-care nurse case management. Of the many combinations of many possible interventions, we would feel comfortable generalizing these results only to interventions whose core elements were similar to the core elements implemented in this protocol, and — as mentioned above — only in settings where the baseline of care was similar to that in this study.

Conclusion

These results indicate no statistically significant association (at $\alpha = 0.05$) between introduction of the new care protocol and either the baseline risk of 30-day all-cause readmission or the trend in that risk. We believe that this is a true null result, and cannot adequately be explained as a result of weaknesses in the study. It is hard to meaningfully compare these results to others reported in the literature due to the heterogeneity of interventions studied in attempts to reduce readmission rates.

For further study, we note that all-cause readmission is a noisy measure, including readmissions due to causes entirely independent of CHF care (and thus biasing results towards the null). More specific outcomes — particularly, readmissions due to CHF, and readmissions due to cardiovascular disease in general — may better measure the effects of a CHF care protocol. We suggest that studies including those outcomes as well as all-cause readmission will be informative.

As well, as new CHF care protocols are implemented, the effectiveness of such measures should be tested; we believe that interrupted time-series analysis with adjustment via multiple logistic regression, as described here, will be appropriate and effective.

CHAPTER 5

TABLES & FIGURES

Tables

Table 1: Classification of study variables

Description	Type
Exposure	
Protocol at discharge	dichotomous
Time (month of study)	continuous
Outcome	
readmission within 30 days, all causes	dichotomous
Covariates	
Age at discharge (years)	continuous
Sex	dichotomous
Race/Ethnicity white black Hispanic other/unknown	categorical (with indicator variables for regression)
Medical charge payor commercial/HMO Medicare Medicaid other	categorical (with indicator variables for regression)
Length of hospital stay (days)	continuous
Length of hospital stay short medium long	categorical (with indicator variables for regression)
Primary diagnosis (25 distinct ICD-9 codes represented)	categorical
Primary diagnosis, binned *(four categories)*	categorical (with indicator variables for regression)
Secondary diagnoses at discharge	Multiple secondary diagnoses present per discharge; link via foreign key to table of secondary diagnoses.
Gagne comorbidity score	continuous
Discharge disposition against medical advice home, self care home, visiting nurse assistance hospice skilled nursing/rehabilitation facility transfer	categorical (with indicator variables for regression)

Table 2: Association between outcome and exposure phase

	Overall N = 2764	**Pre-intervention phase** n = 716	**Intervention phase** n = 2048
Readmissions	21.2% (586)	21.8% (156)	21.0% (430)

Table 3: Risk of readmission by calendar month

minimum	7.1%
1st quartile	16.7%
median	21.1%
Mean	21.2%
3rd quartile	22.9%
maximum	46.4%

Table 4: Associations between continuous covariates and exposure phase

Characteristic	**Overall** N = 2764	**Pre-intervention phase** n = 716	**Intervention phase** n = 2048	**p***
Age, years† (median)	76	75	77	0.015
Length of hospital stay, days (median)	5	4	5	0.140
Gagne score (mean)	5.14	4.75	5.27	< 0.0001

** Mann-Whitney-Wilcoxon test for significance*

† Ages ≥ 90 recorded as 90

Table 5: Associations between categorical covariates and exposure phase

Characteristic	Overall N = 2764	Pre-intervention phase n = 716	Intervention phase n = 2048	p*
Sex				0.340
female	50.0% (1383)	48.5% (347)	50.6% (1036)	
male	50.0% (1381)	51.5% (369)	49.4% (1012)	
Race/ethnicity				0.202
white	73.6% (2035)	74.3% (532)	73.4% (1503)	
black	10.9% (301)	12.2% (87)	10.4% (214)	
Hispanic	14.1% (389)	12% (86)	14.8% (303)	
other/unknown	1.4% (39)	1.5% (11)	1.4% (28)	
Medical charge payor				0.643
commercial insurance/HMO	16.2% (449)	15.1% (108)	16.7% (341)	
Medicaid	5.6% (154)	5.9% (42)	5.5% (112)	
Medicare	76.9% (2126)	77.5% (555)	76.7% (1571)	
other	1.3% (35)	1.5% (11)	1.2% (24)	
Primary diagnosis				< 0.0001
398.x Other rheumatic heart disease	0.9% (24)	3.1% (22)	0.1% (2)	
402.x Hypertensive heart disease	3.8% (105)	5.4% (39)	3.2% (66)	
404.x Hypertensive heart and chronic kidney disease	13.3% (368)	7.3% (52)	15.4% (316)	
428.x Heart failure	82.0% (2267)	84.2% (603)	81.2% (1664)	
Discharge disposition				< 0.0001
against medical advice	0.8% (21)	1% (7)	0.7% (14)	
home, self care	29.2% (807)	38.7% (277)	25.9% (530)	
home, visiting nurse assistance	45.6% (1261)	39% (279)	47.9% (982)	
hospice	2.0% (55)	1.4% (10)	2.2% (45)	
skilled nursing/rehabilitation facility	22.2% (614)	19.8% (142)	23% (472)	
transfer	0.2% (6)	0.1% (1)	0.2% (5)	
Length of stay				0.032
≤ 2 days)	17.5% (483)	19.1% (137)	16.9% (346)	
[3, 8] days	66.1% (1826)	67.3% (482)	65.6% (1344)	
≥ 9 days	16.5% (455)	13.5% (97)	17.5% (358)	

** Fisher's exact test*

Table 6: Durbin-Watson test for autocorrelation, risk of readmission by month

Order	p	
	positive autocorrelation	negative autocorrelation
1	0.254	0.746
2	0.523	0.477
3	0.179	0.820
4	0.256	0.744

Table 7: Unadjusted logistic regression model for association of time, interruption, and interaction time*interruption with log-odds of readmission

