The Use of Inter-Clinician Variation in Measuring Healthcare Performance

Ronald William Buie

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2018

Master's Thesis Committee:

Adam Wilcox, Chair

Anne M Turner

Program Authorized to Offer Degree:

Biomedical and Health Informatics

University of Washington

Abstract

The Use of Inter-Clinician Variation in Measuring Healthcare Performance

Ronald William Buie

Chair of the Supervisory Committee:
Dr. Adam Wilcox, Professor
Department of Biomedical Informatics and Medical Education

To monitor and improve healthcare in the US, providers are required to report healthcare measures as part of regulatory and compensatory systems. However, there are growing concerns that the collection and reporting of these measures may be counter-productive to provider efforts to improve care. Variations in care are known to adversely affect quality, but studies on the relationship of variation within measures and performance of those measures are lacking. We aimed to test if inter-clinician variation of a healthcare measure was associated with performance of that measure and to thereby establish a model for identifying measures that might be more likely associated with opportunities to improve care. We identified the proportion of diabetic patients with blood pressure under control as an important measure. Measure performance and inter-clinician variation was calculated each month and both visual and time series analyses were performed. We found that between 14% and 23% of the performance of our chosen measure was associated with variation between clinicians. This finding suggests that inter-clinician variance of

a measure can be used to help identify measures where opportunities for improvement of clinical processes exist.

TABLE OF CONTENTS

Chapter 1. INTRODUCTION

Variations in the provision of care are associated with poorer health outcomes and increased costs. However, the quality measures regularly collected and reported by providers do not take this variation into account. This may limit the value of these measures in improving quality of care. By determining if there is an association between performance and the inter-clinician variation of a reportable healthcare measure, the study described here aims to explore the possibility of using inter-clinician variation as a tool for identifying a reportable measure's association with quality of care.

The background and significance section discusses the current state of quality improvement efforts in medicine as well as the underlying theory of the variance-quality relationship in healthcare settings. The methods section outlines our scientific approach and methods for evaluating the variance-performance relationship. The results section details the results of conducting our study, including effects on data availability and statistical association. In the discussion section we describe the findings from the research and note limitations. Finally, in the conclusion section, we discuss possible applications and potential next steps for area of research.

1.1 SUMMARY

1.1.1 *Background*

American healthcare faces challenges in providing care that leads to desired health outcomes to the population and is aligned with the latest knowledge. [1-3] Poor quality of care is associated with high rates of harm and wasted resources. [1-3] To address this, quality of care has been made a national priority.[4-6] National efforts to monitor and improve quality of care have resulted in nationwide agencies that regulate and report on the quality of care performed. [4, 7, 8] In order to

perform their respective regulating functions, the Centers for Medicare and Medicaid Services, the Joint Commission, and the National Quality Forum require providers to collect and report measures of quality.

Despite their aim to improve quality, the collection and reporting of these measures is a burden on providers that may impede quality improvement.[4, 9-12] Developing capacity to report these measures requires substantial investments in time and infrastructure on the part of providers.[11] Public reporting is susceptible to gaming and promotes behaviors that can optimize the measure without leading to notable improvements in care.[13-15]

Other industries have successfully improved quality by incorporating approaches to quality improvement that rely on internally motivated and scientifically measured efforts.[16-18] In these approaches, workers close to a process collaborate to identify and execute opportunities for improvement.[5, 17, 19] These approaches rely on stochastic measures to judge the performance of a system, capturing both mean performance and variation from that mean.[16, 17, 20, 21]

Variation in care has been found to be associated with increased costs and worse health outcomes.[4, 9, 17, 18, 20, 22] Variation between clinicians is the second most common source of this variation, behind differences between patients, and may be the most effective type of variation to intervene upon.[23, 24] Intermountain Healthcare and Yale New Haven have successfully focused on clinician level variation to improve outcomes while reducing costs.[25, 26] By studying the relationship between the performance of reportable measures and inter-clinician variation along the measure, it may be possible to identify measures more likely to improve outcomes and reduce costs.

1.1.2 *Methods*

We hypothesized that higher performing processes would exhibit reduced inter-provider variation, and so an inverse relationship exists between inter-provider variance and superior performance of a measure. To test this hypothesis, we identified the AHRQ 2004 unified framework of the quality of a measure[27]. The unified framework identifies measures of quality in the domains of Importance, Scientific Acceptability, Usability, and Feasibility. Our study attempts to address the Importance domain in planning and Scientific Acceptability domain in execution. With the guidance of hospital leadership, we identified National Quality Forum measure 0061 "Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg)" as meeting requirements of the Importance domain.[28, 29] Due to limitations in clinician attribution, we limited our study to the outpatient population.

Our study was set in Harborview Medical Center, a level I trauma center with 413 beds and 258,666 outpatient visits in 2017.[30] The study population included diabetic outpatients from 2011 to 2017. Outpatient data was collected from the UW Clinical Data Warehouse. Per-visit mean systolic and diastolic readings were collected on all diabetic visits during this period. For each visit, an indicator of control status was assigned. We expected mean blood pressure readings to vary by 5 mm/Hg over the year due to seasonal temperature change. [31, 32] We also expected a doubling in our annual study population during 2014 due to the implementation of Meaningful Use Phase 3, where our measure became connected with compensation under the Meaningful Use program. [33] The measure was defined as the proportion of visits with a mean blood pressure below 140/90 mm/Hg each month. To achieve a measure of variation that was not unduly influenced by clinicians who saw only a small number of diabetic patients in a month, inter-

clinician variation was calculated using the over-dispersion parameter of the Quasi-Poisson regression. This provided a measure of inter-clinician variation weighted towards the mean.

We visually analyzed the relationship between measure performance and inter-clinician variation by plotting the measure and its variation on a time series. We statistically modeled the relationship using Auto Regression (AR), Moving Average (MA), and combined AR-MA models.

1.1.3 *Results*

Our data included 216,956 outpatient visits. 14,440 observations were excluded due to missing provider attribution or having outlier blood pressure readings. Our performance measure was built using data from the remaining 200,211 visits. Building our measure resulted in a calculation of the proportion of patients whose blood pressure was under control for each month of the study (N = 84). We then stratified the monthly proportion by clinician and calculated variation as the over dispersion parameter in each month using the Quassi-Poisson regression with number under control as our outcome and the number treated as the predictor.[34]

Visual analysis of our measure showed the expected seasonal change of our measure each year. Our monthly variation did not show an obvious change. We observed an inverse performance-variation relationship in 2014 as a reduction in performance correlated with an increase in variation.

We conducted exploratory analysis to identify parameters for Auto Regression (AR) and Moving Average (MA), and combined ARMA modelling. Models were scored using the Akaike Information Criterion (AIC).[35] All models estimated a significant inverse relationship between measure performance and inter-clinician variation ($P < 0.001$). Our best scoring model, using AR lag 2, estimated that 14% (95% CI 3.7% $P < 0.001$) of the monthly change in our measure inversely correlated to inter-clinician variation.

1.1.4 *Discussion*

In this study we demonstrated a significant ($P < 0.001$) inverse relationship between measure performance and inter-clinician variation suggesting that further research may have merit. The finding of 14% of changes in performance being associated with inter-clinician variation aligns with similar research on sources of variation in care. Meta analyses of these studies estimated clinicians and care teams are associated with 0-18% of the variation in outcomes exclusive of patient satisfaction.[23, 24] Visual analysis of our time series demonstrated an inverse relationship during the implementation of Meaningful Use Phase 3. However, without this major change in performance, visual analysis may not have been compelling. In preparing our data and constructing our measure, data quality issues were evident in the distribution of measured blood pressure and rates of failure to attribute providers. This may have introduced biases into our analysis that may be important when stratifying the measure by clinician. Testing the performance-variation relationship ultimately required building an ARMA model of the time series. While our approach was simpler than some other innovative approaches using process mining or unique statistical models, it is still more complex than we had hoped and may prove a burden to implement by some providers.

Our study had several limitations. Our approach to defining outliers biased our biological data, blood pressures, by forcing a normal distribution on data that likely follows a power law.[36] Due to time and regulatory constraints, we were only able to build and test one measure, where multiple measures would have been preferred. As a limit of time series modelling, it is possible that month-to-month confounding exists that is not captured in our ARMA models. Finally, our approach could not identify causality or directionality to the performance-variation relationship, and so careful interpretation is warranted with this approach.

1.1.5 *Conclusion*

By demonstrating a relationship between inter-clinician variation and the performance of a reportable measure, we provide an initial step in further testing of the use of the performance-variation relationship to identify and utilize reportable measures in quality improvement. The performance-variation relationship may have immediate use in measure selection. Providers optimizing measure selection for current performance might prioritize high performing but low variation measures, may consider high performing but high variation measures, and would likely avoid low performing measures regardless of variation. On the other hand, providers optimizing measure selection for opportunities in quality improvement would prioritize high variation measures, regardless of performance. A provider seeking a compromise between both strategies may prioritize high performing high variation measures and avoid others where possible. The degree of association between variation and performance may attenuate these concerns.

The evaluation of inter-clinician variation has important implications for how measures are used in the future. For an individual provider, low clinician-level variation, or a small correlation between performance and variation, may suggest that a change in a measure is not closely related to care process and so quality improvement efforts should focus on other factors. Aggregated across multiple providers, inter-clinician variation may be used to improve measure development. Applied by regulators to evaluating or compensating individual providers, inter-clinician variation could be used to weight the importance of a measure when evaluating a facility, where measures with high variation may be more heavily weighted.

Future research should test the generalizability and reproducibility of these results with other measures and in other facilities. Other important areas of research include how inter-clinician variation may be most effectively presented to clinicians, administrators, and policy makers, and

implementation research identifying best practices in effectively integrating the use of inter-clinician variation into the clinical environment.

Chapter 2. BACKGROUND AND SIGNIFICANCE

2.1 Defining Quality of Care

Since 1990, the National Academy of Medicine (NAM, formerly the Institute of Medicine (IOM)) has defined quality of care as, "The degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge."[1] Captured in this definition are many key features of quality. Its application is broad ("health services"), it identifies the source of performance goals as being stakeholder preferences and values ("desired health outcomes"), it links processes and outcomes (health services […] increase […] outcomes), it captures randomness of measured outcomes ("increases likelihood of"), and that the provider is responsible for maximally performing within an evolving constraint of technical, medical, and scientific knowledge ("consistent with current professional knowledge").[1] Care that fails in these features, that has a narrow definition of quality, that is not oriented towards the needs of stakeholders, that ignores the link between process and outcome, that does not measure randomness of these outcomes, and where providers fail to pursue an ever-changing knowledge base, is *not* quality care.

2.2 The Costs of Poor Quality Care

Failure to achieve quality presents significant costs. Harm due to substandard inpatient care is, by conservative estimates, responsible for approximately 400,000 deaths per year, making it the 3rd leading cause of death in the nation.[1-3] Estimates of the effect of quality on the mortality in the

outpatient setting would substantially increase this number, but national data on harms caused by outpatient care are not available for the US healthcare system.[37] Wasteful healthcare expenses, in the form of unnecessary services, inefficient delivery, excess administrative costs, inflated prices, poor prevention, and fraud total $750 billion. $395 billion of this is attributable to unnecessary services, inefficiently delivered services, or missed opportunities for prevention.[4] Recent initiatives have helped reduce the rate of unnecessary inpatient mortality but it is expected that further improvement in this area, and in outpatient care, will require more rigorous efforts.[38]

Without drastic changes, the financial costs of our current approach to healthcare and quality improvement is likely to continue to grow.[4] Healthcare expenses as a percentage of Gross Domestic Product (GDP) are expected to increase by 2%, to a total of 19.7%, by 2026.[39] This increase is driven by largely structural factors, including the aging of the population, increases in pharmaceutical costs, and increases in enrollment and patient population under the Affordable Care Act (ACA).[39]

While increasing costs of healthcare are common to many countries, the US is an outlier in the proportion of GDP these expenses consume.[4, 40] Currently estimated at 17.9% of GDP,[41] comparable nations spend roughly half as much on healthcare, with generally superior outcomes.[40] The disproportionate outcomes, despite the increased cost, suggests there are opportunities for further improvement.[42]

2.3 Quality as a National Priority

Promoting quality of care and reducing unnecessary expenditures has been a decades long priority of the NAM.[4-6] In their 2000 report, *Crossing the Quality Chasm*, the NAM set out a vision where all members of the healthcare industry would be tasked with pursuing quality along six aims: safety, effectiveness, patient centeredness, timeliness, efficiency, and equity.[5]

Table 2.1. NAM Six Aims for Achieving Quality in Healthcare

• *Safe*—avoiding injuries to patients from the care that is intended to help them.
• *Effective*—providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit (avoiding underuse and overuse, respectively).
• *Patient-centered*—providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions.
• *Timely*—reducing waits and sometimes harmful delays for both those who receive and those who give care.
• *Efficient*—avoiding waste, including waste of equipment, supplies, ideas, and energy.
• *Equitable*—providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status.

Institute of Medicine, Committee on Quality of Health Care in America. Crossing the Quality Chasm: a new health system for the 21st century. Washington, D.C.: National Academy Press; 2001.

The NAM acknowledged that pursuing these aims would require a cultural shift away from punitive approaches that stifled innovation, and the adoption of aggressive efforts to redefine and measure clinical processes.[5] This would also require increased investments in information technology and improvement of the quality of health and health care performance information available to providers and consumers.[5] In pursuit of these changes NAM envisioned a redefined role for regulators, where information and measures of quality would be defined and widely communicated to responsible parties who would then facilitate a response to this information.[5] The NAM advocated for regulatory incentives and a more collaborative industrial culture that supports and encourages pursuing the quality aims in accordance with these measures.[1]

In the pursuit of this agenda, congress included language in the Affordable Care Act which requires the Department of Health and Human Services (HHS) to establish a National Strategy for Quality Improvement in Health Care.[43] By enacting the strategy, HHS seeks to promote rapid and equitable improvement of the healthcare system through the promotion and incentivization of

best practices among consumers and providers.[7] The National Quality Strategy is an evolving set of strategic objectives centered around three aims:[44]

- Better Care: Improve the overall quality, by making health care more patient-centered, reliable, accessible, and safe.
- Healthy People/Healthy Communities: Improve the health of the U.S. population by supporting proven interventions to address behavioral, social and, environmental determinants of health in addition to delivering higher-quality care.
- Affordable Care: Reduce the cost of quality health care for individuals, families, employers, and government.

