FINAL REPORT

Title: Developing an HIV-Specific Prevention Index Using the Electronic Medical Record

Principal Investigator: Richard T. Meenan, PhD, MPH, MBA, Center for Health Research - Kaiser Permanente, 3800 N. Interstate Ave., Portland, OR 97227. Phone: 503-335-2484. E-mail: richard.meenan@kpchr.org.

Team Members: Michael A. Horberg, MD, MAS, Michael J. Silverberg, PhD, MPH, Chun R. Chao, PhD, Andrew E. Williams, PhD, Mary Ann McBurnie, PhD, Vinutha Vijayadeva, PhD, William J. Towner, MD, Diana Antoniskis, MD, Thomas M. Vogt, MD, MPH, Erin Keast, Wendy Leyden, Beth Tang, Lanfang Xu, Carmen Wong

Organizations: Kaiser Permanente Center for Health Research

3800 N. Interstate Ave., Portland, OR 97227

Mid-Atlantic Permanente Research Institute

2101 East Jefferson Street, 3 West, Rockville, MD 20852

Kaiser Permanente Northern California Division of Research

2000 Broadway, Oakland, CA 94612

Kaiser Permanente Southern California Department of Research & Evaluation

100 South Los Robles Avenue, 2nd Floor, Pasadena, CA 91101

Center for Health Research – Kaiser Permanente

501 Alakawa Street, Suite 201, Honolulu, HI 96817

Inclusive Dates: 4/1/2010-1/30/2013. No-cost extension through 1/31/2014.

Federal Project Officer: Denise Burgess

This research was supported through a grant from the Agency for Healthcare Research & Quality (Grant Award Number: R01 HS017528-01) in response to PA-09-070, AHRQ Health Services Research Projects.

Structured Abstract

Purpose: This study (HIVPI) developed electronic medical record-based quality indices for 14 primary care services associated with standard care for patients with human immunodeficiency virus (HIV): CD4, VL, lipids, BP, flu, DTAP, Pneumovax, Hep A, Hep B, gonorrhea, chlamydia, CRC, PSA, tobacco.

Scope: Data were collected over an 8-year period (2003-2010) by four integrated healthcare organizations covering nearly 16,000 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). Whenever available, performance during the 8-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: HIV/AIDS, electronic medical records, quality of care, health IT, screening, prevention, disease management

PURPOSE

Quality performance measures are being developed that reflect the spectrum of HIV care, and they include testing for and diagnosis of HIV disease, access to care and follow-up, appropriate use of antiretroviral therapy (ART) and opportunistic infection prevention, and outcome measures for patients prescribed ART. The overarching objective of this study was to show where both general health prevention and HIV-related health prevention are deficient and where more systematic interventions to improve health maintenance among HIV-infected patients are needed.

SCOPE

Within developed countries, HIV-related disease is no longer rapidly progressive and invariably fatal but is now an increasingly manageable chronic condition. The use of ART and appropriate infection prophylaxis has contributed greatly to improvements in HIV-related mortality and morbidity, such as reduced incidence of Kaposi's sarcoma, non-Hodgkin's lymphoma, cervical cancer, and opportunistic infections (Keiser et al. 2004, Wood et al. 2006).

However, lives extended by ART create opportunities for other conditions to emerge, such as non-AIDS-related cancers, diabetes, cardiovascular disease, liver failure, renal failure, and chronic obstructive pulmonary disease (Lai and Hardy 2004). The increasing burden of heart disease, cancer, and other chronic diseases means that good-quality care of patients with HIV must include general preventive and management services, such as tobacco cessation counseling and colorectal cancer screening (Kaplan et al. 2002, Aberg et al. 2004). Yet, little is known about either the extent to which such services are currently used within HIV-infected populations or the extent to which their use is associated with the most contemporary ART use and beneficial outcomes (Reinhold et al. 2002, Sheth et al. 2006).