Model				**p***
Change in baseline and trend				0.846
Predictor	**odds ratio**	**coefficient**	**95% CI**	**p**
time (month of study, centered)	0.99	-0.009	-0.036, 0.018	0.508
post implementation of protocol (i)	1.05	0.049	-0.445, 0.554	0.847
time * i	1.01	0.012	-0.018, 0.040	0.437

** Likelihood-ratio test compared to naïve (intercept-only) model, χ^2 on 3 df.*

Table 8: Goodness-of-fit measures for unadjusted logistic regression model for association of time, interruption, and interaction time*interruption with log-odds of readmission

Hosmer-Lemeshow		
number of groups	$\hat{C}$ statistic*	p
6	1.69	0.793
8	1.77	0.940
10	4.96	0.762
12	8.45	0.585
14	14.04	0.298
20	14.61	0.689
le Cessie-van Houwelingen-Copas-Hosmer global goodness of fit		
	$\hat{S}$ statistic	p
	-2.18	0.030
Pearson residuals aggregated by month		
	X^2 statistic†	p
	-2.18	0.210

** χ^2 on (number of groups – 2) df*
† χ^2 on 49 df.

Table 9: Unadjusted logistic regression model for association of interruption with log-odds of readmission

Model				
Change in baseline				
Predictor	**odds ratio**	**coefficient**	**95% CI**	**p**
post implementation of protocol (i)	0.95	-0.047	-0.252, 0.162	0.656

Table 10: Unadjusted logistic regression model for association of time with log-odds of readmission

Model				
Secular trend				
Predictor	**odds ratio**	**coefficient**	**95% CI**	**p**
Time	1.00	-0.001	-0.007, 0.005	0.750

Table 11: Adjusted logistic regression model for association of time and interruption with log-odds of readmission

Model				**p***
Change in baseline and trend, adjusted for covariates				< 0.0001
Predictor	**odds ratio**	**coefficient**	**95% CI**	**p**
post implementation of protocol (i)	1.08	0.080	-0.42, 0.591	0.755
time (month of study, centered)	0.99	-0.013	-0.041, 0.014	0.333
time * i	1.02	0.015	-0.015, 0.045	0.313
age (years)	0.99	-0.013	-0.021, -0.005	0.002
sex				
female	reference			
male	0.80	-0.225	-0.414, -0.036	0.020
race/ethnicity				
white	reference			
black	1.00	-0.001	-0.322, 0.309	0.994
Hispanic	1.47	0.384	0.112, 0.651	0.005
other/unknown	1.42	0.354	-0.429, 1.055	0.345
medical charge payor				
commercial insurance/HMO	reference			
Medicaid	1.80	0.585	0.151, 1.013	0.008
Medicare	1.61	0.475	0.185, 0.775	0.002
other	0.78	-0.252	-1.494, 0.724	0.647
Gagne score	1.09	0.087	0.045, 0.129	< 0.0001

** Likelihood-ratio test compared to naïve (intercept-only) model, χ^2 on 12 df.*

Table 12: Associations between continuous covariates in regression model and month of study (omitting data from ramp-up phase)

Characteristic	Pearson's *r*	Spearman's *ρ*	Kendall's τ
age	0.05	0.05	0.03
Gagne score	0.12	0.11	0.08

Table 13: Associations between categorical covariates in regression model and month of study (omitting data from ramp-up phase)

Characteristic	p*
race/ethnicity	0.070
sex	0.462
medical charge payor	0.799

** Kruskal-Wallis rank-sum test for significance*

Table 14: Goodness-of-fit measures for adjusted logistic regression model for association of time, interruption, and interaction time*interruption with log-odds of readmission

Hosmer-Lemeshow		
number of groups	$\hat{C}$ statistic*	p
6	4.83	0.306
8	4.76	0.575
10	1.86	0.985
12	8.35	0.595
14	6.33	0.899
20	16.89	0.530
le Cessie-van Houwelingen-Copas-Hosmer global goodness of fit		
	$\hat{S}$ statistic	p
	0.13	0.894

** χ^2 on (number of groups – 2) df*

Table 15: Adjusted logistic regression model for association of interruption with log-odds of readmission

Model				**p***
Change in baseline, adjusted for covariates				< 0.0001
Predictor	**odds ratio**	**coefficient**	**95% CI**	**p**
post implementation of protocol	0.91	0.092	-0.301, 0.120	0.392
age (years)	0.99	-0.013	-0.021, -0.005	0.002
sex				
female	reference			
male	0.80	-0.222	-0.412, -0.034	0.021
race/ethnicity				
white	reference			
black	1.00	0.004	-0.325, 0.306	0.981
Hispanic	1.46	0.377	0.106, 0.644	0.006
other/unknown	1.41	0.343	-0.438, 1.043	0.358
medical charge payor				
commercial insurance/HMO	reference			
Medicaid	1.80	0.589	0.155, 1.017	0.007
Medicare	1.60	0.470	0.180, 0.769	0.002
Other	0.76	-0.274	-1.516, 0.702	0.619
Gagne score	1.09	0.086	0.044, 0.127	< 0.0001

Table 16: Adjusted logistic regression model for association of time with log-odds of readmission

Model				**p***
Secular trend, adjusted for covariates				< 0.0001
Predictor	**odds ratio**	**coefficient**	**95% CI**	**p**
time (month of study, centered)	1.00	0.002	-0.008, 0.004	0.433
age (years)	0.99	0.013	-0.021, -0.005	0.002
sex				
female	reference			
male	0.80	0.222	-0.411, -0.033	0.021
race/ethnicity				
white	reference			
black	1.00	-0.004	-0.325, 0.305	0.979
Hispanic	1.46	0.378	0.107, 0.646	0.006
other/unknown	1.41	0.345	-0.437, 1.045	0.356
medical charge payor				
commercial insurance/HMO	reference			
Medicaid	1.80	0.589	0.155, 1.017	0.007
Medicare	1.60	0.471	0.181, 0.770	0.002
other	0.76	-0.270	-1.512, 0.705	0.624
Gagne score	1.09	0.086	0.044, 0.128	< 0.0001

