

Grant Final Report

Grant ID: 5R18HS017016

Using IT to Improve the Quality of Cardiovascular Disease (CVD) Prevention and Management

Inclusive Dates: 09/01/07 - 06/30/10

Principal Investigator:

Andrew E. Williams, PhD* (Inherited from Vogt, early 2010)

Thomas M. Vogt, MD, MPH*

Team Members:

Adrienne C. Feldstein, MD, MPH[†]

MaryAnn McBurnie, PhD[†]

Richard Meenan, PhD[†]

Chuhe Chen, PhD[†]

Yeehwa Daida*

Rosina Everette, MPH*

Weiming Hu[†]

Sharyn M. Nuha*

Mark Schmidt*

Sabina Smith[†]

Aileen Uchida, MPH*

Aleli C. Vinoya*

Carmen Wong, MBA*

Peggy Latare, MD[‡]

Grant Okawa, MD[‡]

Samir Patel, MD[‡]

Gregory Pawlson, MD, MPH[§]

William Vollmer, PhD[‡]

* Kaiser Permanente Center for Health Research, Honolulu, HI

[†] Kaiser Permanente Center for Health Research, Portland, OR

[‡] Kaiser Permanente Hawaii

[§] National Committee for Quality Assurance, Washington, DC

Performing Organization:

Kaiser Permanente Center for Health Research, Hawaii

Project Officer:

Rebecca Roper, MS, MPH

Submitted to:

The Agency for Healthcare Research and Quality (AHRQ)

U.S. Department of Health and Human Services

540 Gaither Road

Rockville, MD 20850

www.ahrq.gov

Abstract

Purpose: This study developed electronic medical record-based quality indices for eleven cardiovascular primary care services. It related physicians' prior index scores to subsequent disease incidence and to care utilization in their patients.

Scope: Data were collected over an 11-year period by two integrated health care organizations covering approximately 750,000 persons in geographically and ethnically diverse populations.

Methods: Two index types were developed for defined (annual) intervals based on observations of defined populations: Prevention Indices (PIs) and Disease Management Indices (DMIs). Performance during the 11-year period was measured against an unvarying standard; an evidence-based guideline for most services. Variation in practice patterns was assessed using descriptive statistics and graphical representation. The association between indices and outcomes was estimated in generalized linear mixed regression models that were adjusted for clustering effects and for confounding by indication.

Results: Longitudinal and cross-sectional variation in practice patterns differed by service type and by organization. Higher DMI scores for blood pressure were associated with lower incident disease and care utilization. The PI for lipid screening was associated with reduced annual outpatient care utilization.

Key Words: electronic medical records, quality of care, health IT, screening, prevention, disease management

The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services of a particular drug, device, test, treatment, or other clinical service.

Final Report

Purpose

This study, renamed Practice Variations and Care Outcomes (PRAVCO; Grant No. R18 HS 17016; PI: Vogt\Williams) addressed goals that were specified in RFA: HS-07-002: ASQ “Enabling Quality through Health IT of improving the quality and safety of health care”. It developed electronic medical record (EMR) based quality metrics (indices hereafter) of primary care services related to the prevention, detection and management of cardiovascular disease (CVD).

It addressed the following areas of interest specified in the RFA:

- Determine the data elements necessary across care settings to create quality measures
- Demonstrate the ability of electronic data systems to provide data for care quality measures across settings
- Demonstrate the value and accuracy of the measures

All work and results described in this report were supported by AHRQ Grant No. R18 HS 1701. The methodological approach did build on previous work funded by the Centers for Disease Control and Prevention, (Grant No: UR5/CCU 917124; PI: Vogt), and the Kaiser Permanente Garfield Memorial Fund (Grant No: 101-9684; PI: Vogt). PRAVCO significantly extended the work accomplished in those projects (Vogt et al, 2004, 2007) in five ways:

1. In order to enhance the ease with which the measures could be exported to other health care organizations, the algorithms it developed were based as much as possible on standardized data elements defined by the trans-institutional data specifications of the HMO Research Network’s Virtual Data Warehouse (VDW) (Hornbrook et al., 2005)
2. It developed and implemented principles for creating quality indices for a new class of services: disease management
3. It developed methods for optimizing the calculation of disease management indices
4. It related indices for these services to annual CVD incidence and health care utilization
5. It identified and implemented methods that minimize confounding due to patients’ preexisting health differences

Data were collected over an 11-year period by two integrated health care organizations covering approximately 750,000 persons in Hawaii and the Pacific North West of the US. The indices were constructed from medical (e.g., procedures, diagnoses, test results) and operational (e.g. membership enrollment dates) data collected by these organizations on the same patients.

EMRs draw together clinically relevant data on a single patient from many disparate data sources. All data on a patient whether medical or operational will be referred to as EMR data hereafter.

Both organizations used customized variations on a CCHIT-certified EMR (Epic's Hyperspace Spring 2007 IU2; November 10, 2008 K- package level) at the end of the observation period. Similar Epic-based EMRs are used widely by integrated health care organizations across the US. The VDW specification harmonizes data extracted from EMRs regardless of the EMR vendor. One site in the study used a different EMR prior to 2005 and had no EMR prior to 2002. Data from prior EMRs and from legacy clinical and operational databases of health care organizations are all harmonized to a common specification in the VDW.

Most of the indices developed in the study measure performance relative to standards defined by evidence-based clinical guidelines. The guidelines used for each service are listed in Table 1 and the corresponding publications are listed in the references. The choice among alternative guidelines for the same service was based on investigators' judgment about the strength of each guideline's evidence basis. A single standard was used across the observation period. Consequently, poorer "adherence" to guidelines prior to their dissemination, is to be expected for services in which guidelines changed during the observation period.

The indices were used to quantify cross-sectional and longitudinal variations in patterns of care among several hundred primary care providers (PCPs) in two integrated health care organizations (the number of PCPs varied by service). Then, analyses determined the association between PCPs' index score defined on current and prior periods and two types of health outcomes in their patients in the current period: CVD incidence and health care utilization.

The indices fall into two broad categories according to the type of service: Prevention Indices (PIs) or Disease Management Indices (DMIs). Analyses of indices' association with subsequent health outcomes were conducted for two purposes:

1. To validate the indices by estimating their predictive validity after controlling for confounding by indication (confounding by indication and the methods for controlling it are described in the methods section)
2. For those guideline-based indices found to be valid, to determine the effect of physician's guideline adherence on their patients' health

The study had two other goals: 1) to develop weighting schemes for the elements used to calculate the indices that enhance their ability to predict related health outcomes; and 2) to identify and implement methods that would minimize the impact of confounding on estimates of the indices' association with health outcomes.

The specific aims were:

Aim 1. Identify practice level primary care variations in preventive care, weight and tobacco smoking management, and selected CVD risk management services.

Aim 2. Determine the associations of quality of preventive care and disease management practices to morbidity and, where possible, to costs of care.

Aim 3. Improve delivery of care.

Scope

Background

Measuring Care Quality with Secondary Data Sources. The frequently cited gap between research and practice is of grave concern to clinicians, patients, and policy makers. It is the principal motivation for many efforts to monitor and improve care on a broad scale using evidence-based practice guidelines. EMR data are secondary data when used in research because they were collected for non-research purposes. EMR data have the potential to open extraordinary new avenues for research and new tools for bridging the gap between research and practice. EMR data are longitudinal and rich in the detail needed to characterize populations, treatments, and outcomes in real world care. In contrast to other secondary data sources that have been used to study care in community settings, EMR data contain observations on individual patients that can be linked together over time with very specific information about the care they receive, whom they receive it from, and how their health varies as a result.

The tantalizing promise of EMR data is that they far surpass claims data and other secondary data sources in the quality of results they make possible in nearly every area of health services research, safety research, and quality improvement including: comparing health outcomes that result from different patterns of care; producing realistic cost estimates for effectiveness comparisons; monitoring secular trends in disease incidence and in the adoption of new treatments; post-marketing surveillance of drug and device safety; defining and comparing treatment implementation strategies; and informing clinical guideline developers about effectiveness in patient subpopulations and across treatment settings.

For this promise to be fulfilled, however, significant methodological advances are required that will establish new principles for defining care quality metrics, disease ascertainment, and that meet other measurement challenges inherent in working with EMR data. The amount and richness of EMR data is not a panacea for everything that ails secondary data sources. Research based on EMR data will be as useful as the underlying data are complete, accurate, and clinically valid.

Cardiovascular Disease. The health services targeted in the study covered the spectrum of CVD care:

- Prevention of CVD through lifestyle change
- Detection and management of chronic conditions that are risk factors for CVD
- Medication management following MI and CHF

CVD is the most frequent cause of death in the US. Advances in measuring practice variation in primary care services that prevent, detect, and manage CVD will help enable those parts of the research and quality improvement agenda for reducing CVD burden in the US that relate to translating research into practice:

- Analyses of care variations and health outcomes in diverse populations to determine what works for whom
- Minimizing unwanted variation in care
- Identifying positive deviance – those clinicians, clinics, and organization who demonstrate consistently superior performance – so that their care processes can be codified and disseminated
- Developing multi-service profile measures that reflect performance quality across service domains
- Informing the development of multilevel interventions by analyzing the separate contributions of and interactions between patient-level, physician-level and higher system-level predictors of variation in care quality

CVD Services

The existence of efficacious CVD prevention, screening, and management primary care services in offers an opportunity to reduce the enormous burden of CVD in the US by measuring practice variation and improving implementation strategies. Where evidence of effectiveness is undeveloped, analyses of practice variation based on reliable metrics can identify positive deviance. The following primary care services, covering the spectrum of CVD care, were examined in the study:

Prevention: Lifestyle Counseling—Physician Counseling for Weight Loss and Tobacco Cessation.

- **Weight Loss:** Though there is insufficient research to determine the efficacy of physician counseling for weight loss in primary care, there is an urgent need to exploit all plausible means of reducing the rising prevalence of obesity. Primary care physicians often fail to identify obesity-related conditions among their obese patients (Melamed et al., 2009). Clinicians' concern about managing obesity has not substantially increased their interest in actively managing their patients' weight (Kristeller & Hoerr, 1997). Longitudinal data on patient's weight is captured routinely. Indices based on EMR data can identify physicians who are successful in reducing weight among obese patients and who might have insights into how to improve counseling in the primary care setting.
- **Tobacco Cessation:** Tobacco smoking is the leading preventable cause of death in the western world. At least 100 randomized trials have confirmed that clinician counseling and advice in the healthcare setting can substantially increase long term smoking cessation rates (Fiore et al 2000). The USPSTF recommends strongly that tobacco advice and counseling is offered to smokers, but does not clarify either the content of the frequency of such advice. A study using EMR data found considerable practice variation in adherence the AHRQ-recommended 5As guidelines for intervening with patients in the medical care setting and that few patients are assessed on readiness to quit, assisted in making quit attempts, or followed-up (Stevens et al, 2005). EMR data offer opportunities

both to explore successful practices in assisting cessation and to relate those practices to morbidity among patients.

Prevention: Risk Factor Detection—Screening for Hypertension, Hyperlipidemia.

- Blood pressure screening: Many randomized and epidemiologic studies have confirmed that behavioral and pharmaceutical interventions for hypertension reduce incident CVD. Screening for hypertension is so thoroughly integrated into the health care process, that routine screening is performed for most persons at most medical office visits, often in considerable excess of the USPSTF recommendations (Vogt et al, 2004). The USPSTF recommends blood pressure screening every two years for healthy persons with no prior conditions relevant to blood pressure.
- Serum Lipid screening: Many clinical trials have proven the efficacy of lipid control among high risk individuals in reducing risk of morbidity and mortality (Grundy et al, 2004). In recent years, recommendations have emphasized measurement of LDL (and HDL which this study did not examine) rather than total cholesterol. Indices based on EMR data can determine the extent to which such screening is being done.

Management: Monitoring and Controlling Plasma Glucose Concentration in Patients with Diabetes Mellitus.