This study explored the feasibility of using the strong electronic medical record and data systems of Kaiser Permanente in identifying the association of general clinical care (as indicated by adherence to published guidelines) with HIV-specific and related outcomes. Our intent was to improve understanding of the influence of primary care practice characteristics on the quality of lifestyle intervention, prevention, and treatment of complications related to HIV/AIDS, morbidity, mortality, and the use of healthcare services. The mechanism for this exploration was the Prevention Index methodology developed at the Center for Health Research (and described below), a straightforward and efficient "person-time"-based approach to the assessment of care quality. The methodology has been tested and validated in the areas of mammography and Pap smears, among other preventive services, and has been applied to clinical areas such as diabetes, cardiovascular disease, and cancer. The Prevention Index methodology applies to such critical issues as patient safety, quality of treatment practices, and relative quality across healthcare systems.

Our objective was to adapt recently developed quality measures to the assessment of the overall quality of care received by HIV-infected adult patients within the Northwest (metro Portland, Oregon), Northern California, Southern California, and Hawaii regions of KP. We used the electronic data available in each region to define a sample of KP members ≥ age 18 who received an incident HIV diagnosis between 1/1/03 and 12/31/07 and who had at least 2 years of continuous membership prior to the incident diagnosis. Nearly 16,000 members across all four regions fit these criteria. We also used the electronic data systems to search for evidence in the medical record of various screening and other programs listed later in this report. Our intent was to use these and other electronic data as necessary to 1) develop the framework of the measures and preliminary estimates; 2) examine the association of the quality of care measures with clinical HIV outcomes; and 3) explore the association of practice variation with HIV quality of care.

METHODS

Because of the strong emphasis on methods in this study, this report describes methods in greater than usual detail. Publications will describe results in further detail.

Aim 1: Develop new composite quality of care measures for use in the management of patients infected with HIV.

Principles of constructing the Prevention (PI) and Disease Management Indices (DMI)

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 his or her data during a defined period.

- Begin with patient-level scores in each interval that, when appropriate, can be rolled up to create PIs and DMIs for higher-level units involved in the care of those patients: individual clinicians, clinics, departments, whole healthcare systems, 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 received a given vaccination during a calendar year). However, the PI scores for these services are still expressed as a percentage when they are rolled up to higher-level units such as clinic-years (e.g., the percent of smokers in a given clinic 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 is inapplicable for 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. 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 used to calculate the index. Both PIs and DMIs are ratios, but 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 (disease "severity"). PIs can be thought of a ratio of line lengths in which the lengths indicate covered and eligible time. DMIs can be thought of as a ratio of areas in which the areas represent ideal and actual levels of control of a clinical value related to disease management. The further away (in an adverse sense) the actual level of control is from the ideal, the greater the perceived disease severity. 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: 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---for example, the USPSTF's current recommendation that average-risk adults be screened for hypertension at least every 2 years.

Exceptions are when no commonly shared evidence-based guideline exists 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: Denominators of PI scores are the amount of time that a person was eligible for a service during a defined period, such as 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 that indicates the occurrence of a disease (e.g., cancer) or 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 healthcare 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. Similarly, in the HIV context, CD4 cell counts or viral load measures are indicators of the success of HIV-related therapy. 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 time period between measurements.

Though PIs are constructed exclusively as a ratio of the time a person was covered for a service to the time he or she was eligible for it, DMIs are constructed as the ratio of how well controlled the relevant clinical value is for that person during the period he or she had the disease. The time parts of these ratios are a concrete quantity (the time a person is observed), but the upper limit of the level of control part of the ratio can be set to any theoretical limit that does not exclude clinically meaningful variation.