** Likelihood-ratio test compared to naïve (intercept-only) model, χ^2 on 10 df.*

Table 17: Study power (by simulation)

γ	OR_γ	ρ	OR_ρ	Power to detect (%) $OR_\gamma \neq 1$	Power to detect (%) $OR_\rho \neq 1$
-0.92	0.4	-0.062	0.94	98	99
		-0.051	0.95	98	96
		-0.041	0.96	98	82
		-0.030	0.97	98	56
		-0.020	0.98	98	30
		-0.010	0.99	98	13
		0.010	1.01	98	10
		0.020	1.02	97	29
		0.030	1.03	96	55
		0.039	1.04	94	76
		0.049	1.05	94	93
		0.058	1.06	96	97
-0.69	0.5	-0.062	0.94	87	99
		-0.051	0.95	84	95
		-0.041	0.96	85	80
		-0.030	0.97	84	56
		-0.020	0.98	85	32
		-0.010	0.99	88	12
		0.010	1.01	85	11
		0.020	1.02	81	31
		0.030	1.03	78	55
		0.039	1.04	75	77
		0.049	1.05	77	90
		0.058	1.06	73	97
-0.51	0.6	-0.062	0.94	62	99
		-0.051	0.95	64	95
		-0.041	0.96	58	82
		-0.030	0.97	64	60
		-0.020	0.98	58	29
		-0.010	0.99	60	14
		0.010	1.01	60	9
		0.020	1.02	53	28
		0.030	1.03	52	53
		0.039	1.04	52	79
		0.049	1.05	47	88
		0.058	1.06	49	95
-0.36	0.7	-0.062	0.94	34	99
		-0.051	0.95	35	96
		-0.041	0.96	35	83
		-0.030	0.97	34	60
		-0.020	0.98	32	29
		-0.010	0.99	34	12
		0.010	1.01	33	12
		0.020	1.02	31	25
		0.030	1.03	32	47
		0.039	1.04	27	74
		0.049	1.05	25	84
		0.058	1.06	23	95

				Power to detect (%)	
γ	OR_γ	ρ	OR_ρ	$OR_\gamma \neq 1$	$OR_\rho \neq 1$
-0.22	0.8	-0.062	0.94	15	98
		-0.051	0.95	18	95
		-0.041	0.96	17	85
		-0.030	0.97	19	56
		-0.020	0.98	17	30
		-0.010	0.99	16	11
		0.010	1.01	15	9
		0.020	1.02	15	27
		0.030	1.03	16	50
		0.039	1.04	14	68
		0.049	1.05	13	88
		0.058	1.06	14	94
0.18	1.2	-0.062	0.94	13	99
		-0.051	0.95	11	97
		-0.041	0.96	14	81
		-0.030	0.97	11	56
		-0.020	0.98	11	27
		-0.010	0.99	16	11
		0.010	1.01	9	9
		0.020	1.02	10	24
		0.030	1.03	8	42
		0.039	1.04	6	64
		0.049	1.05	10	80
		0.058	1.06	10	88
0.26	1.3	-0.062	0.94	21	99
		-0.051	0.95	24	93
		-0.041	0.96	22	82
		-0.030	0.97	19	58
		-0.020	0.98	15	27
		-0.010	0.99	16	11
		0.010	1.01	13	9
		0.020	1.02	17	27
		0.030	1.03	15	42
		0.039	1.04	14	62
		0.049	1.05	13	83
		0.058	1.06	13	88
0.41	1.5	-0.062	0.94	39	100
		-0.051	0.95	40	94
		-0.041	0.96	39	81
		-0.030	0.97	40	55
		-0.020	0.98	36	28
		-0.010	0.99	36	10
		0.010	1.01	30	7
		0.020	1.02	34	25
		0.030	1.03	29	45
		0.039	1.04	26	61
		0.049	1.05	27	74
		0.058	1.06	23	86

				Power to detect (%)	
γ	OR_γ	ρ	OR_ρ	$OR_\gamma \neq 1$	$OR_\rho \neq 1$
0.69	2.0	-0.062	0.94	85	99
		-0.051	0.95	80	94
		-0.041	0.96	80	77
		-0.030	0.97	79	49
		-0.020	0.98	76	25
		-0.010	0.99	74	8
		0.010	1.01	69	7
		0.020	1.02	63	20
		0.030	1.03	62	36
		0.039	1.04	61	57
		0.049	1.05	54	67
		0.058	1.06	50	78
0.92	2.5	-0.062	0.94	95	99
		-0.051	0.95	95	90
		-0.041	0.96	95	74
		-0.030	0.97	95	43
		-0.020	0.98	95	27
		-0.010	0.99	93	11
		0.010	1.01	91	10
		0.020	1.02	88	19
		0.030	1.03	87	33
		0.039	1.04	81	52
		0.049	1.05	78	65
		0.058	1.06	73	75