- Glycated hemoglobin (HbA1c) screening: HbA1c indicates control of blood glucose levels in patients with diabetes mellitus. Higher HbA1c have been associated with CVD, nephropathy, and retinopathy in these patients. The ADA-recommended screening intervals for HbA1c among persons with diabetes is six months if the last HbA1c was < 7% and three months if the last HbA1c was $\geq 7\%$. The management of HbA1c in persons with diabetes is an active research area because of its clinical importance and because of conflicting evidence from clinical trials. EMR data can be used to assess both the timing of HbA1c screening and the level of control of HbA1c.
- Glycated hemoglobin (HbA1c) management: The ADA defines a controlled HbA1c as < 7%. Others bodies recommend different goals. The American College of Endocrinology (ACE) and the American Association of Clinical Endocrinologists (AAACE) recommend < 6.5%, and the American Geriatrics Society recommends < 8% for older adults with a < 5-year life expectancy. Recent studies suggest that control to < 6.5% may be harmful. The ACCORD study was stopped early in 2008 because aggressive management of HbA1c was associated with an increased risk of mortality (ACCORD, 2008).

Management: Controlling Blood Pressure and Serum Lipids.

- Systolic and diastolic blood pressure management: High blood pressure is medically managed as a means of reducing the risk of stroke, congestive heart failure, myocardial infarction, and other cardiovascular diseases. Many clinical trials have proven the efficacy of antihypertensive medication in reducing risk of morbidity and mortality (e.g., Veterans Administration Cooperative Study Group, 1967; Hypertension Detection and

Follow-up Program Cooperative Group, 1979; Neaton et al, 1993; Trafford et al, 1981). Guidelines for managing hypertension usually suggest that systolic blood pressure be maintained below 140 mmHg and diastolic pressure below 90 mmHg. Blood pressure is taken at nearly all primary care visits (an average of 3.8 times per year per adult member). Though the quality of office visit blood pressures is variable and usually not in conformance with National High Blood Pressure Education Program standards of the large number of readings in EMR data partially compensate for this imprecision.

- Serum lipid management: Total serum cholesterol is the sum of low density lipoproteins (LDL), high density lipoproteins (HDL), and very low density lipoproteins (VLDL). CVD risk is associated with higher LDL levels and (to a lesser degree) to VLDL, and inversely associated with higher HDL levels. Indices of LDL management should not be based on total cholesterol. Guidelines recommend a goal of < 120 mg% for persons with high LDL (hyperlipidemia).

Management: Medication Management Following CVD Events.

- Beta blockers: Beta blocker therapy is strongly recommended for all persons who have suffered a myocardial infarction (MI), even those who have contraindications for beta blockers. Although some recommend excluding patients with severely poorly controlled asthma and 2nd and 3rd degree heart block (Borello et al 2003), there is a consensus that there is a net benefit even in those groups with post-MI beta blocker therapy. Pharmacy databases that record the date and days covered by each medication dispensed can be used to indicate the percentage days during a defined period in which these medications were available to patients.
- Angiotensin-Converting Enzyme Inhibitors (ACEIs) & Angiotensin II Receptor Blockers (ARBs): Like beta blockers, ACEIs reduce mortality among patients with a history of CHF and AMI. ACEIs also slow the heartbeat to prevent the heart from getting weaker over time. It can take several months for the full effect of ACEIs to manifest. All patients with CHF should be taking daily ACEI medications indefinitely. ACEIs are contraindicated for persons with renal artery stenosis, angioedema and pregnancy. ARBs are recommended for person with renal artery stenosis and angioedema.

Overview of Prevention Indices and Disease Management Indices

Definitions.

- A *Prevention Index* is a measure of the extent to which a screening or preventive service was delivered to a defined population during a defined interval.
- A *Disease Management Index* is a measure of how effectively a disease or condition was managed in the population defined by the pertinent diagnosis during a defined interval.

Prevention Indices (PIs) and Disease Management Indices (DMIs) are two classes of health service quality measures that can be constructed on whole patient populations using routinely

collected clinical and administrative data. Both are person-time measures that require longitudinal data. The approach developed in this project for constructing these indices is general and flexible. The development of an index for any particular service will be shaped by both data quality and availability issues and by the intended use of the index.

Advantages. The advantages of using EMR data over claims data for defining health service quality indices include:

- Greater specificity in defining the target population (e.g. excluding cancer cases from the defined population targeted by a screening service)
- The ability to link the service to the health outcomes that the service is intended to influence
- The ability to link patients to the specific individuals, clinics, health systems and other entities that performed the service
- The ability to measure more dimensions of the services beyond occurrence (how it was performed, how often, by whom, when in relation to other relevant services, etc.)
- Greater accuracy in determining the intended purpose of the service (e.g. distinguishing between the screening vs. diagnostic uses of the same test)

PCP-Level Focus. The primary focus of this project was on variations in patterns of care delivered by PCPs and the implications of these variations for the health of their patients. The same methods could be used for studies focusing on variation at the patient, clinic, or higher system level. Data in a given year were excluded from analyses for patients without an assigned PCP in that year. Patients who switched PCPs midyear were attributed to the PCP they were assigned to for largest number of days that year.

Hypotheses. The hypothesis for all combinations services and outcomes was that higher PCP-level index scores would be associated with lower rates of incident CVD and health care utilization within their patient panels.

Methods

Because of the strong emphasis on methods development in this study, this report describes methods in greater than usual detail. Results will be described in greater detail in publications.

Aim 1

Identify practice level primary care variations in preventive care, weight and tobacco smoking management, and selected CVD risk management services.

Principals for Index Construction (Aim 1)

The construction of PIs and DMIs is based on a set of common principles. Both PIs and DMIs:

- Are calculated for a defined population over a defined period
- Are ratios in which both the numerator and denominator are constrained by the individual's eligibility for the service and the availability of their data during a defined period
- Begin with patient-level scores in each interval which can be rolled up to create PIs & DMIs for higher level units involved in the care of those patients: Primary Care Physicians (PCPs), clinics, etc.
- Are expressed as a percentage with higher scores indicative of better performance, though perfect scores are not necessarily optimal

Though most PIs and DMIs defined to date are expressed as a percentage, some PIs for services related to lifestyle change are more conveniently expressed as binary values (0 or 1) at the individual patient-period level. These values indicate that an event did or did not occur for that patient during the defined period, e.g. they quit smoking or lost weight during a calendar year. But the PI scores for these services are still expressed as a percentage when they are rolled up to higher level units such as PCP panel-years or clinic-years, e.g., the percent of smokers in a PCP's panel who quit in a year. The PIs for services targeting lifestyle change and the DMIs for medications are atypical in other respects as well. Some of the following description of PIs and DMIs does not apply to them.

Differences between PIs and DMIs. Two features distinguish PIs from DMIs; the measured services' target population and the method of calculation. The first distinction is obvious. The screening and preventive services measured by PIs target patients who are not known to have the relevant disease and who are at more or less average risk for its occurrence. DMIs measure services involved in the management of a disease or a condition among the population of patients with the relevant diagnosis.

The second difference is in the number of parameters that are used to calculate the index. Though both PIs and DMIs are ratios, the numerator and denominator used to calculate PI scores are based solely on time. The numerator and denominator used to calculate DMI scores, however, are products of both time and level of control. PIs can be thought of a ratio of line lengths where the lengths indicate covered and eligible time. DMIs can be thought of as a ratio of areas where the areas represent ideal and actual levels of control of a clinical value related to disease management. The meaning of these terms and the methods for calculating these ratios are explained in detail below.

Numerators Used in the Calculation of PI Scores. The calculation of PI scores requires a clinically meaningful definition of the period of time that a patient should be considered "covered" by a service after it has been performed. PIs quantify the extent to which a service was delivered to a target population in accordance with a guideline-recommended service interval;

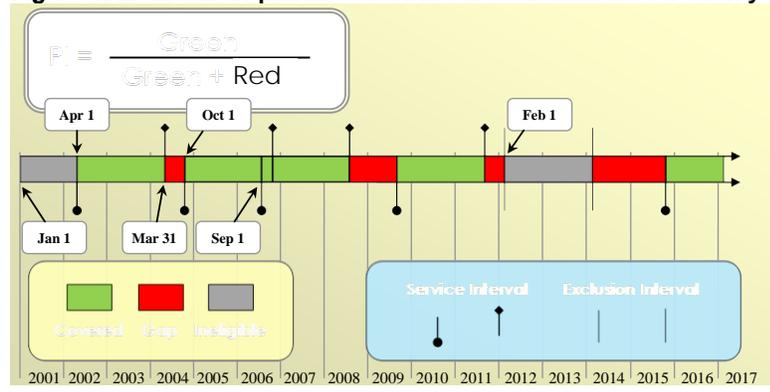
e.g., the USPSTF’s current recommendation that average risk adults screened for hypertension at least every two years. The exceptions are when there is no commonly shared evidence-based guideline for a commonly used service (e.g. counseling obese patients about weight loss) or when the data pertinent to the service are available but impractical to obtain from electronic data sources (e.g. the five ‘A’s for Tobacco cessation in primary care). The numerators of PI scores are the amount of time that a person was covered for a service during a defined period, e.g. 260 days out of a calendar year.

Denominators Used in the Calculation of PI Scores. The denominators of PI scores are the amount of time that a person was eligible for a service during a defined period, e.g. 365 days out of a calendar year. Portions of the defined period are excluded from both the numerator and the denominator if the person becomes either ineligible or unobservable. Periods of ineligibility can be either temporary or permanent. An interval is removed from the numerator and denominator beginning on the date of a diagnosis or procedure which indicates the occurrence of a disease (e.g. cancer) the detection of disease risk (e.g. a biopsy) that excludes the person from the average risk population targeted by the service. Intervals are also excluded if the person cannot be observed because they leave (e.g. switch health care systems) or die.

Numerators Used in the Calculation of DMI Scores. The DMIs for most disease management services use data on the success of a service in addition to the service’s occurrence to construct both the numerator and the denominator of the ratio. Each part of the ratio for most DMIs, in other words, is itself the product of two parameters: *Time* and *Level of Control*. For example, the diastolic and systolic blood pressure values obtained when blood pressure measurements occur are an indicator of the success of hypertension management. They indicate the level of hypertension control. This level of control can vary over time. The numerator for most DMIs is the sum of the products of the amount a clinical value is above goal and the duration of each period of time between measurements.

While PIs are constructed as a ratio of the time a person was covered for a service to the time they were eligible for it, DMIs are constructed as the ratio of how well-controlled the relevant clinical value is for that person during the period they had the disease. While the *Time* parts of these ratios are a concrete quantity the time a person is observed, the upper limit of the Level of Control part of the ratio can be set to on any theoretical limit that doesn’t exclude clinically meaningful variation.

Figure 1. Elements of patient-level PI score calculations for each year of an observation period



Calculating a PI Score. Figure 1 illustrates how to calculate a patient-level PI score for each calendar year of an observation period that starts on Jan 1, 2001 and ends on Feb 1, 2017. In this example, the PI is based on the USPSTF guideline-recommended blood pressure (BP) screening frequency of no more than two years. The hypothetical health plan data used to produce the figure would have to be sufficient to determine that the patient was an average risk adult whose was age-eligible for BP screening during at least some part of the observation period and to capture the dates on which the patient's blood pressure was measured. About half of BP measurements are performed for non-screening purposes (Vogt et al, 2004). It is essential to distinguish between screening and non-screening BP measurements. EMR data can be used to make this distinction algorithmically by identifying previous hypertension or CVD diagnoses and by identifying elevated BP measurements in healthy adults' periods that require frequent observation for a time in order to confirm or refute the presence of hypertension.

The horizontal band in the figure represents the entire observation period. It is subdivided into color-coded categories of person-time used to calculate PI scores. The three categories mark time during which:

- The patient was eligible for the service and was covered (green)
- The patient was eligible for the service but was not covered (red)
- There was insufficient information available to determine whether the patient was covered (gray)

Screening BP measurement dates are marked by lines ending in circles below the band. Lines ending in diamonds above the band mark the end of an interval of covered person-time. Lines extending below and above the band mark the boundaries of intervals of excluded person-time. If the PI score is calculated on calendar years, the score for the patient in each year is the ratio of the green days to the sum of the green and red days in that year. The gray days do not contribute to either the numerator or the denominator.