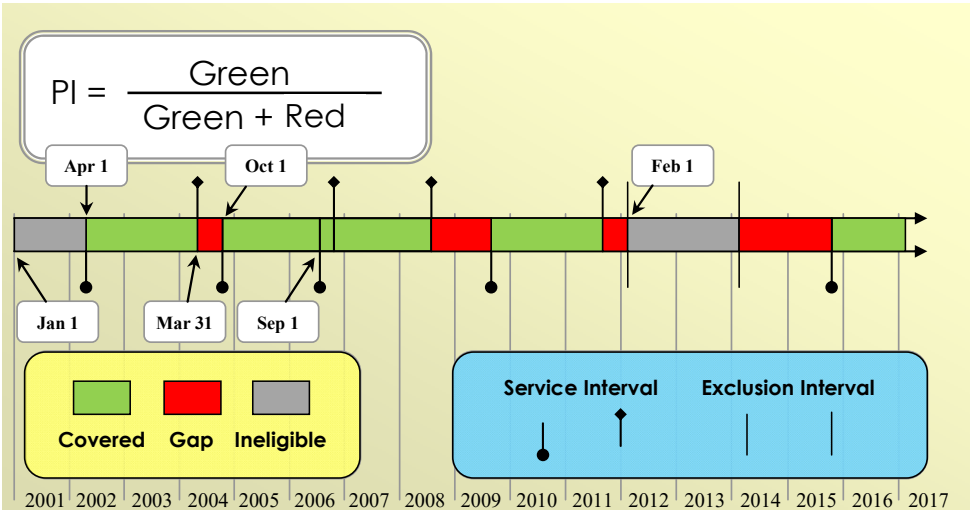


Figure 1. Elements of patient-level PI score calculations for each year of an observation period.

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 2 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 who 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 during periods that require frequent observation for a time in order to confirm or refute the presence of hypertension. Similar issues (and algorithmic solutions) apply as well to services such as colorectal cancer screening (e.g., colonoscopies can be considered preventive or diagnostic depending on the context in which they are administered).

The horizontal band in Figure 1 represents the entire observation period, which 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 contribute to neither numerator nor denominator. The patient with no history of hypertension or CVD enrolled in the plan on the first day of the observation period: January 1, 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 is insufficient 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 2 years of person-time during which she was covered for the service. This period ended on March 31, 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 1, 2004, however. It was 6 months overdue. This red-colored 6-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 2 months before that date on July 30, 2006. This 2-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 1, 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 2-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 2 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 more aggregate levels, such as a clinician's patient panel or a clinic population, it is worth considering whether a minimum number of data points should be required.

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 (severity). 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 is a comparison not 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.

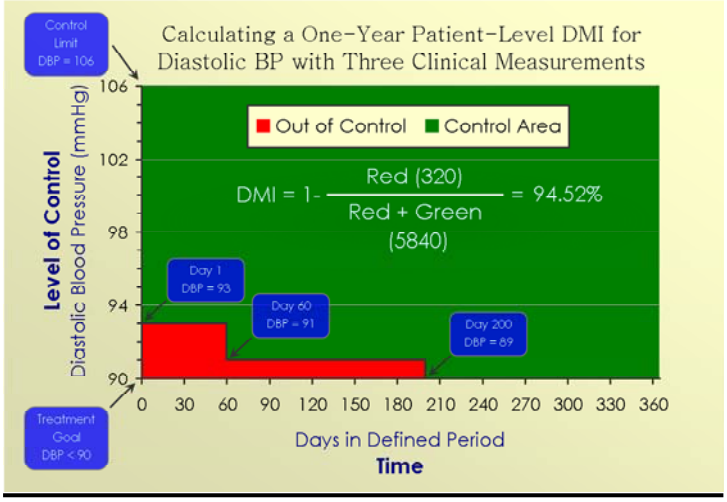


Figure 2. Elements of patient-level DMI score calculations for one year based on three diastolic blood pressure measurements.

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 mean of 91 mmHg. Values greater than three 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 1-year period for which the DMI is to be calculated. It was 3 mmHg above goal on day 1, 1 mmHg above goal 60 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 x 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%.

Using transformed clinical values to optimize a DMI score: There is nothing sacrosanct about the use of raw clinical values to calculate DMI scores. Their numeric representation 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.

Clinical and preventive services evaluated: Table 1 lists the 14 services we attempted to evaluate using the PI/DMI methodology. Some services, such as lab tests or blood pressure measurements, could be evaluated using either the PI (was the test delivered at appropriate intervals?) or the DMI (what was the overall level of control of the clinical metric?).

Table 1. Clinical and preventive services considered

Lab tests
CD4
Viral load
Lipids
Vaccines
Hepatitis A
Hepatitis B
Influenza
Pneumovax
Cancer screening
Colorectal cancer
Prostate cancer
Tobacco status and counseling
Other screening
Hepatitis C
Chlamydia/gonorrhea
Syphilis
Blood pressure

Data requirements for PI or DMI calculation: 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 those data efficiently will vary by the service.

Data extraction: We relied as much as possible on EMR data that had met preliminary quality assurance and standardization by virtue of 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. We also took advantage of the HIV registries maintained at both KPNC and KPSC (KPNW and KPH do not currently have such registries).