Table 18: Study power (by simulation) to detect $OR_\gamma \neq 1$

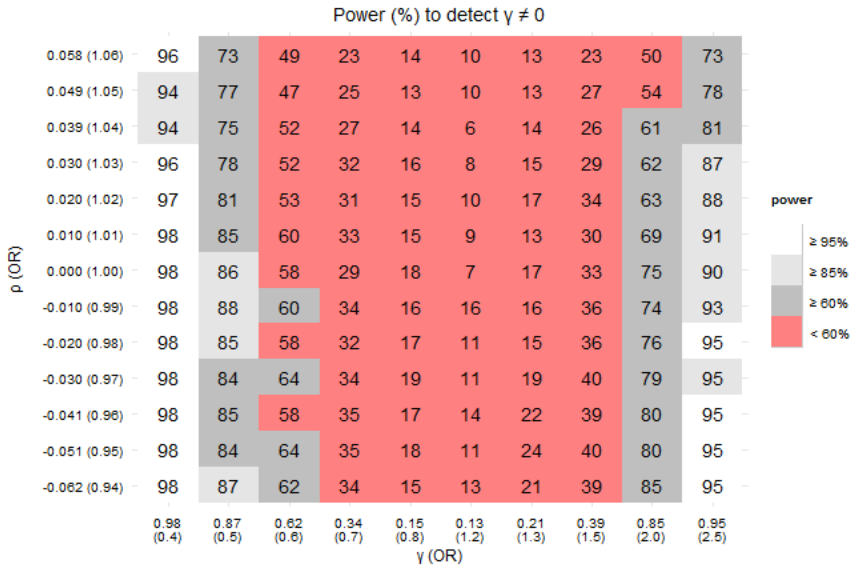
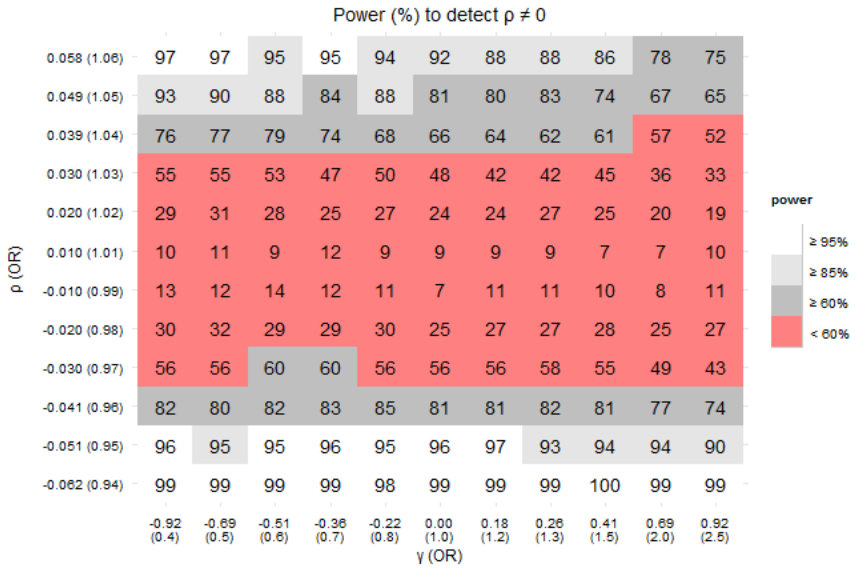


Table 19: Study power (by simulation) to detect $OR_\rho \neq 1$



Figures

Figure 2: Probability of readmission by month of study, with weighted moving average

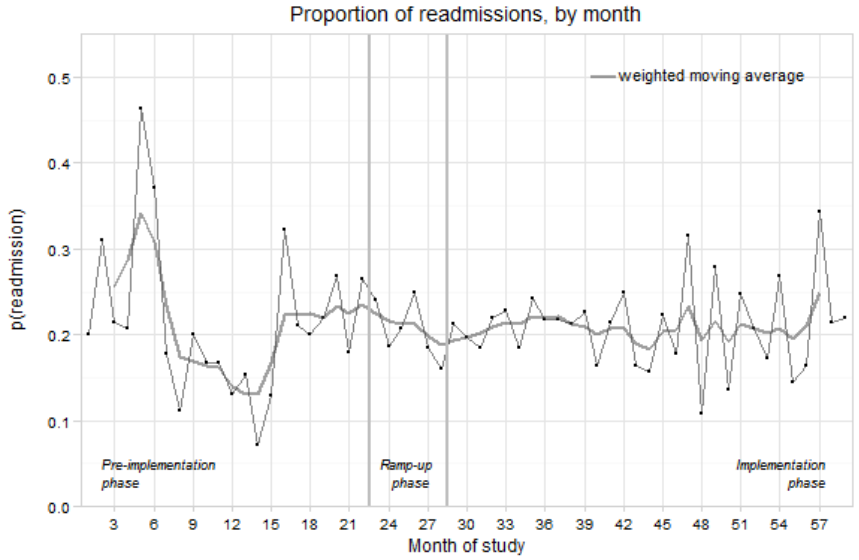


Figure 3: Logit(probability of readmission) by month of study, with weighted moving average