The patient with no history of hypertension or CVD enrolled in the plan on the first day of the observation period; January 1st, 2001. Her first screening BP measurement occurred on April 1, 2002. The gray-colored period between her enrollment and her first BP measurement is excluded from PI calculations because there isn't enough information prior to April, 2002 to determine whether she was due for a BP measurement before then. It is not possible, therefore, to calculate her 2001 PI score.

Her screening BP measurement in April, 2002 initiated a green-colored interval spanning the two years of person-time during which she was covered for the service. This period ended on March 31st, 2004. For 2002 and 2003 she was covered for every day that she was eligible. Her PI score for both years, therefore, is 100%. Her next BP measurement did not occur until October 1st, 2004, however. It was six months overdue. This red-colored six-month gap in coverage was uncovered person-time that decreases her PI score for 2004 to approximately 50%.

Although she was due for her next BP measurement on September 30, 2006, she was screened two months before that date on July 30, 2006. This two-month overlap of covered intervals could be regarded as duplicate coverage and can be captured and quantified during the PI calculation, though it does not contribute to the calculation of the PI score. The "early" BP measurement resets the next due date to July 30, 2008. The marking of covered and uncovered

periods during the remainder of the observation period follow the principles already described with one exception.

On February 1st, 2012, the woman sought treatment in the emergency department for chest pain. She received a diagnostic BP measurement. This initiated an excluded person-time interval because she was temporarily outside the “average risk” population until surveillance determined she should again be regarded as average risk. The length of that excluded interval depends on patient management guidelines. In this example the date of the diagnostic BP measurement initiated a two year interval after which she returned to the regular screening schedule because no further signs of pathology were detected.

Though a PI score could be calculated for 2017 using the two months of available data, it is worth considering whether the sample is adequate to provide a useful estimate. Similarly when patient-level scores are rolled up to physician’s panel of patients or panels are rolled up to clinics etc. it is worth considering whether a minimum number of data points should be required. In this study we required a panel to include a minimum of 30 eligible patients to be the analysis.

Calculating a DMI Score. Intervals of person-time are excluded from DMI score calculations for the same reasons they are excluded from PI calculations. For example, when a patient enrolls in a plan or receives a defining diagnosis part way through a calendar year, only the period subsequent to their enrollment or diagnosis contributes to the calculation of their DMI for that year. For simplicity’s sake, however, the hypothetical data in the following example is for a patient who is eligible and has complete data for the entire year the DMI is calculated on.

The PI calculation was illustrated using a horizontal line that represented a single dimension; time. That calculation reduced to the ratio of covered time to eligible time within a defined period. DMIs measure both time and level of control. Control can be represented as an added vertical dimension that together with the horizontal dimension of time forms an area rather than a line. The ratio for calculating DMIs isn’t a comparison of line lengths but of areas. It quantifies not only how well a clinical value is controlled, but how long it is controlled that well. The analog to the one-dimensional gap in service used in calculating PIs is the area defined by how out of control a clinical value is.

Figure 2. Elements of patient-level DMI score calculations for one year based on three diastolic blood pressure measurements

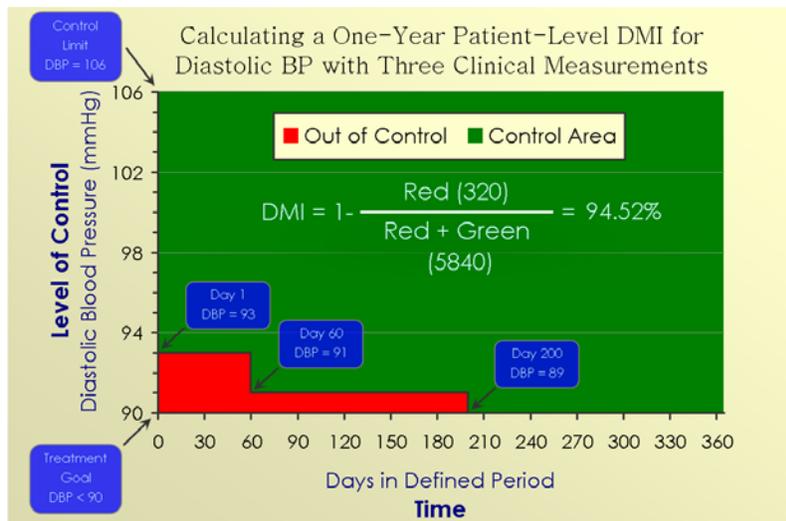
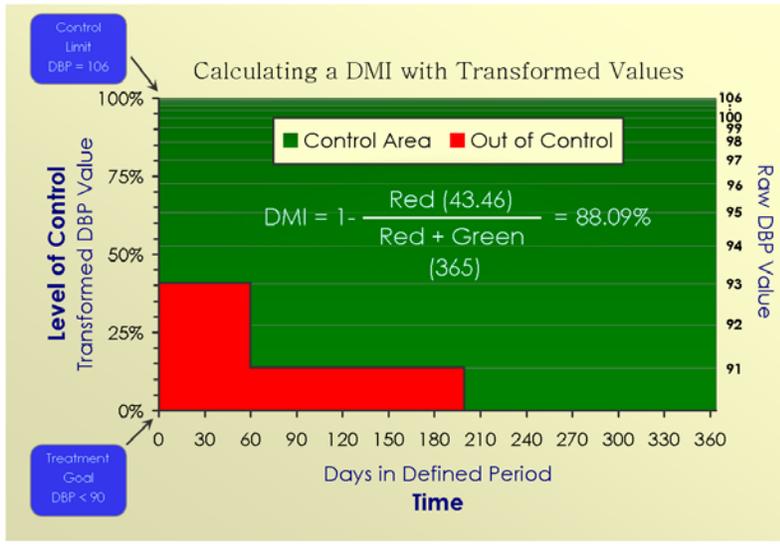


Figure 2 displays one method of calculating a single patient’s DMI for diastolic blood pressure (DBP) for one year using raw BP values. The scale of the control dimension is defined relative to a treatment goal. In this case the treatment goal is 90 mmHg and the height of the control area is three standard deviations of 5 mmHg above a hypothetical patient population of mean of 91 mmHg. Values greater than 3 standard deviations above the mean are treated as equal to three standard deviations above the mean. The horizontal axis represents time and the vertical axis represents the amount that a blood pressure value exceeds the treatment goal. Each BP value above goal initiates a period that is represented by a red rectangle. The width of the rectangle is the number of days between blood pressure measurements. The height is the difference between the measurement and the treatment goal.

In the hypothetical example in Figure 2, a patient’s DBP was measured three times during the one-year period the DMI is to be calculated for. It was 3 mmHg above goal on day 1, 1 mmHg above goal sixty days later, and below goal on day 200. These measurements define two red rectangles. One has an area of 180: 60 days at 3 mmHg above goal. The other has an area of 140: 140 days at 1 mmHg above goal. The sum of those areas is the numerator in the ratio used to calculate the DMI: $180 + 140 = 320$. The denominator is the total eligible days observed multiplied by the control area limit: $365 \times 16 = 5,840$. To make interpretation of the DMI consistent with the PI, the ratio of those two values is subtracted from 1 so that a DMI of 100% indicates complete control during the defined period: $1 - 320/5,840 = 94.52\%$.

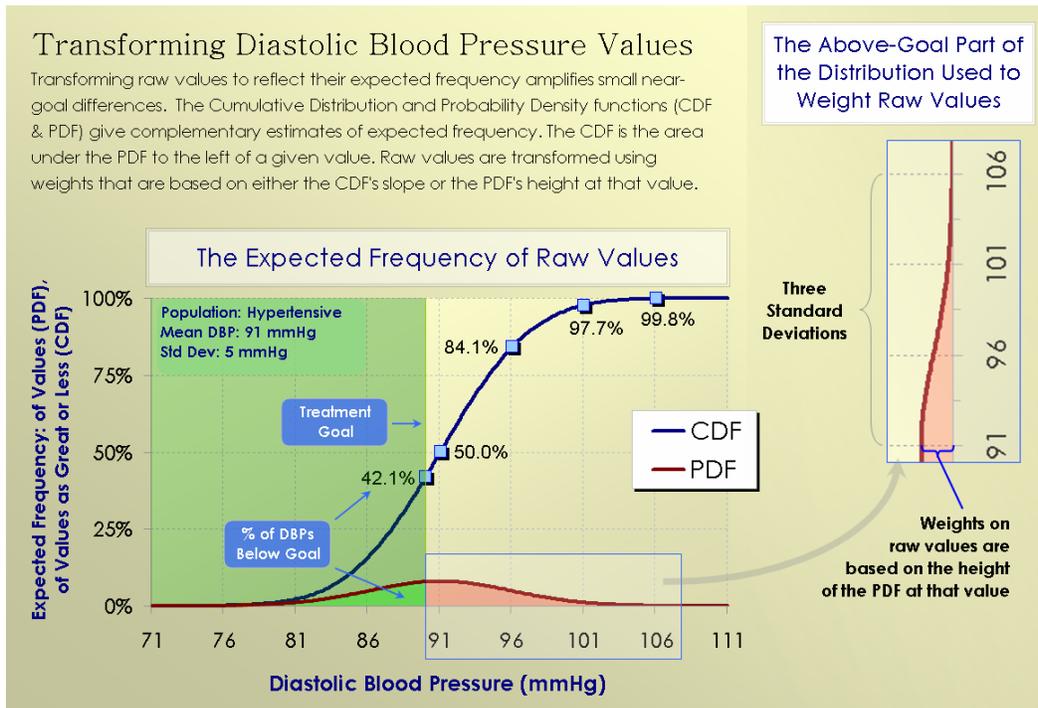
Figure 3. Elements of patient-level DMI score calculations using transformed BP values



Using Transformed Clinical Values to Optimize a DMI Score. Under some conditions it might not be advisable to use raw clinical values to calculate DMI scores. The numeric representation of the clinical values as they are used in the calculation should suit their intended use. If raw values provide too little variability in DMI scores for a given purpose, such as relating DMIs to health outcomes, a transformation should be applied that preserves their rank while increasing their variance. In this project the raw DBP measurements above the USPSTF goal clustered so tightly near the treatment goal that DMIs based on these raw values showed

little variation. Many weighting schemes can increase the DMI's power to predict health outcomes by transforming raw clinical values. Figure 3 shows the effect on the DMI of the transformation that was used in the project.

Figure 4. Using the distribution of raw clinical values in the population to rescale the control dimension of a DMI



A Method for Transforming Raw Clinical Values (Aim 1)

The transformed values used in Figure 3 reflect the probability of observing the raw value. In addition to increasing the variance of DMIs based on near treatment goal DBPs, the use of this transformation reflects the proposition that rare events are less important in population-level analyses of common services related to high prevalence diseases. When the intended use of the DMI dictates a different proposition, alternate transformations should be used.

The raw blood pressure values and other assumptions used to calculate the DMI in Figure 3 are the same as those used in Figure 2. The transformed DBP values used in Figure 3 result from a weighting scheme based on the distribution of above-goal blood pressure values in the population. Two complementary representations of this distribution, the cumulative distribution function (CDF) and the probability density function (PDF), are shown in Figure 4. The above-goal part of the PDF is rotated 90° on the right side of Figure 4. The width of the red shaded area under the PDF gets smaller as raw DBP values increase. The weights for each raw value are proportional to the width of the PDF at that value.

The limiting values of the resulting rescaled control dimension are:

- Lower limit = the value of the CDF for a normal distribution at the treatment goal approximately 0.42.
- Upper limit = the value cumulative distribution function for the normal distribution at 3 SD above the mean: $91 + (3 \times 5) = 106$

The value of the CDF for a normally distributed variable with a mean of 91 and a standard deviation of 5 at 106 is approximately .998 meaning that 99.8% of the distribution is less than or equal to 106.

The rate at which the intervals between the ascending DBP values decrease on the right side Figure 3 is exactly proportional to the slope of the above-goal part of the CDF in Figure 4. In addition to placing weights on the raw values, the transformation rescales them to the half-open unit interval (0, 1] with boundaries defined by the upper and lower limits of the control dimension.