Developed in 1989, the KPNC HIV Registry includes all known cases of HIV infection dating back to the early 1980s (17,630 cumulative KPNC members). Members are initially identified as possibly HIV infected if they have one or more of the following indicators identified from electronic databases: 1) positive HIV antibody test; 2) detectable HIV viral load; 3) CD4/CD8 ratio < 1.0; 4) prescription for any ARV drug; 5) outpatient documentation of HIV infection; 6) HIV/AIDS discharge diagnosis; 7) pathology report for Kaposi's sarcoma or *Pneumocystis carinii* pneumonia; and 8) Centers for Disease Control & Prevention AIDS case report forms, infection control nurse notes, and internal physician reporting. Medical chart confirmation of potential cases is done in KPNC, and the KPNC case ascertainment methodology is estimated to have >95% sensitivity and >99% specificity based on comparisons of the registry and clinician HIV case lists. Data elements maintained in the HIV registries include gender, race/ethnicity, HIV risk group, dates of known HIV and AIDS diagnoses, and date of death.

Established in 2000, the KPSC HIV registry includes all known cases of HIV infection among KPSC members between January 2000 and the present date. In total, 13,392 members have one or more possible HIV indicators, and, of those, 8,314 (62%) meet the minimum criteria to be classified as HIV cases (or presumed cases). Among those, approximately 60% are currently active (10% female). The registry is updated quarterly. Patients are initially identified as possibly infected with HIV if they have one or more indicators from various electronic databases: 1) positive HIV Western blot test; 2) HIV/AIDS confidential case report; 3) one detectable viral load test or two undetectable tests > 90 days apart; 4) CD4/CD8 ratio < 1.0; 5) outpatient diagnosis of HIV infection; 6) hospital discharge diagnosis of HIV infection; 7) prescription filled for PI, NNRTI, NRTI, or FI (excluding lamivudine 100 mg only); or 8) AIDS-defining condition. An HIV case is any patient with a positive HIV Western blot test, or an HIV date or Western blot date in the confidential report. Presumed cases (possible HIV infections) are identified if they meet any of the following conditions: 1) viral load test criteria alone; 2) at least two other indicators from (4) to (7) above except where the only two flags are (1) CD4/CD8 ratio < 1.0 and (2) AIDS-qualifying condition that is only a CD4 count < 200; or 3) any single indicator from (4) to (7) above plus an AIDS date (as the only available information) in the registry.

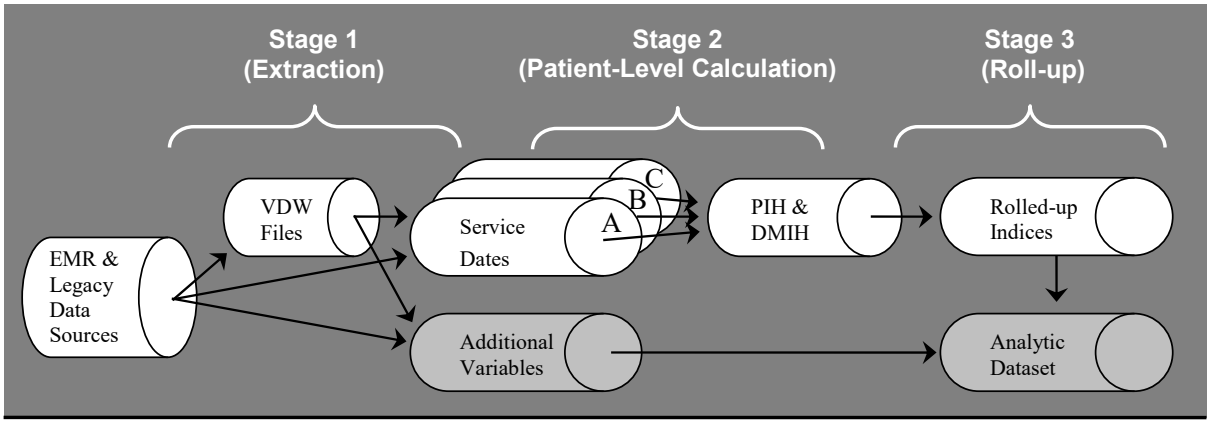


Figure 3. Dataflow leading to the final analytic dataset.