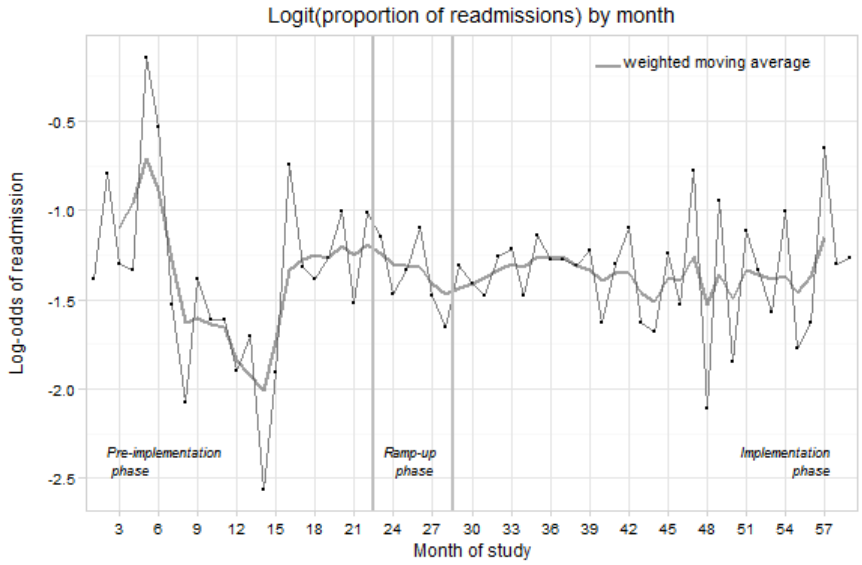


Figure 4: Histogram of probability of readmission, by month, with kernel density

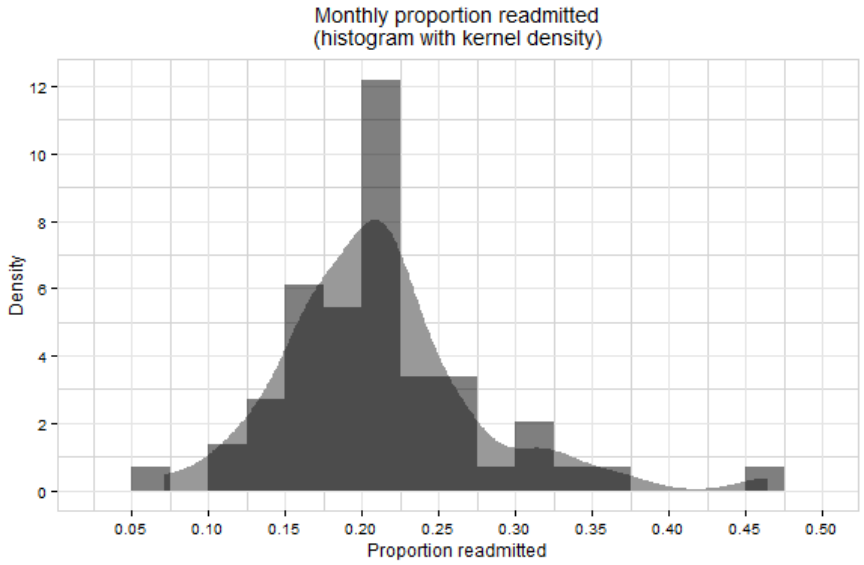


Figure 5: Autocorrelation of probability of readmission by calendar month

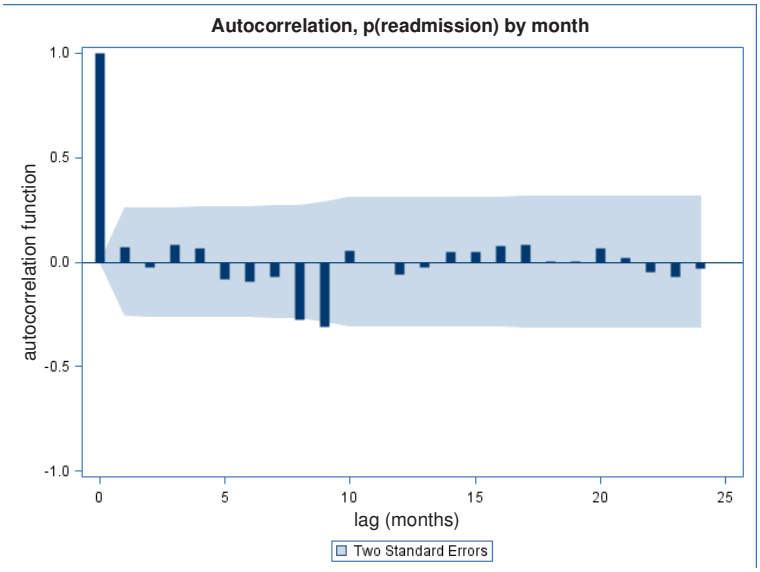


Figure 6: Autocorrelation of probability of readmission by quarter

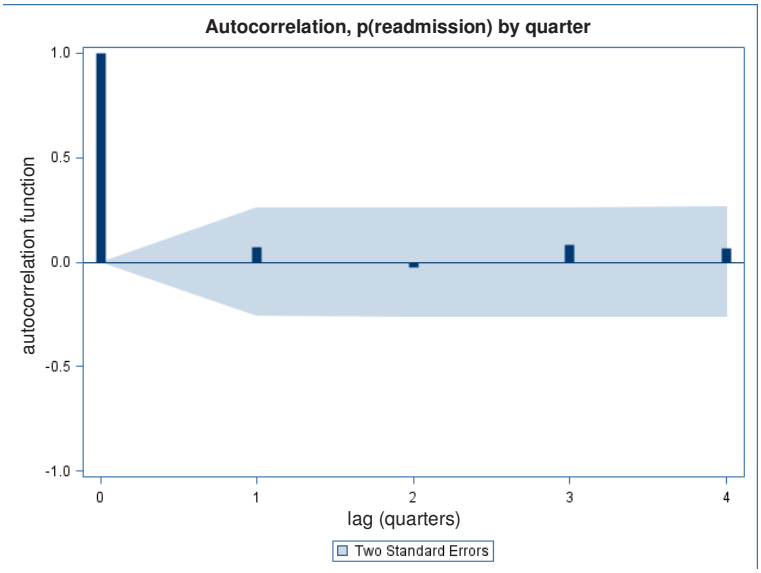


Figure 7: Q-Q normal plot of Pearson residuals, for readmissions aggregated by calendar month, with regression model including *time, i,* and *time*i* as predictors