These aims are enacted through nine policy levers utilized by federal and state entities.[7, 44]

1. Payment: Reward and incentivize providers to deliver high-quality, patient-centered care
2. Public Reporting: Compare treatment results, costs and patient experience for consumers
3. Quality Improvement and Technical Assistance: Foster learning environments that offer training, resources, tools, and guidance to help organizations achieve quality improvement goals
4. Certification, Accreditation, and Regulation: Adopt or adhere to approaches to meet safety and quality standards
5. Consumer Incentives and Benefit Designs: Help consumers adopt healthy behaviors and make informed decisions
6. Measurement of Care Processes and Outcomes: Monitor care delivery and outcomes using nationally consistent measures

7. Health Information Technology: Improve communication, transparency, and efficiency for better coordinated health and health care
8. Evaluation and Feedback: Provide performance feedback to plans and providers to improve care
9. Innovation and Diffusion: Foster innovation in health care quality improvement, and facilitate rapid adoption within and across organizations and communities
10. Workforce Development: Invest in people to prepare the next generation of health care professionals and support lifelong learning for providers

2.4 Measuring and Reporting Quality

Implementing the National Quality Strategy requires the collection and reporting of measures of performance by providers to regulatory, accreditation, and payer organizations.[4, 7, 8] The purpose of the collection and reporting of these measures is to provide valuable information to payers, providers, and regulators about the performance of providers and systems of providers.[4, 44, 45]

2.4.1 *Centers for Medicare and Medicaid Services*

By statute, The Centers for Medicare and Medicaid Services (CMS) requires entities participating in Medicare or Medicaid programs to meet quality assurance reporting requirements. These include, but are not limited to, the Hospital Inpatient Quality Reporting Program, the Hospital Outpatient Quality Reporting Program, and the Quality Payment Program (for clinicians participating in Medicare Part B).[46, 47] Historically, these programs existed as part of quality

assurance of the Medicare and Medicaid programs to prevent fraud and waste [6, 48] and CMS is currently working to align these statutory requirements to the National Quality Strategy.[49]

2.4.2 *Joint Commission*

Closely tied with hospital quality reporting programs is the Joint Commission's (JC) accreditation process. Accreditation (for hospital providers) and certification (for individual departments) by the JC or another approved agency is a requirement for participation in Medicare.[50] The JC, still the accreditor of the majority of hospitals, requires providers to report on a variety of measures as part of the accreditation and certification processes.[51] Their quality improvement initiative, ORYX, aims to closely tie accreditation with their measure reporting requirements and maintains Medicare aligned measures as well as independently designed measures for use by those seeking JC accreditation or certification.[52] Providers are required to select which measures they will report from among designated mandatory and sets of selected measures.[51] While in the past under ORYX the choice of measures and their design was highly flexible, the ORYX measures have become more standardized around those measures that have evidence of improving quality when set to minimum requirements to be used in accreditation or certification.[52]

2.4.3 *National Committee for Quality Assurance*

The National Committee for Quality Assurance (NCQA) collects performance data for use in comparing health plans and systems through its Healthcare Effectiveness Data and Information Set (HEDIS).[53] HEDIS measures are reported by providers within a health group or who are contracted to a health plan that is participating in contributing data to the HEDIS. The resulting aggregate data are sold to organizations (e.g. a business) looking to select plans for their

constituency (e.g. their employees).[53] While generated and reported by providers, HEDIS measures are designed to create a fair comparison of the performance and quality of payer networks and systems rather than to measure the quality of distinct providers within that network or system.[54]

Table 2.2. Purpose of Quality Measures by Agency

Institution	CMS[46, 55, 56]	JC[52]	NCQA[57]
Purpose of Measure	Quality Assurance, Fraud Prevention, Comparative Effectiveness Research	Standardization, Quality Assurance	Selection of Payers and Healthcare Systems By Third Parties
Level of Comparison	Across Hospitals and Clinicians (Outpatient)	Across Hospitals And Departments	Across Hospital Systems and Payers

2.5 Linking Reporting and Compensation

These reporting requirements are tightly linked to compensation and accreditation processes.[16] Across their Quality Reporting Programs, CMS adjusts compensation based on the performance of participants.[25 26] These adjustments take place in the Value Based Purchasing, Medicare Advantage Quality Bonus Payments, Hospital-Acquired Condition Reduction Program, Hospital Readmissions Reduction Program, and the Quality Payment Program.[25-27] While each of these programs have their own rules, full reporting and high performance are required for full compensation. Failure to perform results in reduced compensation, and failure to report can result in reduced compensation or non-participation in the Medicare program.[17 19 25 26 28]

The JC requires annual reporting on measures to the ORYX system as part of accreditation and recertification every 3 years.[21] Providers that are not accredited by the JC or another accreditation agency cannot participate in Medicare and may be at risk of losing their state license.[21]

The NCQA operates the HEDIS with the intent of maintaining a set of measures that are comparable across health plans.[29] CMS, states, and private payers participate in quality reporting to the NCQA as part of the HEDIS.[29 30] Payers, in turn, require their contracted providers to collect and report measures as part of the terms of their contracts where failure to report and failure to perform may constitute a breach of contract between the payer and provider.[31]

2.6 The Burden of Measuring and Reporting on Quality

2.6.1 *Costs*

Reports on the perceptions of hospital leadership and clinical staff suggest that the current measures fail to serve the pursuit of quality and detract from their ability to dedicate resources to process improvement.[4, 9-12] For a hospital to establish the policies and systems required to report on a single measure can take a year of effort and cost over a million dollars.[58] Once established, these measures require continued investment in the form of technology and personel. Massachusetts General Hospital spends 1% of its net patient revenue on the infrastructure necessary to report on 120 established measures.[11] At the same time, providers are faced with increasing costs of operations and reduced payments for services.[59] In spite of these costs, reimbursement systems can effectively penalize providers for attempting to implement programs to improve and measure quality by failing to compensate the preliminary work necessary to establish high quality measures.[4] These factors place strong limits on the capacity of providers to participate in quality reporting systems, with the consensus being that the current approach is unsustainable.[4, 59]

2.6.2 *Misuse*

Despite the expense of reporting quality measures, a majority of providers believe that these measures and their reporting, may not actually improve the quality of care.[13] Providers cite concerns that the measures themselves do not properly measure the intricacies of clinical processes and decision making, that they are subject to manipulation or gaming, and that their reporting and public consumption is more associated with market competition between providers rather than quality improvement.[13-15]

Because reporting requirements are often closely tied with compensation systems, providers are faced with the challenge of prioritizing these measures over other quality improvement priorities.[12] This has resulted in a bias towards meeting external reporting requirements rather than prioritizing quality improvement needs identified by clinicians and staff.[11]

While improving the performance of reportable measures has been associated with measurable gains in inpatient safety[60] there is strong debate as to the value of doing so in primary care where processes have necessarily high variability and patient-driven idiosyncrasy.[8, 11] Reported measures fail to account for when variations in care indicate sound decision making by clinicians, or for socio-economic features of patients that may influence decision making.[8, 61, 62] Where incentives and reporting requirements have not aligned with patient centered or customizable care, clinicians can be forced to either compromise their practice, or face underperformance in reported metrics.[10, 63]

Gaming has been observed in studies of other systems. England's National Health Service was found to have substantial rates of systemic manipulation of statistics where providers found ways to "meet requirements" but "miss the point" by achieving improvements on paper that were

not substantiated in quality research efforts.[10] Evidence of such gaming of reportable statistics also exist in the US system.[63, 64]

Finally, a nationwide study of the effect of Hospital Compare, CMS' public portal of hospital performance in quality and safety measures, found that patients' consumption of care was not affected by the information reported on the website.[65] It is feared that, rather than improving patient awareness and use of care, public reporting is primarily consumed by providers and used to gain competitive advantage.[13, 65]

2.6.3 *Inadequacy*

Even following the near complete deployment of Electronic Health Records and compensation systems for their use, the use of claims data for quality improvement is limiting.[18, 66] Quality improvement requires information about process performance, but such information is rarely part of the health record.[62] Currently, information about process is largely gathered through costly surveys and direct observation.[62] There has been some movement towards establishing effective measures of process around hospital safety, but substantial work is necessary in order to have nationwide standards on relevant to other aims and outcomes.[8]

2.7 Innovations in Quality Improvement in Other Industries

The pursuit of quality and need for safety are not unique to healthcare. Similar approaches relying on the creation of associations and top down regulatory standards were deployed in other industries and also failed to optimize quality when compared to foreign competitors.[16-18]

By the late 70's Japan had begun deploying new forms of quality improvement that provided them an advantage over the US and other countries.[19, 67] As US industries lost market share to

higher quality Japanese techniques, tools, and equipment, US industry leaders began deploying the same and similar models for pursuing quality control and improvement.[16, 18, 19]

In attempting to regain market share, Hewlett Packard conducted internal research that identified a high cost to poor quality. In response they created quality improvement teams at the operating level and trained them to develop pilot programs and test solutions. This resulted in a year-over-year 20% reduction in failure rates.[19] Similarly, Xerox lost 50% of the market to Japanese firms by 1980. In response they set benchmarks against more successful competition and established a companywide culture of prioritizing quality improvement. Quality improvement teams were formed that conducted internal research and changed manufacturing processes, improving quality by 63% and customer satisfaction by 30%. Corning Glass, which already had a long history of exercising similar approaches, aimed to maintain their market share with a renewed commitment to producing quality for customers. Quality improvement teams applied updated techniques to redefine tolerances in production and were able to reduce the costs of manufacture, reducing their break even point from 63% to 55% of sales price.[19]

The models used by these companies have similar features. They depend on building in depth knowledge about processes, modifying the culture of the organization to prioritize scientific approaches to decision making and collaborative approaches to pursuing quality, and using scientific methods of analysis and intervention to incrementally improve performance.[68] In these approaches, knowledge about process is actively gathered from those directly involved with providing and making decisions about a service.[5, 17, 19] Individuals or teams with an in depth knowledge of process are given the freedom to experiment and collaborate which is paired with the responsibility and authority to choose areas to improve and how to measure them.[17, 19] This

activity is established by leadership as the continuous state of the organization so that incremental and continual quality improvement becomes the mission of the organization.[17-19]

2.7.1 *The Role of Variation in Measuring Quality*

2.7.1.1 Common VS Special Variation

Advocates of these approaches stress reliance on stochastic methods to measuring quality.[17, 20, 68] Any measure of performance must capture both the mean performance and the variation from that mean over time in order to successfully describe the performance of the system.[17, 20]

A process is expected to feature random variances in performance that are part of the system as designed.[17, 20] These could be the result of the quality of the equipment, the limits of human performance, or the quality of the inputs to the process.[17] If these structural features are not changed, then the process would be expected to perform similarly, centered around some mean within some range of precision.[17, 68] By identifying this normal variation, deviations from it, or "special variation" can be further identified and singled out as a special event requiring different treatment.[17, 20]

Most important is that the treatment of these two types of variation are different. [17, 20] Special variation is generally a one time or very rare event. [17, 20] It may be an emergent change in fundamental features of the system or environment, or may be a onetime occurrence.[17] In such cases, it may be that no particular response is appropriate, or a sudden major change to the system is required if it is to be stabilized and not collapse.[17] Common variation, on the other hand, is a feature of the system and environment as they are.[17] Understanding that a failure in quality is an expected feature (common variance) it becomes clear that the failure is not the responsibility of any one person, device, or event, and so blaming any individual entity is not an appropriate solution.[17, 18, 68] Common variation can only be addressed by changes to the design

and features of the system.[16, 17] Faced with common variation the correct response becomes to discover and implement effective improvements to the system.[17, 18, 68]

2.7.1.2 The Importance of Targeting Variation

A measure's performance can be improved by targeting precision or by targeting mean performance.[17, 20, 22] These approaches tend to favor improving precision over mean performance.[17, 20, 22] One reason these approaches favor precision is that a highly precise result, even if not of a desirable mean, is predictable, and so generally more manageable.[17, 20, 22] This increase in manageability through stability, even if underperforming, allows other actors reliant on that process to compensate, and so still effectively manage their own quality.[16, 17, 21] A classic example is the superior utility of an inaccurate but precise train (e.g. exactly 20 minutes late every day) over an accurate but imprecise train (e.g. arrives at 2pm, plus or minus one hour).[17] Another reason to favor reducing variance over improving mean performance is that the sources of variance in a measure are often inputs or features that are themselves variant.[16, 17, 21] The variance of a measure's performance can be reflective of the precision of its components, providing insight into the process that is not available when considering mean performance alone.[16, 17, 20, 21]

2.8 Measuring Variation In Healthcare

We find a strong relationship between variations in performance and worse health outcomes [4, 9, 18] and increased costs of healthcare.[69, 70] Meta-analyses of sources of variation in care suggest that differences between patients are the major source of variation in outcomes, seconded by differences between clinicians.[23, 24] While cited as the dominant source of variation, differences between patients are not under the direct control of providers, and, even where they can be managed, are likely best managed at the clinician level.[8, 24] Wennberg's research into regional

variations in healthcare found strong positive associations between regional variations in provider performance and the costs of care.[23, 24] This associations was primarily driven by differences in availability of care and the practice styles of clinicians and persisted when controlling for differences between communities and patients.[70] Studies of the results of variance at the clinician level have found it associated with adverse changes in performance of care processes[25, 71, 72], costs[73], harms[73], and patient satisfaction[24, 74] suggesting that efforts to identify and reduce variance at this level may be highly effective in improving the quality of care.

We have examples of providers successfully using measures of clinician-level variation to improve quality. For over 25 years Intermountain Healthcare has pursued internal process improvement efforts focused on identifying and reducing inappropriate clinician level variance.[25] In their "shared baseline" approach, deviations from an established clinical guideline are measured and the variance reported to clinicians. The effect of this is an informative feedback loop that prompts clinicians to identify and reduce unwarranted deviations from the guideline in their practice. The result has been substantial cost savings and improved outcomes that have made this approach a standard part of their quality improvement efforts.[25]

In a similar effort, Yale New Haven focused on reducing provider variation in physician medication orders.[26] Their work focused on aligning the language and communication frameworks of clinicians and finance specialists so that clinicians would receive meaningful information of the degree and effects of variation in their practices.[26] This led to an increase in correct use of medications and a reduction in costs, improving outcomes and saving money for the provider.[26]

2.9 Exploring the Measure – Variance Relationship

Based on its association with quality, and as exemplified in the success of Intermountain Healthcare, the pursuit of quality in healthcare may be better served if providers focus on reducing clinician level variation. However, providers are faced with limited resources and mandatory reporting requirements of measures that do not take this variation into account.[73] By studying the relationship between performance on reportable measures and clinician level variation, it may be possible to identify reportable measures that are more likely to improve outcomes and reduce costs. More broadly, applications of such research may improve the design and implementation of quality measures by providers, payers, and regulators.

Chapter 3. METHODS

We hypothesized that higher performing processes will exhibit reduced inter-provider variation, and so an inverse relationship exist between inter-provider variance and superior performance of a measure. By testing this hypothesis, we hoped to provide proof-of-concept and identify if provider-level variation of a reportable measure is a topic meriting further research.