We extracted the three types of data listed in Stage 1 above to calculate PI and DMI scores. Most of these data were extracted from either a Virtual Data Warehouse (VDW) that harmonizes data from disparate sources at all participating sites or for KPNC and KPSC, from their local HIV registries. We rolled up the patient-level data and combined these with other variables to create the analytic dataset.

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. Variables that were out of range were winsorized, and 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. These standardized requests were then executed locally at each participating site, and data were sent through secure file transfer to the KPNW Center for Health Research for compilation and validation. Standardized queries and tabular reports were developed to assess data integrity of the preliminary and of the patient-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 3.

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 large variety of 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 ensure their validity and appropriateness. Table 2 lists the rules we applied to each service regarding appropriate service interval, length of prior eligibility required, and permanent and temporary exclusions.

Single lab values or other measurements, such as CD4 count values or HIV RNA levels, 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 healthcare 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 was counted as a single encounter.

Our final study population totaled 15,950 KP HIV-infected members across all four data sites. Eligible members were age 18 or above at 1/1/03 or enrollment date and had to be enrolled 24+ months between 1/1/03 and 12/31/10, with 3-month eligibility gaps patched. Utilization data were collected between 1/1/01 and 12/31/10 to allow for observation periods before 2003. Our study population of KP HIV-infected members was 90% male, 56% White, 14% Black, 18% Hispanic, and 3% Asian; note, however, that 24% of eligible members were of unknown race and 49% were of unknown ethnicity, because such fields were not populated in the KP system during our study period.

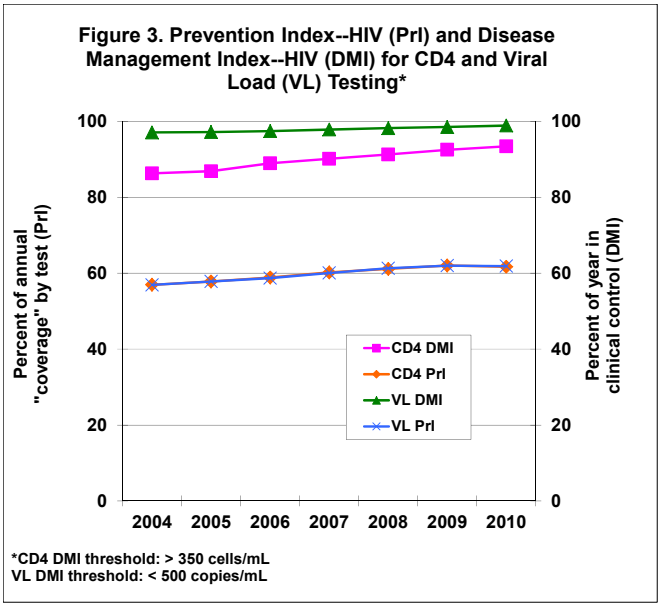
Table 2. List of permanent and temporary exclusions by service