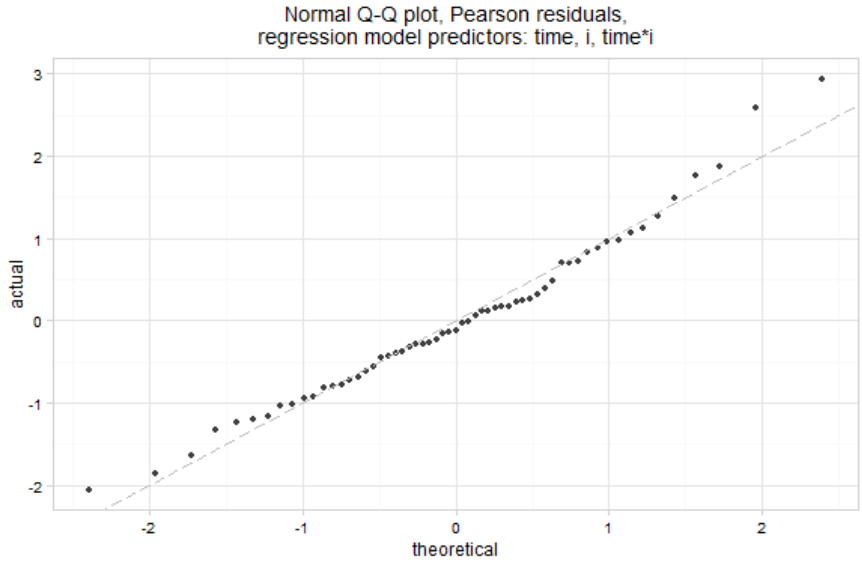


Figure 8: Logit(probability of readmission) by month of study, pre-implementation phase, with loess best fit and best-fit line

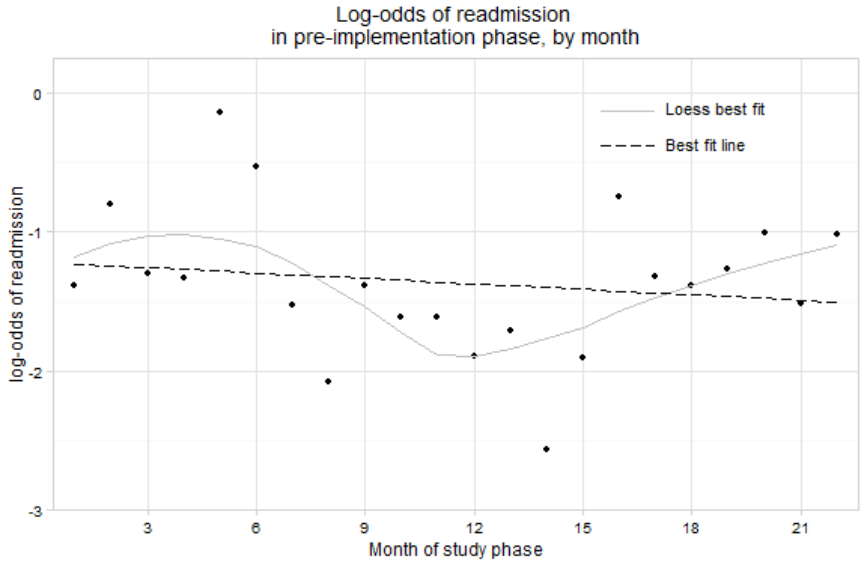


Figure 9: Logit(probability of readmission) by month of study, post-implementation phase, with loess best fit and best-fit line

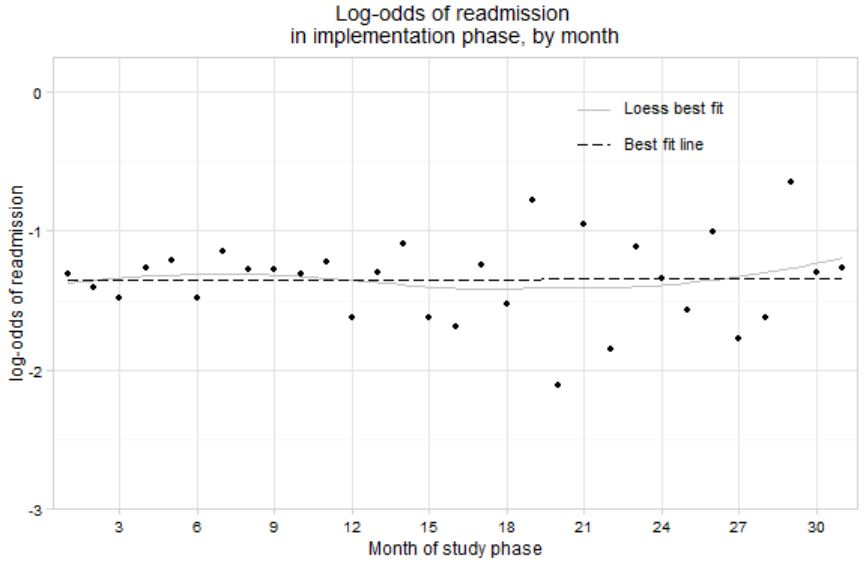
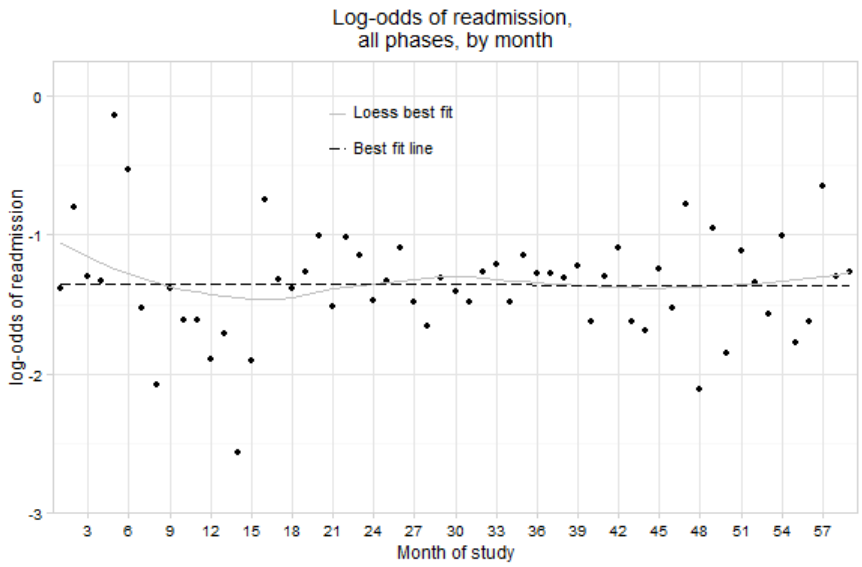