3.1 Framework of A Quality Measure

In order to be broadly useful and achieve external validity, a measure and its method of construction should meet industrial criteria. Multiple frameworks exist for identifying the appropriateness of a measure for use in comparing providers and reporting to regulatory and quality measuring agencies. In 2004 the AHRQ analyzed measure evaluation frameworks used by the NQF, AHRQ, National Healthcare Quality Report (NHQR), JC, and the NCQA in order to develop a synthesized set of evaluation criteria usable by the larger hospital quality

community.[27] This unified framework evaluates a measure along 4 domains: importance, scientific acceptability, usability, and feasibility, for its use in reporting across hospitals.

Importance captures the degree to which improvement of the measure is expected to positively affect the quality of care and is under the control of the provider. Scientific acceptability captures the degree to which the belief of the measure's relationship to quality is evidence based, and the precision with which the measure and its expected effect are defined and observable. Usability captures the degree to which the measure's performance can be analyzed and communicated for decision making purposes. Finally, feasibility captures the degree to which the measure can be reliably implemented and maintained over time. Each of these domains have subpoints that further define them and reflect the source frameworks.

Table 3.1. AHRQ Integrated Framework of Measure Quality Demonstrating Combined Features of the AHRQ, NQF, JC, and NCQA Frameworks for Quality of Reportable Measures for Use by Their Organizations

Importance	• Assesses an important leverage point for improving quality; significant to target audiences; impact on health • Opportunity for improvement, considerable variation in quality of care exist. • Aspect of quality is under provider or health system control • Should not create incentives or rewards to improve without truly improving quality of care
Scientific Acceptability	• Relationship to quality is based on scientific evidence • Well defined and precisely specified • Valid, measures the intended aspect of quality; accurately represents the concept being evaluated; data sources are comparable • Adequate proportion of total variation is explained by provider performance and amount of variation in measurement is small after provider performance and patient characteristics are taken into account • Reliable, producing the same results a high proportion of time in the same population

	• Precise, adequately discriminating between real differences in provider performance and reasonable sample size exists to detect actual differences; captures all possible cases and bias related to case exclusion or limited data are minimal. • Risk adjustment is adequate to address confounding bias
Usability	• Effective (understandable and clear) presentation and dissemination strategies exist • Statistical testing can be applied to communicate when differences in performance levels are greater than would be expected by chance • Has been used effectively in the past and/or have high potential for working well with other indicators currently in use • • Compelling content for stakeholder decision-making
Feasibility	• Consistent construction and assessment of the measure • Feasible to calculate; benefits exceed financial and administrative burden of implementation • Confidentiality concerns are addressed • • Audit strategy can be implemented, quality of data are known

Fraser RD. Guidance for using the AHRQ Quality Indicators for hospital-level public reporting or payment. In: Agency for Healthcare Research and Quality, editor. Rockville, MD: Agency for Healthcare Research and Quality,; 2004.

This study attempted to address features of the Importance and Scientific Acceptability domains of the AHRQ's unified framework for our chosen measure. We attempted to meet features of the Importance domain through the selection of an NQF approved measure and with consultation from hospital leadership in which measure to select. Some Scientific Acceptability domain features were addressed in the planning of the study, and others were the subject of the analysis. In the background and planning of the study we have attempted to address our approach's 'relationship to quality', 'precise definition', and 'validity'

We intended to assess the 'proportion of total variation explained by provider performance', 'reliability', and 'precision' by testing our hypothesis. This was done using the stochastic models outlined below.

In the context of the AHRQ's framework, 'risk adjustment' refers to hospital level adjustments necessary to improve comparability across the system. This was not appropriate for our study, and instead we modeled visit level adjustments to improve comparability across providers.

3.2 Defining a Measure

We worked with hospital leadership to identify the measure of blood pressure control status of diabetic II patients within the hospital as an appropriate measure. Based on the NQF measure 0061 "Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg)", control of blood pressure in the diabetes I and II population has been identified by the federal government as an important outcome for reducing the rate of cerebrovascular incidents and heart disease in this population.[28, 29] This has prompted its inclusion in multiple regulatory programs. Because of this, HMC is required to report measure as part of participation in federal and private payer programs.

This measure was identified as meeting multiple criteria of the importance domain.

- 'Assesses an important leverage point for improving quality' – the measure is an active priority for the hospital and used in current quality improvement efforts
- 'Opportunity for improvement' – the hospitals has performed better in the past and monthly performance is various
- 'Aspect of quality is under provider or health system control' – The measure's improvement is the responsibility of the provider and clinicians, and is closely connected to medication and case management.[28, 75]

For our study, the measure of 'Diabetics under control' was defined with a numerator of all outpatient visits with a diagnosis of diabetes Type I or II whose systolic blood pressure reading

was below 140 mm/Hg and diastolic reading below 90 mm/Hg and a denominator of all diabetic outpatient visits, without exclusions.

We were advised by hospital leadership that attribution of clinicians to inpatient visits would not be achievable. Once admitted, the patient record is assigned a single clinician. However, during their stay, multiple clinicians may rotate into the patient's care team. Lack of documentation on these clinicians would make it impossible to accurately associate a measure with a specific clinician. Because of this we limited our study to the outpatient population, where attribution was feasible.

3.3 SETTING

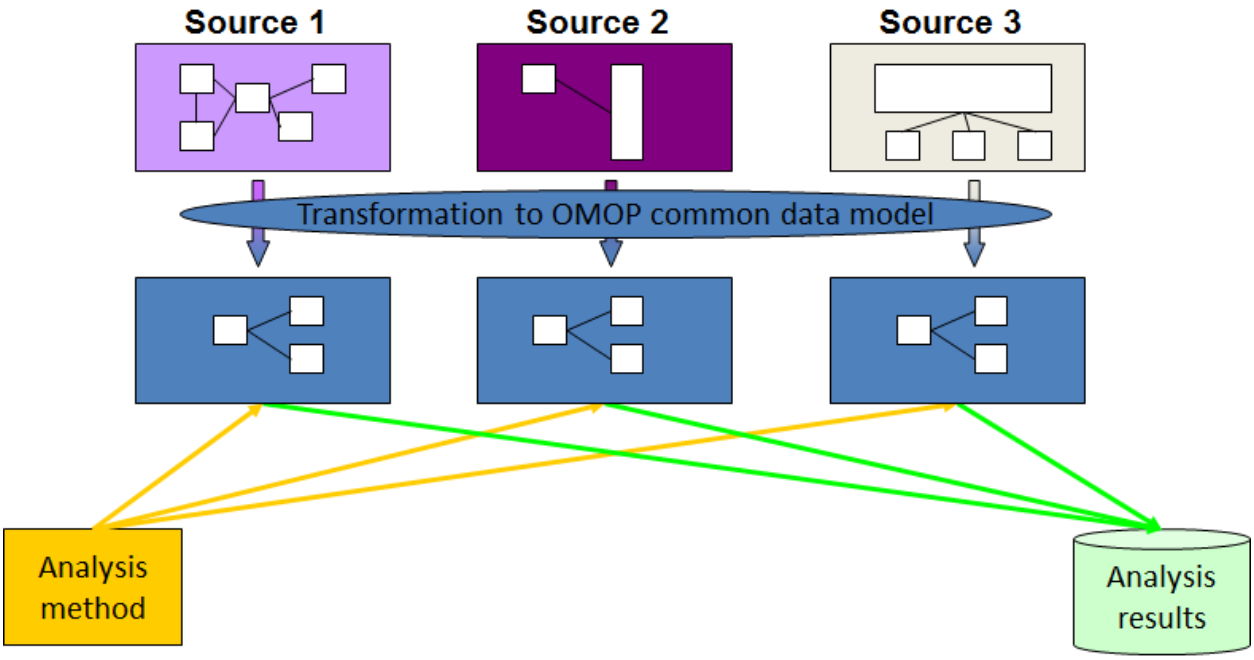
3.3.1 *Harborview Medical Center*

Harborview Medical Center (HMC) is a 413 bed state hospital and the only designated Level 1 trauma center in Washington.[30] As part of its mission to provide comprehensive care to King County, HMC maintains a broad spectrum of outpatient primary and specialty care services providing 258,666 outpatient visits in 2017.[76] The hospital is owned by King County and managed by the University of Washington, who operates the information technology infrastructure used by the hospital.[30]

3.3.2 *Clinical Data Warehouse*

Outpatient records from HMC are warehoused within a Microsoft Amalga powered Clinical Data Warehouse (CDW).[77] The CDW lacks a unified schema and similar data are fragmented across multiple tables and databases. Because of this, querying data from the Amalga CDW would have required substantial time and expertise.

At the time of this study, there was an effort headed by Dr. Adam Wilcox to migrate the data warehouse to a schema that is compliant with the Observational Health Data Sciences and Informatics (OHDSI) Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). The OMOP model provides ease of accessibility and generalizability of data access and manipulation through alignment of clinical observations to a standardized schema and nomenclatures.[78] By aligning data from multiple departments into a unified and publicly defined schema, the location of essential information could be identified and queried within the limited scope of our research.



Observational Health data Sciences and Informatics. OMOP Common Data Model. 2018.

Figure 3.1. Diagram of An Implementation of the OMOP Common Data Model Showing Efficiency Gained in Analysis by Aligning Multiple Data sources to the Same Data Model

3.4 POPULATION

The study population included outpatients with a diagnosis of diabetic I or II prior to or during the study period of January 1st, 2011 to December 31st, 2017. Observations were of mean per-visit blood pressure readings. Observations were collected for each visit where blood pressure was taken. Only visits where blood pressure was measured were considered.

3.4.1 *Effect of Meaningful Use Phase 3*

From 2013 to 2015 the proportion of cases to the outpatient population doubled from 8% to 17% with most of this increase occurring in 2014. This increase corresponded with the start of meaningful use phase 3 which linked collection of our measure to federal compensation and reporting requirements.[33] It was hypothesized that this increase in population represents additional collection efforts of clinicians and departments in order to bring the hospital into compliance with the new law.

3.4.2 *Seasonal Change in Blood Pressure*

Studies of blood pressure readings have found that an individual's mean blood pressure rises and falls by 5 mm/Hg throughout the year.[31, 32] This change is physiological and occurs with changes in ambient temperature, rising in the colder months and lowering in the warmer months.[31] We expected to see this reflected in our patient level data.

This physiological change was not controlled for in our measure definition nor its implementation. Because of this, we expected that the performance of our measure may rise and fall seasonally as patients within 5 mm/Hg of being in control or out of control changed status.

3.5 Data Acquisition

Individual visits were queried from the UW ITS OMOP CDM compliant database. The variables collected were the visit ID, visit year, visit month, systolic blood pressure, diastolic blood pressure, patient ID, and provider ID. For visits where multiple blood pressure readings were taken, the mean of these measures was used. Visits without a provider ID, with measures more than 3 standard deviations from the mean, or where the diastolic and systolic were less than 20 points apart were excluded. An indicator variable of 1 if meeting our definition of under control was then assigned to each visit, and an aggregate monthly measure of the proportion of visits under control per month of the study was calculated. This monthly calculation of the proportion of patients whose blood pressure was under control constituted an approximation of the measure traditionally reported by the hospital.

3.6 Estimation of Variation

For each month, patient visits were stratified by clinician, and the proportion of patients under control calculated per clinician per month. It was identified that within a given month, many clinicians may see 5 or fewer cases, while some may see over 20. Traditional variance would weight these clinicians evenly and so an alternative to traditional variance was sought. Following statistical consult with the University of Washington Department of Biostatistics we decided to substitute the over-dispersion parameter of the Quasi-Poisson regression in place of traditional variance. This parameter was used in order to achieve a weighted estimate of variation that would be less affected by providers with a smaller numbers of cases within the month.

The Poisson regression is a common regression for modeling count data. However, the Poisson regression assumes that the mean rate of events and the variance from that mean are equal.

As an alternative to the traditional Poisson regression, the Quasi-Poisson regression is similar but relaxes the mean-variance assumption by allowing the variance to differ from the mean. Our Quasi-Poisson regression followed the model:

$$\log\left(\frac{E[control_t]}{population_t}\right) = \theta$$

Where control is the number of patients under control, population is the total number seen, and t is the t^{th} provider.

Within the Quasi-Poisson regression variance from the mean is a linier function. The degree to which the variance differs from the mean is reported as the overdispersion parameter.[34] The variance model follows:

$$Var[control_t] = \phi E[control_t]$$

Where ϕ is the overdispersion parameter.

3.7 Visualization of Mean-Variance Relationship

Initial analysis included visualization of the measure-variance relationship on a time series chart. Qualitative observations were made annotating key features and the degree to which a measure-variance relationship was observable. These observations explored hypothesized relationships between features of our data and inter-clinician variation. If our hypothesis were correct, we would generally have expected to see inverse movements of or performance measure and inter-provider variation. However, because our seasonal change in blood pressure is environmental and universal to the patient population, we would not have expected provider performance to vary seasonally.

3.8 Statistical Modeling of Mean-Variance Relationship

To satisfy the Scientific Acceptability domain, stochastic time series analyses was applied using auto regression (AR) $[a]$, moving average (MA) $[b]$, and combined ARMA$[c]$ models, following by Akaike information criterion (AIC)[35] testing for best fit. General forms of the models used are:

$$[a]\ X_t = c + \sum_{i=1}^{p} \varphi_i X_t - i + e_t$$

$$[b]\ X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

$$[c]\ X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Where t is time, φ are autocorrelations in our models, and θ are moving averages in our models.

Exploratory analysis was conducted using visualizations of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). These were used to identify potential periods of autocorrelation that can be controlled for using auto regression and moving average parameters. Multiple models were tested for best fit, including controlling for seasonal variation, annual correlation, and our structural change in 2014.

3.9 Approval and Privacy

Expedited approval for this study was obtained from the University of Washington Human Subjects Division Institutional Review Board. Analysis of patient and provider identifiers was performed within the UW Information Technology Services servers and accessed through the

hospital system's secure virtual private network. Data exported for consultation with biostatisticians was de-identified according to Section 164.514(a) of the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.