Procedurally, the transformation algorithm is implemented in two steps. In step one, the value of the CDF at goal is subtracted from the CDF of any above-goal raw value. For example, the CDF for a raw DBP value of 92 is approximately 0.58. Subtracting the CDF at goal (0.42) gives approximately 0.159, or about 16% above goal. Raw values that are at or below goal are simply replaced with zeros in this step.

In step two, the weighted value remaining after this subtraction is multiplied so that it covers a range from 0 up to but not including 1. Without any subtraction, the full range of a CDF for a normally distributed variable is 0 to 1 and is interpretable as the percentiles of the distribution. To retain this interpretation, the range of the above-goal CDF values has to be stretched using a multiplier that is proportional to the size of the part that was subtracted.

For example, if the mean and the treatment goal were both 90 mmHg, exactly 50% of the distribution would be above goal since the normal distribution is perfectly symmetrical about the mean. Therefore, the multiplier needed to transform the above-goal portion of the CDF back into percentiles after subtracting the bottom half would be 2 ($50\% \times 2 = 1$). When the population mean and the treatment goal are not the same, this multiplier is given by the inverse of the proportion of the distribution that exceeds the treatment goal: $1/(1-\text{CDF at goal})$.

Using the values in the DBP example, the multiplier is equal to $1/(1-0.42) \approx 1.73$. Multiplying the difference (0.159) between the CDF for the raw DBP of 92 and CDF at goal by this number yields a transformed value of 27.4%. In contrast, the raw value is only 2/16 or 12.5% above goal.

Technically, the upper limit of this rescaled control dimension does not include 1 because the CDF for the normal distribution reaches 1 at infinity. In practice this will have no impact and transformed values exceeding some limit near 1 can be rounded up if a lowest DMI of 0% representing a completely uncontrolled condition is desired.

The choice of 3 SD ensures that > 99% of the distribution of above-goal is captured by this range. The exact percentage below that limit will be a function of the difference between the mean and the treatment goal. Rescaling the control axis using an upper limit defined by any other number of standard deviations greater than 3 will have almost no effect on the transformation.

Data Requirements. The particular method of calculating a given PI or a DMI varies with the data that are available in a given setting at a given point in time. The kinds of data needed can be classified by the uses to which they are put in calculating the indices:

- Defining the relevant patient population
- Capturing the occurrence of the service
- Excluding periods of time during which patients are not eligible for the service
- Excluding periods of time during which data on the patient are unavailable
- Capturing an indicator of the level of control of a clinical parameter

The quality and kinds of data available and the feasibility of obtaining it efficiently will vary by the service.

Data Extraction. This project sought to gain efficiency by relying as much as possible on EMR data that had met preliminary quality assurance and standardization by virtue of its inclusion within the HMO Research Network’s Virtual Data Warehouse (VDW). Not all data required for the study were available in the VDW, and additional quality assurance analyses and development were required on data that were in the VDW.

Figure 5. Model of data flow

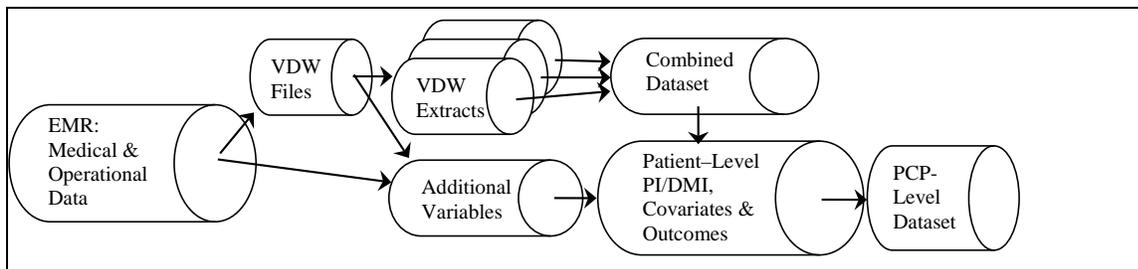


Figure 5. Multiple extracts from the VDW sources were concatenated to create a combined dataset. Patient-level PI & DMI scores and other calculated variables were created using the combined dataset and additional VDW and EMR data. PCP-Level PI & DMI scores, covariates, and outcome were calculated from the patient-level dataset.

Figure 5 illustrates the data flow from extraction through index construction to create the patient-level and the final PCP-level analytic datasets. Data on practice variation was calculated from patient-, PCP-, and clinic-level data at each data-contributing site. Hypothesis tests were run on the PCP-level dataset which combined data from both sites and included calculated panel-year characteristics used in the development of propensity scores, covariates, panel-level incidence rates, and panel-level health care utilization rates.

Data Development. Data in the analytic datasets were analyzed for completeness, consistency with expected frequency, abrupt changes over time, plausible value range, and logical consistency with other values. Variable that were out of range were windsorized, invalid or logically inconsistent values were not included. Standardized data extraction request forms that specified all data elements and code values were developed and used for each service. Standardized queries and tabular reports were developed to assess data integrity of the preliminary and of the patient-level and PCP-level datasets. The dependence of the PI and DMI variables in particular on many other variables required frequent revision of extraction programs at some stage within the dataflow depicted in Figure 5.

Table 1. Index type; service; defined population, defined period, years of data from each site; types of data, guideline and average number of observations per year for each PI and DMI

Index Type	Service	Defined Population	Defined Period	Observed Pop/Period	Source data types	Guideline	Average Obs per Year
DMI	Diastolic Blood Pressure	Age ≥ 21; 2+ ICD-9 codes for Hypertension	1 Year	Site A: 2007-2008; Site B: 1998-2008	ICD-9, CPT, HCPCS, BP from EMR	USPSTF: DBP <90 mmHg	PT: 90,017 PCPs: 230
DMI	Systolic Blood Pressure	Age ≥ 21; 2+ ICD-9 codes for Hypertension	1 Year	Site A: 2007-2008; Site B: 1998-2008	ICD-9, CPT, HCPCS, BP from EMR	USPSTF: SBP <140 mmHg	PT: 90,017 PCPs: 230
DMI	Serum Lipids	Age ≥ 21; 2+ ICD-9 codes for Hyperlipidemia	1 Year	Site A: 1998-2008; Site B: 1998-2008	ICD-9, CPT, HCPCS, Lipids from EMR	ATP III: LDL <130 mg/dL	PT: 37,205 PCPs: 311
DMI	HbA _{1c}	Age ≥ 21; 2+ ICD-9 code indications of Diabetes Mellitus	1 Year	Site A: 2005-2008; Site B: 1998-2008	ICD-9, CPT, HCPCS, HbA _{1c} from EMR	ADA: < 7.0%	PT: 23,802 PCPs: 231
DMI	Beta Blockers	Age ≥ 21; ICD-9 code indication of prior MI	1 Year	Site A: 1998-2008; Site B: 1998-2008	ICD-9, CPT, HCPCS, Rx Fills from Pharmacy	No Usable Guideline: Continuous medication availability	PT: 13,659 PCPs: 376
DMI	ACEI & ARBs	Age ≥ 21; ICD-9 code for prior CHF	1 Year	Site A: 1998-2008; Site B: 1998-2008	ICD-9, CPT, HCPCS, Rx Fills from Pharmacy	No Usable Guideline: Continuous medication availability	PT: 16,150 PCPs: 173
PI	BP	Age ≥ 21; No prior ICD-9 code indications of Hypertension or CVD	1 Year	Site A: 2005-2007; Site B: 1998-2008	ICD-9, CPT, HCPCS, BP from EMR	USPSTF: Screen annually	PT: 294,559 PCPs: 461
PI	Serum Lipids	Age ≥ 21; No prior ICD-9 codes for Hyperlipidemia or CVD	1 Year	Site A: : 1998-2008; Site B: 1998-2008	ICD-9, CPT, HCPCS, Lipids from EMR	ATP III/ USPSTF: Screen every 5 yrs	PT: 16,129 PCPs: 352
PI	HbA _{1c}	Age ≥ 21; 2+ ICD-9 codes for Diabetes Mellitus; No prior ICD-9 codes for CVD	1 Year	Site A: 2007-2008; Site B: 1998-2008	ICD-9, CPT, HCPCS, HbA _{1c} from EMR	ADA: Screen every 6 months if last HbA _{1c} < 7%; every 3 months if HbA _{1c} ≥ 7%	PT: 45, 475 PCPs: 303

Table 1. Index type; service; defined population, defined period, years of data from each site; types of data, guideline and average number of observations per year for each PI and DMI (continued)

Index Type	Service	Defined Population	Defined Period	Observed Pop/Period	Source data types	Guideline	Average Obs per Year
PI	Weight	Age ≥ 21; BMI ≥ 30 on last measurement in prior year; No prior ICD-9 codes for CVD	1 Year	Site A: 2007-2008; Site B: 1998-2008	Weight from EMR	No Usable Guideline: % of obese patients who lost weight	PT: 81,640 PCPs: 271
PI	Tobacco	Age ≥ 21; Active smoker status at last assessment in prior year; No prior ICD-9 codes for CVD	1 Year	Site A: 2007-2008; Site B: 1998-2008	Smoking status from EMR	No Usable Guideline: % of smoking patient who quit	PT: 69,351 PCPs: 296

In several cases, new primary sources in the EMR data had to be developed to obtain valid values. It is worth emphasizing the vital role in the construction of these datasets of experts who have comprehensive knowledge of the vast variety data sources they draw upon and the history of changes to those sources within an organization.

Variable Definition, Coding, and Clinical Validation. When possible, published criteria for using diagnosis codes, procedure codes, or other EMR data sources were used to define case ascertainment. The ICD-9, CPT, HCPCS codes, lab values, and definitions for caseness and other calculated variables were reviewed by investigators with clinical expertise, and by individuals with knowledge of institutional coding practices to assure their validity and appropriateness. Because of the chronic nature of CHF events, the number hospital days with a primary discharge of CHF was used for case ascertainment.

Single lab values or other measurements, such as blood pressure readings, that are used in clinical algorithms to ascertain caseness, can have poor predictive value. Despite intense resources devoted to redundant coding and other quality assurance efforts, diagnoses are occasionally not recorded. For both these reasons, we required two diagnoses of hypertension, hyperlipidemia, and diabetes to define caseness. EMR encounter codes that specify the encounter setting were used to determine health care utilization frequency. Outpatient visits included codes used for all direct patient contact (in-person or by telephone) with treatment specialists (e.g., physicians, physical therapist, mental health, dietician, etc.). Because of occasional redundancy for encounters in source data, any number of outpatient visits or emergency department visits during a defined period were counted as a single encounter. Emergency department encounters were tracked manually until early 2007 at site A and were of insufficient integrity to include in analyses.

The two medication adherence indices (the DMIs for Beta Blockers and ACEI & ARBs) are modified versions of a continuous medication availability measure developed in the prior work of Vollmer and his group for the PEANUT study (NHLBI Grant No R01 HL083433; PI: Vollmer; Vollmer et al., 2007). Vollmer consulted with the PRAVCO investigators regarding which of several continuous medication availability (CMA) measures was best suited to the purposes of this study and Chen adapted the algorithm.

For each care quality index, Table 1 lists the index type; the service measured; the defined population, defined period, and the years from each data-contributing site used in constructing

the index; the types of data used, the clinical guideline that performance was measured against; and the average number of observations per year that contributed to analyses.

Variation in practice patterns was determined through descriptive statistics and graphs that documented cross-sectional and longitudinal variation in PI scores and outcomes at the patient, PCP, and clinic level in each data-contributing site. Standard tables tabulating, numbers of eligible patients, variation in PI scores and outcomes were developed and used for each service.

Aim 2

Determine the associations of quality of preventive care and disease management practices to morbidity and, where possible, to costs of care.

Hypotheses. The hypothesis for all services was that higher quality care as measured by the PI or DMI for the service would be associated with lower CVD incidence and lower health care utilization within PCPs' patient panels.

Lag-Time between Care and Outcome. We assumed that each year's rates of CVD incidence and utilization in a panel of patients would reflect the care they received from their PCP both in that year and in one or more prior years. For most services, therefore, index scores and outcomes used in analyses of hypothesis tests were defined on two-year periods. For management and screening of serum lipids and for tobacco cessation counseling, indices were defined on a two-year period but incident cases were defined on a four-year period with twice the weight placed on the middle two years (Figure 6). The choice of the number of years to define incidence on was based on the estimated lag between effectively delivered care and CVD outcomes.