Figure 10: Logit(probability of readmission) by month of study, all study months, with loess best fit and best-fit line



APPENDIX:
ICD-9 DIAGNOSIS CODES USED TO DETERMINE CHF DISCHARGE DIAGNOSIS

Description	Code
Heart failure	428.x
Congestive heart failure, unspecified	428.0
Left heart failure	428.1
Systolic heart failure (.2x, x = 0 – 3)	428.2
Diastolic heart failure (.3x, x = 0 – 3)	428.3
Combined systolic and diastolic heart failure (.4x, x = 0 – 3)	428.4
Rheumatic heart failure	398.91
Malignant Hypertensive heart disease with HF	402.01
Benign Hypertensive heart disease with HF	402.11
Unspecified Hypertensive heart disease with HF	402.91
Malignant Hypertensive heart and chronic kidney disease stage I–IV, or unspecified with HF	404.01
Malignant Hypertensive heart and chronic kidney disease stage V or ESRD with HF	404.03
Benign Hypertensive heart and chronic kidney disease stage I–IV, or unspecified with HF	404.11
Benign Hypertensive heart and chronic kidney disease stage V or ESRD with HF	404.13
Unspecified Hypertensive heart and chronic kidney disease stage I - IV, or unspecified with HF	404.91
Unspecified Hypertensive heart and chronic kidney disease stage V or ESRD with HF	404.93
Cardiomegaly	429.3

REFERENCES

1. Heidenreich PA, Albert NM, Allen LA, et al. Forecasting the Impact of Heart Failure in the United States A Policy Statement From the American Heart Association. *Circ. Heart Fail.* 2013;6(3):606–619.

2. Pazos-Lopez P, Peteiro-Vazquez J, Carcia-Campos A, et al. The causes, consequences, and treatment of left or right heart failure. *Vasc. Health Risk Manag.* 2011;7:237–254.

3. Bui AL, Horwich TB, Fonarow GC. Epidemiology and risk profile of heart failure. *Nat. Rev. Cardiol.* 2011;8(1):30–41.

4. Hobbs R. Clinical burden and health service challenges of chronic heart failure. *Br. J. Gen. Pract.* 2010;60(577):611–615.

5. June 2007 Report to the Congress: Promoting Greater Efficiency in Medicare - Jun07_EntireReport.pdf. (http://www.medpac.gov/documents/Jun07_EntireReport.pdf). (Accessed November 10, 2013)

6. Medicare's Hospital Readmission Reduction Program FAQ. (http://www.acep.org/Legislation-and-Advocacy/Practice-Management-Issues/Physician-Payment-Reform/Medicare-s-Hospital-Readmission-Reduction-Program-FAQ/). (Accessed October 3, 2013)

7. Brown RS, Peikes D, Peterson G, et al. Six Features Of Medicare Coordinated Care Demonstration Programs That Cut Hospital Admissions Of High-Risk Patients. *Health Aff. (Millwood)*. 2012;31(6):1156–1166.

8. Bradley EH, Curry L, Horwitz LI, et al. Hospital Strategies Associated With 30-Day Readmission Rates for Patients With Heart Failure. *Circ. Cardiovasc. Qual. Outcomes*. 2013;6(4):444–450.

9. Silverstein MD, Qin H, Mercer SQ, et al. Risk factors for 30-day hospital readmission in patients >=65 years of age. *Proc. Bayl. Univ. Med. Cent.* 2008;21(4):363–372.

10. Jaarsma T, van der Wal ML, Lesman-Leegte I, et al. Effect of moderate or intensive disease management program on outcome in patients with heart failure: Coordinating study evaluating outcomes of advising and counseling in heart failure (coach). *Arch. Intern. Med.* 2008;168(3):316–324.

11. Del Sindaco D, Pulignano G, Minardi G, et al. Two-year outcome of a prospective, controlled study of a disease management programme for elderly patients with heart failure. *J. Cardiovasc. Med.* 2007;8(5):324–329.

12. Thompson DR, Roebuck A, Stewart S. Effects of a nurse-led, clinic and home-based intervention on recurrent hospital use in chronic heart failure. *Eur. J. Heart Fail.* 2005;7(3):377–384.

13. Lappé JM, Muhlestein JB, Lappé DL, et al. Improvements in 1-Year Cardiovascular Clinical Outcomes Associated with a Hospital-Based Discharge Medication Program. *Ann. Intern. Med.* 2004;141(6):446–W–88.

14. Naylor MD, Brooten DA, Campbell RL, et al. Transitional Care of Older Adults Hospitalized with Heart Failure: A Randomized, Controlled Trial. *J. Am. Geriatr. Soc.* 2004;52(5):675–684.

15. Kimmelstiel C, Levine D, Perry K, et al. Randomized, Controlled Evaluation of Short- and Long-Term Benefits of Heart Failure Disease Management Within a Diverse Provider Network: The SPAN-CHF Trial. *Circ. Sept. 14 2004*. 2004;110(11):1450–1455.

16. Kasper EK, Gerstenblith G, Hefter G, et al. A randomized trial of the efficacy of multidisciplinary care in heart failure outpatients at high risk of hospital readmission. *J. Am. Coll. Cardiol.* 2002;39(3):471–480.