Chapter 4. RESULTS

4.1 POPULATION

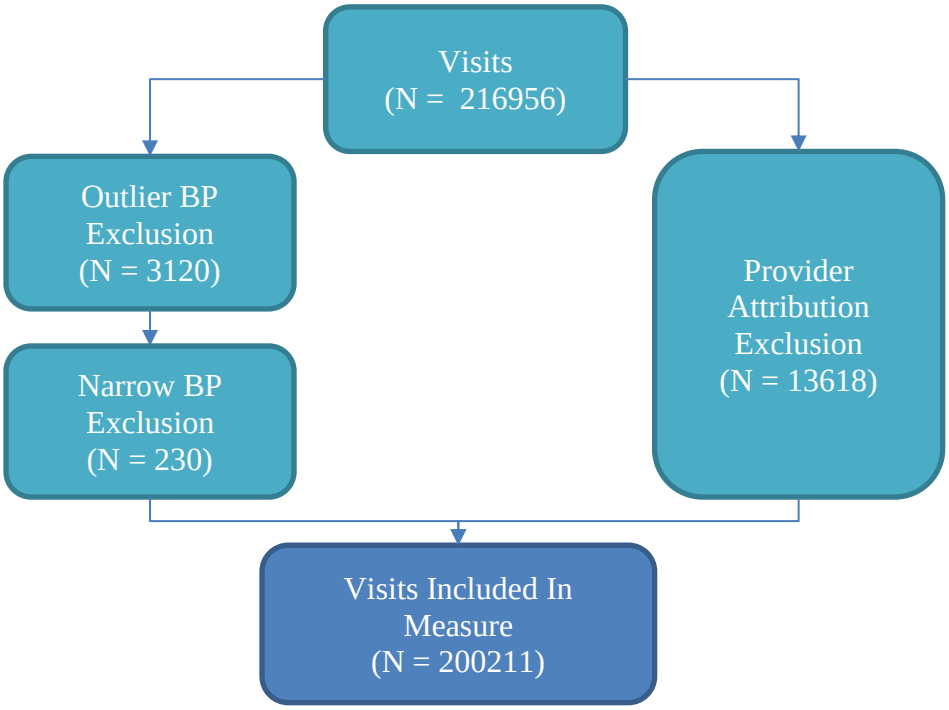


Figure 4.1. Chart of Study Population and Exclusions by Type

A retrospective analysis of the relationship between performance of a healthcare measure and inter-clinician variation of the measure was performed using data available within the University of Washington Medical Center Clinical Data Warehouse. Data were retrospectively collected on 216,956 diabetic outpatient visits to Harborview Medical Center from January 2011 through December 2017. Data was limited to those visits where a complete blood pressure reading was on

record. For visits where more than one blood pressure was found, the average of the systolic and diastolic values was recorded.

4.1.1 *Exclusions*

Application of our exclusion criteria resulted in 3,350 observations being rejected for having either a blood pressure with less than a 20 point difference between systolic and diastolic readings (N = 230) or being greater than 3 standard deviations from the mean systolic (75 - 190 mm/Hg) or diastolic (37 - 112 mm/Hg) reading in our population (N = 3,120). Because our full analysis requires stratification by providers, observations without provider attribution at the visit level were also excluded (N = 13,618). There was overlap between these exclusions, so that only 14,440 total observations were excluded from analysis.

Table 1 shows the numbers of patients, providers, and visits used to build our measure, as well as mean blood pressure readings, and visits rates.

Table 4.1. Numbers and Rates of Visits, Providers, and Patients Across All Years (2011 through 2017)

Subject	Total	Mean (SD)
Visits	200211	
Providers	1736	
Patients	16349	
Visits Per Month		2383 (1142)
Visits Per Patient		12 (20)
Providers Per Month		317 (182)
Monthly Visits Per Provider		8 (10)

4.1.2 *Blood Pressure Readings*

Before excluding outliers and narrow blood pressures, our mean blood pressure reading was 132/75 (SD 19 and 13 mm/Hg.) The range of systolic observations was 5 mm/hg to 448 mm/Hg, and diastolic from 0 mm/Hg to 174 mm/Hg. These ranges include values that are physiologically implausible and should be excluded. We identified two types of exclusions for blood pressure readings, outliers and narrow readings. Outliers were defined as observations more than 3 standard deviations from the mean diastolic and mean systolic readings. Included ranges were 75 to 190 mm/Hg systolic and 37 to 112 mm/Hg diastolic. Narrow blood pressures were defined as any observation where systolic and diastolic readings were within 20 points of each other. After excluding outliers, our mean observed blood pressure readings were 132 systolic and 74 mm/Hg diastolic, with respective ranges of 75-186 mm/Hg and 40-110 mm/Hg, representing more physiologically plausible ranges.

Table 4.2. Demonstrating Effect of Exclusions on Achieving Normal Ranges in Mean Blood Pressures.

Exclusion Type	N	Pre Mean (SD)	Post Mean (SD)	Pre Min - Max	Post Min - Max
Systolic Outlier	1507	132 (19)	132 (18)	5 - 448	75 - 186
Diastolic Outlier	1129	75 (13)	74 (12)	0 - 174	40 - 110

4.2 BUILDING A MEASURE

Our measure of proportion of diabetics whose blood pressure was under control was constructed from our encounter level data and calculated per month. This resulted in 84 monthly time series observations of the number of visits where the patient's blood pressure was under control, the total

number of visits in the month, and the proportion (number under control divided by the total number of visits).

To calculate the inter-provider variation of our measure, creation of the same measure of performance was calculated separately for each provider. This resulted in a set of measures for each month, one for each provider who saw at least one patient that month. The variation on these measures within a given month was estimated using the over-dispersion parameter from a Quasi Poisson regression. This regression used the number under control as our outcome and the number of visits as our predictor. This was performed for each month, resulting in another 84 months of time series observations.

Table 4.3. Mean and Standard Deviation of Performance Measure and Measure of Variation, Demonstrating Expected Ranges of Each.

Measure	Mean (SD)
Performance Measure	0.67 (0.03)
Variation Measure	0.39 (0.05)

4.3 Visual Time Series Analysis

Visual analysis of our performance-variation relationship was conducted in 3 stages. First visualizing the original measure as might be performed within the hospital. Then visualizing our measure of variation. Finally, a visualization combining the two was observed. In testing our hypothesis, we intended to observe 3 features: the inverse relationship between measure performance and variation month to month, the existence of this relationship during a major event where it would be expected (implementation of Meaningful Use Phase 3 (MU3), and the lack of

relationship during events when it would not be expected (population wide seasonal changes in blood pressure.)

Key features were apparent within a visualization of the time series. Each year appeared to feature an increase and then decrease in performance, coinciding with seasonal changes in temperature. Within Year 4 (Months 37 to 48) there appeared to be a decrease in performance coinciding with our expected changes during the enaction of MU3.

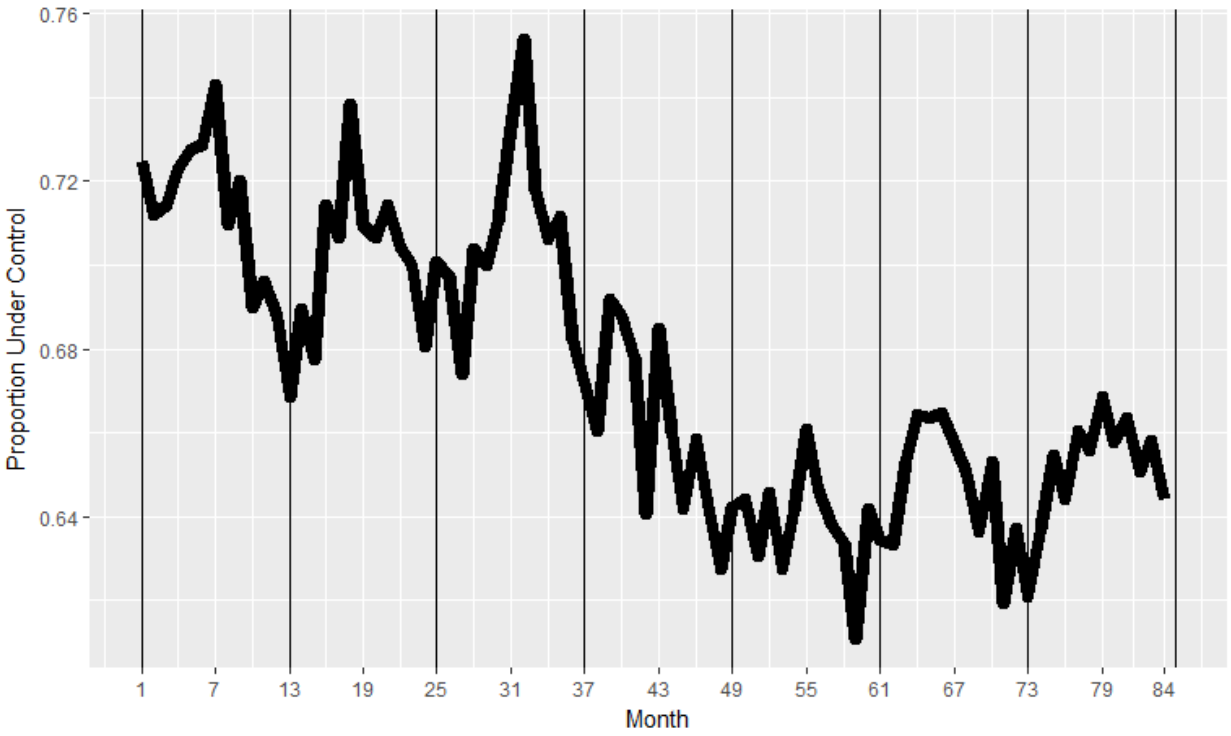


Figure 4.2. Monthly Chart of Measure Performance Demonstrating Obvious Seasonal Trend, And Change in Mean Performance During Year 4 (Month 37 to 48).

Visualizing the inter-clinician variation of our measure we saw a possible increase in variation in Year 4 (Month 37 to 48), coinciding with the implementation of MU3. A seasonal trend was not

readily apparent, suggesting that this trend does not vary between clinicians and is ubiquitous across our patient population.

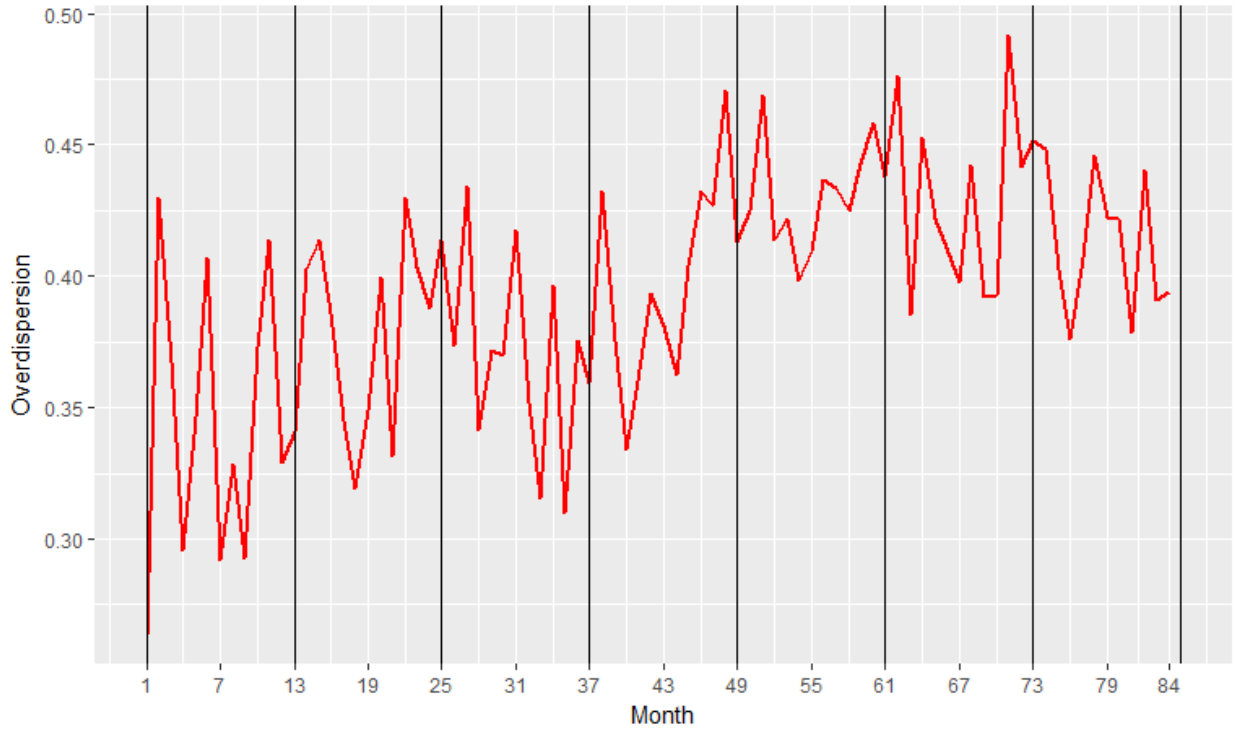


Figure 4.3. Monthly Chart of the Measure of Variation Showing Noisy Process with Non Obvious Seasonal Trend, and Obvious Change in Mean Performance During Year 4 (Months 37 to 48)

We created a single chart demonstrating our performance-variation relationship by centering the Y axis of the overdispersion parameter to the mean performance of our performance measure. The hypothesized inverse performance-variation relationship was apparent in the inverse changes from before to after the implementation of MU3 in Year 4 (Months 37to 48). However, the same relationship was not apparent month to month.

The seasonal change in performance that was apparent in our measure of performance was not apparent in our measure of variation. This aligns with our hypothesis that seasonal effects would not increase inter-clinician variation.

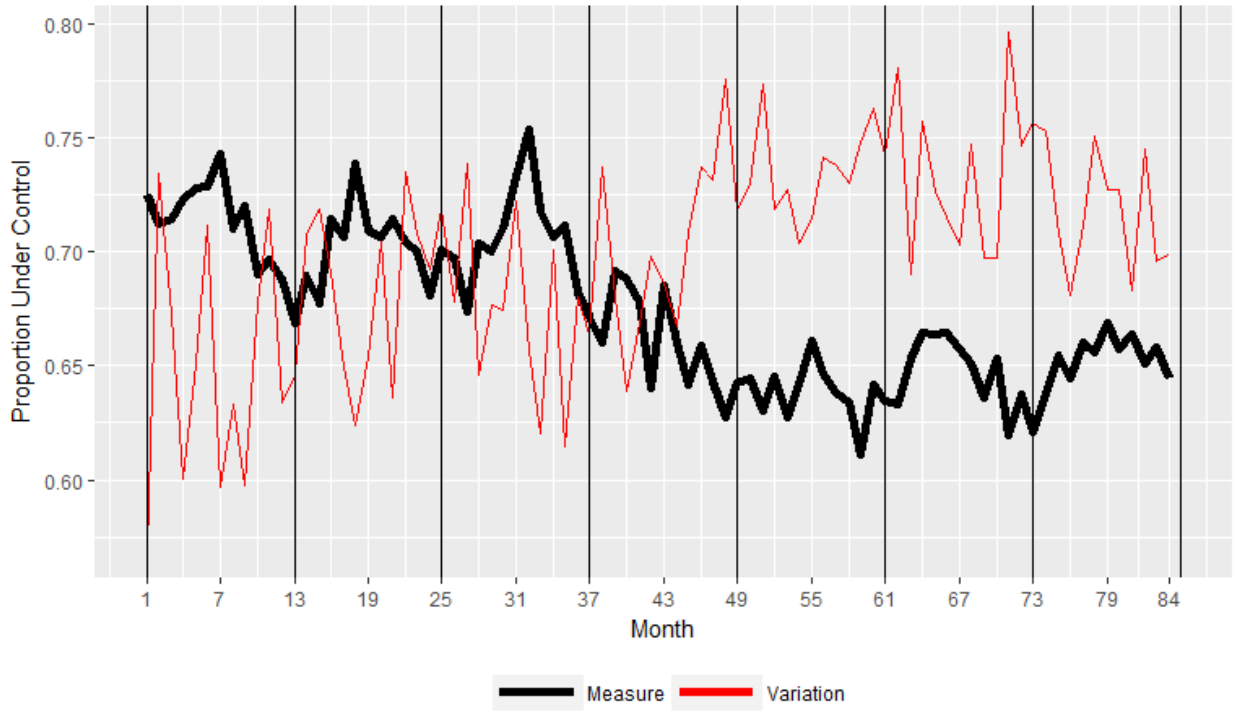


Figure 4.4. Overlay of Measure of Variation on Measure Performance Demonstrating Observable Inverse Relationship in Year 4 (Months 37 to 48) and No Apparent Relationship Between Seasonal Changes in Performance and Inter-Clinician Variation.

