Title
Evaluating Quality Improvement Interventions: Strengthening Causal Inference with Observational Data

Author
Prasad, Priya April

Publication Date
2016

Peer reviewed|Thesis/dissertation

Evaluating Quality Improvement Interventions: Strengthening Causal Inference with Observational Data

by

Priya April Prasad, M.P.H

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Epidemiology and Translational Sciences

[redacted]

Dedication and Acknowledgements

There are several people to whom I owe a debt of gratitude for supporting me throughout my life, education, and career.

My sincerest thanks go to my dissertation committee, Ralph Gonzales, Stephen Shiboski, Lydia Zablotska, and Nathaniel Gleason, as well as Margaret Handley, who served on my Qualifying Exam committee. One of the hallmarks of my experience in the Epidemiology & Translational Science program has been drawing on your knowledge, expertise, and commitment to mentorship and it was an honor working with each of you.

Nat, I am so grateful for the opportunities you have given me as my sponsor to learn, grow, and apply my academic training to the complex and ever-changing landscape of quality improvement in the era of electronic medical records. Look forward to decades of collaboration stemming from our work together!

My thanks go to the faculty and staff of the Department of Epidemiology & Biostatistics, in particular Robert Hiatt, Maria Glymour, and Thomas Newman, for your counsel and guidance over the last several years. Becoming a student again after 7 years in the workforce certainly took some adjustment but the curriculum you developed for our program enriched my experience and has provided me with the toolbox to continue my service as an epidemiologist. I have been so inspired by my fellow Epidemiology & Translational graduate students who I am thrilled to have cultivated lasting friendships with, particularly my cohort-mates, Yea-Hung Chen, Cecily Miller, Heidi Moseson, and Eugénie Poirot. Problem set and study parties along with regular sanity checks were appreciated throughout the years!

My decision to pursue a PhD would not have been a reality without the support and encouragement of Theoklis Zaoutis and Susan Coffin. You gave me confidence to know that something bigger was out there for me to explore through your faith in my abilities, your guidance and mentorship, your encouragement of my professional development, your leadership… Leaving the research group, the family you created, was one of the most challenging decisions of my life but I knew you were behind me, rooting for me from afar.

To Kateri Leckerman, my dear friend who we lost too soon. The soul of this dissertation was formed in our office, our hide-away, as we scoured the literature for ways to conduct our analyses and lived our lives back to back and together for those wonderful years at The Children's Hospital of Philadelphia… wanting to do meaningful work, wanting to make a difference in people's lives, wanting to show that what we were doing mattered and was valid, in both data and in mission. I carry you with me during each study, each analysis, and every line of code I write. I promise to continue the work we started together over a decade ago in your honor.

I feel so blessed to have had my parents to guide me throughout my life. Krishna and Anandhi Prasad, you have been my constants, supporting me, offering advice when needed or just providing a listening ear as I traveled north, south, east and west… home is always wherever you are. Thank you for having faith in me and investing in my education and my character from day one.

To my husband, Jeremi, our journey and our love has spanned cities, states, countries and continents. You have stayed true to your promise at the blinking red stoplight late at night on Lancaster Ave; you have shown me great things. Thank you for encouraging me to pursue this degree and for being my sounding board when I was in the thick of it. I love you, our beautiful family, and the life we are building. You are my good luck charm and my best friend.

And to my Kiran. You have borne witness to the most exhilarating and most challenging moments of the last 3 years. You were there with me, kicking and hiccupping, as I studied for my Qualifying Exam, listening to album after album with me (I am convinced this is why music is such an important part of your life to this day!), arriving within hours of turning in my exam. It is a distinct joy to be your mother and to watch you grow and become a gracious and loving human being, the kind of boy who will share a morsel of food before taking a bite himself. I hope that you will continue to blossom into a person who will explore, create, and fight for what he believes in, lifting up those around him who are unable to find their voice. The work I do is in your name, to secure a brighter future for the life you will lead. Always know that giving you life gave me life.

Evaluating Quality Improvement Interventions:

Strengthening Causal Inference with Observational Data

Priya April Prasad

Abstract

As innovations in healthcare delivery systems and electronic medical records (EMR) data capture develop, the methods employed to evaluate interventions disseminated to improve the quality and efficiency of patient care should evolve as well. Healthcare quality improvement (QI) interventions can be evaluated at the level of the individual or the group and a key initial step in designing an evaluation plan is to determine the desired level of inference. Before data are even collected, a comprehensive analytic plan should be developed to accurately measure outcomes and exposures, taking into account the source, fidelity, and complexity of the data available for inclusion and the assumptions that must be fulfilled when assessing the effect of the intervention. Through my dissertation work I have explored these themes and how they relate to the development of a responsible QI evaluation in three separate domains.

The first chapter of my dissertation focuses on assessing the effect of an intervention at the individual level to identify and manage sepsis, a syndrome which causes significant morbidity and mortality. In a retrospective cohort study of adults discharged from the University of California San Francisco (UCSF) Medical Center with severe sepsis or septic shock between 2012 and 2014, the adjusted risk of mortality was estimated using Poisson regression for a binary outcome variable and an adjusted number needed to treat (NNT) was calculated. The analysis revealed that the UCSF sepsis bundle was associated with a 31% decreased risk of in-hospital mortality across hospital units (adjusted incidence rate ratio 0.69, 95% Confidence Interval (CI) 0.53, 0.91) after robust control for confounders and risk adjustment and the adjusted NNT of 15 (CI 8, 69) provides a reasonable and achievable goal to observe measureable improvements in outcomes for patients diagnosed with SS/SS.

In the second chapter of my dissertation, I review the causal inference framework and how it relates to the evaluation of quality improvement interventions that use aggregate outcome and exposure

data in the setting of interrupted time-series analysis. The reader is presented with an analysis plan diagram along with detailed guidance on how to design a robust quality improvement evaluation. The manuscript focuses on the different strategies that can be employed when modeling the effect of an intervention measured at the group level and the reader is presented with a concrete example from a multidisciplinary intervention implemented at UCSF Medical Center to decrease the use of packed red blood cell transfusions.

The third chapter of this work focuses on developing a metric to measure timely access to ambulatory specialty care. Using data from nearly 60,000 patients who sought primary care at UCSF Medical Center between 2013 and 2015, I explored associations between population-level weekly ambulatory specialty care access defined three ways and the rate of population-level weekly poor health outcomes, including emergency department visits, inpatient encounters, and mortality. Specialties of interest included cardiology, hematology, neurology, otolaryngology and head and neck surgery, and urology. Based on unadjusted Poisson analysis, there were correlations identified between poor outcomes and timely access to care in some specialties and the results provide a springboard for future exploration of metric performance and adjustment.

As a body of work, my dissertation illustrates the breadth of the field of healthcare QI, providing evidence to support the continued evolution of robust methods for evaluation of interventions that will improve the quality, safety, and value of healthcare delivered nationally and globally.

Table of Contents

List of Tables

Chapter 1:

Chapter 2:

Chapter 3:

List of Figures and/or Illustrations

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 1: Relationship between a sepsis intervention bundle and in-hospital mortality among hospitalized patients: a retrospective analysis of real-world data

Priya A. Prasad, Erica R. Shea, Stephen Shiboski, Mary C. Sullivan, Ralph Gonzales, and David Shimabukuro

INTRODUCTION

Sepsis is a systemic response to infection which can lead to tissue damage, organ failure, and death. The incidence of severe sepsis has increased in recent years in the United States[1] while the mortality has decreased and ranges from 15% to 30%[1,2].

In 2001, Rivers et al. demonstrated significant reduction in septic shock mortality through the use of early goal-directed therapy[3]. Based on these findings, in 2002 the Surviving Sepsis Campaign (SSC) was formed to increase sepsis awareness among clinicians and to provide evidence-based guidelines on sepsis control and management. The SSC, in collaboration with the European Society of Intensive Care Medicine (ESICM), created a standardized bundle of interventions to manage sepsis early in the illness course and tested it worldwide[4]. The University of California, San Francisco (UCSF), received funding from the California Delivery System Reform Incentive Program to deploy an intervention bundle similar to the SSC and track bundle compliance among the patients diagnosed with severe sepsis or septic shock (SS/SS).

While data from randomized controlled trials demonstrate the protective effect of the SSC's resuscitation bundle[5], the aim of our study was to explore the relationship between in-hospital mortality and a minimally invasive SS/SS bundle within a retrospective observational cohort. Using methods developed for the robust analysis of observational data, we provide evidence supporting the practical application of a quality improvement initiative to address outcomes related to SS/SS.

METHODS

Study Population

Subjects were adult patients (age ≥ 18 years) discharged between January 1, 2012 and December 31, 2014 and who received an International Classification of Diseases, 9th Revision (ICD-9) diagnosis code of Severe Sepsis (995.92) or Septic Shock (785.82) at UCSF, a quaternary care academic medical center. Patients were excluded if they were admitted with a "Do Not Resuscitate/Do Not Intubate" (DNR/DNI) order, if their status changed to DNR/DNI at any time between admission and within a day of sepsis presentation, or if they were transferred from another institution with SS/SS present on admission (POA) or developed SS/SS within 24 hours of admission. If a subject had multiple admissions with a diagnosis of SS/SS during the study period, each discharge was included independently in the analysis.

Major Study Definitions

SS/SS time of presentation:

Time of SS/SS presentation was defined as the time at which two signs of systemic inflammatory response syndrome (SIRS)[6] and one sign of organ failure in the presence of a known or suspected infection were identified through chart review. Chart abstraction entailed an extensive review of provider progress notes and nursing documentation, along with vital sign and laboratory data. All cases were first reviewed by a registered nurse quality analyst. Complex cases were escalated to the medical director of the Sepsis Program and/or the emergency department (ED) physician champion for further review. Furthermore, the cases found to be non-compliant were sent to the front-line teams for feedback and review.

Bundle compliance: Elements of the intervention bundle appear in Table 1. Of note, the UCSF bundle did not include measurement of central venous pressure or central venous oxygen saturation through a central venous catheter. Compliance with each bundle element was assessed starting from the time of SS/SS presentation based on medical record review. Bundle compliance was represented as a binary variable indicating that all eligible elements were satisfied (i.e. care was considered bundle compliant), In addition, individual bundle elements were evaluated to investigate impacts on mortality.

Mortality: In-hospital mortality was determined through manual chart abstraction.

Data collection

Data on bundle compliance, discharge disposition, and demographics were collected using the UCSF electronic medical record (EMR). Clinicians (nurses and physicians) conducted a standardized chart review of every study subject to confirm SS/SS diagnosis and determine the time of SS/SS presentation.

The Vizient Clinical Database/Resource Manager (CDB/RM) was used to identify subject comorbidities based on administrative data. Vizient membership includes 117 not for profit academic hospitals and more than 300 affiliated community hospitals in 42 US states. Approximately 110 academic hospitals and 190 affiliate community hospitals participate in the CDB/RM. Elements extracted included diagnosis codes with POA designation, procedure codes, admission/discharge severity of illness and risk of mortality calculated using 3M's APR-DRG groupers, and Medicaid/Medicare as primary or secondary payor.

Immunocompromised status and malignancy were identified using Vizient service lines (Appendix A). End stage renal disease (ESRD) (ICD-9 585.6) and congestive heart failure (CHF) (ICD-9 428.0) POA were identified using ICD-9 codes. SS/SS was counted as POA if the time of presentation based on manual symptom review was within 24 hours of hospital admission or ED triage time, whichever was earlier. The study was approved by the UCSF Committee on Human Research, received a waiver of documentation of informed consent, and adheres to the EQUATOR guidelines.

Statistical analysis

Chi-squared tests were used to evaluate associations between the in-hospital mortality outcome, bundle compliance, and other categorical variables describing demographic and admission characteristics (Table 2). Differences in the distribution of patient age between groups defined by the bundle compliance and in-hospital mortality variables were summarized using medians and interquartile ranges (IQR) and with Mann-Whitney rank-sum tests. Multiple predictor regression models were constructed to identify independent risk factors for in-hospital mortality. Models included the bundle compliance indicator

variable and adjusted for additional predictors with marginal associations with the outcome significant at the $p< 0.20$ level. For ease of interpretation, we elected to use Poisson regression for the in-hospital mortality outcome with robust standard errors[7] as the exponentiated regression coefficients from this approach have an adjusted IRR interpretation. Fitted models were also used to estimate the marginal risk difference comparing mortality risk in the two bundle compliance groups, as well as the number needed to treat (NNT) to save one life.

A propensity score model was developed to relate bundle compliance to confounders of the association with in-hospital mortality. Categorical variables for quintiles of propensity score were then generated and the degree of balance achieved by the estimated scores in the distribution of confounders between those who were bundle compliant and those who were bundle non-compliant was assessed. We then graphically compared the distributions of propensity scores between groups to assess the degree of balance and overlap. As a confirmatory analysis, we included the propensity score quintiles as categorical variables in the Poisson regression to obtain the marginal risk ratio and risk difference for the estimated effect of bundle compliance on in-hospital mortality. Finally, we used the estimated propensity scores to define inverse probability of treatment weights and incorporated them into the Poisson regression to obtain the marginal risk ratio and risk difference for the estimated effect of bundle compliance on mortality.[8]

We explored potential interactions between bundle compliance and other predictors included in the adjusted regression model. We performed stratified analyses by diagnosis code status using separate regressions as the two included sepsis codes reflect distinct disease processes and we wanted to identify whether the bundle differentially affected these two populations.

All analyses were conducted using Stata 13 (College Station, TX).

RESULTS

Study cohort

During the three year study period there were 69,582 adult discharges and 1,844 discharges with an ICD-9 code for SS/SS at UCSF (incidence of 27 per 1,000 discharges). Of those, 1,029 discharges

(956 unique patients) met the inclusion criteria for bundle compliance assessment. Those who were excluded from assessment were more likely to have expired at discharge (35% v. 18%, p <0.00001) and were statistically significantly older (median 66 years v. 64 years, p=0.0001). All patients had the severe sepsis code documented and 446 patients also had the septic shock code documented. The majority of patients presented with SS/SS in the ED (703, 68%), followed by the ICU (137, 13%), acute care units (128, 12%), and transitional units (61, 6%).

Relationship between bundle compliance and in-hospital mortality

Among the 1,029 study subjects, 742 were bundle compliant (72%). Those who were bundle compliant were less likely to die based on bivariable analysis (14% v. 27% mortality, p<0.0001)(Figure 1). There were 181 total subjects (18%) who died during the study period with a median time to death of 9 days (IQR 4 to 20 days), with no significant difference in time to death between those who were bundle compliant or non-compliant (median of 9 days versus 10 days, respectively, p=0.3065). Factors marginally associated with in-hospital mortality upon bivariable analyses can be found in Table 2. Overall bundle compliance was significantly associated with decreased mortality on bivariable analysis. Further, all individual bundle elements except vasopressor administration were also significantly associated with decreased in-hospital mortality in the bivariable analysis.