Figure 6. PCP index timeline

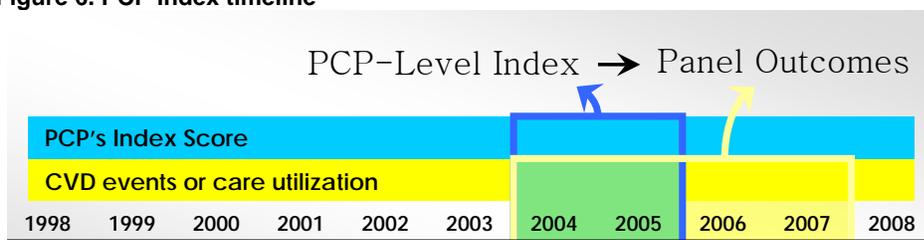


Figure 6. For most services, indices and outcomes were defined on two-year periods. For three services, serum lipid management and screening, and tobacco cessation counseling, CVD incidence was defined on a four-year period.

Analytic Strategy. Hypotheses were tested on PCP-level data pooled from both sites across all years using generalized linear mixed models (GLMM). GLMM were chosen because of their advantages in analyzing data that exhibit within-cluster correlation, nonconstant variability and outcomes that are not normally distributed (Wolfinger, 1993). Generalized linear mixed models consist of the following: (a) linear predictors, (b) a monotonic mapping of predictors to the mean of the data, and (c) an outcome that is distributed as any of the family of exponential distributions – Beta, Binomial, Gamma, Normal, Lognormal, Poisson etc.. The models used to

tests hypotheses for each service specified a Poisson distribution for incident outcomes and a log link function between predictor and outcome variables. The number of eligible patients in a panel-year was used as the offset value. Maximum likelihood was used to estimate fixed effects. Distributional and collinearity assumptions were checked graphically prior to running tests. Parameter estimates for the fixed effect represent the change in CVD incidence or health care utilization for a 10% increase in the PI or DMI score for the service.

Clustering. The covariance structure of data within of levels clustering units were: (a) panel-years within years PCPs, and (b) PCPs within clinics. The first type stemmed from the fact that many PCPs contributed multiple panel-years to the analysis. We assumed that the second type would arise because working in the same clinic would naturally produce within-clinic clustering of practice patterns. The effects of clinic-level clustering were small and had no meaningful impact on other parameter estimates. We therefore modeled the data according to a observation year within PCP structure. We modeled within PCP covariance as unstructured using a Cholesky parameterization.

Random Effects. The hypotheses we tested were not concerned with rates of change in incident events. We chose, therefore, to test our hypotheses by fitting a multilevel or “random effects” model of the effects of model parameters on the intercept (population average) for each set of within PCP observations. In generalized linear mixed models that adjusted for clustering effects by modeling the covariation of observations within patients, of patients within PCPs panels, and of PCPs within clinics. Outcomes observed in each year were predicted by the quality indices defined on the prior year except for the Tobacco the Lipid screening services where the DMIs were defined over the prior three years combined.

The focus on PCP-level care patterns determined the choice of statistical models. Hazard or other time-to-event models were inappropriate because individual patients who were initially disease-free were not the unit of analysis. Data from persons not assigned to PCPs are excluded from analyses. Post hoc analyses were done to assess the impact of including these data.

Covariate Balance. Confounding by indication occurs when estimates of the effect of some treatment or attribute are biased by differences in the patient populations who received the treatment or who received care from an entity with that attribute. This source of bias is one of the primary obstacles to analyses of observational data. Without random assignment to treatments or entities, it should be assumed that the patient populations being compared in the analysis are unequal for any of a variety of reasons that might have an effect on the probability of the health outcome: they are older, sicker, have less access to care, are less adherent to care regimens, etc. In this study’s analyses, the challenge was to determine how to control for the confounding effects of preexisting differences among panels when estimating the association between indices and outcomes. Panels of older, sicker patients will be have higher morbid and utilization outcomes, but they will also receive more attention from their PCPs and thereby drive up the PCP’s index score. The collinearity between these panel attributes and the predictor means that including them as covariates in the model will bias the estimated association between index and outcome.

There are two primary approaches to addressing the possibility of confounding by indication when analyzing observational data: Instrumental variables and propensity scores. An instrumental variable determines treatment while being otherwise unrelated to the treatment

outcome. Thus, randomized treatment assignment is itself an instrumental variable. But variables that determine treatment or a care-related attribute of interest while being otherwise unrelated to the treatment outcome are rare and there are no good methods for determining whether a chosen instrumental variable has functioned as intended.

Propensity scores are a means of achieving the balance that randomization would achieve on at least those variables that can be measured. In contrast to standard covariates, they are unlikely to produce estimation problems from collinearity with the predictor. They are developed by using potential confounders to predict the treatment condition or care-related attribute of interest in a regression model. Propensity scores are the probability of being assigned to a treatment group or of receiving care from someone with that attribute given each individual's combination of potential confounders.

The propensity scores used in analyses in this study were developed using the following attributes of each panel-year prior the year(s) on which the quality indices was defined: 1) mean patient age; 2) mean Charlson Comorbidity Index (an inpatient diagnosis-based measure the total number of 16 high mortality associated conditions a patient has [Charlson et al, 1987]); 3) percent of male patients; 4) percent of patients with a diagnosis of diabetes; 5) percent of obese patients; 6) percent of patients who have chronic kidney disease; and 6) total patients in the panel. Outcome variables were removed from calculation of the Charlson Comorbidity Index and variables used to define population were removed from propensity score development where appropriate. The propensity score was used as the single covariate in the statistical models used to test hypotheses for each service.

Exploratory analyses were done to investigate the relative importance of contributors to propensity scores using standardized scatter-plot and cross-tabulations of mean PI or DMI scores within propensity score quintiles.

Aim 3: Improve Delivery of Care

The study intended to: A) determine whether PI & DMI quality scores were associated with CAHPS random surveys of patient satisfaction; and B) to feedback PI & DMI scores back to clinicians and observe whether the feedback led to changes in the performance levels or the CAHPS scores. Although initially informed otherwise, we later determined that we could not get CAHPS scores with specific patient identities because the company that kept the data refused to release identity information. Also, the timeline of the study was insufficient to have feedback and monitor subsequent performance changes. Instead, we worked with our consultant, Dr. Gregory Pawlson, of the NCQA, who has proposed additional testing of these measures with the view to adopting them as HEDIS measures when they change to EMR based measures.

Results

Practice Variation

The practice variation in PCP-level DMI and PI scores are plotted in figures 7 and 8. The box and whisker plots mark the 25th, 50th, and 75th percentile of PCP-level index scores defined

on each individual year in the observation period. Whisker lengths were calculated as the interquartile range multiplied by 1.5 with ceiling values of 100% and floor values of 0%. Observed variations at site A are on the left and those at site B are on right side of each figure. Services are grouped by type.

Figure 7. Variation in PCP-level DMI scores

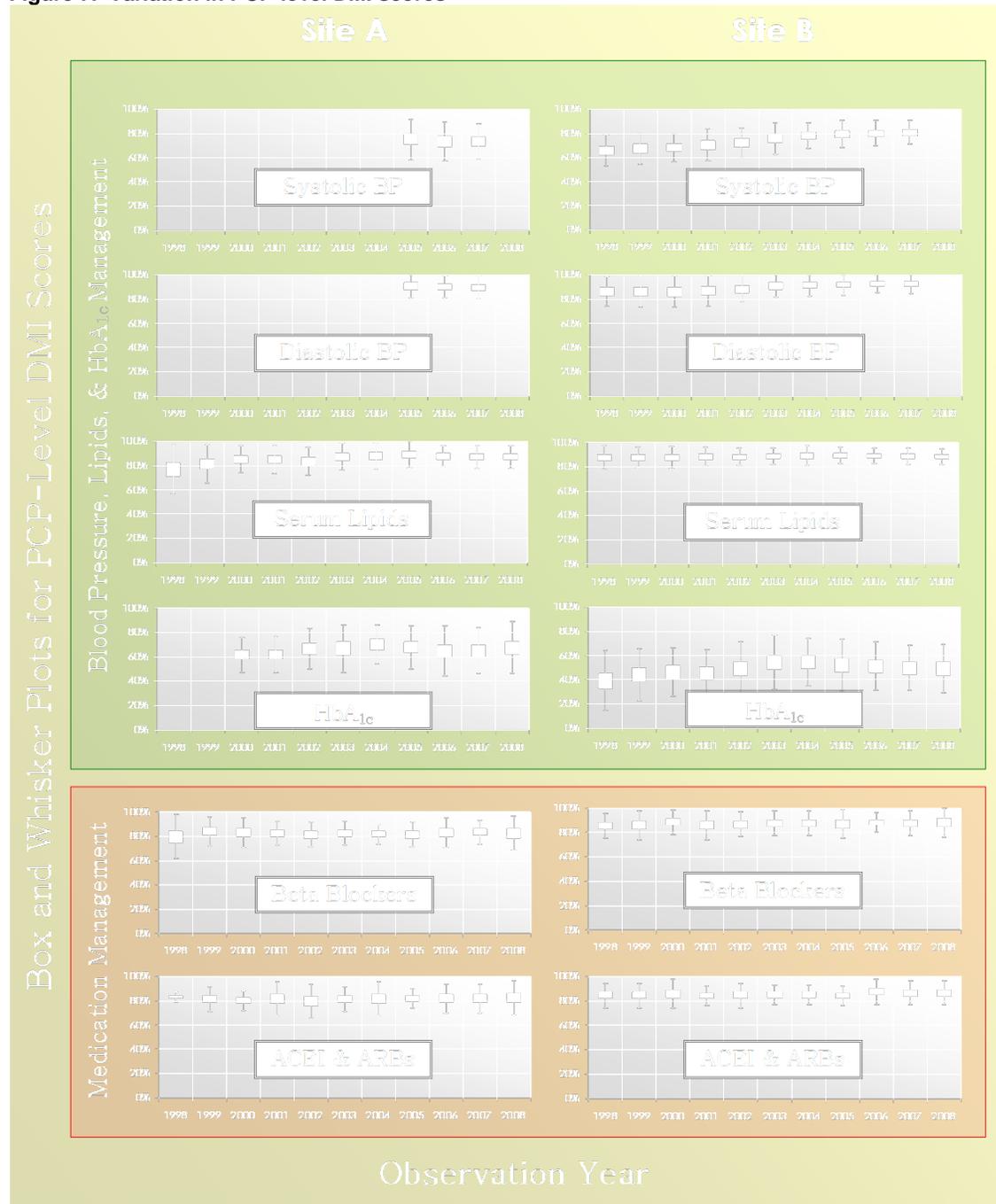


Figure 7. Distribution of DMI scores for management of blood pressure, serum lipids and HbA_{1c} and for medication management. Box plots mark the 25th 50th and 75th percentile of the distribution of the PCP-level DMI score in each year. Whisker length is the interquartile range multiplied by 1.5.