4.4 Time Series Analysis

4.4.1 *Parameterizing ARMA Models*

To more accurately measure the relationship between the performance of the measure and inter-clinician variation, we explored and constructed several time series analyses. We first explored the possible usefulness of such models by deconstructing our measure of performance. The time series

deconstruction allows us to visually separate seasonal trends from noise and structural trends in the data. Deconstructing our time series showed expected features of the data, including our sharp decrease in performance in Year 4 (Months 37 to 48) and our possible seasonal trend across all years.

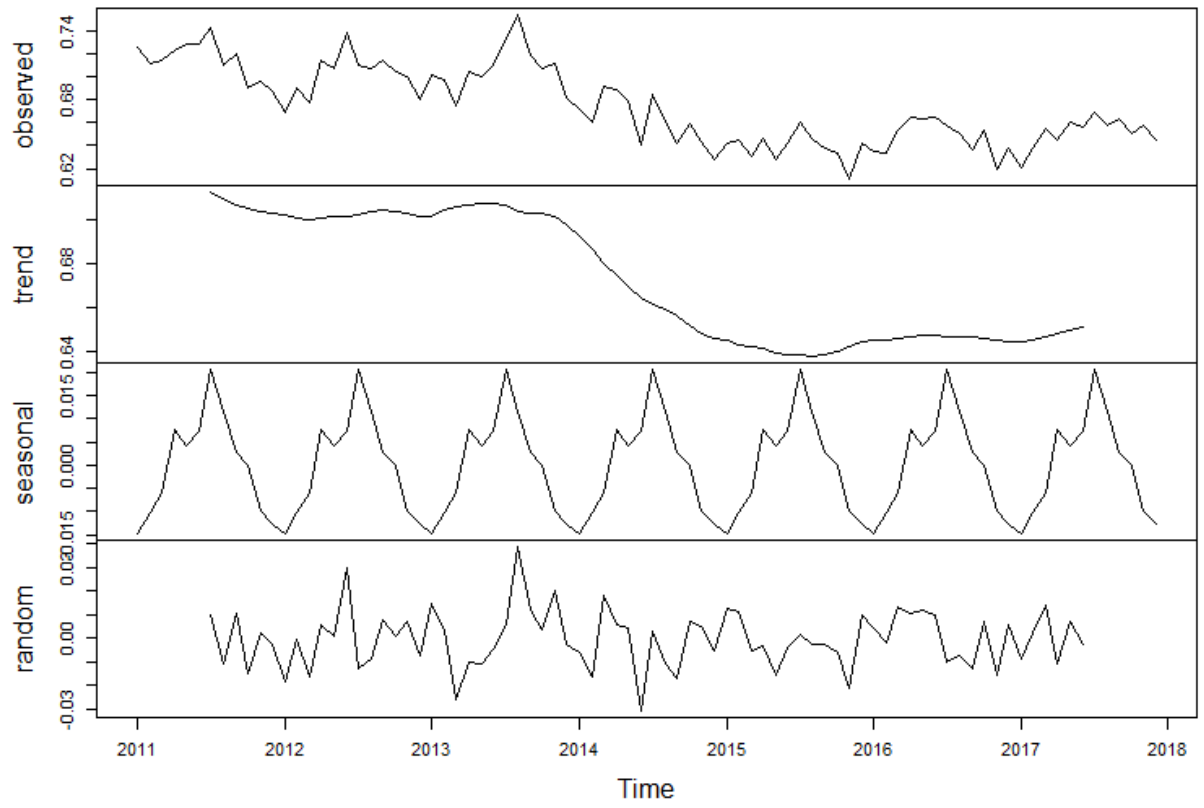


Figure 4.5.Time Series Decomposition Demonstrating Original Time Series, Structural Trend of the Series, Season Pattern of the Series, and Unaccounted for Monthly Changes in the Series Used to Visualize Major Characteristics of the Series.

Demeaning our data at June 2014 (Month 42) resulted in an apparently stationary time series. Stationarity of the demeaned time series was tested using the Phillips-Perron Unit Root Test (P = 0.01) and an Augmented Dickey-Fuller test (P = 0.01). These results suggested that if we

controlled for changes in this period, our data would be appropriate for ARMA modeling. A variable indicating pre-June 2014 (Month 42) status was applied to the time series.

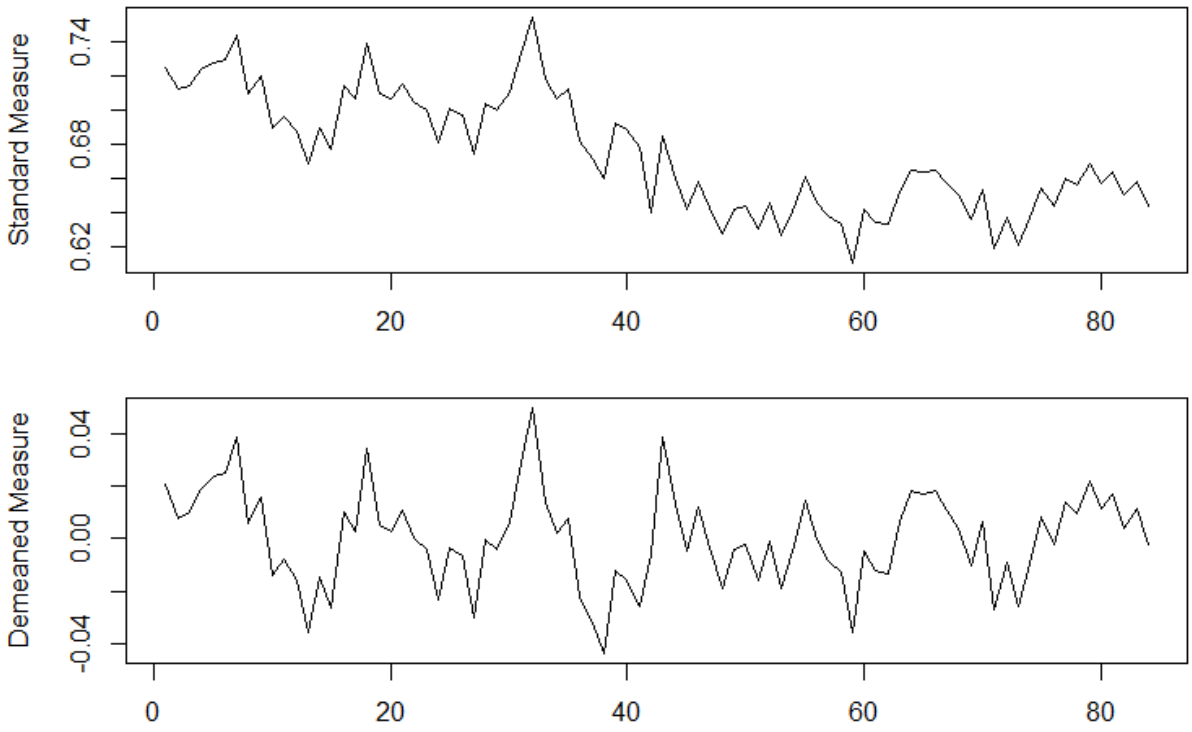


Figure 4.6. Time Series Plots Comparing of Normal and Demeaned Time Series Showing Non-Stationarity of the Original Series and Stationarity of the Demeaned Series.

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots were used to identify characteristics of our data and further parameterize our ARMA models. Our ACF plot showed a gradual decrease in correlation suggesting autocorrelation across the time series. Because this time series was non-stationary, an additional ACF was applied to the demeaned time series. Plotting the ACF of the demeaned time series resulted in an ACF plot reflecting possible auto-correlation during time periods of lag one and two. This suggested that models including auto regressions of lags one or two months should be included in our analysis.

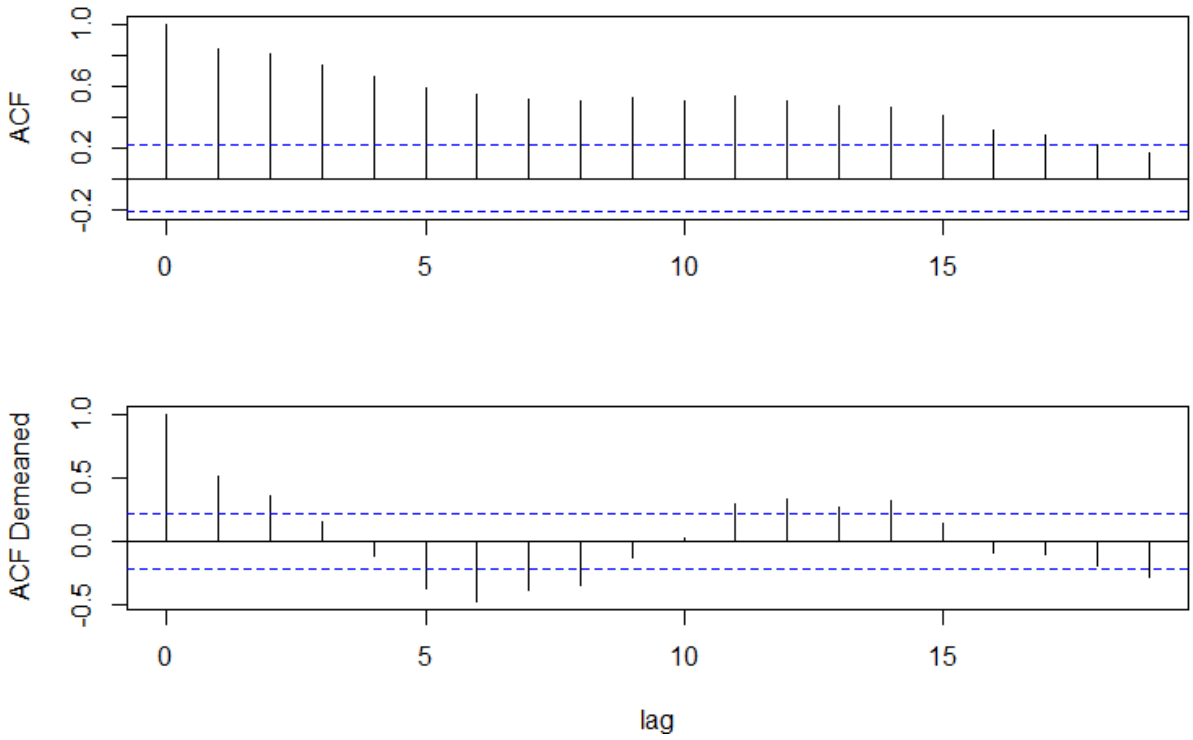


Figure 4.7. Plots of Time Series Autocorrelation Functions Showing Potential Autocorrelation Across Past Two Years in Our Normal Data and Autocorrelation at Lags 1 and 2 Using Data Demeaned to June 2014 (Month 42) That Could Be Adjusted for Using an Auto Regression of 1 or 2 and an Indicator Variable of Pre June 2014 Status in Our Time Series Model.

PACF controls for intervening months and so was applied to the normal time series. This plot showed possible autocorrelation at lags of one and two months. This suggested that models including a moving average of 1 and 2 should be included in our analysis.

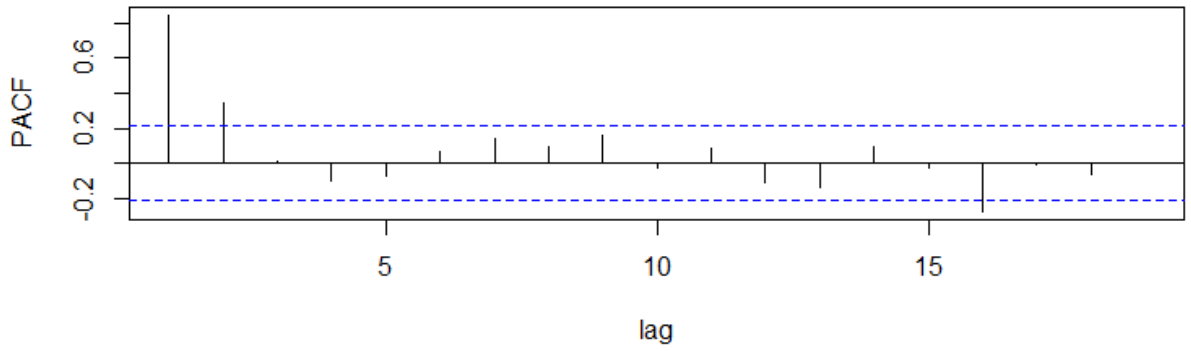


Figure 4.8.Plot of Time Series Partial Autocorrelation Function Showing Possible Autocorrelation in Time Periods 1 and 2 That Could Be Adjusted for Using A Moving Average of 1 or 2 in Our Time Series Model

4.4.2 *ARMA Modelling*

Multiple models were developed to test the relationship between performance of our measure and inter-provider variation. These models included none, one, and two period auto regressions, along with one, and two period moving averages. Additionally, because our measure was affected by seasonal changes in blood pressure, we hypothesized that models including a 12-month seasonal lag should also be included. In all models we controlled for if our data were from before June 2014 (Month 42) or not. AIC was used to identify the best fitting model.