The adjusted risk of in-hospital mortality was an estimated 31% lower among those who experienced complete bundle compliance compared to those who did not (IRR 0.69, CI 0.53, 0.91), when controlling for SS/SS presentation in the ED, SS/SS POA, age, admission severity of illness and risk of mortality, Medicaid/Medicare payor status, immunocompromised host status, and CHF POA (Table 3). The adjusted marginal risk difference was 6.8% (CI 1.4%, 12.1%) and the adjusted NNT was 15 (CI 8, 69).

There were no significant issues of balance or overlap identified when mean values of propensity scores within propensity score quintiles were compared between those who were bundle compliant and bundle non-compliant (Table 4). In addition, no significant areas of overlap were identified with a graphical comparison of the distributions of propensity scores between groups (Supplemental Figure 1),

indicating that our adjusted estimates were adequately supported by the data collected. When propensity score quintiles were included in the Poisson regression as categorical variables, the marginal IRR was 0.70 (CI 0.53, 0.93) and the marginal risk difference was 6.6% (CI 1.2%, 12.0%). When propensity scores were included in the Poisson regression as inverse probability of treatment weights, the marginal IRR was 0.72 (CI 0.53, 0.98) and the marginal risk difference was 5.8% (CI 0.07%, 11.6%).

There was no statistically significant difference in the relationship between in-hospital mortality and the bundle among patients documented with the severe sepsis code alone (adjusted IRR 0.73, CI 0.46, 1.17); however, there was a trend towards a decreased risk among those who had the septic shock code documented and received bundle compliant care (adjusted IRR 0.72, CI 0.52, 1.01).

There were no statistically significant interactions ($p<0.20$) identified between bundle compliance and the other factors in the multivariable model. Of note, there was no statistically significant interaction between bundle compliance and sepsis POA (interaction term $p = 0.686$).

Other Factors Independently Associated with Mortality in the Study Cohort

Patients who presented with SS/SS were significantly less likely to die (adjusted IRR 0.55, CI 0.32, 0.92) when compared to those who developed SS/SS after hospital admission (Table 3). Age was also a significant predictor of mortality in the study cohort (adjusted IRR 1.13 per 10 year increase in age, CI 1.03, 1.24).

DISCUSSION

Based on observational data from a retrospective cohort of SS/SS patients who received care at UCSF, complete bundle compliance with a minimally invasive sepsis intervention bundle was associated with a 31% decreased risk of in-hospital mortality, the adjusted in-hospital mortality risk difference between those who did and did not experience complete bundle compliance was 6.8% (CI 1.4%, 12.1%), and the adjusted NNT to save one life was 15 individuals (CI 8, 69). In addition, those who had an SS/SS diagnosis POA experienced a 45% decreased risk of in-hospital mortality after controlling for confounders.

Our findings demonstrate that those who experienced complete compliance with a sepsis care bundle implemented as a quality improvement initiative were less likely to expire, which is important in the setting of poorly-understood epidemiological trends in SS/SS incidence and mortality. The increasing incidence of severe sepsis has been accompanied by steadily declining mortality rates but it remains unclear how much of the decline in mortality is due to improved sepsis recognition and treatment versus the effect of increased coding capture which floods the severe sepsis denominator with less severely ill patients[9,10]. At our institution the incidence of ICD-9 coded SS/SS was 27 per 1,000 discharges. Angus et al. identified 3 cases per 1,000 population and 2.26 cases per 100 hospital discharges when utilizing infection and organ failure ICD-9 codes[11].

Results of the IMPreSS trial of 1,794 ED and ICU patients showed that sepsis bundle compliance led to 36% lower odds of in-hospital mortality when controlling for ICU admission, sepsis code status, location of diagnosis, APACHE II score, and country (odds ratio 0.64, CI 0.47, 0.87)[4]. The findings from our study are comparable but include data from all hospital units. The SSC found that compliance with all resuscitation measures was associated with a 21% decrease in odds of in-hospital mortality when controlling for Sepsis Severity Score and participation in a sepsis resuscitation bundle awareness campaign[12]. The results from the evaluation of our bundle which did not include monitoring of central venous pressure or central venous oxygen saturation through a central venous catheter are consistent with those published in other trials that have questioned the benefit of goal-directed therapy[13-15].

Our study's adjusted NNT is larger than those reported in previous analyses. Cardoso et al. calculated an NNT of 6 among a cohort of ICU patients in Portugal[16] while Otero et al. combined data from several different sepsis control programs to calculate a summary NNT of 5 with individual programs reporting NNTs between 3 and 11[17]. The strength of our NNT is that it was calculated using observational data from all hospital units, not just ICUs or EDs, which provides evidence to support the use of sepsis care bundles in broader hospital settings.

SS/SS POA was associated with a decreased risk of in-hospital mortality in our cohort, which is a novel finding that warrants further study. In addition, we did not identify a significant interaction between

bundle compliance and SS/SS POA. There have been two other recent studies which have examined differences in patient outcomes when SS/SS is POA versus hospital-acquired (HA) and report findings consistent with our work. Page et al. found patients with HA sepsis had a greater risk of inpatient mortality than patients with community acquired sepsis.[18] Similarly, Jones et al. found HA sepsis was associated with a higher mortality than POA sepsis[19]. While prior studies utilized administrative POA indicators to determine sepsis status, time of SS/SS presentation and SS/SS POA status was calculated based on clinician review. When comparing manual review of dates as the gold standard to the administrative SS/SS POA flag in our dataset, we found that the sensitivity of the administrative flag was 96% and the specificity was 73%, a discrepancy that may have led to the differences in findings in previous studies.

It is unclear why SS/SS POA survival is significantly improved compared to HA SS/SS. The majority of patients with SS/SS POA present in the ED where teams are generally more experienced with sepsis recognition and had access to critical resources and treatment protocols. We were unable to ascertain blood culture results in our sepsis cohort but another potential explanation to explore is whether the pathogen mix differs in virulence between those who present to the hospital with sepsis and those who develop it in an inpatient unit which could have implications for antibiotic stewardship and development of antimicrobial protocols applied in different clinical settings.

Although every attempt was made to minimize bias and confounding in this observational study, our findings should be interpreted with some limitations in mind. This is a single center study which may not be generalizable in other settings and populations. Based on the workflow of the UCSF electronic record, mortality status was known at the time of compliance evaluation leaving the evaluation susceptible to bias. However, the bundle elements were timestamp-driven which makes it unlikely that data collection would be affected by outcome status. While chart reviews were conducted for all subjects coded with SS/SS, it is unknown how many SS/SS shock patients were missed because of documentation that was not sufficient to code them as such. Previous studies found severe sepsis patients identified via

coded data are more severely ill than severe sepsis patients whose records do not include the codes[20]. The Vizient database was used to query ICD-9 diagnosis and procedure codes, as well as admission severity of illness and risk of mortality that are based on the 3M APR-DRG groupers. If there were any inconsistencies in coding of ICD-9 diagnoses present on admission, severity of illness and risk of mortality scores may have been affected. Our study excluded 815 patients who were either transferred to UCSF with SS/SS POA and may or may not have received bundle elements at an outside institution or DNR/DNI because elements of the intervention bundle may not have been consistent with patients' goals of care. Therefore, our results may not be generalizable to the transfer and DNR/DNI population. While an adjusted sensitivity analysis of the relationship between bundle compliance and mortality could not be conducted due to limitations in the data, the unadjusted IRR for bundle compliance would be 0.38 (95% CI 0.31, 0.46) if all were counted as bundle non-compliant and if all were counted as bundle compliant, the unadjusted IRR for bundle compliance would be 1.03 (95% CI 0.83, 1.27). Finally, we also elected to analyze each discharge independently therefore if there was an effect of clustering at the patient level we did not include it in our models.

In conclusion, we found that a simple sepsis resuscitation bundle was associated with a decreased risk of in-hospital mortality within a cohort of SS/SS patients at UCSF. Given the changing epidemiology of SS/SS and the demonstrated effectiveness in randomized controlled trials as well as observational studies like ours, future work should focus on determining the attributable cost and length of stay saved by patients who experience complete compliance with sepsis intervention bundles. In addition, the contribution of each bundle element to the risk of mortality should be assessed with rigor in real-world settings to target resources towards interventions which will be most efficient at improving care in all corners of the world.

ACKNOWLEDGEMENTS

The authors would like to thank Donna Hays, Manager of Clinical Documentation Improvement, UCSF Medical Center, Herb Dye, Lead Clinical Documentation Specialist, UCSF Medical Center, and

Christopher Fee, MD, Professor of Emergency Medicine, UCSF Medical Center, for their contributions to the design of this study.

References

1. Stoller J, Halpin L, Weis M, et al. Epidemiology of severe sepsis: 2008-2012. *J Crit Care.* 2016;31(1):58-62.
2. Gaieski DF, Edwards JM, Kallan MJ, Carr BG. Benchmarking the incidence and mortality of severe sepsis in the United States. *Crit Care Med.* 2013;41(5):1167-1174.
3. Rivers E, Nguyen B, Havstad S, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med.* 2001;345(19):1368-1377.
4. Rhodes A, Phillips G, Beale R, et al. The Surviving Sepsis Campaign bundles and outcome: results from the International Multicentre Prevalence Study on Sepsis (the IMPreSS study). *Intensive Care Med.* 2015;41(9):1620-1628.
5. Lu Y, Zhang H, Teng F, Xia WJ, Sun GX, Wen AQ. Early Goal-Directed Therapy in Severe Sepsis and Septic Shock: A Meta-Analysis and Trial Sequential Analysis of Randomized Controlled Trials. *J Intensive Care Med.* 2016.
6. Levy MM, Fink MP, Marshall JC, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Crit Care Med.* 2003;31(4):1250-1256.
7. Zou G. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol.* 2004;159(7):702-706.
8. Vittinghoff E, Glidden DV, Shiboski S, McCulloch CE. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models.* Second ed: Springer Science+Business Media, LLC; 2012.
9. Gohil SK, Cao C, Phelan M, et al. Impact of Policies on the Rise in Sepsis Incidence, 2000-2010. *Clin Infect Dis.* 2016;62(6):695-703.
10. Lagu T, Rothberg MB, Shieh MS, Pekow PS, Steingrub JS, Lindenauer PK. Hospitalizations, costs, and outcomes of severe sepsis in the United States 2003 to 2007. *Crit Care Med.* 2012;40(3):754-761.

11. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med.* 2001;29(7):1303-1310.

12. Levy MM, Rhodes A, Phillips GS, et al. Surviving Sepsis Campaign: association between performance metrics and outcomes in a 7.5-year study. *Intensive Care Med.* 2014;40(11):1623-1633.

13. Mouncey PR, Osborn TM, Power GS, et al. Trial of early, goal-directed resuscitation for septic shock. *N Engl J Med.* 2015;372(14):1301-1311.

14. Peake SL, Delaney A, Bailey M, et al. Goal-directed resuscitation for patients with early septic shock. *N Engl J Med.* 2014;371(16):1496-1506.

15. Yealy DM, Kellum JA, Huang DT, et al. A randomized trial of protocol-based care for early septic shock. *N Engl J Med.* 2014;370(18):1683-1693.

16. Cardoso T, Carneiro AH, Ribeiro O, Teixeira-Pinto A, Costa-Pereira A. Reducing mortality in severe sepsis with the implementation of a core 6-hour bundle: results from the Portuguese community-acquired sepsis study (SACiUCI study). *Crit Care.* 2010;14(3):R83.

17. Otero RM, Nguyen HB, Huang DT, et al. Early goal-directed therapy in severe sepsis and septic shock revisited: concepts, controversies, and contemporary findings. *Chest.* 2006;130(5):1579-1595.

18. Page DB, Donnelly JP, Wang HE. Community-, Healthcare-, and Hospital-Acquired Severe Sepsis Hospitalizations in the University HealthSystem Consortium. *Crit Care Med.* 2015;43(9):1945-1951.

19. Jones SL, Ashton CM, Kiehne LB, et al. Outcomes and Resource Use of Sepsis-associated Stays by Presence on Admission, Severity, and Hospital Type. *Med Care.* 2016;54(3):303-310.

20. Whittaker SA, Mikkelsen ME, Gaieski DF, Koshy S, Kean C, Fuchs BD. Severe sepsis cohorts derived from claims-based strategies appear to be biased toward a more severely ill patient population. *Crit Care Med.* 2013;41(4):945-953.

Table 1.1. University of California San Francisco sepsis resuscitation bundle elements

Intervention Bundle Element	**Specifications**
Lactate Level	Blood lactate level drawn
Blood Culture	Blood culture drawn prior to initiating antibiotics
Antibiotic Administration	a) Initiation of broad spectrum antibiotics within 3 hours of sepsis presentation in the Emergency Department OR b) Initiation of broad spectrum antibiotics within 1 hours of sepsis presentation in an Inpatient Hospital Unit
Fluid Administration	If the patient was hypotensive or had a lactate level >4 mmol/L, starting an intravenous fluid bolus
Vasopressor Administration	If the patient remained hypotensive following fluid administration, starting an intravenous vasopressor

Table 1.2. Bundle, demographic and clinical characteristics of 1,029 patients diagnosed with severe sepsis or septic shock (SS/SS) stratified by mortality status

Characteristic	**Alive (n=848)**	**Died (n=181)**	**p-value**
Bundle Compliance Elements			
Perfect bundle compliance (binary)	638 (75%)	104 (58%)	<0.0001
Lactate compliant	793 (94%)	156 (86%)	0.001
Blood culture compliant	792 (93%)	154 (85%)	<0.0001
Antibiotics compliant	758 (89%)	134 (74%)	<0.0001
Fluids compliant, if applicable (n=705)	521 (92%)	118 (84%)	0.004
Vasopressors compliant, if applicable (n=445)	294 (86%)	94 (91%)	0.158
Admission Characteristics			
Age at admission (median, IQR)	63 (52, 73)	67 (56, 79)	0.006
Male gender	473 (56%)	97 (54%)	0.591
White race	403 (48%)	85 (47%)	0.891
Admission severity of illness			0.001
Minor or Moderate	100 (12%)	24 (13%)	
Major	341 (40%)	47 (26%)	
Extreme	407 (48%)	110 (61%)	
Admission risk of mortality			<0.0001
Minor	59 (7%)	11 (6%)	
Moderate	110 (13%)	32 (18%)	
Major	300 (35%)	30 (17%)	
Extreme	379 (45%)	108 (60%)	
Medicaid/Medicare (primary or secondary)	705 (83%)	160 (88%)	0.079
SS/SS present on admission	664 (78%)	102 (56%)	<0.0001
Immunocompromised status			0.009
None	752 (89%)	157 (87%)	
Malignancy	37 (4%)	17 (9%)	
Other immunocompromised condition	59 (7%)	7 (4%)	
End stage renal disease present on admission	37 (4%)	11 (6%)	0.321
Congestive heart failure present on admission	117 (14%)	39 (22%)	0.008
SS/SS diagnosis in emergency department	613 (72%)	90 (50%)	<0.0001

Table 1.3. Poisson models for independent risk factors for death in a cohort of severe sepsis or septic shock (SS/SS) patients who had bundle compliance assessed

Variable	**Risk of Death IRR (95% CI) with binary bundle compliance exposure***
Compliance with all eligible bundle elements	0.69 (0.53, 0.91)
SS/SS present on admission	0.55 (0.32, 0.92)
Age scaled by 10 years	1.13 (1.03, 1.24)
Admission severity of illness	
Minor or Moderate	REF
Major	1.00 (0.59, 1.71)
Extreme	1.62 (0.84, 3.14)
Admission risk of mortality	
Minor	REF
Moderate	1.44 (0.78, 2.68)
Major	0.72 (0.33, 1.55)
Extreme	1.37 (0.62, 3.01)
Medicaid as primary/secondary payor	1.27 (0.82, 1.96)
Immunocompromised status	
No immunocompromise	REF
Malignancy	1.41 (0.88, 2.24)
Other immunocompromised condition	0.71 (0.40, 1.26)
Congestive heart failure present on admission	1.31 (0.96, 1.78)
SS/SS presentation in the emergency department	0.78 (0.46, 1.31)

*IRRs and 95% CIs estimated with Poisson regression and assumed robust variance.