17. Krumholz HM, Amatruda J, Smith GL, et al. Randomized trial of an education and support intervention to preventreadmission of patients with heart failure. *J. Am. Coll. Cardiol.* 2002;39(1):83–89.

18. Anderson C, Deepak B v., Amoateng-Adjepong Y, et al. Benefits of Comprehensive Inpatient Education and Discharge Planning Combined With Outpatient Support in Elderly Patients With Congestive Heart Failure. *Congest. Heart Fail.* 2005;11(6):315–321.

19. Sochalski J, Jaarsma T, Krumholz HM, et al. What Works In Chronic Care Management: The Case Of Heart Failure. *Health Aff. (Millwood)*. 2009;28(1):179–189.

20. Whellan DJ, Hasselblad V, Peterson E, et al. Metaanalysis and review of heart failure disease management randomized controlled clinical trials. *Am. Heart J.* 2005;149(4):722–729.

21. Göhler A, Januzzi JL, Worrell SS, et al. A Systematic Meta-Analysis of the Efficacy and Heterogeneity of Disease Management Programs in Congestive Heart Failure. *J. Card. Fail.* 2006;12(7):554–567.

22. Holland R. Systematic review of multidisciplinary interventions in heart failure. *Heart*. 2005;91(7):899–906.

23. Roccaforte R, Demers C, Baldassarre F, et al. Effectiveness of comprehensive disease management programmes in improving clinical outcomes in heart failure patients. A meta-analysis. *Eur. J. Heart Fail.* 2005;7(7):1133–1144.

24. Phillips CO, Wright SM, Kern DE, et al. Comprehensive discharge planning with postdischarge support for older patients with congestive heart failure: A meta-analysis. *JAMA*. 2004;291(11):1358–1367.

25. Gwadry-Sridhar FH, Flintoft V, Lee DS, et al. A systematic review and meta-analysis of studies comparing readmission rates and mortality rates in patients with heart failure. *Arch. Intern. Med.* 2004;164(21):2315–2320.

26. McAlister FA, Lawson FM., Teo KK, et al. A systematic review of randomized trials of disease management programs in heart failure. *Am. J. Med.* 2001;110(5):378–384.

27. Jaarsma T, Wal MHL van der, Hogenhuis J, et al. Design and methodology of the COACH study: a multicenter randomised Coordinating study evaluating Outcomes of Advising and Counselling in Heart failure. *Eur. J. Heart Fail.* 2004;6(2):227–233.

28. Jaarsma T, Veldhuisen DJ van. When, how and where should we "coach" patients with heart failure: The COACH results in perspective. *Eur. J. Heart Fail.* 2008;10(4):331–333.

29. Dahlström U. Frequent non-cardiac comorbidities in patients with chronic heart failure. *Eur. J. Heart Fail.* 2005;7(3):309–316.

30. Havranek EP, Masoudi FA, Westfall KA, et al. Spectrum of heart failure in older patients: Results from the national heart failure project. *Am. Heart J.* 2002;143(3):412–417.

31. Philbin EF, DiSalvo TG. Prediction of hospital readmission for heart failure: development of a simple risk score based on administrative data. *J. Am. Coll. Cardiol.* 1999;33(6):1560–1566.

32. Quan H, Li B, Duncan Saunders L, et al. Assessing Validity of ICD-9-CM and ICD-10 Administrative Data in Recording Clinical Conditions in a Unique Dually Coded Database. *Health Serv. Res.* 2008;43(4):1424–1441.

33. Gagne JJ, Glynn RJ, Avorn J, et al. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J. Clin. Epidemiol.* 2011;64(7):749–759.

34. Elixhauser A, Steiner C, Harris DR, et al. Comorbidity Measures for Use with Administrative Data. *Med. Care*. 1998;36(1):8–27.

35. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J. Chronic Dis.* 1987;40(5):373–383.

36. Yourman LC, Lee SJ, Schonberg MA, et al. Prognostic Indices for Older Adults: A Systematic Review. *JAMA J. Am. Med. Assoc.* 2012;307(2):182–192.

37. Hosmer DW, Hosmer T, Le Cessie S, et al. A Comparison of Goodness-of-Fit Tests for the Logistic Regression Model. *Stat. Med.* 1997;16(9):965–980.

38. Allison P. Measures of Fit for Logistic Regression. In: *SAS Global Forum 2014*. Washington, DC: 2014 (Accessed August 26, 2014)(http://www.statisticalhorizons.com/wp-content/uploads/GOFForLogisticRegression-Paper.pdf). (Accessed August 26, 2014)

39. Team RC. R: A Language and Environment for Statistical Computing. Vienna, Austria: 2013.(http://www.R-project.org/)

40. SAS Software. Cary, NC, USA: SAS Institute Inc; version 9.4.

41. R Core Team. foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ... 2014.(http://CRAN.R-project.org/package=foreign)

42. Wickham H. ggplot2: elegant graphics for data analysis. Springer New York; 2009.(http://had.co.nz/ggplot2/book)

43. Harrell FE. Hmisc: Harrell Miscellaneous. 2014.(http://CRAN.R-project.org/package=Hmisc)

44. Zeileis A, Hothorn T. Diagnostic Checking in Regression Relationships. *R News*. 2002;2(3):7–10.

45. Kohl M. MKmisc: Miscellaneous functions from M. Kohl. 2013.(http://www.stamats.de)

46. Harrell FE. rms: Regression Modeling Strategies. 2014.(http://CRAN.R-project.org/package=rms)

47. Combined Comorbidity Score for Claims Data. *Brigh. Womens Hosp. Div. Pharmacoepidemiol. Pharmacoeconomics*. (http://www.drugepi.org/dope-downloads/#Combined Comorbidity Score for Claims Data). (Accessed September 4, 2014)

48. Hoenig JM, Heisey DM. The Abuse of Power. *Am. Stat.* 2001;55(1):19–24.