Figure 8. Variation in PCP-level DMI scores

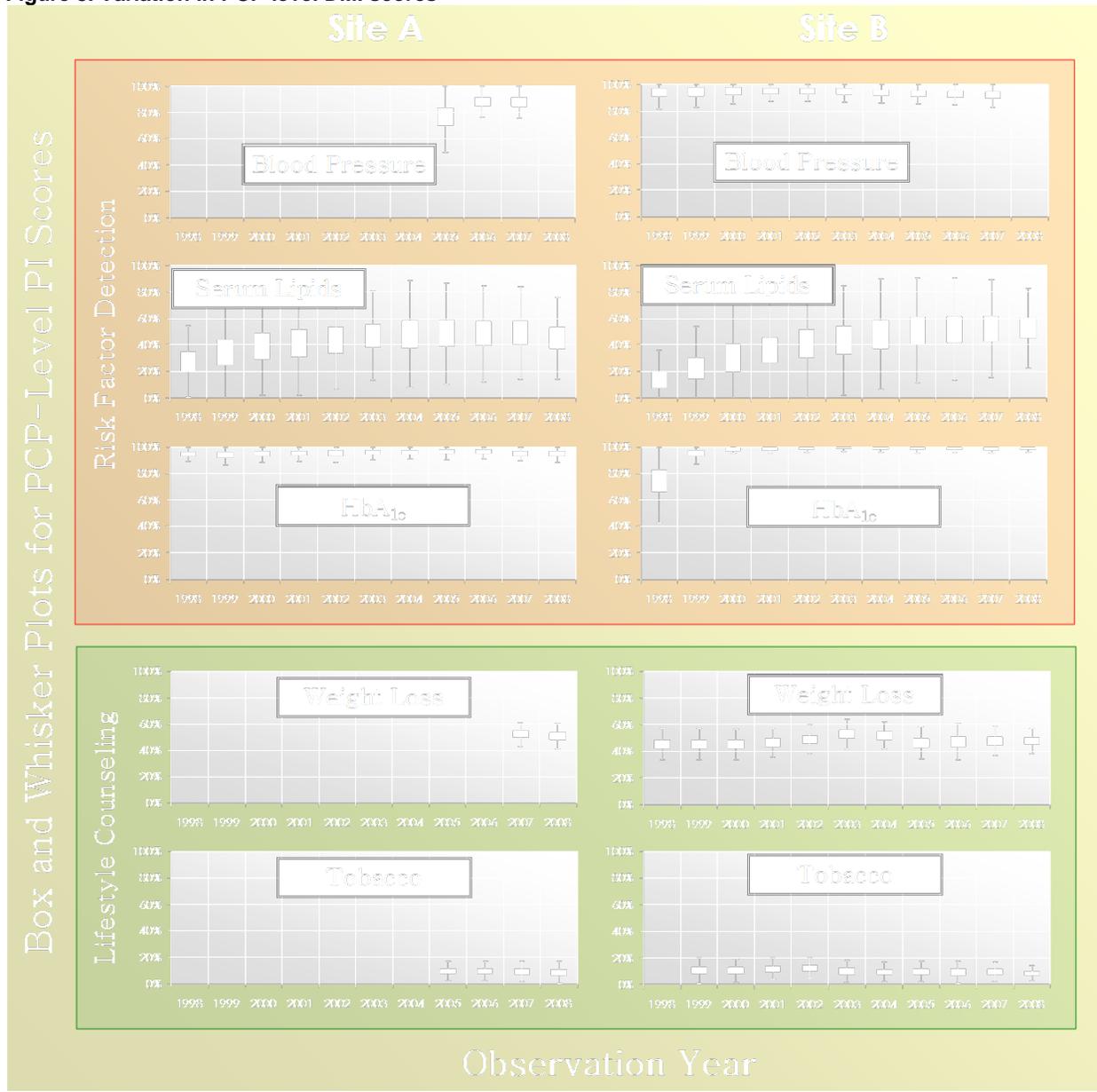


Figure 8. Distribution of PI scores for blood pressure, lipid, and HbA_{1c} screening and for weight-loss and tobacco cessation. Box plots mark the 25th, 50th, and 75th percentile of the distribution of the PCP-level PI score in each year. Whisker length is the interquartile range multiplied by 1.5.

HbA_{1c} screening is different than other screening services, however, because it is used to manage rather than detect a condition. It is grouped with the prevention and screening services as a PI because of how it is calculated rather than because of its function: it is calculated based on recommended time intervals, while the DMI for HbA_{1c} is calculated based on time and level of control of HbA_{1c}.

Cross-sectional variations in practice patterns were largest for serum lipid screening and smallest for HbA_{1c} screening. The median index scores were highest for HbA_{1c} screening and lowest for tobacco cessation counseling. Longitudinal variation was greatest for lipid screening

and for HbA_{1c} management. The smallest longitudinal variations were observed in both medication management services and HbA_{1c} screening. For most services, the median index values and the amount of cross-sectional and longitudinal variation were similar across sites. The greatest cross-site variation in median scores was in HbA_{1c} management. The DMI for HbA_{1c} was consistently higher at site A. There were larger cross-sectional and longitudinal variations and lower median index scores in blood pressure screening at site A than at site B.

Cross-service comparisons in performance level and in the amount and kind of variation have to be understood in the context of the differences between the services themselves, the algorithms for constructing the index for them, and the relationship between the index and health outcomes. Differences in algorithms have been reviewed above. The adequacy of EMR data to capture the relevant indicators of service delivery and service quality also varies. A key factor in interpreting the index for any service is its association with health outcomes.

Associations between Indices and Health Outcomes

Figure 9 plots the effects for the PCP-level indices for all services for each CVD incidence outcome. The effects for health care utilization are plotted in figure 10. Effects less than 0 indicate a reduction in incident disease or care utilization. The size of the effect represents the expected change in incidence or utilization for every 1,000 patients in a PCP’s panel associated with a 10% increase in the index score for the service listed on the horizontal axis. Effects for DMIs are grouped to the left of those for PIs in both figures. The point estimates are the effects found for the service index using GLMM after propensity score adjustment. The error bars are the standard errors of the effect estimates.

Figure 9. PI & DMI association with incident CVD

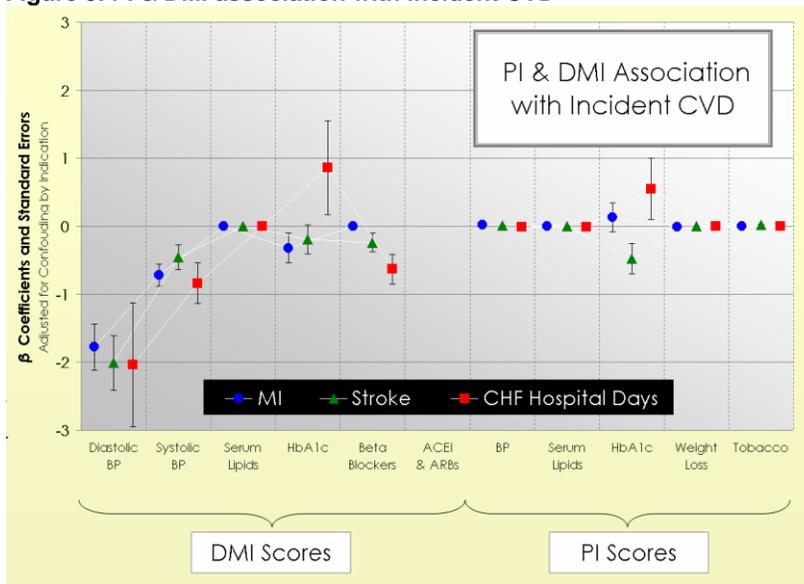


Figure 9. PCP-level DMI and PI score effect estimates for each CVD incident outcome after adjustment for confounding by indication. Error bars are the standard errors for the effect estimates. Effect estimates represent the change in CVD incidence rate per 1,000 eligible panel members for a 10% increase in the PI or DMI score.

In Figure 9, the effects for indices are plotted using blue circles for MI outcomes, green triangles for stroke, and red squares for CHF hospital days. In Figure 10, blue circles represent indices effects for outpatient care and green triangles represent indices effects for emergency department visits.

Figure 10. PI & DMI association with health care utilization

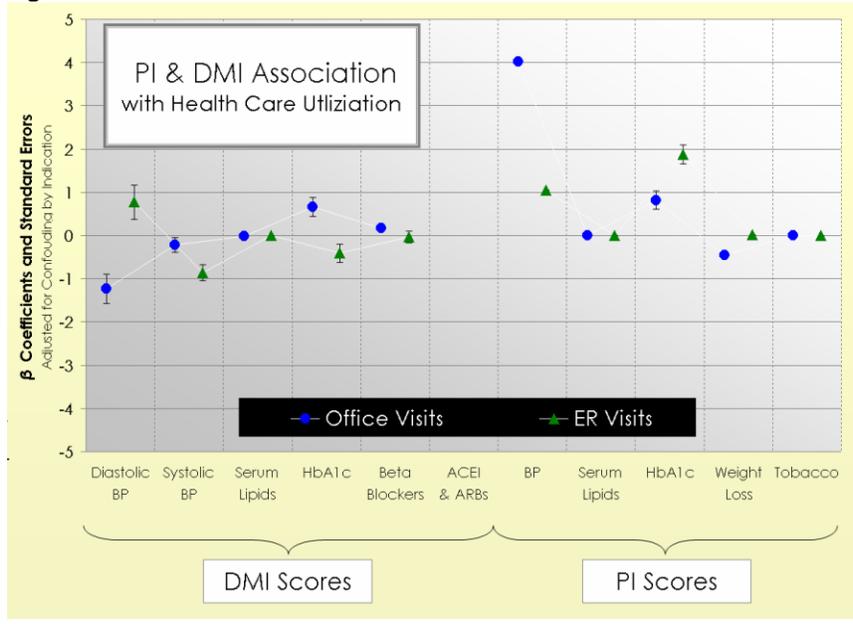


Figure 10. PCP-level DMI and PI score effect estimates for each category of health care utilization after adjustment for confounding by indication. Error bars are the standard errors for the effect estimates. Effect estimates represent the change in of health care utilization rate per 1,000 panel members for a 10% increase in the PI or DMI score. ACEI/ARB results excluded due to size relative to other effects.

Table 2 gives the same effects and standard errors in numerical form along with the p values from the GLMM. Effects with a p value < .05 are in bold faced font. Some of the most interesting results were found for management of systolic and diastolic blood pressure. These results are described in detail to illustrate the general approach. Other results are described briefly.

Table 2. Hypothesis test results. Effect estimates, standard errors, and p values for the association between each care quality index and each CVD incidence and annual health care utilization outcome (Effect estimates represent change in outcome per 1000 eligible panel members for a 10% increase in the PCP-level PI or DMI score. Effect estimates were adjusted for confounding by indication using a single propensity score as a covariate)

Table 2a. CVD Risk factor detection and management. Defined population: Dx = no hypertension

Service	Outcome	β	SE	p
BP Screening	MI	0.025	0.005	< .0001
BP Screening	Stroke	0.009	0.005	.10
BP Screening	CHF	-0.006	0.006	.29
BP Screening	Outpt Visits	4.020	0.471	< .0001
BP Screening	ER Visits	1.041	0.806	.20

Table 2b. CVD Risk factor detection and management. Defined population Dx = hypertension

Service	Outcome	β	SE	p
Diastolic BP Management	MI	-1.778	0.342	< .0001
Diastolic BP Management	Stroke	-2.009	0.399	< .0001
Diastolic BP Management	CHF	-2.036	0.909	< .05
Diastolic BP Management	Outpt Visits	-1.232	0.180	< .0001
Diastolic BP Management	ER Visits	0.769	0.133	< .0001
Systolic BP Management	MI	-0.719	0.161	< .0001
Systolic BP Management	Stroke	-0.455	0.181	< .01
Systolic BP Management	CHF	-0.839	0.299	< .01
Systolic BP Management	Outpt Visits	-0.220	0.093	< .05
Systolic BP Management	ER Visits	-0.864	0.312	< .01

Table 2c. CVD Risk factor detection and management. Defined population: Dx = no hyperlipidemia

Service	Outcome	β	SE	p
Serum Lipid Management	MI	0.002	0.006	.71
Serum Lipid Management	Stroke	-0.002	0.008	.81
Serum Lipid Management	CHF	-0.009	0.009	.33
Serum Lipid Management	Outpt Visits	-0.0004	0.001	.46
Serum Lipid Management	ER Visits	-0.002	0.001	< .001

Table 2d. CVD Risk factor detection and management. Defined population: Dx = hyperlipidemia

Service	Outcome	β	SE	p
Serum Lipid Management	MI	0.001	0.001	.35
Serum Lipid Management	Stroke	0.0002	0.0004	.65
Serum Lipid Management	CHF	-0.00001	0.00003	.60
Serum Lipid Management	Outpt Visits	-0.010	0.004	< .05
Serum Lipid Management	ER Visits	-0.008	0.006	.18

Table 2e. CVD Risk factor detection and management. Defined population: Dx = diabetes mellitus