Table 1.4. Mean values of covariates included in the bundle compliance propensity score model stratified by compliance status and propensity score quintile

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
SS/SS POA					
Non-compliant	0.107	0.55	1	1	1
Compliant	0.141	0.646	1	0.994	1
Age in years					
Non-compliant	64.1	60.7	63.3	63.3	52.3
Compliant	60.8	61.6	66.2	67.7	62.7
Admission Severity of Illness					
Non-compliant	1.08	1.35	1.72	1.14	1.3
Compliant	1.36	1.43	1.55	1.13	1.41
Admission Risk of Mortality					
Non-compliant	1.62	2.1	2.59	2.59	2.18
Compliant	1.93	2.21	2.24	2.53	2.25
Immunocompromised					
Non-compliant	0.273	0.3	0.438	0	0.0455
Compliant	0.271	0.52	0.15	0	0.0165
Congestive Heart Failure POA					
Non-compliant	0.19	0.262	0.156	0.0313	0
Compliant	0.212	0.331	0.26	0	0.00549
In Emergency Department at Time of Presentation					
Non-compliant	0	0.3	1	1	1
Compliant	0	0.496	1	1	1
Medicaid/Medicare Payor					
Non-compliant	0.76	0.9	0.75	0.813	0.864
Compliant	0.729	0.874	0.746	0.891	0.956

Figure 1.1. Study population for the evaluation of a sepsis resuscitation bundle at the University of California, San Francisco (UCSF)

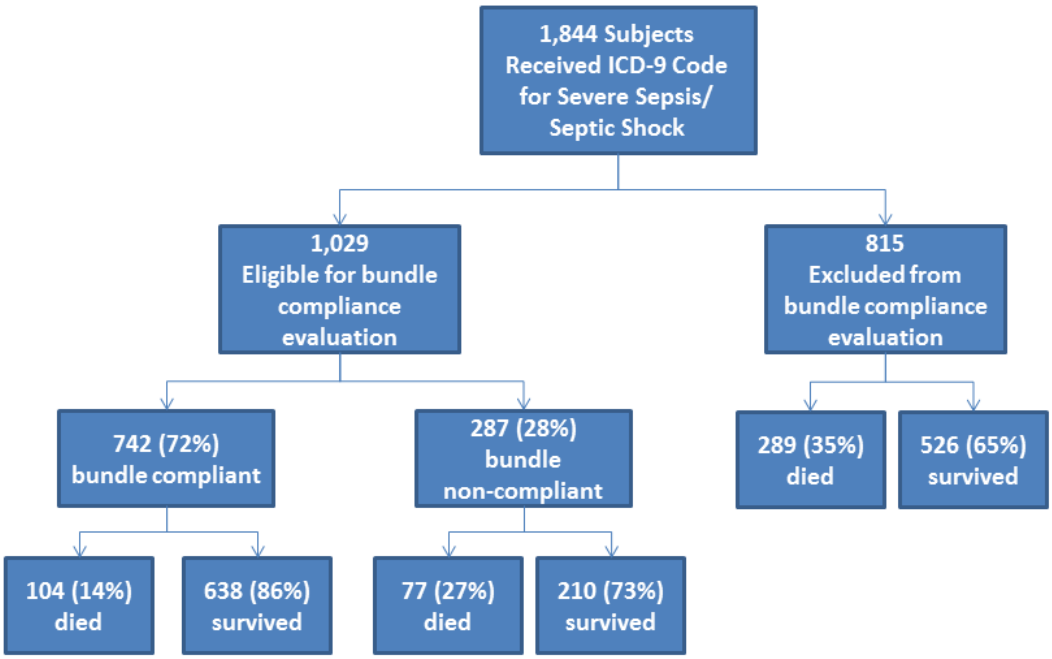
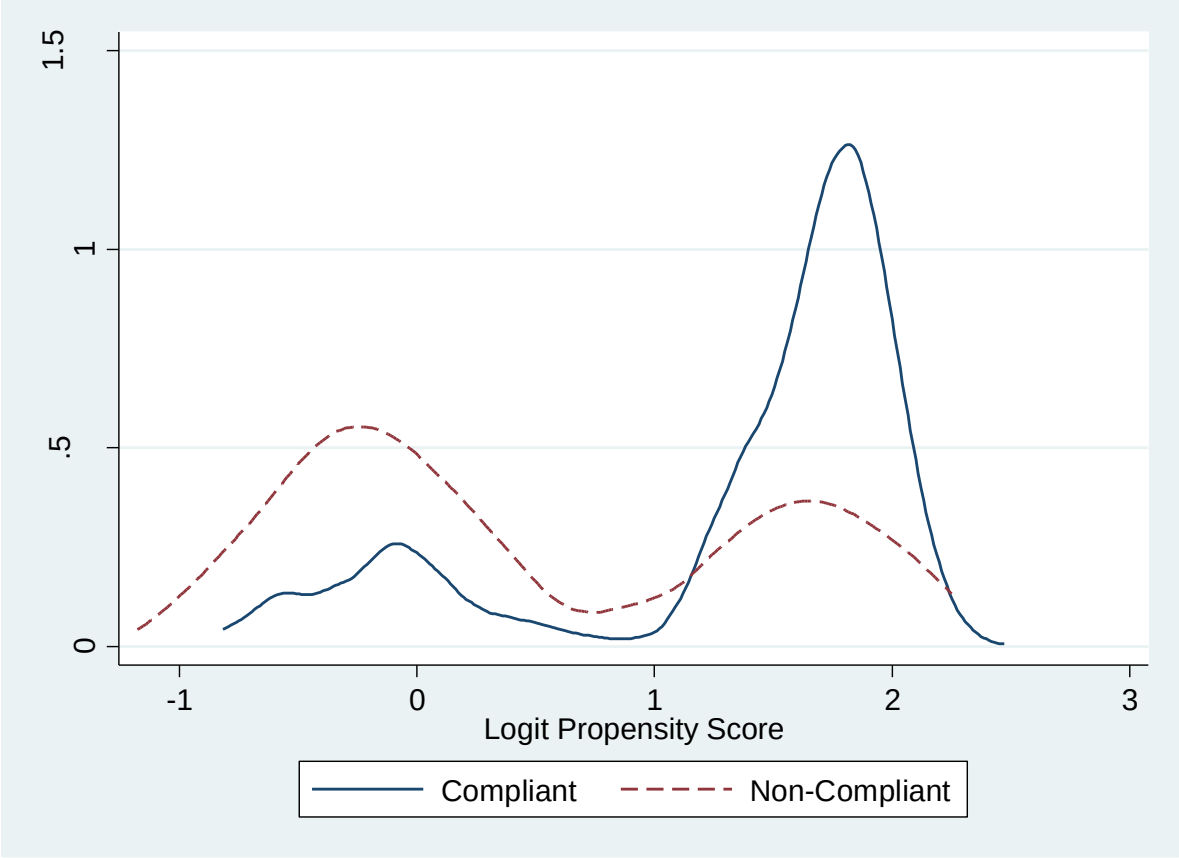


Figure 1.2. Propensity Scores in Bundle Compliant and Bundle Non-Compliant Subjects



Chapter 2: Applying causal inference methods to quality improvement initiatives implemented at the group level: Narrative review of a research methodology

Priya A. Prasad, Steven Shiboski, and Ralph Gonzales

INTRODUCTION

Increased attention has been paid in the recent literature to conducting observational studies with an eye towards causal inference and strengthening internal validity. Causal inference methods emphasize the requirement for clear statement of assumptions needed to draw causal inferences from existing data. The causal inference framework has allowed a more precise language to evolve to describe factors that may confound and mediate the relationship between exposure and outcome, particularly when exposures, outcomes and covariates are being measured over time.[1] By encouraging investigators to explore these issues prior to performing a study, appropriate analyses can be specified in the design stage of an evaluation.

The goal of this work was to provide a review of causal inference methodology and its application to the evaluation of quality improvement (QI) interventions using interrupted time-series analyses (ITSA) for evaluations where data are only available at the aggregate level and no patient or individual level data are available, no individual aggregate outcome and exposure data are collected. In addition to a review of literature on causal inference and ITSA, we provide a real world example illustrating the concepts discussed, evaluating the effects of an intervention applied at a group level aimed at decreasing packed red blood cell (PRBC) transfusions.

SUMMARY OF RELEVANT LITERATURE

Historically, it has been challenging to publish evaluations of QI initiatives in the peer-reviewed literature because the methods used often vary from those traditionally applied when studying the etiology of disease and the findings of QI studies are perceived to be unique to the institution under evaluation.[2] Because it is imperative that QI evaluations are designed to provide robust evidence on the utility of interventions, QI leaders and stakeholders have developed the SQUIRE guidelines. These guidelines provide frontline clinicians and local champions with tools to effectively design and present work that is

both internally valid and allows the investigator to assess the causal relationship between the intervention and the outcome of interest.[3]

Causal Inference and Quality Improvement

In QI evaluation, the ultimate goal is to show that there is a causal relationship between the interventions implemented and the outcomes that are observed in the target population. The causal effect of an intervention for a particular individual or unit of analysis is defined as the contrast between the outcomes observed in the presence and absence of intervention. In the causal inference framework, both outcomes are referred to as "potential" or "counterfactual" but only one of them is actually observed. Because the causal effect is typically not observable at the individual level, an intervention effect is summarized as the average causal effect, defined as the difference between population average potential outcomes with and without the intervention.[1]

Assumptions required for causal inference with observational data

Under specific assumptions, estimates from observational studies match causal effects. To derive valid causal effect estimates from observational data, we must assume that the treatment is consistently defined and that there is no interference or contamination in outcomes between units of observation. Another key assumption is that potential outcomes are independent of intervention assignment. Another name for this condition is *exchangeability* : the potential outcomes for treated units are exchangeable with the potential outcomes of untreated units, even though the actual outcomes of treated and untreated units may differ. Exchangeability is generally satisfied in randomized controlled trials, but is more of a concern for observational studies, including most QI evaluations. In these situations, a weaker version of exchangeability must apply for valid causal inference: intervention assignment is conditionally independent of the potential outcomes conditional on the values of observed confounding covariates. The weaker version of exchangeability implies that valid inference can be drawn when all confounders are correctly and fully accounted for. As a result, investigators need to ensure that all such variables are identified during the design stage, measured during the data collection phase, and controlled for in

analyses. A final practical requirement for valid estimation of causal effects is that the distribution of variables controlled for in the analysis should be overlapping in the intervention and non-intervention groups. If this assumption is not satisfied, estimated effects may be due in part to extrapolation where few data exist, affecting the credibility of results.[1, 4]

When it seems impossible to fully account for all potential confounders, an alternative approach is to contrast the unit of intervention against itself, and let each unit serve as its own control, for example, in a pre-post comparison. In this setting, the exchangeability assumption relates to the unit's outcomes change after intervention only due to the introduction of the intervention. When multiple observations are available leading up to the introduction of the intervention and afterwards, this is referred to as an interrupted time-series analysis (ITSA).

Interrupted Time-Series Analysis

QI interventions are often implemented at the group level at a particular time rather than randomized within individuals, limiting information about intervention effects to before/after comparisons of outcomes. In addition, outcome, exposure, and covariate data for QI evaluations are often only available in an aggregate group-level form. ITSA is a common method used to analyze the resulting data, quantifying the effect of the "interruption" by comparing average outcomes observed before and after the intervention.[5] In its simplest form, a time-series involves repeated measurements of an outcome within a unit of analysis (i.e., individuals, clinical units, hospitals, facilities) over time.[6, 7] Figure 2.1 provides a plot of the time-series from our real-world example. Key elements of an ITSA include assessment of the pattern the effect takes (level and slope), how well the pattern holds moving forward in time, and whether there is an immediate effect following the interruption.[8] There are strengths of ITSA for QI evaluation and research. First, by modeling observed outcomes prior to and following the intervention, one can account for underlying secular and cyclical trends[6, 9, 10] and ITSA methods can be applied in situations where only aggregate data exist.[7, 9] ITSA provides an intuitive display of data for policy makers and administrators as well.[7] A phenomenon to consider with ITSA is *autocorrelation,* where neighboring values of the outcome may be more highly related than those farther apart in time.

Ignoring autocorrelation when present can lead to spurious inferences about intervention effects from resulting analyses.[6, 9, 11]

Causal Inference and ITSA: Assumptions and Causal Quantities

Given that detailed descriptions of requirements for valid inferences from ITS designs are provided in a number of references,[6, 8, 12, 13] we provide a brief summary here. Adapting the counterfactual framework for causal inference introduced above to apply to an ITSA design requires that the key assumption of exchangeability holds, even though outcomes observed in the absence and presence of the intervention are separated in time. In particular, ITSA approaches assume that potential outcomes can be estimated using the trajectory of the observed data prior to the implementation of the intervention.[12] Another major assumption which allows for formal inference to be drawn from ITSA results is the *stationarity* of the distribution of the outcome. An outcome is stationary if its distribution is the same regardless of the point in time at which we elect to observe it, a concrete example being white noise. By ensuring stationarity of the data, there can be some element of replication in a time series which allows for formal causal inference.[6] As mentioned above, assumptions of autocorrelation should also be considered when appropriate.