Table 4.4. AR and MA Models Excluding 12 Month Seasonal Lag Demonstrating An Inverse Relationship With High Significance of Inter Clinician Variation in All Models Along With High Significance of Controlling for Meaningful Use Phase 3, High Significance of Controlling for Structural Break, and Similar AIC Scores as Models with Seasonal Lag Including Highest Scoring Model AR(2)

	Proportion under-control				
	AR(0)	AR(1)	AR(2)	MA(1)	MA(2)
	(1)	(2)	(3)	(4)	(5)
ar1		0.439***	0.611***		
		(0.113)	(0.113)		
ar2			0.324**		
			(0.105)		
ma1				0.302***	0.347**
				(0.088)	(0.110)
ma2					0.220*
					(0.097)
intercept	0.746***	0.727***	0.739***	0.734***	0.728***
	(0.022)	(0.018)	(0.026)	(0.020)	(0.019)
Inter-provider variance	−0.236***	−0.184***	−0.140***	−0.205***	−0.189***
	(0.051)	(0.044)	(0.037)	(0.046)	(0.043)
Pre June 2014	0.043***	0.040***	−0.016	0.043***	0.042***
	(0.005)	(0.007)	(0.017)	(0.005)	(0.006)
Observations	84	84	84	84	84
Akaike Inf. Crit.	−437.304	−450.518	−454.154	−445.573	−448.351
Notes:	*p<0.05; **p<0.01; ***p<0.001				

Table 4.5. ARMA Models Excluding A 12 Month Seasonal Lag Demonstrating An Inverse Relationship With High Significance of Inter Clinician Variation in All Models Along With High Significance of Controlling for Meaningful Use Phase 3, But Loss Of Significance of Controlling For Structural Break, and Similar AIC Scores as Models without Seasonal Lag Demonstrating Limited Significance But Slight Improvement in Scores In Most Models By Including Seasonal Lag

	Proportion under-control			
	ARMA(1, 1)	ARMA(2, 1)	ARMA(1,2)	ARMA(2, 2)
	(1)	(2)	(3)	(4)
ar1	0.955***	0.611	0.945***	1.041
	(0.037)	(0.320)	(0.043)	(0.675)
ar2		0.324		−0.093
		(0.300)		(0.657)
ma1	−0.328**	0.000	−0.338**	−0.433
	(0.117)	(0.356)	(0.123)	(0.676)
ma2			0.137	0.174
			(0.127)	(0.288)
intercept	0.740***	0.739***	0.737***	0.737***
	(0.026)	(0.026)	(0.026)	(0.026)
Inter-provider variance	−0.149***	−0.140***	−0.136***	−0.135***
	(0.038)	(0.038)	(0.039)	(0.040)
Pre June 2014	−0.013	−0.016	−0.015	−0.015
	(0.019)	(0.019)	(0.019)	(0.019)
Observations	84	84	84	84
Akaike Inf. Crit.	−453.167	−452.154	−452.262	−450.278
Notes:	*p<0.05; **p<0.01; ***p<0.001			

Table 4.6. AR and MA Models Including A 12 Month Seasonal Lag Demonstrating An Inverse Relationship With High Significance of Inter Clinician Variation in All Models Along With High Significance of Controlling for Meaningful Use Phase 3 and Similar AIC Scores as Models without Seasonal Lag Demonstrating Limited Significance But Slight Improvement in Scores In Most Models By Including Seasonal Lag

	Proportion under-control			
	AR(1) seasonal lag	AR(2) seasonal lag	MA(1) seasonal lag	MA(2) seasonal lag
	(1)	(2)	(3)	(4)
ar1	0.421***	0.377**		
	(0.112)	(0.132)		
ar2		0.217		
		(0.129)		
ma1			0.284**	0.341**
			(0.087)	(0.112)
ma2				0.225*
				(0.093)
sar1	0.211	0.225*	0.217*	0.230*
	(0.109)	(0.109)	(0.110)	(0.108)
intercept	0.729***	0.726***	0.733***	0.729***
	(0.018)	(0.018)	(0.019)	(0.018)
Inter-provider variance	−0.186***	−0.171***	−0.200***	−0.187***
	(0.042)	(0.043)	(0.045)	(0.042)
Pre June 2014	0.040***	0.034**	0.042***	0.041***
	(0.007)	(0.013)	(0.006)	(0.006)
Observations	84	84	84	84
Akaike Inf. Crit.	−452.155	−453.352	−447.343	−450.671
Notes:	*p<0.05; **p<0.01; ***p<0.001			

A significant inverse relationship between the performance of our measure and variation between providers was found in all models (P < 0.001). A 100% change in our measure was associated with a 14% to 24% inverse change in inter-provider variation depending on the model. This effect was attenuated in our lowest (best) scoring model, AR(2), which incorporated only a two period autoregression function. Our indicator of pre-June 2014 status is significant in most of our models, suggesting that its inclusion was important in modelling our time series. Four of our models included a seasonal lag of 12 months (ssr1). Controlling for seasonal lag was statistically insignificant in our AR(1) model, and only slightly significant in our other models.. However, the seasonally adjusted AR(1), MA(1), and MA(2) models outperformed their non-seasonally adjusted

variants providing mixed results to the importance of including seasonal lag in our time series models.

Chapter 5. DISCUSSION

5.1 Demonstration of Inverse Measure-Variation Relationship

Our finding of an inverse relationship between the performance of a measure and inter-clinician variation aligns with what we would expect if performance were tied to inputs and processes at the clinician level.[17, 79] We selected a measure based on the recommendation of hospital leadership. This recommendation was based on the measure's importance to the hospital's current quality improvement efforts and an expectation that the measure may be more likely to demonstrate the hypothesized inverse relationship between performance and inter-provider variation if any such relationship existed. Therefore, it is not entirely surprising that some relationship was found. Our choice of a measure likely to demonstrate the expected relationship provided an opportunity for testing the potential worth of applying this method to other measures in the future. A negative result, despite these ideal circumstances, would have suggested that our general theory was not viable or that our methods were unsound.

5.2 Degree of Measure-Variation Relationship

The relationship between clinician level variation and measure performance was comparable to other studies of sources of variation in outcomes of care. Literature reviews of studies analyzing variation in outcomes or performance have found between 0% and 18% of this variation to be associated with the clinician or care team level, exclusive of patient satisfaction which was observed as high as 83.9%.[23, 24] Our observations, ranging from 14% to 23.6%, extend beyond

this range, but our best fitting model fits within this range at 14%. This suggests that our results may be reasonable. However, the studies reviewed did not look specifically at clinician level variation of a reportable measure and so they may not be comparable.

5.3 VISUAL ANALYSIS

Visualization of measure performance revealed a noisy process. Month to month and year over year changes were not easily interpretable. However, the presence of an obvious drop in performance during Year 4 (Months 37 to 48) presented an opportunity for gross visual analysis of the relationship between performance and inter-clinician variation.

Our data contained a structural break in Year 4, when collection of our measure was linked to compensation under Meaningful Use. This resulted in an obvious change in performance that could be contrasted with a possible change in variation. The first three and last three years of our study do not exhibit this change. When viewing a time series of Years 5 through 7 (Months 49 to 84) no performance-variation relationship is apparent. Without the structural break caused by the introduction of MU3, a visual analysis may have not been compelling in suggesting any relationship or providing actionable information.

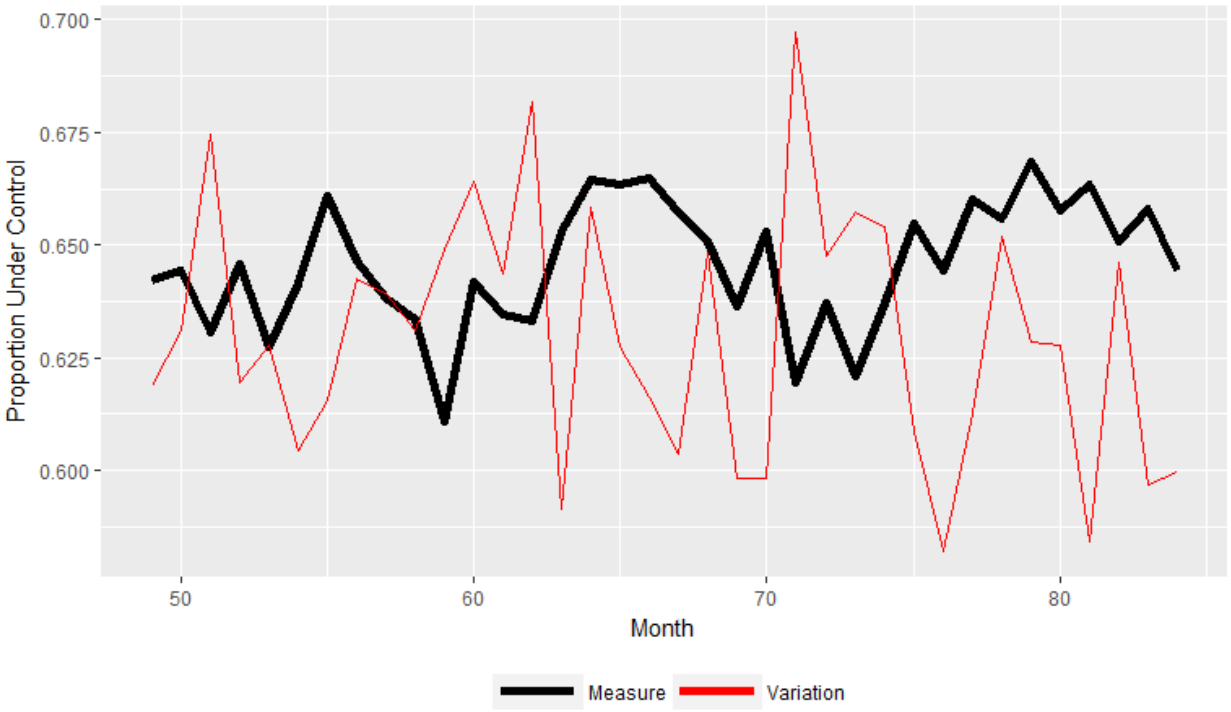


Figure 5.1. Visualization of Measure – Variation Relationship from 2015 to 2017 (Months 49 to 84) Showing Lack of Obvious Inverse Relationship Due to Noise in Variation.

The communication of variation was essential to Intermountain Healthcare's use of inter-clinician variation in order to prompt improvements in care.[25] Similarly, Yale New Haven's work focused on the challenges of communication between internal finance auditors and clinical staff in order to achieve actionability on the audits performed by the hospital's financial department.[26] This suggests that efforts to communicate inter-clinician variation in the selection of measures and quality improvement projects must consider the sensitivity of the method deployed, and that a simple visualization of the performance-variance relationship may be insufficient for detecting or communicating results outside of high magnitude changes in the performance-variation relationship.

5.4 DATA QUALITY

The secondary use of EHR and claims data is a contested topic. Such data is often of questionable quality, and not structured for secondary uses.[8, 80-82] The data are primarily collected for billing and case management, resulting in biases of availability relative to these uses. The availability and quality of primary data collection are also affected by human error, changes in clinical processes and changes in information system design. Outliers, structurally idiosyncratic data, and high rates of missing data were found in our analysis of data quality.

We chose to define outliers at three standard deviations, assuming a normal distribution. This resulted in the exclusion of 1.19% (N = 2384) of our data and more normal distributions of blood pressure. This is more than would be expected by chance (N = 541).

Table 5.1. Calculated Ranges of Systolic and Diastolic Blood Pressures to Three Standard Deviations from the Mean, the Expected Number Based on 3 Standard Deviations of a Normal Distribution of 216956 Observations, and the Number of Diastolic, Systolic, and Combined Observations Actually Excluded Indicating Reasonable but Elevated Data Loss.

Measure	Min	Max	N Predicted	N Observed
Diastolic	37	112		1129
Systolic	75	190		1507
Any			541	2384 (1.19%)

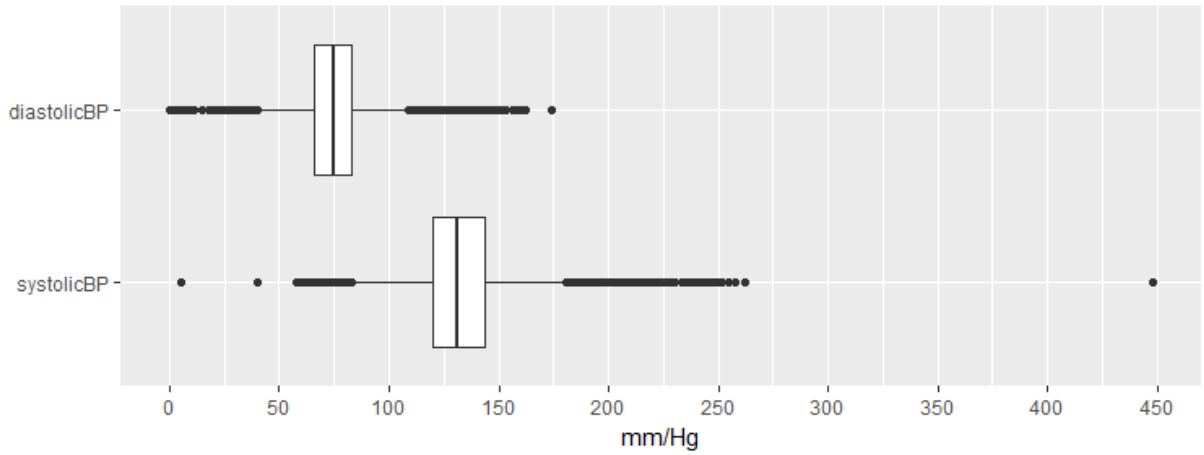


Figure 5.2. Boxplots Depicting Mean (line in box), Interquartile Range (ends of box), 1.5 Times the Interquartile Range (whiskers), and Extreme Points (dots), of Blood Pressure Measures Showing Outliers at 0 mm/Hg and Above 200 mm/Hg.

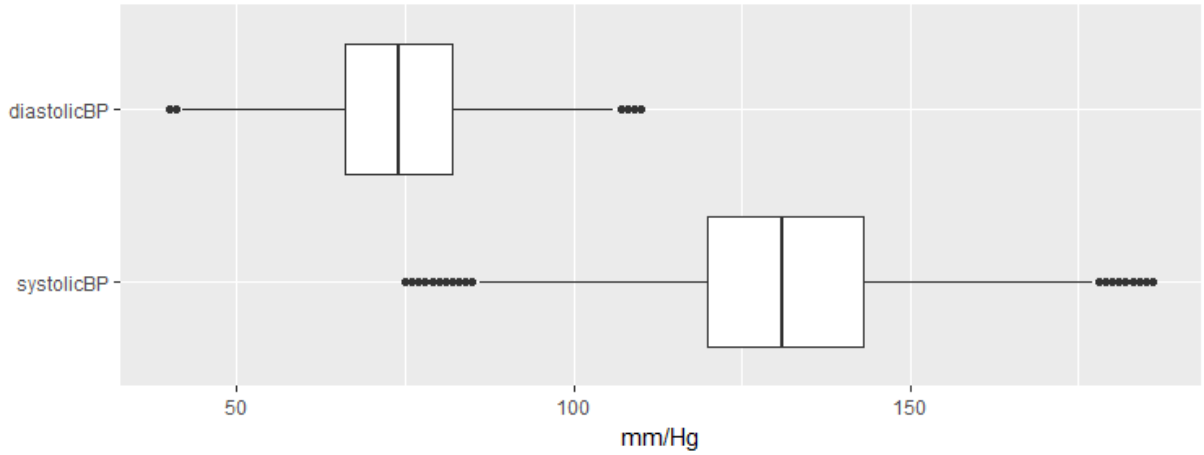


Figure 5.3. Boxplots Depicting Mean (line in box), Interquartile Range (ends of box), 1.5 Times the Interquartile Range (whiskers), and Extreme Points (dots), of Blood Pressure Measures After Excluding Outliers and Narrow Measures Showing Distributions Within Realistic Norms.