Service	Outcome	β	SE	p
HbA _{1c} Screening	MI	0.131	0.212	.54
HbA _{1c} Screening	Stroke	-0.474	0.225	< .05
HbA _{1c} Screening	CHF	0.551	0.450	.22
HbA _{1c} Screening	Outpt Visits	0.814	0.315	< .05
HbA _{1c} Screening	ER Visits	1.875	0.325	< .0001
HbA _{1c} Management	MI	-0.320	0.222	.15
HbA _{1c} Management	Stroke	-0.413	0.313	.18
HbA _{1c} Management	CHF	1.271	0.590	< .05
HbA _{1c} Management	Outpt Visits	0.658	0.177	< .001
HbA _{1c} Management	ER Visits	-0.411	0.293	.16

Table 2f. Post CVD event medication management. Defined population: Dx = prior MI

Service	Outcome	β	SE	p
Beta Blocker Adherence	MI	-0.007	0.004	.33
Beta Blocker Adherence	Stroke	-0.005	0.005	.76
Beta Blocker Adherence	CHF	-0.002	0.007	.95
Beta Blocker Adherence	Outpt Visits	0.161	0.054	< .01
Beta Blocker Adherence	ER Visits	-0.028	0.117	.81

Table 2g. Post CVD event medication management. Defined population: Dx = prior CHF

Service	Outcome	β	SE	p
ACEI/ARB Adherence	MI	-56.343	40.479	.16
ACEI/ARB Adherence	Stroke	-41.990	44.980	.35
ACEI/ARB Adherence	CHF	-68.667	63.090	.28
ACEI/ARB Adherence	Outpt Visits	-42.350	33.537	.21
ACEI/ARB Adherence	ER Visits	45.760	40.866	.26

Table 2h. CVD-related lifestyle counseling. Defined population: obesity (body mass index \geq 30)

Service	Outcome	β	SE	p
Weight Loss	MI	-0.007	0.004	.33
Weight Loss	Stroke	-0.005	0.005	.76
Weight Loss	CHF	-0.002	0.007	.95
Weight Loss	Outpt Visits	-0.463	6.857	.33
Weight Loss	ER Visits	0.020	0.003	< .0001

Table 2i. CVD-Related lifestyle counseling. Defined population: active tobacco smoking status

Service	Outcome	β	SE	p
Tobacco Cessation	MI	0.003	0.009	.72
Tobacco Cessation	Stroke	0.021	0.012	.07
Tobacco Cessation	CHF	-0.001	0.017	.94
Tobacco Cessation	Outpt Visits	0.005	0.005	.37
Tobacco Cessation	ER Visits	-0.006	0.006	.34

Dx= Diagnosis; MI= Myocardial Infarct; CHF= Congestive Heart Failure; ER= Emergency department; BP= Blood Pressure. Outpt= Outpatient. CHF outcome= hospital days not incident events. Effects with a p value $<$.05 are in bold face font. Except where the variable was related to the population definition or outcome, propensity scores were developed using the site and the following panel-year attributes: mean age; mean Charlson Comorbidity Index; percent male; percent with diabetes; percent obese; percent with chronic kidney disease; total patients eligible for the service. ACEI/ARB results were doubly adjusted for propensity scores and the same variables as separate covariates. Indices for all services were defined on two-year periods. Outcomes were defined on two-year periods except for serum lipid screening and management and tobacco cessation counseling which were defined on four-year period with doubly weighted middle years.

DMI for Diastolic and Systolic Blood Pressure

Figure 11 displays the practice variation that was observed in the management of systolic BP and diastolic BP (SBP & DBP) and CVD incidence rates site B from 1999 to 2008. Data from site A, available for 2006-2008, are similar but not shown. DMI scores were calculated on a period including the labeled year and the prior using the weighting scheme described above. Each graph displays the distribution of 2-year DMI scores for all PCPs in a year.

The SBP DMIs are on the left side of Figure 11 and those for DBP are on the right. Incident MI, stroke, and CHF hospital days per 10,000 panel members are shown in the heights of the yellow, black, and blue bars on the left side of each SBP graph.

Figure 11. PCP-level DMI for blood pressure at Site B

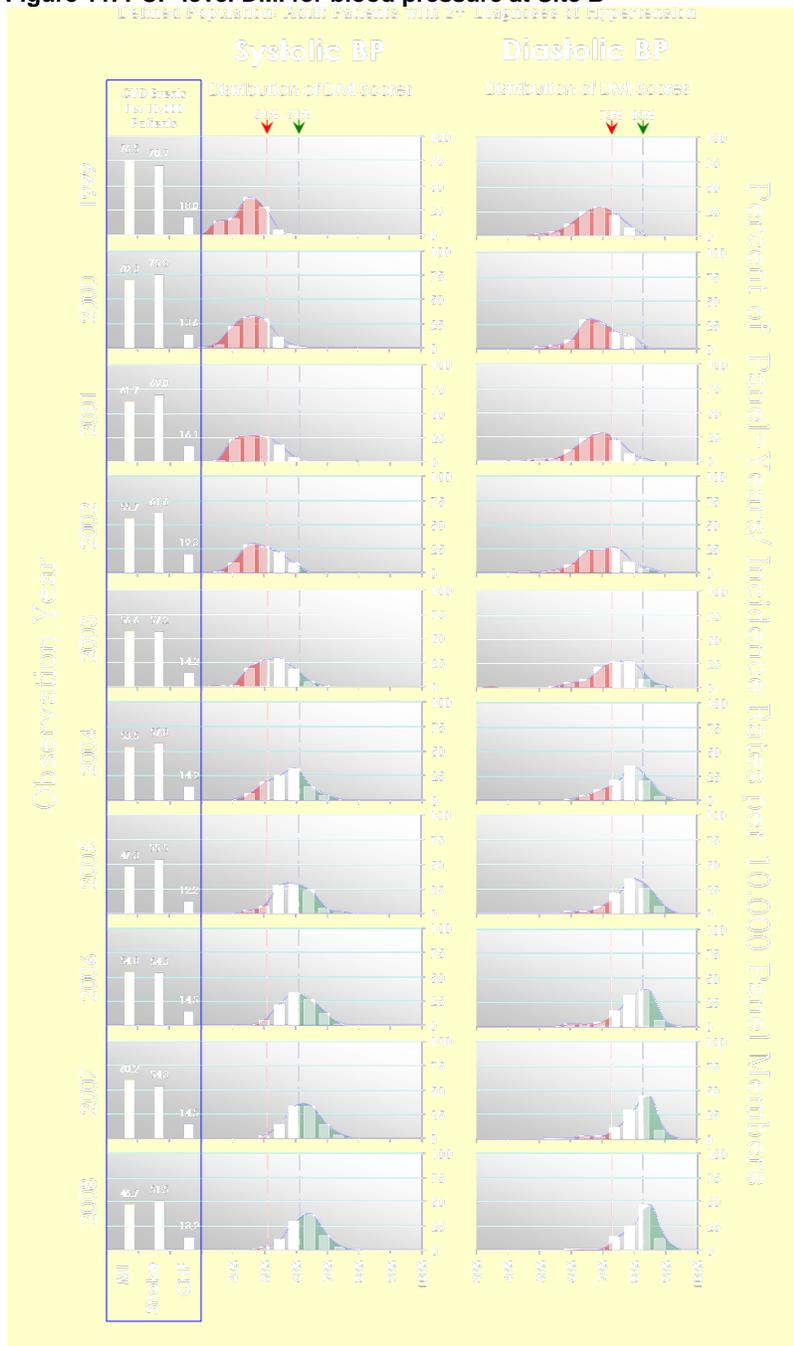


Figure 11. PCP-level DMI distributions and incident CVD rates at site B. DMI scores were calculated over the labeled year and prior year combined using weighted BP measurements. Red and green shaded areas are below or above a 10% span centered on the overall median. Incidence is per 10,000 patients.

The dashed red and green vertical lines in each graph mark a 10% span in DMI scores centered on the overall DMI median: 51%-61% for SBP and 73%-83% for DBP. Areas below this span are shaded red and those above it are shaded green. This view of the data facilitates the interpretation of the effects for the DMIs in the regression models.

The effect estimates for each quality index (see Table 2, Figures 10 & 11) represent the expected change in health outcomes for a 10% increase in the service index after adjustment for confounding by indication. The 10% span in the graphs for each DMI in Figure 11 shows both the improvement relative to one's peers that an increase of 10% represents in each year, and the implication of the change in DMI scores that the entire population of PCPs exhibited over the observation period.

To illustrate the meaning of the regression model estimates in terms of patient's incident CVD, consider two DMI scores for systolic BP for a hypothetical PCP. His first DMI score, constructed using data from the years 2000 to 2001 was 50%. The second from the years 2001 to 2002 improved to 60%. The effect estimate for the DMI for systolic BP was 0.72 for incident MIs. This means that 0.72 fewer patients would be expected to have an MI event for every 1,000 members of his panel in 2001 and 2002. By extension, 72 fewer patients would be expected to have an MI event in those years in every 1,000 members of 100 PCP's panels if they all made the same improvement.

An average of 141 PCPs contributed to each DMI score distribution in Figure 8. Note the longitudinal change in the percent of the DMI distribution above and below the 10% span each year in the entire population of PCPs. The percent of PCPs whose 2-year DMIs were in the red shaded areas decreased from 89% in 1999 to 2% in 2008 for SBP and from 69% to 5% for DBP. The change in the percent of PCPs whose DMIs were in the green shaded areas above the 10% span each year was also dramatic going from 0% in 1999 to 61% in 2008 for SBP and from 0% to 51% for DBP.

Figure 12. Distribution of longitudinal change in individual PCPs

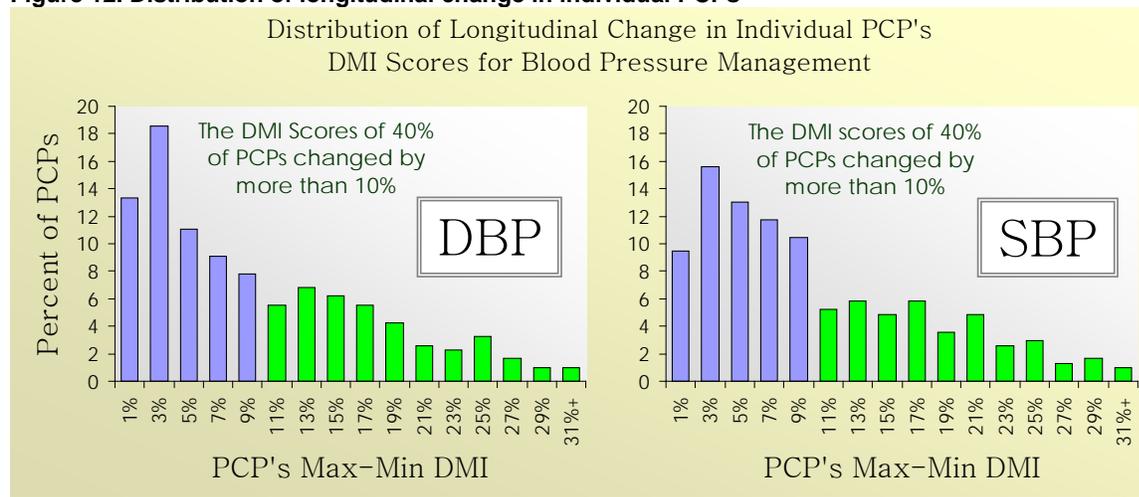


Figure 12. The change in individual PCP's DMI scores for blood pressure management in adults with diagnosed hypertension.

The data in Figure 11 are repeated cross-sectional views that show change in all PCPs. Figure 12 shows the distribution of change in individual PCP's 2-year DMI scores for DPB and SBP during the observation period. The categories on the X axis are the absolute value of the difference in each PCP's maximum 2-year DMI score minus their minimum 2-year DMI score over the entire observation period at both sites. Bars heights give the percent of PCPs who score changed by the amount in each category. As expected, PCPs who were observed for more years tended to have higher absolute change. As is obvious from figure 11, most change was improvement.