One of the advantages of ITSA is that it is possible to account for the effects of the passage of time in the estimation of causal quantities. By adding nuance to the interpretation of the average causal effect, the average treatment history effect that can be calculated in an ITSA takes into account the effect of the treatment at a given time but also the effect of the treatment at lagged values of time. The contemporaneous effect of treatment can also be derived and can be considered the "blip effect." The measure of contemporaneous effect of treatment would allow the investigator to determine the immediate effect of switching all individuals from no treatment to treatment at a specified time.[13]

Time-dependent confounding in time series studies

An area of recent focus in epidemiologic research has been on how to handle a special case of confounding where the confounder is a risk factor for the outcome and a predictor of subsequent

exposure, a phenomenon called time-dependent confounding (TDC).[4] Studies in pharmacoepidemiology often involve TDC because the effects of drug treatment may be dependent on time and factors associated with the outcome may be affected by previous drug treatment as well. In these situations, marginal structural models can reliably estimate the causal effects of time-varying exposures when a system includes TDC that are affected by prior treatment.[14] Although it is beyond the scope of the present paper, methods have been developed to incorporate TDC into ITSA.[13]

Threats to Validity in ITSA

Similar to other analytic approaches for data from observational studies, ITSA does have some internal and external threats to validity. Because satisfying the assumptions for causality in this setting can be difficult, critics suggest that ITSA and other quasi-experimental study designs are subject to biases and confounding that are not present in randomized controlled trial designs and the causal claims made from ITSA should be interpreted with caution.[15] Frameworks that apply to ITSA stipulate that an intervention is merely a potential cause of the effect, therefore its use as a justification for causal inference has been called into question[16] as there could be other causes and unmeasured confounding that are unaccounted for in the analysis, particularly when only a single group is followed before and after an intervention.[17] Other potential threats to validity in single group ITSAs include instrumentation bias, where the method for ascertaining the outcome may change during the study period, and selection bias if the characteristics of the single group change over the study period. Several of these threats could be addressed by including a control group in the analysis.[8, 17]

Real-World Example: Effect of an Intervention to Decrease Packed Red Blood Cell Transfusions

For the remainder of this work we will use a real-world example to illustrate the analytic approaches presented. Following is a description of this example.

Description of the intervention:

The Caring Wisely "Transfuse One" intervention was implemented at the hospital service level in October 2013 in three hospital units: Orthopedics, Neurosurgery, and Hospital Medicine (excluding

Hematology-Oncology) services at the University of California at San Francisco, an academic medical center. The intervention included a physician, nurse and trainee education and awareness campaign that included updated practice guidelines and recommendations from the American Association of Blood Banks. A tree map showing the transfusion volume and mean pre-transfusion hemoglobin levels for hospital services and individual physicians was developed to demonstrate to target audiences how services compared with one another.

Data Collection and Analysis

The RBC transfusions team proposed to reduce total RBC transfusions by at least 5% for adult non-intensive care unit patients over the course of one year, with a focus on transfusing one unit at a time based on pre-transfusion hemoglobin levels. Aggregate outcome data were collected over a 2-year period, from July 1, 2012 to June 30, 2014. Non-intervention hospital units that served adult patients were included as the comparator group for the intervention. We considered two outcomes of interest: the count of total PRBC units transfused to determine the overall effect of the intervention at the level of the hospital and the average PRBCs transfused per discharge to determine the effect the intervention had over the hospital census to control for changing patient volume. The primary exposure of interest was the "Transfuse One" intervention. Additional data elements included median patient age, proportion of patients who were male and proportion of patients with a primary payor of Medicaid/Medicare, measured averaged across the intervention units. All analyses were conducted in Stata 13 (College Station, TX). Details of the statistical analysis appear throughout the remainder of the manuscript and the study will be referred to as the "CW-PRBC Evaluation."

Developing an Analysis Strategy Using Aggregate QI Data

In order to draw causal inference from observational data in any type of analysis the final analytic model must be correctly specified, meaning that outcomes, exposures, and covariates are appropriately defined. The following sections provide a summary of proposed steps that should be taken when developing an analysis plan for a QI evaluation that involves allocating an intervention at a group level. A schematic for the development of an analysis plan can be found in Figure 2.2. For the following

discussion, one must assume that an intervention is applied at the group level, the outcome is measured at the group level and the outcome is measured over time. It should be noted that, while multiple unique situations may fall outside of the scope and inclusion of the diagram, Figure 2.2 is meant to provide a general guide for the analysis plan.

Defining the outcome of interest and covariates

The primary and secondary outcomes in a QI evaluation study will drive the analysis plan and ultimately the utility of the evaluation. First the broader outcome of interest should be considered, such as utilization or satisfaction and should take into account the target audience. While often the availability of data is the factor which dictates the outcome selected, when there are options available an investigator should consider measures that will be directly affected by the intervention and measures for which there is the ability and the statistical power to detect a change.[18] The investigator should then sketch a causal model for the effect of the intervention on the broad outcome that includes all covariates thought to have an effect on the system under study in order to address the assumptions of no unmeasured confounding required for causal inference.[14, 18]

Application to the CW-PRBC Evaluation: In our study, the outcome of interest was both total PRBC units transfused (count) and mean PRBC units transfused per discharge (continuous).

Specifying a model for analysis based on the type of outcome

Once the investigator identifies the primary outcome, a statistical model based on the distribution of the outcome data can be selected to analyze the time series.

Continuous outcome data: Linear regression is used with continuous data and allows the investigator to observe how the average value of the outcome varies with the manipulation of other variables. The assumptions of linear regression include that average outcomes are linearly related to predictors, and that the errors are independent and follow a standard normal distribution with constant variance.[1] Alternate versions of the standard linear model that account for longitudinal outcome measures[19] such as the autoregressive integrative moving averages (ARIMA) model (discussed in detail below) and random effects regression are also potential candidates to consider in this setting.

Count outcome data: *Poisson and negative binomial regression models* apply to outcomes that can be expressed as counts, such as car accidents, number of hospitalizations, or number of PRBC transfusions, as in our real-world example. Both models typically assume that the logarithm of the average outcome is a linear function of the predictor variable(s) of interest. While the Poisson model specifies equality for the outcome mean and variance, this is relaxed in the negative binomial model, making the latter preferred in situations where *overdispersion* is suspected.[20] Integer-valued autoregressive (INAR) Poisson regression is another modeling strategy that is useful when the mean of the counts is low and it is necessary to control for serial correlation of data, which may occur when the time scale used is cut into shorter intervals, such as days or weeks.[21]

Binary/categorical outcome data: *Logistic regression* is the most common regression model used for binary outcomes, and in the case of rare outcomes (e.g. outcome odds <0.10, the estimated odds ratios may provide a close approximation to the relative risk.[22] Logistic regression output is easily interpretable and is present in standard statistical packages.[1] Logistic regression techniques have been extended to cover analyses of ordinal outcome data, such as answers to scaled survey questions,[23] as well as nominal, non-ordered categorical outcome data, such as sites of infection.[24] *Poisson regression* can also be applied to binary outcomes and yields coefficients that have a relative risk interpretation, provided that robust standard errors are used for inference.[25]

Application to the CW-PRBC Evaluation: In our example, we elected to model the total PRBC units transfused using Poisson regression for count outcomes and the mean PRBC units transfused per discharge with linear regression for continuous outcomes.

Selecting an inference strategy

After selecting an appropriate model, variability of resulting estimates needs to be assessed to allow valid inferences. A number of alternate approaches for this have been proposed[1] , and below we discuss those most applicable to ITSA:

Autoregressive integrative moving averages model for autocorrelation (ARIMA):

When autocorrelation is confirmed and there are enough time points at which the continuous outcome is measured, a formal approach to time-series modeling may be employed with the ARIMA model. The ARMAX model, an extension of the ARIMA model, allows the investigator to include covariates. Developed by Box and Jenkins,[26] the ARIMA model allows the investigator to control for trend as well as seasonality. The underlying assumption of an ARIMA model is that errors are normally distributed.[21] There are three elements that must be specified in an ARIMA model: the order of autocorrelation observed in the data, the level of differencing required to achieve stationarity, and the order required for the moving averages portion of the model.[27] The general rule of thumb for ARIMA modeling is that there should be 20 observations pre-intervention and 20 observations post-intervention but the investigator should ensure that complete cycles that are appropriate to the system being evaluated are represented within the range of observations recorded.[28] Other references note that 100 repeated measurements are necessary to draw the appropriate inference from ITSA.[8] Although ARIMA modeling has traditionally been used for count outcomes in ITSA, the assumption of normal distribution of errors is often violated.[21]

Newey-West: Newey-West standard error estimates provide a robust alternative to ARIMA methods to account for autocorrelation, and apply to of the family of generalized linear models, (e.g. linear regression, Poisson regression, etc.).[29]

Robust: Robust standard errors are used when an investigator is unsure if the modeling strategy selected accounts correctly for the correlation structure of the data under study. After estimates are computed using the default correlation structure of the model, the within-subject residuals of the observed data are used to calculate robust standard errors. These robust standard errors will be valid if all other aspects of the model are properly specified and the sample is adequately large.[1]

Clustered: Clustered standard errors should be used when there is more than one event per subject or unit of observation.[1] Clustered standard errors should be used in situations where the clusters are small in size and when there are several clusters of data, otherwise the cluster variable should be analyzed as a fixed effect or standard errors should be bootstrapped.[30]

Boostrap: Bootstrap standard errors provide an alternative to other approaches when methods have not been developed to calculate standard errors for a particular model, or if the underlying assumptions of a selected model are clearly violated.[1]

Application to the CW-PRBC Evaluation: In our study, we elected to take a conservative approach and used robust standard errors for both the Poisson model of log mean PRBC units transfused and the linear model of mean PRBC units transfused per discharge.

Selecting a strategy to model the effect of the intervention

Given the nature of implementing unit-wide QI interventions, instead of viewing the intervention itself as a treatment in the classical sense it is often more appropriate to view time as the primary exposure of interest in the analysis. While time can be incorporated into the model in several different ways, we propose three options: a pre-post intervention indicator, an interaction between the pre-post indicator and continuous time, and restricted cubic splines.

Pre-post intervention indicator: In the event that the investigator expects an immediate and sustained effect of the intervention, an indicator could be created specifying a period prior to the intervention ("pre") and a period following the intervention ("post"). Using a pre-post indicator the effect of the intervention would be averaged during each period and if graphed over time would look like a step down at the point of intervention, in the absence of any other covariates.

Interaction between pre-post indicator and time: By including the pre-post intervention indicator in the model as well as incorporating it into an interaction term with continuous time, it is possible to track whether time modifies the effect of the intervention. This strategy allows the investigator to incorporate both a jump in the rate of the outcome (discontinuity) and a change over time.

Restricted cubic splines: Splines can be used if the investigator would like to allow for flexibility in the effect of the intervention over time. Splines permit heterogeneity of the effect of time on the outcome.[13] Cutpoints (knots) can be placed at pre-specified time-points within the distribution of the predictor and the effect of the predictor is modeled as cubic polynomials between knots. Beyond the final knot the

effect of the predictor is constrained to be linear. The relationship between the outcome and the predictor is smoothed at the knots.[31]

Application to the CW-PRBC Evaluation: Our analysis modeled the effect of the intervention with all three strategies in the Poisson model of log mean PRBC units transfused using robust standard errors and the linear model of mean PRBC units transfused per discharge using robust standard errors.

Drawing causal inference from the specified model

Once all assumptions outlined above have been met and the model of outcome, exposure and covariates has been properly specified, it is possible to draw causal inference from the resulting effect estimates.

RESULTS OF THE CW-PRBC EVALUATION

There were 14,712 discharges during the baseline period of July 2012 through September 2013 and 9,030 discharges during the intervention period of October 2013 through June 2014. Demographic characteristics appear in Table 2.1. Discharges with Medicaid/Medicare as the payor were more common during the intervention period ($p<0.0001$).

Among the intervention hospital units, the log mean count of PRBC units transfused was significantly lower during the intervention period when compared to the baseline period (IRR 0.87, 95% CI 0.77, 0.97) when controlling for proportion male and proportion Medicaid/Medicare, if the effect of the intervention was modeled using the pre-post intervention indicator (Table 2.2, Figure 2.3). However, when the effect of the intervention was modeled using the pre-post intervention indicator and time interaction as well as time modeled as a restricted cubic spline, there was no difference in the log mean count of PRBC units transfused when comparing the intervention period to the baseline period.

Among the intervention hospital units, 0.09 fewer PRBC units were transfused per discharge during the intervention period when compared to the baseline period (95% CI -0.15, -0.03) when controlling for proportion male and proportion Medicaid/Medicare, if the effect of the intervention was modeled using the pre-post intervention indicator (Table 2.3, Figure 2.3). When the effect of the intervention was modeled using the pre-post intervention indicator and time interaction, 0.003 fewer PRBC units were transfused per discharge during the intervention period when compared to the baseline

period (95% CI -0.004, -0.001) but the interaction term was not significant. When time was modeled as a restricted cubic spline, there was no difference in the mean PRBC units transfused when comparing the intervention period to the baseline period.

Although the log mean count of PRBC units transfused during the intervention period was no different than during the baseline period in the non-intervention units using all three modeling strategies for the effect of the intervention, patients on Medicaid/Medicare received significantly more blood than those who were not, regardless of modeling strategy (Table 2.2). For each 10% increase in the proportion of patients discharged as Medicaid/Medicare as the primary payor, the log mean count of PRBC units transfused increased 7.68 times (95% CI 1.83, 31.36) when controlling for the effect of the intervention and the proportion male, based on the model including the pre-post intervention indicator and time interaction. In addition, for each 10% increase in the proportion of patients discharged as Medicaid/Medicare as the primary payor, the mean PRBC units transfused per discharge increased by 1.63 units (95% CI 0.55, 2.70) when controlling for the effect of the intervention and the proportion male, based on the model including the pre-post intervention indicator and time interaction (Table 2.3).

DISCUSSION

In this manuscript we provided a summary of the relevant literature on the application of the causal inference framework to the evaluation of QI interventions, presented a schematic for the development of an analysis plan for QI evaluations that involve measurement of aggregate outcomes over time, and used a real-world example to illustrate the development of an ITSA analysis plan.

There are some important reasons to utilize aggregate data models. Aggregate data are more easily gathered and can be cost-effective, particularly in resource-limited settings. Aggregate data analyses are also most appropriate if the effect of an intervention is being assessed at the level of the group and individual level comparisons will not be made. For example, if a quality improvement intervention is deployed at the level of the unit then it may be most appropriate to evaluate the effect of that intervention at the unit level. However, if the goal is to apply an intervention to an individual and draw inference about the effect of the intervention in that individual, as is the case in most epidemiologic

studies, aggregate data models may be subject to the ecological fallacy where the conclusions drawn from the group-level data may not be generalizable at the individual level. There have been modeling strategies developed to address the issue of ecological fallacy in aggregate data, including the two-phase design and the hybrid design.[32]

When it comes to selecting an approach to model the effect of the intervention, the investigator must think critically about the inference that is expected to be drawn from the result. As we have described, often a binary pre-post indicator variable is used to model the effect of the intervention. While this is an intuitive strategy, there could be statistical limitations to this approach. By forcing the effect of the intervention to be averaged across the pre- and post- time periods, the granularity in the rate of the outcome over the course of the baseline and intervention periods may be lost.[28] In addition, secular trends are ignored and any effect that is observed is assumed to be due to the intervention, which could lead to over or under-estimations of the intervention effect. Adding an exchangeable control group which exhibits similarities to the intervention population and conducting between-group analyses of the outcome could make the pre-post evaluation more robust but it can be challenging to identify the appropriate control group.[33] While our real-world example did include a comparator group in the analysis, it was comprised of all units that did not participate in the intervention so the inferences that can be drawn from our results are limited. In situations where the investigator believes that time modifies the effect of the intervention, using interactions between the intervention and time or modeling time as a spline may be more appropriate and allow for greater flexibility and nuance to be reflected in the intervention effect estimates.