Additionally, even excluding these outliers, there was apparent bias in how data were collected. Spikes in data availability were observed on even numbers and at tens values near physiological norms (70 and 80 mm/Hg diastolic, and 130 mm/Hg systolic).

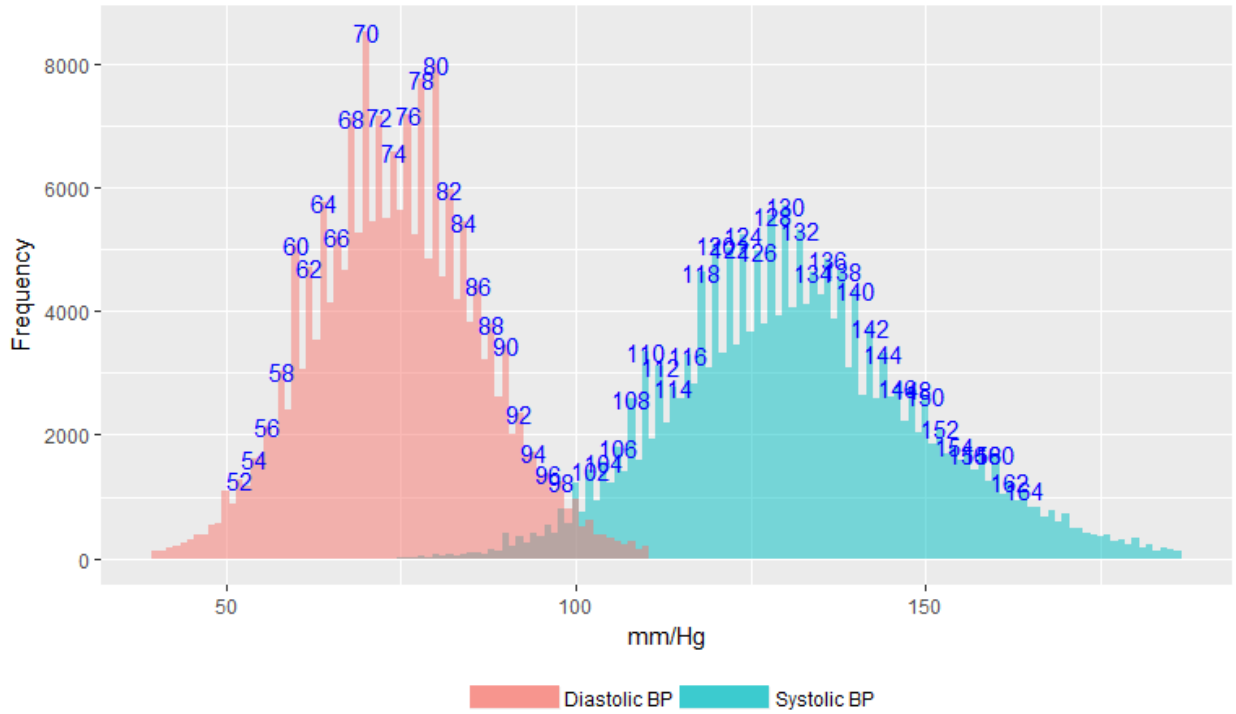


Figure 5.4. Histogram Depicting Distribution of Diastolic (red) and Systolic (blue) Blood Pressure Records From 2011 Through 2017 After Exclusion of Outlier and Narrow Measures Showing Effects of Technological and Human Biases in Data Collection.

It is hypothesized that these are due to technological (in the case of even numbers) and human (in the case of physiologically expected numbers) biases in how data were collected.

Finally, provider attribution was limited. Known challenges in attributing providers among the inpatient population led us to limit our study to the outpatient population. Among outpatients, provider attribution was still not perfect, resulting in 13,618 (6.3%) of our data being excluded from analysis. We cannot assume that this missingness is unbiased, as it may be due to technological characteristics of certain offices, or behaviors of certain providers.

The smaller patient volumes used when calculating at the clinician level may make these biases uniquely important compared to the same measure at the provider/facility level. Our findings suggest that further work may be necessary in the area of improving and validating the quality of EHR data for use in reporting clinician-level variation.

5.5 Seasonal Variation

We observed seasonal variation in the blood pressures recorded by our patient population. This is an expected phenomenon documented in the literature.[31, 32] However, this feature was not identified in our planning meetings with hospital leadership. We also did not find mention or accounting of this phenomenon in our review of literature regarding the use of this reportable measure in quality improvement.[28, 33] This suggests that the measure as currently deployed does not account for seasonal changes in blood pressure.

Studies of the effect of seasonal variation in blood pressure scores found it robust to intervention[32]. Because of this, it may be that the standard to be met within the measure should optimally vary throughout the year to adjust for this effect. Failure to adjust for seasonal changes in blood pressure may result in providers periodically failing to meet the measure based primarily on this environmental factor rather than any changes in clinical processes.

Because the seasonal effect is ubiquitous among the patient population and not heavily responsive to normal care, we would not expect inter-clinician variation of our measure to vary seasonally. To the extent that clinician level variation is related to performance and not confounded by seasonal effects, its reporting provides a more accurate picture of the effect of clinician performance on quality than the measure alone.

5.6 Analytical Approach

We had hoped to demonstrate a lightweight and non-resource-intensive approach towards achieving value in healthcare statistics. During the initial planning and exploration for carrying out the study, it was expected that a relatively simple analysis would be sufficient. Early approaches explored included visual time series analysis, scoring of month to month correlation,

change point analysis, and the use of control charts of provider performance. Of these approaches, only visualization of the time series was executed in the final analysis. The others were deemed insufficient to answer the research question and it was ultimately decided that an ARMA model would be best.

It may be that simplicity is not a viable option in pursuing methods to effectively measure clinical processes. More sophisticated approaches to measuring care processes are in development. Živaljević, et. al. explored the development of a parametric statistical model designed to describe hierarchical relationships of the components of clinical processes. Provided with EHR data related to the process, their outcome is a quantified measure of quality from 0 to 1.[83] Alternatively, non parametric approaches based on process mining allow for discovery of the process based on the data available.[84, 85] These approaches can describe sequences of events as pathways, and identify rates of deviations past a designated threshold as deviations from quality.[86] Our approach differs from these in that it is technologically unsophisticated.

This provides some advantage. The techniques and tools applied in our approach been in use for years, and so it is more likely that in-house expertise would be available to implement them. Furthermore, our approach is closely tied to performance measures as they currently exist, and so may have more immediate applications.

5.7 LIMITATIONS

Significant limitations affect the internal and external validity of our study. The ranges of blood pressure kept within our study are physiologically plausible, and so face validity of these exclusions may be acceptable, but our cutoff was arbitrary and the means of applying it were unsophisticated. Additionally, our approach for defining outliers was based on a normal distribution and yet, as biological data, it is likely that blood pressure readings follow a power law

distribution.[36] These factors may adversely affect the internal validity of our results, and properly accounting for non-normal distributions may require more sophisticated approaches depending on the subject being measured.

Due to constraints on time and access to information, our study was limited to one measure at one institution. This greatly limits generalizability. It would have been ideal to test our hypothesis against several more measures, including ones where a positive result was not highly expected. However, doing so was outside the scope of this project.

Because our analysis relied on modelling a time series, we cannot exclude the possibility that we failed to do so adequately. Our methods of exploration and multiple models were presented in order to address this, but it is possible that confounding effects extend beyond the month to month changes predicted in our model. Models featuring additional potential confounders, such as time of day, weather, and patient demographics, may have allowed for discovery and control of further inter-correlation.

It is worth noting that our approach only measures correlation and cannot identify causality or directionality to the relationship. Background reading suggested that the major direction of the relationship is that inter-provider variation affects measure performance. However, it is theoretically possible for variation to be due to changes in our measure, such as some providers responding to reports of the measure and others failing to. Such an example would also be an autocorrelated process. A more rigorous approach would include primary data collection to better capture potential causal factors of inter-provider variation.

Finally, clinician-level variation may exist due to factors beyond the control of the clinicians. In some cases, the performance-variation relationship may still be useful information, such as being able to identify when inputs to care processes, such as machines, staff, or information,

unexpectedly vary between clinicians. Proper investigation could prompt improvements in infrastructure or availability of resources. At other times information of inter-clinician variation may not be helpful, such as capturing inter-clinician variation between primary care practices that should ideally vary. In such a case, an attempt to reduce variation by aligning practices could be harmful to outcomes. Application of this approach does not limit the need to carefully consider what are the possible sources of variation and if they should be reported.

Chapter 6. CONCLUSION

By demonstrating the existence of an inverse relationship between inter-clinician variation and the performance of a reportable healthcare measure this study attempted to explore if clinician-level variation of a measure might be a useful indicator of a measure's relationship to quality. Demonstrating an association between performance and inter-clinician variation provides an initial step in addressing broader concerns around improving the use of reported quality measures in hospital quality improvement.

Our quantification of both the degree and strength of the relationship of measure performance to inter-clinician variation provides insight into how much, if any, improvement may be maximally expected by quality improvement efforts addressing clinician activities or inputs. To the extent that this approach is generalizable there may be potential for improving the selection, use, and design of reportable measures for quality improvement within a facility or care process.

The use of inter-clinician variation to identify measures with high levels of clinician level variation may be immediately useful in measure selection. The current trend is towards expanded flexibility in allowing providers to select from a variety of reportable measures across assigned categories. This poses a challenge where the provider must identify measures that they expect to

be able to perform well in and that may benefit the provider and their patient population by reducing costs and improving the quality of care. Currently, providers rely on observation of the mean performance of a proposed measure and on expert opinion from relevant departments and QI personnel in making such decisions, however the introduction of inter-clinician variation may further inform measure selection. Generally, performance in care is dominated by factors at the patient level and secondarily at clinician levels [23] and opportunities for the provider to improve the performance of a measure may be largely with those measures expressing high inter-clinician variation.

This may have important implications in strategies for the selection of measures. For example, if a provider is seeking to choose measures in which they can immediately perform well, they might target ones with low variation as well as high performance. If variations in performance are not sufficient to cause underperformance, selection of high performing measures with high variation may also be acceptable. However, high inter-clinician variation would suggest that inputs to, or execution of, care processes were not stable across the organization and that there may be risk of underperforming in the future. Low performing measures with or without high inter-provider variation would likely be avoided in this strategy.

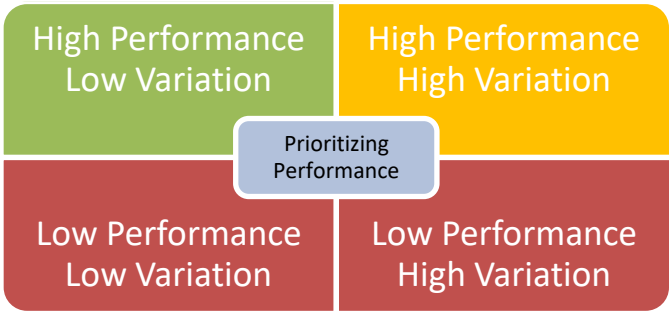


Figure 6.1. Depicting Prioritization of Performance-Variation Relationship When Selecting Measures Optimizing Current Performance Highest Preference Indicated in Green and Lowest in Red

Where a provider is selecting measures for their alignment with opportunities for quality improvement, inter-provider variation could take priority over mean measure performance. In this case, a measure that is high performing, and with high variation would still be an ideal candidate for adoption, whereas, without information on inter-provider variation, such a measure may be ignored. Conversely, a low performing measure with low variation would be a less ideal candidate for selection, whereas without information on inter-provider variation such a measure may have been preferred alongside other measures

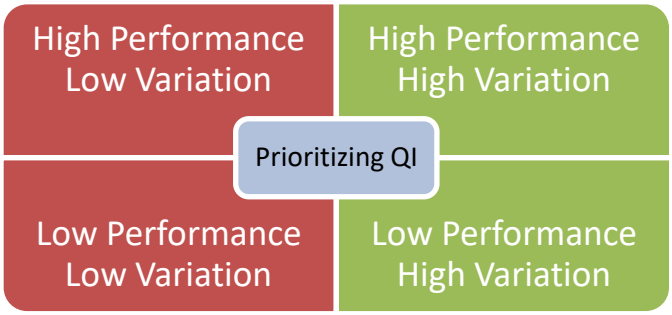


Figure 6.2. Depicting Prioritization of Performance-Variation Relationship When Selecting Measures Optimizing Opportunities for Quality Improvement Highest Preference Indicated in Green and Lowest in Red

The inclusion of variation suggests possible alignment between these two strategies. The provider seeking opportunities for quality improvement will show preference for high variation measures. Meanwhile, the performance seeking provider will prefer high performance measures. These two strategies overlap with the selection of high performance and high variation measures.

Notably, the performance seeking provider may have a stronger preference for high performing, low variation measures (top left box in Figure 6.1). Because of this, a strategy focused on high performing high variation measures could represent a suboptimal strategy for the

performance seeking provider, indicated by the yellow box in Figure 6.1. In a hybrid approach, this may be mitigated by engaging in strategies to reduce variation and improve the quality of these measures, turning them into high performing low variation measures. In a less risk averse environment, the provider may also target low performing high variation measures in the hopes that they can be sufficiently improved to a high performing state. In all strategies, low performing low variation measures would be the least preferred because clinician level interventions to improve performance of these measures may be least likely to result in adequate performance.

The degree of association between variation and performance may attenuate these concerns. If the association is large, it may be expected that even small, and potentially easier, reduction in variation can cause large improvement in quality. On the other hand, in measures where inter-provider variation is only slightly associated with measure performance, we might not expect significant change in the measure even with significant reductions in inter-clinician variation.

Beyond measure selection, the addition of observations of inter-clinician variation has important implications for how performance may be considered in the future. In current practice, a measure is considered underperforming or high performing based on monthly or periodic mean performance. However, this approach cannot differentiate if changes in performance are due to factors intrinsic or external to the care process. A change occurring in a measure characterized by low inter-clinician variation, or where the measure-variation association were small, could suggest that the change is less likely to respond to intervention at the clinician level and other areas should be investigated. The opposite is not necessarily true. A measure with high inter-clinician variation or where the measure-variation association were large may still change due to other causes.

Aggregated across multiple facilities or providers, inter-provider variation may further inform the validity of a measure for national reporting. Measures with high clinician level variation in

more types of facilities may be more appropriate for general use and as required measures. By linking measure reporting to inter-provider variation of the measure, we may further improve how measures are reported. Implementation of clinician-level variation and its benchmarking across facilities may provide a scheme for weighting the value of a measure in compensation and quality comparing schemes. Measures for which a provider has low clinician-level variation may be down weighted or no longer be required to report by the provider.

In considering the use of inter-clinician variation we have avoided defining exactly what constitutes "high" and "low" variation. Answering this is both a matter of strategy and further research.