These statistically and clinically significant decreases are not as readily apparent from the change in raw incident CVD rates across years (See the left side of each SBP graph in Figure 11). The rates shown in figure 11 are multiplied by 10 to give incidence per 10,000 panel members. While the annual incidence rates drop steadily for stroke, they fluctuate for MI and CHF after an initial drop from 1999 to 2000. There was significant confounding by indication for the DMI for DBP in particular. Higher DMI scores were strongly positively correlated with incident MI, stroke, and CHF. After propensity score adjustment, the association was in the expected direction.

DMI for Serum Lipids

Increase in the quality of lipid management as measured by the DMI were not associated with change in incident CVD. It was associated with small decreases in health care utilization that was statistically significant for outpatient visits but not for emergency department visits. The failure to find an association with incident CVD given the strong epidemiologic and moderate clinical trial evidence that one exists, suggests the need for an alternative measurement strategy. There are several possible reasons that the expected results were not observed:

1. Insufficient variability in DMI scores
2. Insufficient lag time between measured service performance and health outcomes
3. Large differences in efficacy of the various approaches to lipid management in terms of their impact on CVD risk.
4. An inappropriate LDL standard (e.g., the LDL standard is too low and more aggressive treatment might show an effect).

Both sites had very high and similar PCP-level DMI scores for serum lipids that remained within a narrow range for the entire observation period. There was somewhat greater variation in site A. Transformed values that inflate near goal differences as described above might improve the predictive power of the index.

DMI for HbA_{1c}

The DMI for HbA_{1c} effects for MI, stroke and ER visits were not statistically significant. The effects were statistically significant for both CHF hospital days and office visits but were strongly in an unexpected. HbA_{1c} DMI scores were substantially higher at A than at site B

during the entire observation period. Practice variation was also wider in at site B. There was substantial variation at both sites over the observation period.

DMI for Beta Blockers and ACEI/ARBs

The PCP-Level DMI scores for beta blockers and for ACEI/ARBs measured continuous medication availability in patients with a prior history of MI and CHF respectively. Inpatient stay days and days prior to the initial MI or CHF diagnosis during the defined interval were excluded from both the numerator and denominator. Little longitudinal variation was observed at either site for either class of medications. For beta blockers, except for the first year in site A the inter-quartile range was consistently near 4% to 6% over time: from 82% to 88% at site A and from 85% to 89% at site B. ACEI/ARB scores were equally consistent over time and remained within an similarly narrow and slightly lower range at site A: inter-quartile range covered approximately 6% across all years from approximately 80% to 86%. Greater adherence to Beta Blocker in patients with a history of MI was associated with an increase in office visit suggesting the need for an improved propensity score. Greater adherence was not associated with improved CVD or other care utilization outcomes in either class of medication.

PI for BP Screening

The PI for BP screening was associated with a very small but statistically significant increase in MI incidence and a larger statistically significant increase in office visits. This suggests that the need for more adequate control of confounding by indication. Scores in site B were somewhat higher than site A which only had data available from 2005 to 2007. Data from 2005 at site A were excluded from analyses because of integrity problems.

PI for Serum Lipid Screening

There was wide variation in the PI scores for lipid screening both within and across years (Figure 8). The effects were in the expected direction but not statistically significant for stroke, CHF, and outpatient visits. Higher PI scores were associated with a small but statistically significant reduction in ER visits. The 2008 HEDIS LDL screening scores (which include total cholesterol) for sites A and B, respectively, were 87 and 91. The 2008 PI scores which include only LDL levels were 28 for site A and 42 for site B. In a previous study (Vogt 2007), the 2002 L-PI including total cholesterol screening for site A was 52 while this analysis, re-applying the methodology, but excluding total cholesterol tests and including only those with an LDL measure gave a score of 25. By inference, about a little less than half of the lipid screening being done involves a separate measurement of the LDL levels.

PI for HbA_{1c}

The effects for the PI for HbA_{1c} were statistically significant and in the expected direction for stroke. The effects were in the wrong direction and statistically significant for both care utilization outcomes. Scores were consistently extremely high and showed very little cross-sectional or longitudinal variation.

PI for Weight Loss

The variations PCP-level PI score for weight loss were striking. In site B, 60 to nearly 80 percent of obese patients under age 65 with no serious co-morbidities in the highest-performing practice lost weight in each year (data not shown). The PI scores of PCPs with the lowest scores ranged from about 20 to 40 percent with the mean just below 50 percent, and the middle three quintiles clustered tightly between about 42 percent to about 50 percent. In Site A, for the two years for which the PI for weight loss could be calculated (2007-08), the scores ranged from a low of just above 30 percent to a high of 69 percent (data not shown). There was a small but statistically significant increase in ER visits associated with higher PI scores for weight loss.

PI for Tobacco Cessation

The PCP-level PI score for tobacco varied from as many as 30 percent of their smoking patients quitting over a single year to zero. Not all of these quits are permanent and the measure as defined in this study was not predicated on sustained cessation for periods exceeding one year. Further investigation is needed in two areas: 1) the variation in duration of smoking across practices and the association of that variation with outcomes (among the subset of individuals with very long follow-up), and 2) the identification of what high-performing physicians are doing to support cessation among their patients. The PI for tobacco cessation was not significantly associated with changes in incident CVD or health care utilization.

Conclusions

Validity of PI and DMI methods – Inappropriate care has many causes. Some are difficult for health systems to address: inadequate resources, poor quality clinical guidelines, inaccurate or nonspecific diagnostic tests. Other causes can be addressed. Failure to receive recommended care may result from organizational deficiencies, clinician failure to recommend appropriate services, or from patient refusal to follow recommendations. In settings where EMR data are accessible to providers and patients, the PI and DMI evaluate care based on the same information available to the parties who are accountable for care. Quality indices based on that information are guaranteed a basic functional validity. Several of the indices developed in the project show sufficient promise to warrant additional development.

References

ADA (American Diabetes Association). Treatment of hypertension in adults with diabetes. *Diabetes Care* 2003;26, Suppl 1:580-82.

(The) ACCORD Study Group. Effects of Combination Lipid Therapy in Type 2 Diabetes Mellitus. *NEJM*, 2010; 362:1563-1574.

Amery A, Birkenhager W, Brixko R, Bulpitt C, Clement D, Deruyttere M, De Schaepdryver A, Collery C, Ragard R, Forette F, Forte J, Hamdy R, Henry JF, Joossens JV, Leonette G, Lund-Johansen P, O'Malley K, Petrie JC, Strasser T, Tuomilehto J, Williams B. Efficacy of antihypertensive drug treatment according to age, sex, blood pressure, and previous cardiovascular disease in patients over the age of 60. 1986, *The Lancet*;589-92.

- ATP III Guidelines Report of the National Cholesterol Education Program (NCEP) of the National Heart, Lung, and Blood Institute. (USDHHS, NIH Pub. No 01-3305; May, 2001)
- Breslow NE, Clayton DG. Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*. 1993. 88; 9–25.
- Borello F, Beahan M, Klein L, et al. 2003, *Rev Cardiovas Med*;3:Suppl3:S13-S24.
- Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373-383
- Fiore MC, Bailey WC, Cohen SJ, Dorfman SF, Goldstein MG, Gritz ER, Heyman RB, Jaen CR, Kottke TE, Lando HA, Mecklenburg RE, Mullen PD, Nett LM, Robinson L, Stitzer ML, Tommasello AC, Villejo L, Wewers ME. Treating tobacco use and dependence: a clinical practice guideline. U.S. Department of Health and Human Services (2000). Available at: <http://www.surgeongeneral.gov/tobacco>.
- Grundt SM, Cleeman JI, Bairey Merz CN, Brewer HB Jr, Clark LT, DB Hunninghake, Pasternak RC; Smith SC Jr, Stone NJ, for the Coordinating Committee of the National Cholesterol Education Implications of Recent Clinical Trials for the National Cholesterol Education Program Adult Treatment Panel III Guidelines. *ProgramCirculation*. 2004;110:227-239.
- Hornbrook MC, Hart G, Ellis JL, Bachman DJ, Ansell G, Greene SM, Wagner EH, Pardee R, Schmidt MM, Geiger A, Butani AL, Field T, Foucyzi H, Miroshnik I, Liu L, Diseker R, Wells K, Krajenta R, Lamerato L, Duda CN. Building a virtual cancer research organization. 2005, *JNCI*;30:12-25.
- Laino C. Aggressive Treatments Don't Help Diabetes Patients. 2010, WEBMD: <<http://diabetes.webmd.com/news/20100315/aggressive-hypertension-therapy-for-diabetes-patients>>.
- Hansson L, Lindholm LH, Ekblom T, Dahlöf B, Lanke J, Schersten B, Wester P-O, Hedner T, de Faire U, for the STOP-Hypertension-2 study group. 1999, *The Lancet*; 354:1751-56.
- Hypertension Detection and Follow-up Program Cooperative Group. Five-year findings of the Hypertension Detection and Follow-up Program. I. Reduction in mortality of persons with high blood pressure, including mild hypertension. 1979, *JAMA*; 242:2562-71.
- Joint National Committee. The seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. The JNC 7 report. *JAMA*. 2003;289:2560-2572.
- Kristeller JL, Hoerr RA. Physician attitudes toward managing obesity: Differences across six specialty groups. 1997, *Prev Med* 26:542-49.
- Liang KY, Zeger SL, Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*. 1986. 73; 13–22
- McCullagh P, Nelder JA. *Generalized Linear Models*, Second Edition, London: Chapman and Hall. 1989
- Melamed O, Nakar S, Vinker S. Suboptimal identification of obesity by family physicians. 2009, *Amer J Managed Care* 15:619-624.
- Neaton JD, Grimm RH, Jr, Prineas RJ, Stamler J, Grandits GA, Elmer PJ, Cutler JA, Flack JM, Schoenberger JA, McDonald R, Lewis CE, Liebson PR, for the Treatment of Mild Hypertension Study Research Group. 1993, *JAMA*; 270:713-24.
- Rosenbaum P, Rubin D. Estimating causal effects from large data sets using propensity scores. 1997, *Ann Intern Med*;127:757-63.
- Schwartz MB, Chambliss HO, Brownell KD, Blair SN, Billington C. *Obes Res*. 2003 Sep;11(9):1033-9.
- SHEP Cooperative Research Group. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension. Final results of the Systolic Hypertension in the Elderly Program (SHEP). 1991, *JAMA*;265:3255-64.
- Trafford JAP, Horn CR, O'Neal H, McGonicle R, Halford-Maw L, Evans R. Five year follow-up effects of treatment of mild and moderate hypertension. 1981, *Brit Med J*; 282:1111-13. Stevens VJ, Solberg LI, Quinn VP, Rigotti NA, Hollis JA, Smith KS, Zapka JG, France E, Vot T, Gordon N, Fishman P, Boyle RG. Relationship between tobacco control policies and the delivery of smoking cessation services in nonprofit HMOs. 2005; *J Natl Can Inst Moonograph* No. 35:75-79
- U.S. Preventive Services Task Force (USPSTF). The Guide to Clinical Preventive Services 2009. *Recommendations of the US Preventive Services Task Force*. Rockville, MD: Agency for Healthcare Research and Quality; 2009.
- U.S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2009;151:716-726.
- U.S. Preventive Services Task Force. Screening for high blood pressure: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2007;147-783-786.

Vogt TM, Aickin M, Ahmed F. The Prevention Index: Using Technology to Improve Quality Assessment. *Health Services Research*, 2004;39:511-30.

Vogt TM, Aickin M, Feldstein AC, Hu WR, Uchida AR. Electronic Medical Records and Prevention Quality: The Prevention Index. *Amer J Preventive Medicine*, 2007;33:291-296.

Vollmer WM, Kirshner M, Peters D, Drane A, Stibolt T, Hickey T, Tom GI, Buist AS, O'Connor EA, Frazier EA, Mosen D. Use and impact of an automated telephone outreach system for asthma in a managed care setting. *Am J Manag Care*. 2006 Dec;12(12):725-33.

Wolfinger R, O'Connell M. Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*. 1993. 4; 233-243.

List of Publications and Products

Manuscripts have been submitted reporting the results found for the DMI for blood pressure and another on the indices for weight loss and tobacco. Several other manuscripts are in preparation.