The results of our real-world analysis showed that the conclusions drawn from the data were significantly different depending on the analytic strategy that was employed to model the effect of the intervention. When the pre-post intervention indicator was used to model the effect, there was a significant decrease in the total PRBC units transfused and the mean PRBC units transfused per discharge between the baseline and intervention periods. However, when an interaction with time was included or time was accounted for using restricted cubic splines, there was no significant intervention effect; the

observed rates of PRBC transfusion in our study were no different than what would have been expected from the baseline trend.

ARIMA models and other formal time-series methods should be considered when there is confirmed autocorrelation between data points. In addition, these methods should only be considered when there are a sufficient number of outcome measurements to identify the model which is affected by the amount of error present in the data, the effect of periodicity, and the number of lags that should be accounted for in the analysis.[8]

A major issue that comes up in quasi-experimental designs such ITSA is clustering at varying levels. As mentioned earlier in this manuscript, while it is possible to account for clustering in the standard error calculation or to include the clustered variable in the model as a fixed effect if the cluster sizes are large, another strategy to employ is a mixed effects model. A mixed effects model includes both random effects and fixed effects and is useful in situations where longitudinal repeated measures data are available for analysis. As an example, in our study, while our outcome was measured on the aggregate group level of the intervention units, there were multiple discharge attendings that provided care for multiple patients over the study period. Standard models are only able to accommodate clustering at one level, when more than one level of clustering is identified, mixed effect models are required.

Beyond ITSA, there are other study designs that can be considered when an intervention is being applied at the level of the group. Cluster-randomized trials involve randomization of exposure at the community level. Benefits of community-level randomization are maximized in situations where the number of intervention groups randomized is large and analyses should account for clustering.[34] While the internal validity afforded by conducting RCTs is attractive, it may be impractical to implement some of these strategies in a healthcare system where providers regularly communicate, leading to potential contamination between experimental and control groups. Issues of contamination could exist whether we randomize at the provider or practice level. Additionally, accounting for clustering could make the analysis more cumbersome to carry out and the required sample size for the study may be prohibitive, causing additional analytic complications.[35] The use of a stepped-wedge design can improve the internal

validity of a quasi-experimental design. In this design, each observational unit receives the intervention after a baseline period of observation but the implementation of the intervention is randomly staggered over time. This methodology is particularly appropriate when an intervention is efficacious in one population and is being implemented within other populations to assess effectiveness of the program at the population level. Additionally, the analysis of data from a stepped-wedge design is quite flexible because each observational unit contributes control and intervention observations so between and within-group comparisons can be made and more complex inferences can be drawn from the data[36]. By ensuring that there are multiple observation units, multiple observations before and after intervention, and by staggering the implementation of the intervention over time, investigators can maximize inferences drawn from the data.[35, 36]

We have outlined an algorithm to developing a correctly specified model when data are available in aggregate and inferences are to be drawn at the group level and have guided the reader through the application of the algorithm to a QI evaluation. As described above with real data from the CW-PRBC Evaluation, the approach taken to model the effect of the intervention had a significant impact on the conclusions that could be drawn from the analysis and investigators should be vigilant about appropriately incorporating the effects of time and baseline trends in evaluations. In conclusion, using our algorithm in conjunction with the guidance provided within the text, we believe investigators can develop QI evaluation plans which are robust and hypothesis-driven, particularly when modeling the effect of the intervention.

ACKNOWLEDGEMENTS

The authors would like to thank Christopher Moriates, MD, Catherine Lau, MD, Delene Johnson, CLS, Kathy Curcione, RN, Jeffrey Lam, RN, and Chuck McCulloch, PhD, for their contributions to the study.

References

1. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. *Regression Methods in Biostatistics Linear, Logistic, Survival, and Repeated Measures Models Preface*, 2005.
2. Davidoff F, Batalden P. Toward stronger evidence on quality improvement. Draft publication guidelines: the beginning of a consensus project. *Qual Saf Health Care* 2005, **14**(5)**:** 319-325.
3. Ogrinc G, Mooney SE, Estrada C, Foster T, Goldmann D, Hall LW, *et al.* The SQUIRE (Standards for QUality Improvement Reporting Excellence) guidelines for quality improvement reporting: explanation and elaboration. *Qual Saf Health Care* 2008, **17 Suppl 1:** i13-32.
4. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000, **11**(5)**:** 550-560.
5. Morgan GA, Gliner JA, Harmon RJ. Quasi-experimental designs. *Journal of the American Academy of Child and Adolescent Psychiatry* 2000, **39**(6)**:** 794-796.
6. Diggle PJ. *Time Series A Biostatistical Introduction*. Oxford Science Publications: Oxford, England, 1990.
7. Penfold RB, Zhang F. Use of interrupted time series analysis in evaluating health care quality improvements. *Academic pediatrics* 2013, **13**(6 Suppl)**:** S38-44.
8. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin Company: Boston, MA, 2002.
9. Cook TD, Cambell DT. *Quasi-Experimentation Design & Analysis Issues for Field Settings*. Houghton Mifflin Company: Boston, MA, 1979.
10. Ramsay CR, Matowe L, Grilli R, Grimshaw JM, Thomas RE. Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. *International journal of technology assessment in health care* 2003, **19**(4)**:** 613-623.
11. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *Journal of clinical pharmacy and therapeutics* 2002, **27**(4)**:** 299-309.

12. Morgan SL, Winship C. *Counterfactuals and Causal Inference Methods and Principles for Social Research Introduction*, 2015.

13. Blackwell M, Glynn AN. How to Make Causal Inferences with Time-Series Cross-Sectional Data.

14. Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part III. *Value Health* 2009, **12**(8)**:** 1062-1073.

15. Li LC, Moja L, Romero A, Sayre EC, Grimshaw JM. Nonrandomized quality improvement intervention trials might overstate the strength of causal inference of their findings. *Journal of Clinical Epidemiology* 2009, **62**(9)**:** 959-966.

16. Eichler M. Granger causality and path diagrams for multivariate time series. *Journal of Econometrics* 2007, **137**(2)**:** 334-353.

17. Linden A. Challenges to validity in single-group interrupted time series analysis. *J Eval Clin Pract* 2016.

18. Coster WJ. Making the Best Match: Selecting Outcome Measures for Clinical Trials and Outcome Studies. *Am J Occup Ther* 2013, **67**(2)**:** 162-170.

19. Davidian M, Giltinan DM. *Nonlinear Models for Repeated Measurement Data*, vol. 62. CRC Press: Boca Raton, FL, 1995.

20. Verbeke G. Regression Analysis of Count Data. A. C. Cameron and P. K. Trivedi (eds.), Cambridge University Press, Cambridge, 1998. No. of pages: xv+404. Price:£17.95. ISBN 0-512-63567-5. *Statistics in Medicine* 2001, **20**(7)**:** 1143-1144.

21. Quddus MA. Time series count data models: an empirical application to traffic accidents. *Accid Anal Prev* 2008, **40**(5)**:** 1732-1741.

22. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol* 1987, **125**(5)**:** 761-768.

23. McCullagh P. Regression models for ordinal data. *Journal of the royal statistical society Series B (Methodological)* 1980**:** 109-142.

24. Engel J. Polytomous Logistic Regression. *Statistica Neerlandica* 1988, **42**(4)**:** 233-252.

25. Zou G. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol* 2004, **159**(7)**:** 702-706.

26. Box GEP, Jenkins GM, Reinsel GC, Ljung GM. *Time Series Analysis: Forecasting and Control*, Fifth edn. John Wiley & Sons, Inc.: Hoboken, New Jersey, USA, 2015.

27. Diggle PJ. *Time Series: A Biostatistical Introduction*. Oxford University Press: New York, USA, 1990.

28. Crabtree BF, Ray SC, Schmidt PM, O'Connor PJ, Schmidt DD. The individual over time: time series applications in health care research. *J Clin Epidemiol* 1990, **43**(3)**:** 241-260.

29. Newey WK, West KD. A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 1987, **55**(3)**:** 8.

30. Cameron AC, Gelbach JB, Miller DL. Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics* 2008, **90**(3)**:** 414-427.

31. Steenland K, Deddens JA. A practical guide to dose-response analyses and risk assessment in occupational epidemiology. *Epidemiology* 2004, **15**(1)**:** 63-70.

32. Haneuse S, Bartell S. Designs for the combination of group- and individual-level data. *Epidemiology* 2011, **22**(3)**:** 382-389.

33. Grimshaw J, Campbell M, Eccles M, Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. *Fam Pract* 2000, **17 Suppl 1:** S11-16.

34. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Lippencott Williams & Wilkins: Philadelphia, PA, 2008.

35. McCulloch CE. Analytical and Design Issues in Translational Research. 2013. p. Epi 245 Lecture.

36. Handley MA, Schillinger D, Shiboski S. Quasi-experimental designs in practice-based research settings: design and implementation considerations. *Journal of the American Board of Family Medicine : JABFM* 2011, **24**(5)**:** 589-596.

Table 2.1. Demographic characteristics of discharges during baseline and intervention periods

	Intervention Units			Non-Intervention Units		
Characteristic	Baseline (7/12-9/13) n = 14,712	Intervention (10/13-6/14) n = 9,030	p-value*	Baseline (7/12-9/13) n = 14,145	Intervention (10/13-6/14) n = 8,484	p-value*
Age in years	60 (46, 71)	60 (46, 71)	0.8838	49 (33, 62)	49 (33, 63)	0.528
Male gender	7,316 (50%)	4,575 (51%)	0.161	5,719 (40%)	3,335 (39%)	0.095
Medicaid/Medicare	9,338 (63%)	6,021 (67%)	<0.0001	6,701 (47%)	4,084 (48%)	0.265

*Within-group p-value

Table 2.2. Model Estimates for Total PRBC Units Transfused

Intervention Modeling	**Covariate**	**Intervention** IRR (95% Robust CI)	**Non-Intervention** IRR (95% Robust CI)
Pre-post indicator	Intervention Indicator	0.87 (0.77, 0.97)	0.97 (0.85, 1.09)
	Proportion Medicaid/Medicare	0.53 (0.15, 1.86)	3.65 (0.92, 14.48)
	Proportion Male	2.09 (0.63, 6.95)	3.01 (0.60, 15.04)
Pre-post indicator-Time Interaction	Intervention Indicator	0.99 (0.99, 1.00)	1.00 (0.99, 1.00)
	Intervention indicator x time	0.99 (0.99, 1.00)	1.00 (0.99, 1.00)
	Proportion Medicaid/Medicare	0.83 (0.19, 3.57)	3.91 (0.94, 16.36)
	Proportion Male	1.74 (0.53, 5.71)	2.99 (0.58, 15.36)
Restricted Cubic Spline	Baseline spline	0.87 (0.73, 1.04)	1.04 (0.85, 1.28)
	Intervention spline	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)
	Proportion Medicaid/Medicare	0.54 (0.15, 2.00)	4.04 (0.99, 16.34)
	Proportion Male	2.07 (0.65, 6.67)	3.23 (0.66, 15.89)

Table 2.3. Model Estimates for Mean PRBC Units Transfused per Discharge

Intervention Modeling	**Covariate**	**Intervention** Mean PRBC Units Transfused per Discharge (95% Robust CI)	**Non-Intervention** Mean PRBC Units Transfused per Discharge (95% Robust CI)
Pre-post indicator	Intervention Indicator	-0.09 (-0.15, -0.03)	-0.03 (-0.14, 0.07)
	Proportion Medicaid/Medicare	-0.23 (-0.93, 0.46)	1.02 (-0.19, 2.22)
	Proportion Male	0.30 (-0.40, 1.00)	0.40 (-1.09, 1.88)
	Constant	0.56 (-0.02, 1.15)	0.27 (-0.52, 1.06)
Pre-post indicator-Time Interaction	Intervention Indicator	-0.003 (-0.004, -0.001)	0.00003 (-0.003, 0.002)
	Intervention indicator x time	-0.001 (-0.004, 0.002)	-0.003 (-0.008, 0.002)
	Proportion Medicaid/Medicare	0.24 (-0.54, 1.03)	1.12 (-0.13, 2.36)
	Proportion Male	0.18 (-0.50, 0.89)	0.40 (-1.12, 1.92)
	Constant	0.41 (-0.18, 1.00)	0.23 (-0.63, 1.09)
Restricted Cubic Spline	Baseline spline	-0.09 (-0.18, 0.004)	0.06 (-0.11, 0.23)
	Intervention spline	-0.0001 (-0.005, 0.004)	-0.005 (-0.01, 0.002)
	Proportion Medicaid/Medicare	-0.22 (-0.96, 0.51)	1.15 (-0.09, 2.38)
	Proportion Male	0.30 (-0.39, 0.99)	0.49 (-0.97, 1.95)
	Constant	0.56 (-0.04, 1.17)	0.17 (-0.61, 0.95)

Figure 2.1. Total PRBC units transfused by week between July 1, 2012 and June 30, 2014 including the baseline linear trend (red) extrapolated through the intervention period (green dashed)

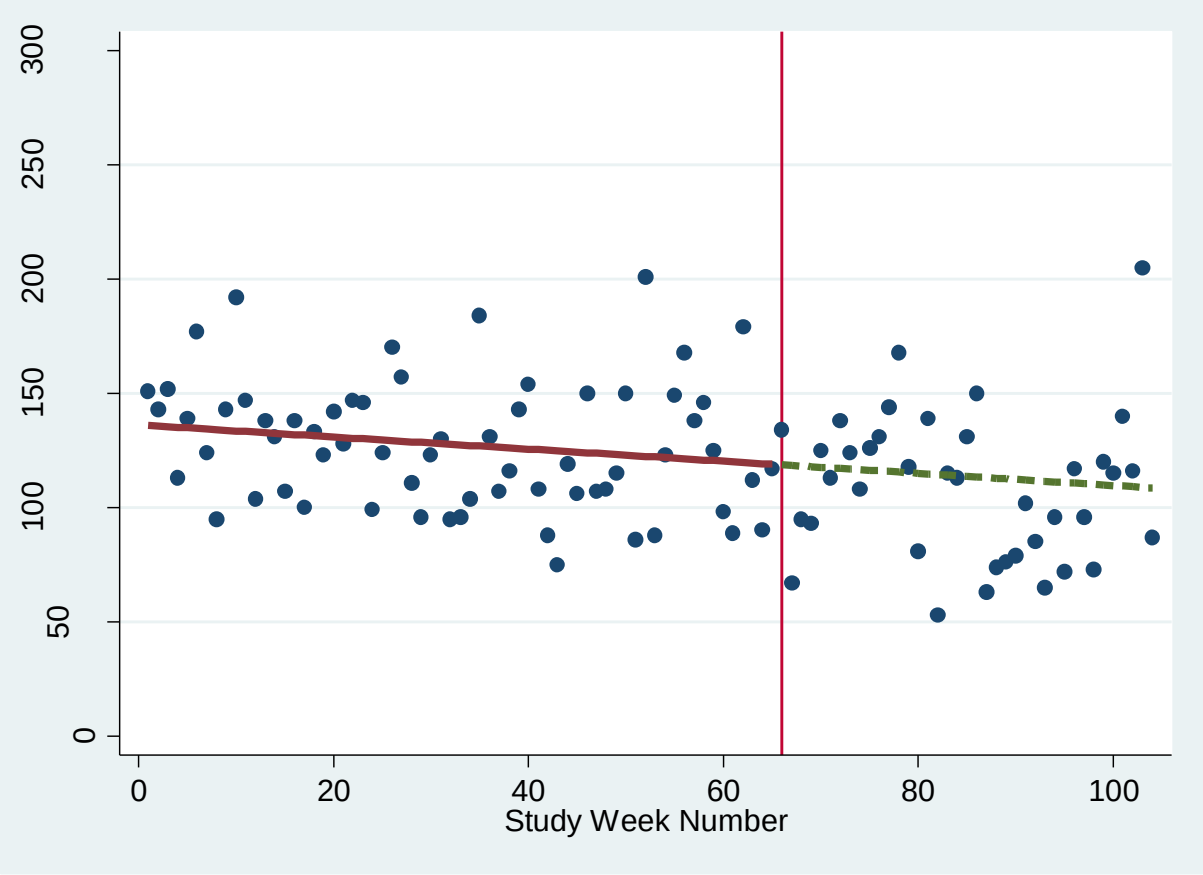


Figure 2.2. Analysis plan diagram for a QI evaluation allocating an intervention at a group level