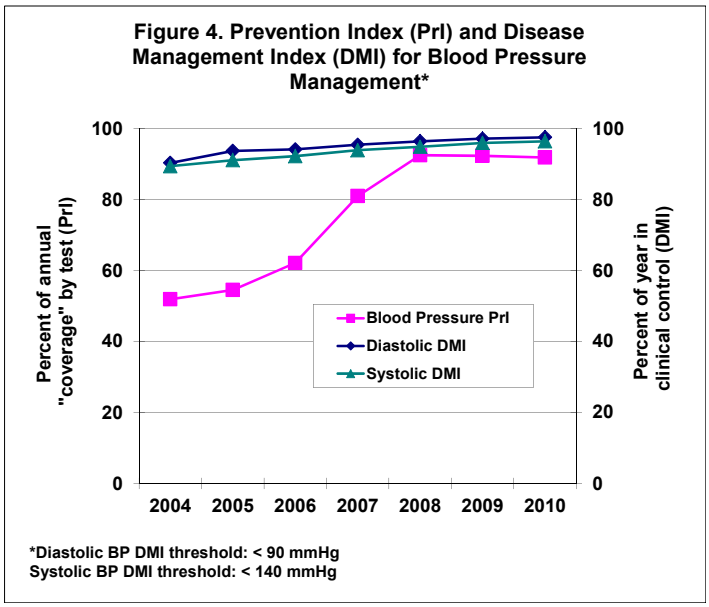
	Maximum interval between services	Mos. prior elig.	Permanent exclusions	Temporary exclusions
CD4				
All, age 18+	at least twice annually; two measures at least 60 days apart	12	None	None
Viral load (HIV RNA)				
All, age 18+	at least twice annually; two measures at least 60 days apart	12	None	None
Lipid screening				
Male, age 35-65; Female, age 45-65	Annual	12	Diabetes CVD CHF	Chest pain Syncope Leg pain
Hepatitis A vaccine				
All, age 18+	Once, then 6-12 months	12	Prior Hep A infection	None
Hepatitis B vaccine				
All, age 18+, negative for HBV antigen/antibod	3X in 6 months	12	Chronic HBV infection	None
Influenza vaccine				
Age 65+; high risk, age 18-64	Annual	12	None	None
Pneumococcal vaccine				
Age 65+; high risk, under age 65	One time	12	None	None
Colorectal cancer screening (various)				
All, age 50-80	FOBT, 1 year	12	CRC	Rectal bleeding
Data extracted 40+	Sigmoidoscopy, 5 years	60	Polyp w/o removal	Polyp removal
Age range 50-75	Colonoscopy, 10 years	120	Polyposis	Abdominal, gastric, colon pain
	Barium enema, 2 years	24	Ulcerative colitis Crohn's disease	
Prostate cancer screening (PSA)				
Male, age 50+	Annual	12	Prostate cancer	Abnormal PSA test PSA within 90 days of previous Tests recurring in 9 months
Tobacco status and counseling				
All, age 18+	Check smoking status every 5 years Counsel smokers yearly	12	Nonsmokers	None
Hepatitis C screening				
All, age 18+	At least once; annually if IDU or MSM as HIV risk	12	None	None
Chlamydia/gonorrhea screening				
Female, age 14-40; Male HIV+, age 18+	Annual	12	None	None
Syphilis screening				
All, age 18+	Annually if MSM, at least once if female	12	Heterosexual male (but at least once)	Heterosexual male (but at least once)
Blood pressure screening				
All, age 18+	2 years	24	Hypertension dx/rx	Chest pain ER, hospital, urgent care BPs

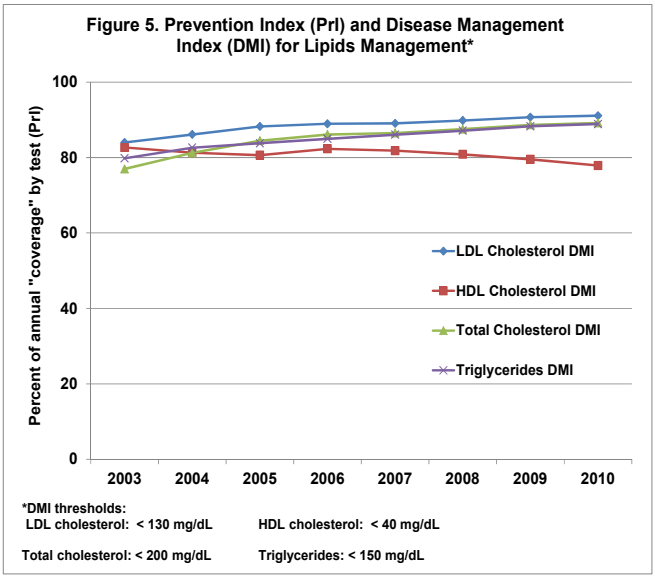
RESULTS

LAB-BASED MEASURES

The following figures illustrate trends in the performance of the PI and/or DMI across multiple services and years. As Figures 3-5 indicate, between 2003 and 2010, the quality of care of KP members with HIV within the four data sites was very good in terms of level of control of various clinical measures: CD4 count, viral load, blood pressure management, and lipids. For all such measures, the associated mean patient-level DMI was well above 80%, meaning that, for a given calendar year, the average amount of time that HIV-infected patients were in control of a given clinical measure was above 80% (292 days).