Future research should test the reproducibility of these results against more measures and environments. Key to the usefulness of this approach is an understanding of the characteristics of measures, processes, and systems where the relationship holds true and can be utilized. Measures where an inverse association is found in some facilities but not in others may suggest differences in how the measure was implemented, or that there are fundamental differences in care processes between the two groups of providers. The execution of similar and identical studies in other environments and on other measures can identify the conditions and environment for which the inverse performance-variation relationship is robust.

To be useful, inter-provider variation must be effectively presented to decision makers. It is likely that clinicians making care decisions will require different presentations of the data than administrators and regulators. Identifying useful aggregations and visualizations of the performance-variation relationship for these users may constitute different avenues of user experience research.

Ultimately implementation research should be conducted to test the effectiveness of reporting inter-clinician variation on improving care. The implementation of systems capable of collecting and reporting a stratified measure constitutes an increased cost to providers and to the healthcare system. As has been seen in currently reported measures, it is possible for a measure to be scientifically valid, but not useful in improving care for a variety of reasons. Research on the effect and best practices in implementing measures utilizing inter-provider variation may help avoid this pitfall.

BIBLIOGRAPHY

[1] Institute of Medicine, Committee on Quality of Health Care in America. To Err Is Human: Building a Safer Health System. In: Kohn LT, Corrigan JM, Donaldson MS, editors.: Institute of Medicine; 2000.

[2] Lieber JB. Medical Error is The Third Leading Cause of Death. 2015.

[3] Makary MA, Daniel M. Medical error-the third leading cause of death in the US. BMJ. 2016;353:i2139.

[4] Institute of Medicine, Committee on the Learning Health Care System in America. Best Care at Lower Cost Path to Continuously Learning Health Care in America. Washington, D.C.: National Academies Press,; 2013. p. 1 online resource (437 p.

[5] Institute of Medicine, Committee on Quality of Health Care in America. Crossing the Quality Chasm: a new health system for the 21st century. Washington, D.C.: National Academy Press; 2001.

[6] Institute of Medicine. Medicare: A Strategy for Quality Assurance, Volume I. In: Lohr KN, editor. Medicare: A Strategy for Quality Assurance: Volume 1. Washington (DC)1990.

[7] Agency for Healthcare Research and Quality. National Quality Measures Clearinghouse. https://www.qualitymeasures.ahrq.gov/search?q=&f_Endorser_String=National+Quality+Forum&page=12018.

[8] Young RA, Roberts RG, Holden RJ. The Challenges of Measuring, Improving, and Reporting Quality in Primary Care. Ann Fam Med. 2017;15:175-82.

[9] Shortell SM, Bennett CL, Byck GR. Assessing the Impact of Continuous Quality Improvement on Clinical Practice: What It Will Take to Accelerate Progress. Milbank Quarterly. 1998;76:593-624.

[10] Bevan G, Hood C. What's measured is what matters: Targets and Gaming in the English Public Health Care System. Public Administration. 2006;84:517-38.

[11] Meyer GS, Nelson EC, Pryor DB, James B, Swensen SJ, Kaplan GS, et al. More quality measures versus measuring what matters: a call for balance and parsimony. BMJ Qual Saf. 2012;21:964-8.

[12] Cook DA, Pencille LJ, Dupras DM, Linderbaum JA, Pankratz VS, Wilkinson JM. Practice variation and practice guidelines: Attitudes of generalist and specialist physicians, nurse practitioners, and physician assistants. PLoS One. 2018;13:e0191943.

[13] Goff SL, Lagu T, Pekow PS, Hannon NS, Hinchey KL, Jackowitz TA, et al. A Qualitative Analysis of Hospital Leaders' Opinions About Publicly Reported Measures of Health Care Quality. The Joint Commission Journal on Quality and Patient Safety. 2015;41:169-AP7.

[14] Powers BW, Navathe AS, Chaguturu SK, Ferris TG, Torchiana DF. Aligning incentives for value: The internal performance framework at Partners HealthCare. Healthc (Amst). 2017;5:141-9.

[15] Lindenauer PK, Lagu T, Ross JS, Pekow PS, Shatz A, Hannon N, et al. Attitudes of hospital leaders toward publicly reported measures of health care quality. JAMA Intern Med. 2014;174:1904-11.

[16] Deming WE. The need for change. The Journal for Quality and Participation. 1994;17:30-1.

[17] Deming WE. Out of the crisis. Cambridge, Mass.: Cambridge, Mass. : Massachusetts Institute of Technology, Center for Advanced Engineering Study; 1986.

[18] Berwick DM. Continuous Improvement as an Ideal in Healthcare. New England Journal of Medicine. 1989;320:53-6.

[19] Garvin DA. Managing quality : the strategic and competitive edge. New York : London: New York : Free Press ; London : Collier Macmillan; 1988.

[20] Deming WE. On Probability as a Basis for Action. The American Statistician. 1975;29:146-52.

[21] Deming WE. On Some Statistical Aids toward Economic Production. Interfaces. 1975;5:1-15.

[22] Berwick DM, James B, Coye MJ. Connections between Quality Measurement and Improvement. Medical Care. 2003;41:1130-138.

[23] Fung V, Schmittdiel JA, Fireman B, Meer A, Thomas S, Smider N, et al. Meaningful variation in performance: a systematic literature review. Med Care. 2010;48:140-8.

[24] Stolzmann KL, Meterko M, Shwartz M, Young GJ, Pekoz EA, Benzer JK, et al. Accounting for Variation in Technical Quality and Patient Satisfaction: The contribution of Patient, Provider, Team, and Medical Center. Medical Care. 2010;48:6760682.

[25] James BC, Savitz LA. How Intermountain trimmed health care costs through robust quality improvement efforts. Health Aff (Millwood). 2011;30:1185-91.

[26] Allegretto S, Michelson D. Looking at Variation Through a Quality Lens. www.hfma.org: Healthcare Financial Management Association; 2014.

[27] Fraser RD. Guidance for using the AHRQ Quality Indicators for hospital-level public reporting or payment. In: Agency for Healthcare Research and Quality, editor. Rockville, MD: Agency for Healthcare Research and Quality,; 2004.

[28] Grossman A, Grossman E. Blood pressure control in type 2 diabetic patients. Cardiovasc Diabetol. 2017;16:3.

[29] Cerda M, Diez-Roux AV, Tchetgen ET, Gordon-Larsen P, Kiefe C. The relationship between neighborhood poverty and alcohol use: estimation by marginal structural models. Epidemiology. 2010;21:482-9.

[30] University of Washington Medical Center. About Harborview. 2018.

[31] Quan L, Dong J, Li Y, Zao L. The effectiveness of intensive nursing care on seasonal variation of blood preassure in patients on peritoneal dialysis. 2011.

[32] Hermann JM, Rosenbauer J, Dost A, Steigleder-Schweiger C, Kiess W, Schofl C, et al. Seasonal Variation in Blood Pressure in 162,135 Patients With Type 1 or Type 2 Diabetes Mellitus. J Clin Hypertens (Greenwich). 2016;18:270-8.

[33] Centers for Medicare and Medicaid Services. Stage 3 Program Requirements for Providers Attesting to their State's Medicaid Promoting Interoperability (PI) Programs. 2017.

[34] Hoef JMV. QUASI-POISSON VS. NEGATIVE BINOMIAL REGRESSION_ HOW SHOULD WE MODEL OVERDISPERSED COUNT DATA? Ecoloogy. 2007;88:2766-72.

[35] Burnham KP, Anderson DR. Multimodel Inference:Understanding AIC and BIC in Model Selection. Sociological Methods & Research. 2004;33:261-304.

[36] West BJ. Where medicine went wrong rediscovering the path to complexity. Hackensack, NJ: Hackensack, NJ : World Scientific; 2006.

[37] Slawomirski L, Aurasen A, Klazings N. The Economics of Patient Safety: Strengthening a value-ased approach to reducing patient harm at national level. Organisation for Economic Co-operation and Development,; 2017.

[38] Agency for Healthcare Research and Quality. National Scorecard on Rates of Hospital-Acquired Conditions 2010 to 2015: Interim Data From national Efforts To Makes Health Care safer. Rockville, MD: Agency for Healthcare Research and Quality; 2016.

[39] Centers for Medicare and Medicaid Services. National Health Expenditure Data. April 17 2018 ed. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/index.html2018.

[40] Organisation for Economic Co-operation and Development. Health at a Glace 2015: OECD Indicators. Paris: OECD Publishing; 2015.

[41] Centers for Medicare and Medicaid Services. National Health Expenditure Data. 2018.

[42] Dzau VJ, McClellan MB, McGinnis JM, Burke SP, Coye MJ, Diaz A, et al. Vital Directions for Health and Health Care: Priorities From a National Academy of Medicine Initiative. JAMA. 2017;317:1461-70.

[43] Department of Health and Human Services. National Strategy for Quality Improvement in Health Care. 2011.

[44] Agency for Healthcare Research and Quality. About the National Quality Strategy. Rockville, MD: Agency for Healthcare Research and Quality; 2017.

[45] Centers for Medicare and Medicaid Services. CMS Quality Strategy 2016. In: Centers for Medicare and Medicaid Services, editor.2016.

[46] Centers for Medicare and Medicaid Services. Hospital Quality Initiative. 2017.

[47] Centers for Medicare and Medicaid Services. What To Report. 2018.

[48] Marjoua Y, Bozic KJ. Brief history of quality movement in US healthcare. Curr Rev Musculoskelet Med. 2012;5:265-73.

[49] Centers for Medicare and Medicaid Services. Meaningful Measures Hub. 2017.

[50] Centers for Medicare and Medicaid Services. Hospital Inpatient Quality Reporting Program. 2018.

[51] The Joint Commission. Accreditation Guide for Hospitals. In: The Joint Commission, editor.2018.

[52] Joint Commission. Specifications Manual for National Hospital Inpatient Quality Measures Version 5.3. 2018.

[53] National Committee for Quality Assurance. HEDIS & Performance Measurement. 2017.

[54] National Committee for Quality Assurance. HEDIS Measure Development Process. 2018.

[55] Centers for Medicare and Medicaid Services. MIPS Overview. 2018.

[56] Centers for Medicare and Medicaid Services. Hospital Value-Based Purchasing. In: Services HaH, editor.2017.

[57] National Committee for Quality Assurance. HEDIS & Performance Measurement. 2018.

[58] Burns J. HEDIS Is the Hassle That Became a Habit. Measuring Care2017.

[59] Agency for Healthcare Research and Quality. AHRQ Conference on Health Care Data Collection and Reporting. 2006.

[60] Agency for Healthcare Research and Quality. Saving Lives and Saving Money: Hospital-Acquired Conditions Update. Rockville, MD2015.

[61] Selby JV, Schmittdiel JA, Lee J, Fung V, Thomas S, Smider N, et al. Meaningful Variation in Performance What Does Variation in Quality Tell Us About Improving Quality? Medical Care. 2010;48:133-9.

[62] Cantiello J, Kitsantas P, Moncada S, Abdul S. The evolution of quality improvement in healthcare: patient-centered care and health information technology applications. Journal of Hospital Administration. 2016;5.

[63] Kiernan F, Buggy DJ. What's measured matters: measuring performance in anaesthesia. British Journal of Anaesthesia. 2015;114:869-71.

[64] Saver BG, Martin SA, Adler RN, Candib LM, Deligiannidis KE, Golding J, et al. Care that Matters: Quality Measurement and Health Care. PLoS Med. 2015;12:e1001902.

[65] Safavi KC, Dai F, Gilbertsen TA, Schonberger RB. Variation in surgical quality measure adherence within hospital referral regions: do publicly reported surgical quality measures distinguish among hospitals that patients are likely to compare? Health Serv Res. 2014;49:1108-20.

[66] Lawson EH, Louie R, Zingmond DS, Brook RH, Hall BL, Han L, et al. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. Ann Surg. 2012;256:973-81.

[67] If Japan Can... Why Can't We? NBC White Paper Series: The Deming Institute; 1980. p. 1:16:38.

[68] Berwick DM. Controlling Variation in Health Care: A Consultation from Walter Shewhart. Medical Care. 1991;29:1212-25.

[69] Juran Institute. Reducing the costs of poor-quality health care through responsible purchasing leadership.(Brief Article)(Statistical Data Included). 2002. p. 3.

[70] Wennberg J, Gittelsohn A. Variations in Medical-Care among Small Areas. Sci Am. 1982;246:120-&.

[71] Traub SJ, Chang YH, Grieger R, Lokareddy S, Patterson J, Tolson H, et al. 187 Inter-Physician Variability in Emergency Department Length of Stay for Discharged Patients. Annals of Emergency Medicine. 2015;66.

[72] Del Piccolo L, Mazzi M, Saltini A, Zimmermann C. Inter and intra individual variations in physicians' verbal behaviour during primary care consultations. Soc Sci Med. 2002;55:1871-85.

[73] Xu T, Makary MA, Al Kazzi E, Zhou M, Pawlik TM, Hutfless SM. Surgeon-Level Variation in Postoperative Complications. J Gastrointest Surg. 2016;20:1393-9.

[74] Cutler D, Skinner J, Stern AD, Wennberg D. Physician Beliefs and Patient Preferences: A New Look At Regional Variation in Health Care Spending. Nber Working Paper Series. Cambridge, MA 02138: National Bureau of Economic Research; 2013.

[75] Weber MA, Schiffrin EL, White WB, Mann S, Lindholm LH, Kenerson JG, et al. Clinical Practice Guidelines for the Management of Hypertension in the Community A Statement by the American Society of Hypertension and the International Society of Hypertension

Journal of Clinical Hypertension. 2014;16:14-26.

[76] University of Washingont Medical Center. Harborview's Mission. 2018.

[77] Welton NJ. The University of Washington electronic medical record experience. J Med Libr Assoc. 2010;98:217-9.

[78] Observational Health Data Sciences and Informatics. OMOP Common Data Model. 2018.

[79] Wennberg JE. Time to tackle unwarranted variations in practice. BMJ. 2011;342.

[80] Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. BMJ. 2003;326:1070.

[81] Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Washington, DC). 2016;4:1244.

[82] Cruz-Correia R, Rodrigues P, Freitas A, Almeida F, Chen R, Costa-Pereira A. Data Quality and Integration Issues in Electronic Health Records2009.

[83] Živaljević A, Mitrović Ž, Petković M. Conceptual and mathematical model for quality improvement in health care. The Service Industries Journal. 2013;33:516-41.

[84] Weerdt JD, Caron F, Vanthienen J, Baesens B. Getting a Grasp on Clinical Pathway Data: An Approach Based on Process Mining. 2013.

[85] Baker K, Dunwoodie E, Jones RG, Newsham A, Johnson O, Price CP, et al. Process mining routinely collected electronic health records to define real-life clinical pathways during chemotherapy. Int J Med Inform. 2017;103:32-41.

[86] Huang Z, Dong W, Ji L, Yin L, Duan H. On local anomaly detection and analysis for clinical pathways. Artif Intell Med. 2015;65:167-77.