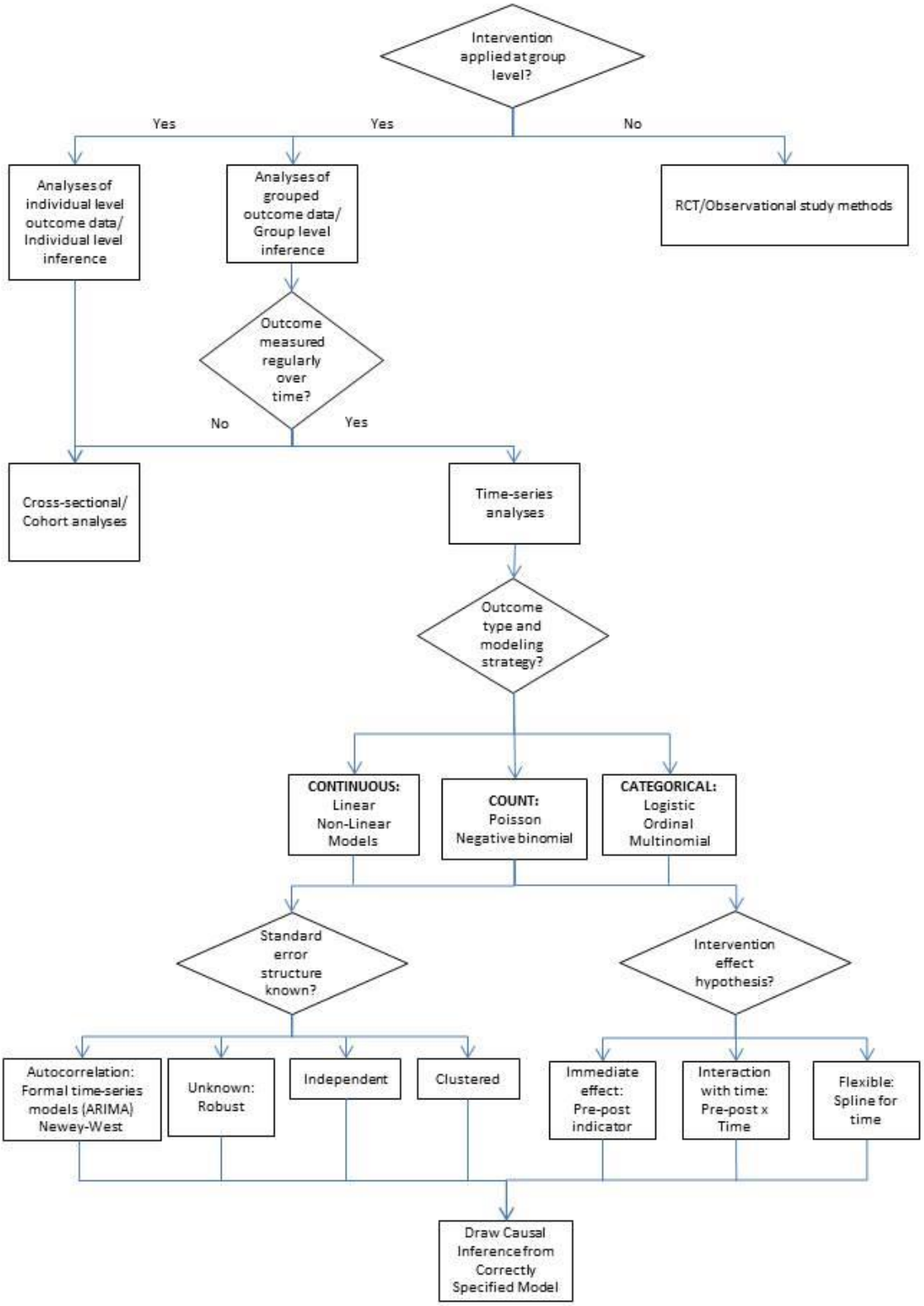
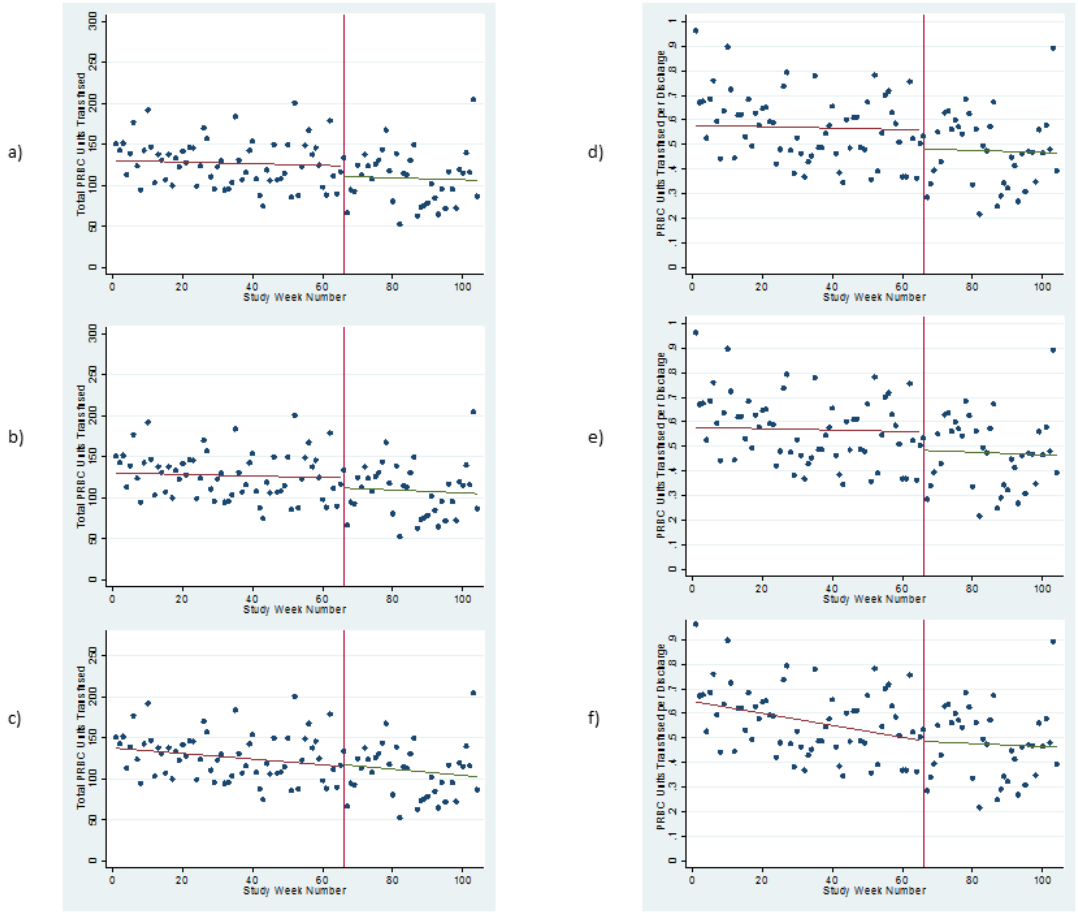


Figure 2.3. Interrupted time-series analysis of the effect of an intervention to decrease packed red blood cell transfusions at the University of California San Francisco Medical Center with varying intervention modeling strategies employed



Legend: Figures 2.3a-c include graphs of total PRBC units transfused at the group level by study week. Figures 2.3d-f include graphs of mean PRBC units transfused per discharge at the group level. Graphs a and d model the intervention with a pre-post intervention indicator which would allow for assessment of change in outcome level between the baseline and intervention periods, graphs b and e model the intervention using a pre-post intervention indicator-time interaction which would allow for assessment of change in the outcome level as well as slope comparing the baseline to the intervention period, and graphs c and f model the intervention through a restricted cubic spline for time as a proxy for intervention and allows for assessment of difference in slope between the baseline and intervention periods.

Chapter 3: The effect of timely access to ambulatory specialty care on healthcare outcomes: a population level analysis

Priya A. Prasad, Lydia Zablotska, Steven Shiboski, Ralph Gonzales, and Nathaniel Gleason

INTRODUCTION

As the proportion of the United States population aged over 65 years grows and the Affordable Care Act expands medical coverage, focus has been drawn to improving access for both primary and specialty care services. [1, 2] It is estimated that by 2025, the demand for primary care services will increase by 14% and the need for specialty care services will grow by 15 to 30% depending on the specialty and the state.[1] In an effort to bridge these gaps in access to ambulatory specialists, innovations in care delivery, such as telemedicine, have evolved.[3, 4]

While prior research has demonstrated that geographical access to ambulatory specialty care affects the rate of health outcomes,[5] little is known about whether *timely* access to ambulatory specialty care impacts measurable patient outcomes. Institutions often strive to offer patients access to ambulatory specialty care within a short window of referral, but determining a meaningful metric of access time can be challenging because of scheduling workflow and patient preferences.[6]

The goal of our study was to develop various metrics to define specialty care access time and to explore potential associations between these metrics and poor patient outcomes, including emergency room visits, hospitalizations, and mortality, all stratified by specialty. We hypothesized that decreased access would correspond to an increased rate of poor outcomes.

METHODS

Source Cohort

The source cohort for this study was patients who sought primary and/or ambulatory specialty care at the University of California San Francisco (UCSF). The UCSF primary care (UCSF PC) population was defined as patients who attended at least one in-person office visit in UCSF Primary Care between January 1, 2013 and December 31, 2015.

Study Population

The study population selected from the source cohort for the proposed analyses included members of the UCSF primary care source cohort who were either seen in a given specialty or received a referral order to the given specialty between January 1, 2015 and December 31, 2015.

Specialties of interest for evaluation of timely access

We focused our analysis on five UCSF medicine specialties for which we believed it would be reasonable to identify an outcome that could be attributable to the referral within the year of referral. These specialties included cardiology, hematology, neurology, otolaryngology and head and neck surgery (OHNS), and urology.

Outcomes of Interest

Emergency department visits for all causes as well as inpatient encounters for all causes occurring at UCSF were identified using hospital billing data. All-cause mortality was ascertained from the electronic medical record (EMR).

Access Time Definitions

Definition #1: UCSF Primary Care Population new patient visit access time

The median weekly new patient visit (NPV) access time was calculated using arrived new patient visits based on billing Evaluation & Management codes (99201-99205, 99241-99245, 99385-99387, and 90791-90792) occurring between January 1, 2015 and December 31, 2015, matching visits to referral orders placed between July 1, 2014 and December 31, 2015 and occurring most proximally within 6 months of the confirmed new patient visit. Referral orders were no longer eligible for inclusion if they occurred greater than 6 months before a new patient visit to maintain consistency with the UCSF referral order authorization system. Outcomes were tracked for patients in the UCSF PC population who either received a referral order or had an arrived office visit at the given specialty between January 1, 2015 and December 31, 2015.

Definition #2: Total visits occurring in a specialty

For this analysis, all specialty care visits arrived by members of the source cohort (UCSF PC and non-UCSF PC patients), regardless of reason, were included from January 1, 2015 to December 31, 2015. We calculated the access to specialty care as the total number of office visits occurring weekly in the given specialty (specialty-level variable). The goal of this metric was to provide a global measure of visit access. Our hypothesis was that during weeks where the visit counts were low, providers were either out

of office for personal reasons or for other commitments related to serving as a provider in an academic medical center. Outcomes were tracked for patients in the UCSF PC population who either received a referral order or had an arrived office visit at the given specialty between January 1, 2015 and December 31, 2015.

Definition #3: Total UCSF primary care population new patient visits occurring in a specialty

For this analysis, all arrived new patient specialty care visits for the UCSF PC population were included from January 1, 2015 to December 31, 2015. We calculated the access to specialty care as the total number of new patient office visits occurring weekly in the given specialty (specialty-level variable). The goal of this metric was to provide a global measure of new patient visit access. Our hypothesis was that during weeks where the new patient visit counts were low, providers were either out of office for personal reasons or for other commitments related to serving as a provider in an academic medical center. Outcomes were tracked for UCSF PC population patients who received a referral order to the given specialty between January 1, 2015 and December 31, 2015.

Data collection

The electronic medical record (EMR) was queried to identify all arrived primary care visits, arrived specialty care visits, and referral orders placed by patient, date, and specialty. Data collected for each primary care encounter included patient demographics (MRN, age, gender, race, ethnicity), encounter identification number, date, location, visit provider, payor, and the first five recorded ICD-9 discharge codes. Data collected for each specialty care encounter included patient demographics (MRN, age, gender, race, ethnicity), encounter identification number, date, location, specialty mapped based on encounter department (Epic DEP), visit provider, payor, and the first five recorded ICD-9 discharge codes. Data collected for each referral order included patient demographics (MRN, age, gender, race, ethnicity), referral order identification number, date, referring department, referring provider, authorizing provider, specialty mapped based on description, and the first five recorded ICD-9 discharge codes associated with the referral order. Emergency department admissions, inpatient admissions, and total remittance were captured using billing data. Because the data were collected for an ongoing quality

improvement initiative, the study received exemption from the UCSF Human Research Protection Program Institutional Review Board.

Statistical analysis

Our study was conducted using data aggregated at the level of the week and the specialty. Population characteristics, weekly access using each definition, and weekly rates of outcomes were summarized using frequencies and proportions for count and categorical data and medians and interquartile ranges for continuous data. Poisson regression was used to identify associations between our varying definitions of access and the count of the summary outcome measure, which included ED encounters, hospitalizations, and death. Confidence intervals and p-values were derived using robust standard errors to account for any autocorrelation in the weekly time series or any other nuances in the observed distribution of outcomes.

RESULTS

Population characteristics:

During the study period, there were 59,245 patients who attended at least 1 in person PC office visit included in the UCSF PC population, 33,834 patients who were in the UCSF PC Population and who received at least 1 specialty referral order or arrived for at least 1 specialty visit at UCSF in 2015 (access definition 1 and 2 population), and 26,111 members of the UCSF PC population patients who received at least 1 specialist referral order in 2015 (access definition 3 population). Demographics of the populations appear in Table 3.1.

Specialty level analyses of access time metrics

Summaries for each of the access metrics and outcomes by population can be found in Table 3.2. In addition, Figures 3.1-3.5 provide a graphical display of the weekly access measures and outcomes.

Cardiology: Based on unadjusted Poisson models with robust standard errors, there was an association between median weekly access time and outcomes. For each day increase in cardiology weekly NPV access time, the rate of outcomes in the population increased by .04% ($p = 0.005$) (Table 3.3). There was no association between weekly outcomes and either total weekly visits or weekly new patient visits to

cardiology. For each single visit increase in NPV in the previous week, the rate of weekly outcomes in the population in the current week decreased by 0.08% ($p = 0.015$) (Table 3.4).

Hematology: Based on unadjusted Poisson models with robust standard errors, there was an association between median weekly NPV access time and outcomes. For each day increase in hematology weekly NPV access time, the rate of outcomes in the population decreased by 0.03% ($p < 0.001$) (Table 3.3). There was no association between weekly outcomes and either total weekly visits or weekly new patient visits to hematology, nor was there an unadjusted association between outcomes and any of the lagged access definitions (Table 3.4).

Neurology: Based on unadjusted Poisson models with robust standard errors, there was no association between weekly outcomes and weekly NPV access time, total weekly visits, or weekly new patient visits to neurology (Table 3.3). For each day increase in neurology weekly access time in the previous week, the rate of weekly outcomes in the population decreased by 0.04% ($p = 0.025$) (Table 3.4).

OHNS: None of the unadjusted Poisson models with robust standard errors revealed a significant association between population outcomes and the current week's access metrics (Table 3.3) or the previous week's access metrics (Table 3.4).

Urology: Based on unadjusted Poisson models with robust standard errors, there was an association between median weekly NPV access time and outcomes. For each day increase in hematology weekly NPV access time, the rate of outcomes in the population decreased by 0.04% ($p < 0.001$) (Table 3.3). There was no association between weekly outcomes and either total weekly visits or weekly new patient visits to urology, nor was there an association between outcomes and any of the lagged access definitions (Table 3.4).

DISCUSSION

In this manuscript we have outlined three different strategies to measure access to specialty care at UCSF Medical Center including time from referral to new patient visit, total visits to a specialty, and total new patient visits to a specialty. We then determined whether an association existed between the metrics and population level all-cause outcomes.

Our models revealed a few interesting relationships that warrant further exploration. Results from our unadjusted analysis of the cardiology specialty showed that for each 10 day increase in median NPV access time, the rate of outcomes increased by 4%, a finding which supported our initial rationale that increased time between referral and visit would be associated with poorer patient outcomes. When we included the lagged access metrics which represented the previous week's access, we found that there was a decrease in the rate of population-level outcomes in the index week for each unit increase in the access the previous week in neurology for NPV visit access time and in cardiology for total NPVs. One potential hypothesis for this finding is that staff and clinicians may prioritize scheduling patients who are at risk of a poor outcome quickly after periods of known poor access.

Although little data exists relating patient outcomes to ambulatory specialty care access, studies have been published demonstrating that increased wait time until surgery is associated with adverse health outcomes and worsening of symptoms.[6] Beyond the potential risk of adverse outcomes related to delayed access, studies have been published addressing additional consequences of longer wait times. In a case-control study conducted at Banner University Medical Center between March and October 2014, investigators found that the adjusted odds of missing a gastroenterology appointment increased by 14% for every 10 day increase in time between referral and scheduled appointment (per day, 95% CI 1.01-1.02).[2] The psychological toll that is placed on a patient during the time between referral and diagnosis has also been explored. In studies of surgery wait time, uncertainty and powerlessness were also identified as consequences of poor access.[6]

Our data show that the median weekly access time exceeded the UCSF institutional target of two weeks in all specialties evaluated except cardiology, findings which are not unique to our institution. Researchers at St. Paul's Hospital in Vancouver, Canada found that the mean wait time from referral to GI office visit was 63 days and 59% of individuals who were referred to GI for endoscopy for colorectal cancer experienced wait times greater than the length recommended by the Canadian Association of Gastroenterology Wait Time Consensus Group. When alarm signs for colorectal cancer were present, the mean wait time from referral to endoscopy was 86 days, much greater than the recommended 60 days.[7]

Improving access to care has become a priority for health systems in the US and abroad and with this increased attention, strategies have been developed to address gaps in timely access to specialty care. Members of our investigative team developed an electronic consult system at UCSF that is integrated into the electronic medical record workflow at UCSF. The eConsult platform supports the work of both the PCP and the specialist and reduces the need for unnecessary specialty office visits which can be appropriately managed in primary care. Based on our program evaluation, after implementation of the eConsult program at UCSF, referrals to specialty care for office visits decreased by 19% between the baseline (July 2011 through June 2012) and intervention periods (July 2012 through May 2013).[8] Telemedicine is another strategy that has been proposed to maximize the access to specialty care. In a study of patients with head and neck cancer in the Veterans Health Administration (VHA) system, subjects were given the opportunity to select in-person surgery consultation or a telemedicine alternative which included biopsy at a remote location, review by experts of all aspects of the medical record at the Palo Alto Veterans Affairs, and a 30-minute audiovisual teleconference with the patient. The telemedicine consultation option saved patients and the VHA over $19,000, 600 hours of travel, 14.5 metric tons of carbon dioxide emissions, over 1,600 miles of travel per patient and $900 in travel expenses per patient.[9] A recently proposed stepped wedge cluster randomized controlled trial will test the implementation of the Specific & Timely Appointments for Triage (STAT) program in 8 community and subacute outpatient settings in Australia and New Zealand. The STAT method involves providers protecting weekly appointment times in their schedules to manage new referrals based on a review of current supply and demand. The providers have the opportunity to prioritize new and follow-up cases based on the relative difficulty of management and triage is done at the point of care as opposed to through a complex triage system.[10] One study showed that time from referral to first appointment was 4 days fewer after implementation of the STAT method in an outpatient physiotherapy clinic.[11] While implementation of innovations in care delivery has begun, it remains to be seen what affect these strategies will have on timely access to specialty care and the impact on patient outcomes.

There are limitations to our study. We restricted our cohort to the UCSF primary care population under the assumption that these patients would be more likely to seek specialty care and experience outcomes of interest within the UCSF healthcare system. However, it is possible that patients in the UCSF primary care population presented to institutions outside our system and these outcomes would not be included in our study. The outcomes included were all-cause emergency department encounters, inpatient hospitalizations, and mortality. With the current design we were unable to ascertain whether the outcomes were in fact related to the requirement for specialty access. Our metric for NPV access time uses the time between referral order and arrived visit, with no-shows being excluded and there is no adjustment for patient rescheduling. Because we wanted to perform a global assessment of the access metrics on population level outcomes, we did not adjust for any health system, specialty, provider or patient characteristics, which may leave our results subject to bias.

Our manuscript provides a platform for describing our three access metrics but there is much future work to be done. In addition to more formally assessing the performance of each of our access time metrics, next we will explore whether time modifies the effect of access on outcomes given that the data were collected over a one year period. In order to control for cyclical and secular trends, additional years of data could be added and more traditional time-series analysis methods, such as autoregressive integrated moving average analyses, could be employed. In addition, relevant confounders of the association between outcomes and access should be identified and included in these models to obtain a clearer picture of the true effect of timely access on outcomes. And lastly, while our analysis approaches access and outcomes from the population level, it may be more relevant to study the effect of access at the individual level for future analyses.

ACKNOWLEDGEMENTS

The authors would like to thank Michael Wang, Ariana Afshar, John Hillman, PhD, and Phil Lesko for their contributions to this manuscript.

REFERENCES

1. Dall TM, Gallo PD, Chakrabarti R, West T, Semilla AP, Storm MV. An aging population and growing disease burden will require a large and specialized health care workforce by 2025. *Health Aff (Millwood)* 2013, **32**(11)**:** 2013-2020.

2. Shrestha MP, Hu C, Taleban S. Appointment Wait Time, Primary Care Provider Status, and Patient Demographics are Associated With Nonattendance at Outpatient Gastroenterology Clinic. *J Clin Gastroenterol* 2016.

3. Caffery LJ, Farjian M, Smith AC. Telehealth interventions for reducing waiting lists and waiting times for specialist outpatient services: A scoping review. *J Telemed Telecare* 2016, **22**(8)**:** 504-512.

4. Winpenny EM, Miani C, Pitchforth E, King S, Roland M. Improving the effectiveness and efficiency of outpatient services: a scoping review of interventions at the primary-secondary care interface. *J Health Serv Res Policy* 2016.

5. Graves BA. Integrative literature review: a review of literature related to geographical information systems, healthcare access, and health outcomes. *Perspect Health Inf Manag* 2008, **5:** 11.

6. Fogarty C, Cronin P. Waiting for healthcare: a concept analysis. *J Adv Nurs* 2008, **61**(4)**:** 463-471.

7. Janssen RM, Takach O, Nap-Hill E, Enns RA. Time to Endoscopy in Patients with Colorectal Cancer: Analysis of Wait-Times. *Can J Gastroenterol Hepatol* 2016, **2016:** 8714587.

8. Gleason N, Prasad PA, Ackerman S, Ho C, Monacelli J, Wang M, *et al.* Adoption and impact of an eConsult system in a fee-for-service setting. *Healthc (Amst)* 2016.

9. Beswick DM, Vashi A, Song Y, Pham R, Holsinger FC, Rayl JD, *et al.* Consultation via telemedicine and access to operative care for patients with head and neck cancer in a Veterans Health Administration population. *Head Neck* 2016, **38**(6)**:** 925-929.

10. Harding KE, Watts JJ, Karimi L, O'Reilly M, Kent B, Kotis M, *et al.* Improving access for community health and sub-acute outpatient services: protocol for a stepped wedge cluster randomised controlled trial. *BMC Health Serv Res* 2016, **16**(a)**:** 364.

11. Harding KE, Bottrell J. Specific timely appointments for triage reduced waiting lists in an outpatient physiotherapy service. *Physiotherapy* 2016, **102**(4)**:** 345-350.

Table 3.1. Characteristics of the UCSF Primary Care Population and the Cohorts Assessed for Outcomes to Explore Access Definitions

	UCSF Primary Care Population* N=59,245	Definition 1&2 Outcomes Population[#] n=33,834	Definition 3 Outcomes Population[$] n=26,111
Age in Years (median, IQR)	46.8 (33.4, 61.1)	51.1 (36.6, 64.1)	52.0 (37.9, 64.4)
Male	24,328 (41%)	13,219 (39%)	10,554 (40%)
Race			
White	28,395 (48%)	16,848 (50%)	12,888 (49%)
Asian	12,459 (21%)	7,122 (21%)	5,375 (21%)
Black/African American	4,690 (8%)	2,907 (9%)	2,390 (9%)
Hawaiian/Pacific Islander	1,203 (2%)	545 (2%)	437 (2%)
American Indian	127 (<1%)	77 (<1%)	62 (<1%)
Other	8,139 (14%)	4,548 (13%)	3,592 (14%)
Unknown	4,232 (7%)	1,787 (5%)	1,367 (5%)
Hispanic ethnicity	4,842 (8%)	2,822 (8%)	2,254 (9%)

*Patients with at least one visit to UCSF Primary Care between January 1, 2013 and December 31, 2015
[#]Members of the UCSF Primary Care Population who had at least one referral order or arrived visit to the specialty of interest during 2015.
[$]Members of the UCSF Primary Care Population who had at least one referral order to the specialty of interest during 2015.

Table 3.2. Summary of Access and Outcomes for Specialties of Interest

	Average Median New Patient Visit Access Time in Days: Definition 1	Average 2015 Total Visits (Range): Definition 2	Average 2015 New Patient Visits (Range): Definition 2	Average Outcomes (Range): Definition 1 & 2 Population[#]	Average Outcomes (Range): Definition 3 Population[$]
Cardiology	28 (11,66)	294.3 (43,370)	19.5 (5,39)	47.5 (31,76)	19.7 (12,36)
Hematology	52.5 (12,125)	345.7 (197,429)	4.6 (1,11)	16.4 (11,22)	8.2 (4,13)
Neurology	38.0 (11,65)	581.7 (80,788)	22.2 (4,40)	45.8 (29,61)	29.1 (20,42)
OHNS*	38.5 (9.5,85)	307.1 (79,417)	25.4 (1,41)	27.8 (15,42)	21.2 (11,34)
Urology	32.3 (13,60)	302.7 (145,372)	17.5 (10,30)	24.8 (13,38)	13.3 (5,19)

*OHNS, otolaryngology and head and neck surgery

[#]Members of the UCSF Primary Care Population who had at least one referral order or arrived visit to the specialty of interest during 2015.

[$]Members of the UCSF Primary Care Population who had at least one referral order to the specialty of interest during 2015.

Table 3.3. Poisson Regression Point Estimates for Relationship between Different Definitions of Access and Outcomes

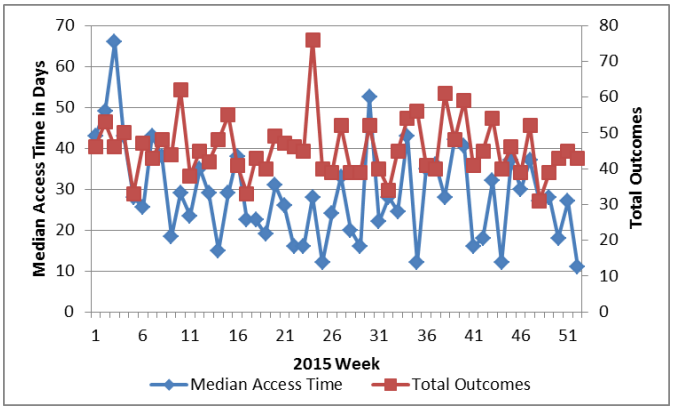
	New Patient Visit Access Time (IRR, 95% CI); p-value	Total Visits (IRR, 95% CI); p-value	Total New Patient Visits (IRR, 95% CI); p-value
Cardiology	1.004 (1.001, 1.007); 0.005	1.001 (0.999, 1.001); 0.111	1.004 (0.996, 1.013); 0.328
Hematology	0.997 (0.996, 0.999); <0.001	1.000 (0.999, 1.001); 0.756	1.001 (0.971, 1.033); 0.933
Neurology	1.001 (0.998, 1.005); 0.448	1.000 (1.000, 1.001); 0.332	0.995 (0.988, 1.002); 0.147
OHNS*	1.001 (0.998, 1.004); 0.401	1.000 (1.000, 1.001); 0.156	1.005 (0.999, 1.012); 0.107
Urology	0.996 (0.992, 0.999); 0.018	0.999 (0.998, 1.001); 0.310	0.998 (0.984, 1.012); 0.762

Table 3.4. Poisson Regression Point Estimates for Relationship between the Previous Week's Access Defined Three Ways and Outcomes

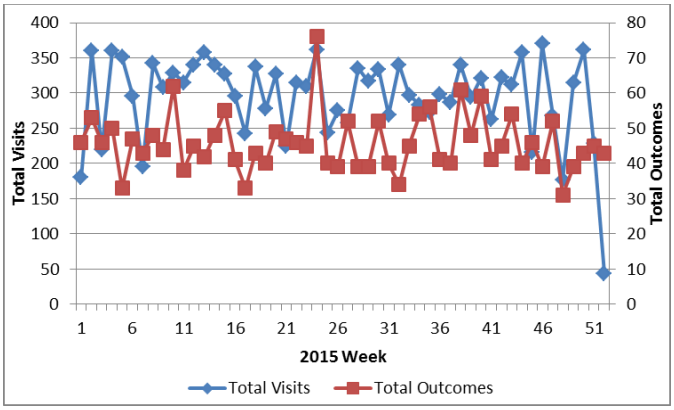
	New Patient Visit Access Time (IRR, 95% CI); p-value	Total Visits (IRR, 95% CI); p-value	Total New Patient Visits (IRR, 95% CI); p-value
Cardiology	0.998 (0.993, 1.002); 0.427	1.000 (0.999, 1.001); 0.803	0.992 (0.986, 0.998); 0.015
Hematology	1.001 (0.999, 1.003); 0.219	1.000 (0.999, 1.001); 0.721	0.990 (0.959, 1.022); 0.538
Neurology	0.996 (0.993, 0.999); 0.025	1.001 (0.999, 1.001); 0.771	1.001 (0.994, 1.008); 0.772
OHNS*	1.000 (0.997, 1.002); 0.768	0.999 (0.998, 1.000); 0.087	1.000 (0.992, 1.007); 0.994
Urology	0.996 (0.991, 1.001); 0.136	0.999 (0.998, 1.001); 0.249	0.991 (0.974, 1.009); 0.346

Figure 3.1. Weekly Access and Population Outcomes for Cardiology, 2015

i) Access Definition #1: Median Weekly Access Time



ii) Access Definition #2: Total Visits



iii) Access Definition #3: Total New Patient Visits

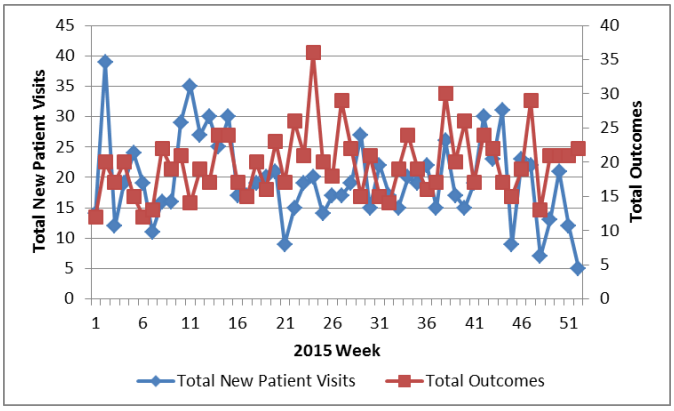
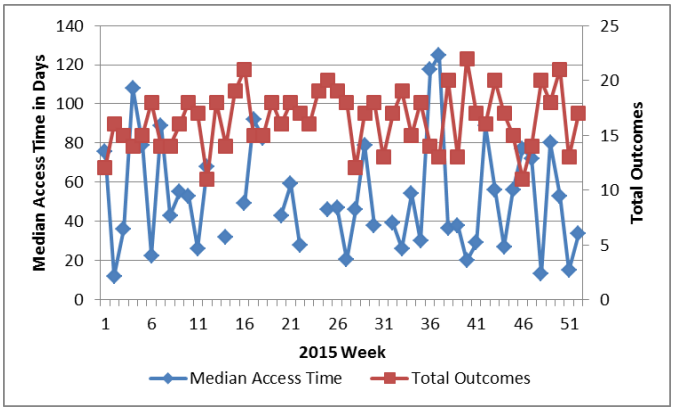
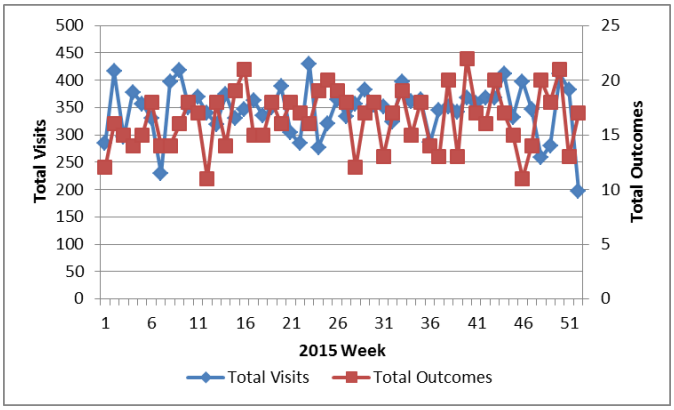


Figure 3.2. Weekly Access and Population Outcomes for Hematology, 2015

i) Access Definition #1: Median Weekly Access Time



ii) Access Definition #2: Total Visits



iii) Access Definition #3: Total New Patient Visits

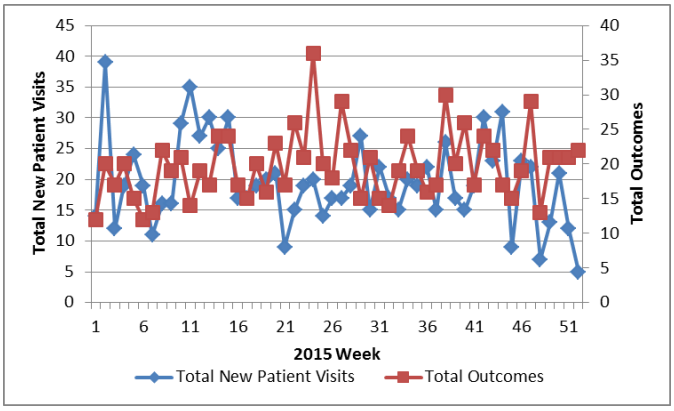
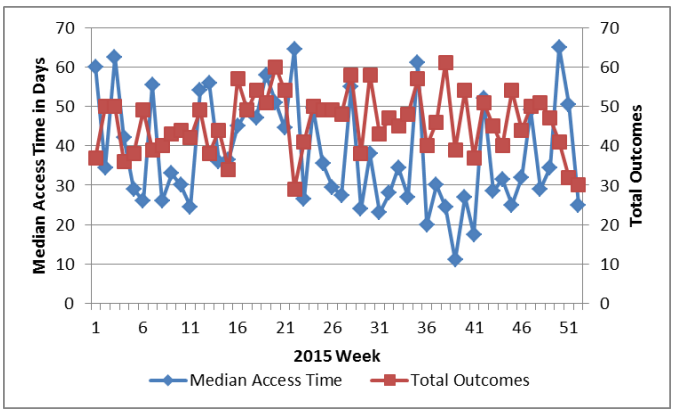
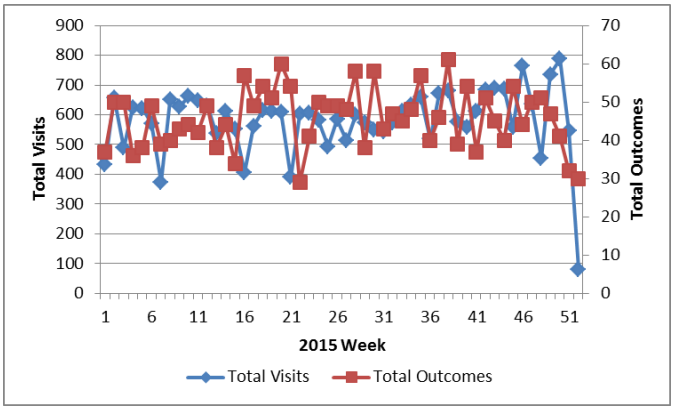


Figure 3.3. Weekly Access and Population Outcomes for Neurology, 2015

i) Access Definition #1: Median Weekly Access Time



ii) Access Definition #2: Total Visits



iii) Access Definition #3: Total New Patient Visits

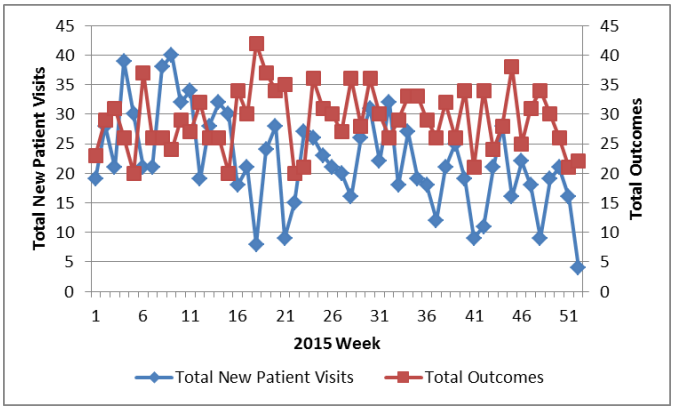
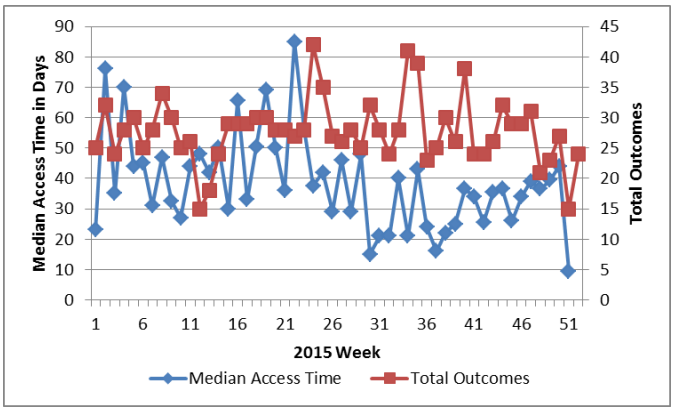
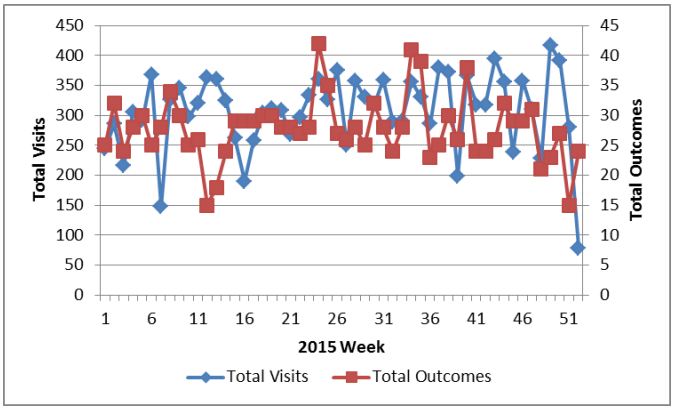


Figure 3.4. Weekly Access and Population Outcomes for Otolaryngology and Head and Neck Surgery (OHNS), 2015

i) Access Definition #1: Median Weekly Access Time



ii) Access Definition #2: Total Visits



iii) Access Definition #3: Total New Patient Visits

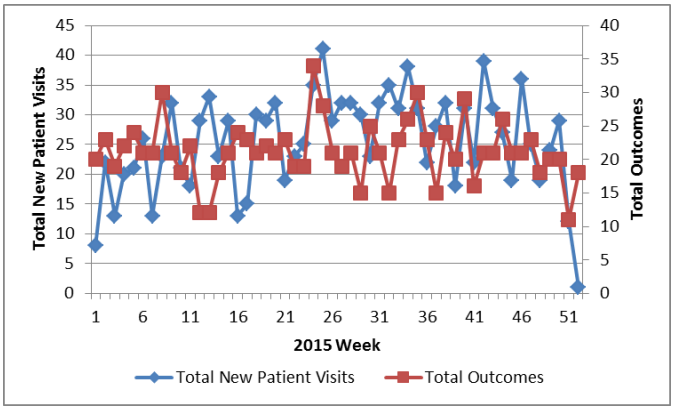
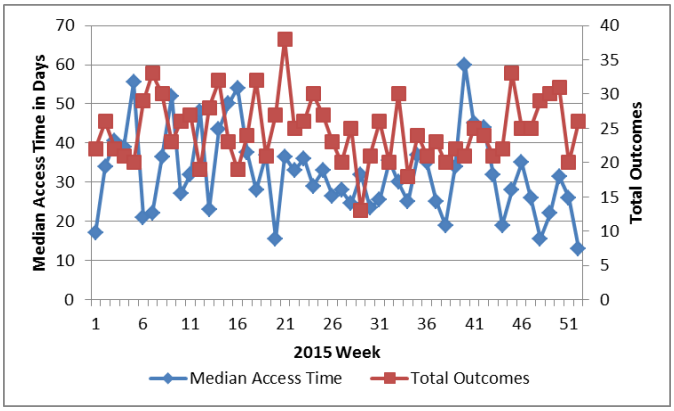
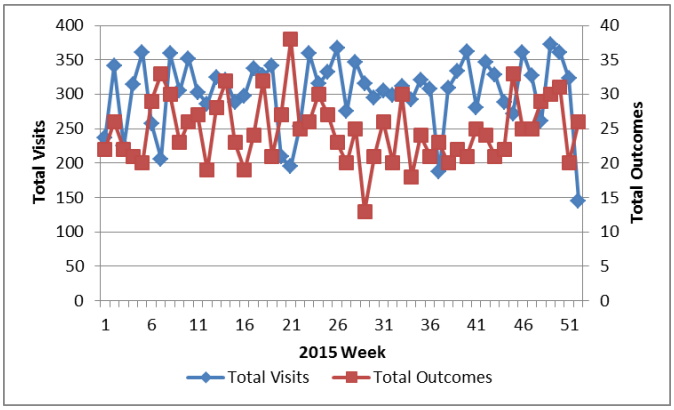


Figure 3.5. Weekly Access and Population Outcomes for Urology, 2015

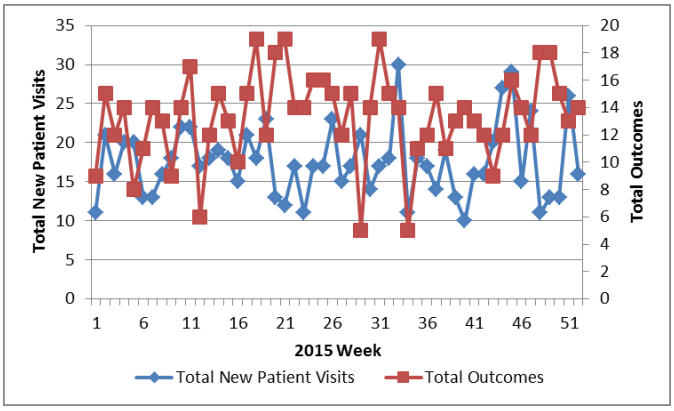
i) Access Definition #1: Median Weekly Access Time



ii) Access Definition #2: Total Visits



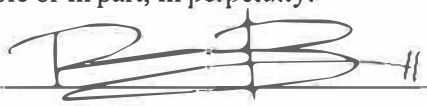
iii) Access Definition #3: Total New Patient Visits



Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature  Date 12/14/16