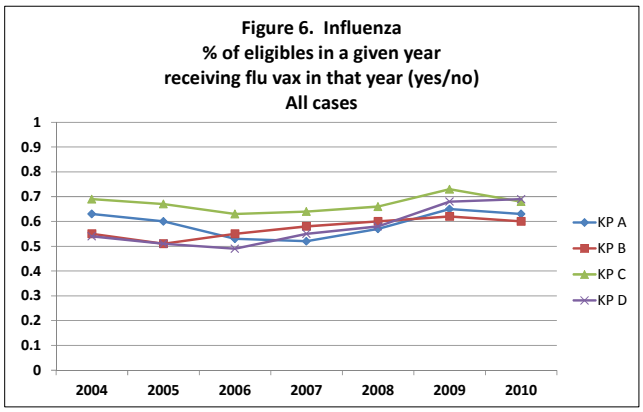


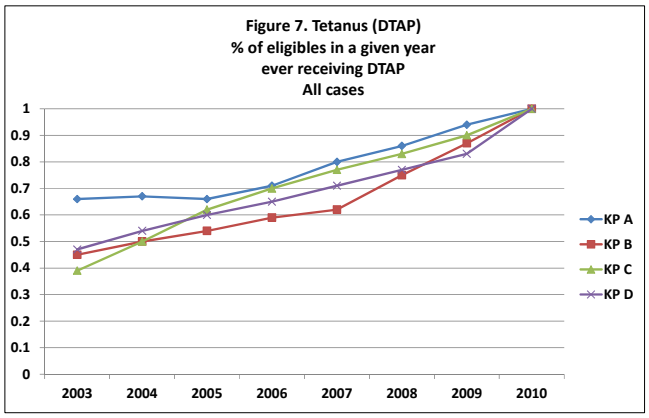


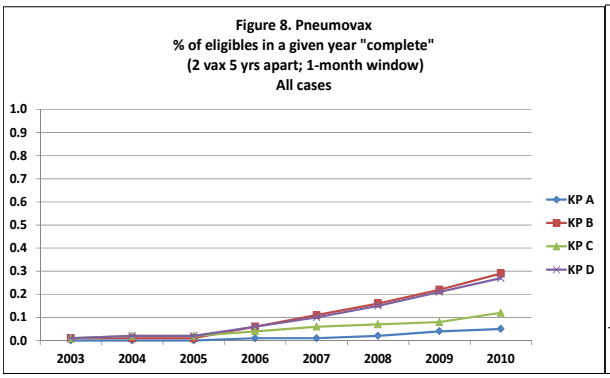


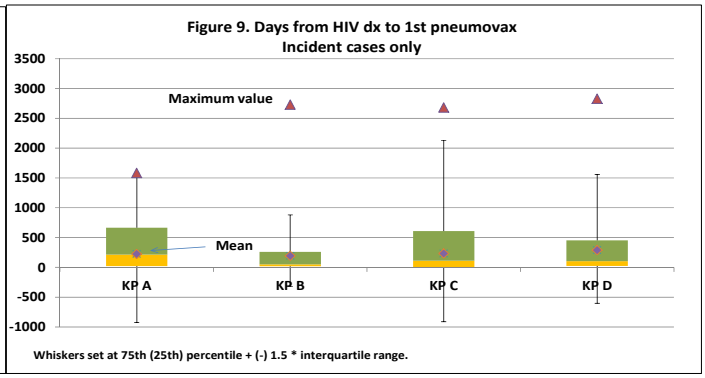
VACCINES

Figures 6-11 describe the experience of the participating sites between 2004 and 2010 in terms of administering recommended vaccinations to their HIV-infected patients. We initially considered applying the PI methodology directly to vaccination services, but we determined that it was more straightforward to apply a simple yes/no metric (i.e., how many eligible patients received a given vaccination in a given year?). Figure 6 indicates that, for annual influenza vaccine, approximately 60% of eligible patients at each site received an recommended annual "flu shot." Over time, most HIV patients in KP received a tetanus shot (Figure 7). Relatively few patients were "complete" for pneumococcal vaccine (Figure 8), but this is at least in part due to a relatively narrow window of time (1 month) for the second vaccination to occur and be counted as "on time." Also, the recommended 5-year interval between vaccinations very likely reduced the number of eligible patients in our population who could be recognized as "complete" under any circumstances over our 2003-2010 study period. Figure 9 shows that across sites, the average number of days from incident HIV diagnosis to first pneumovax was 250-300 days over the study period.

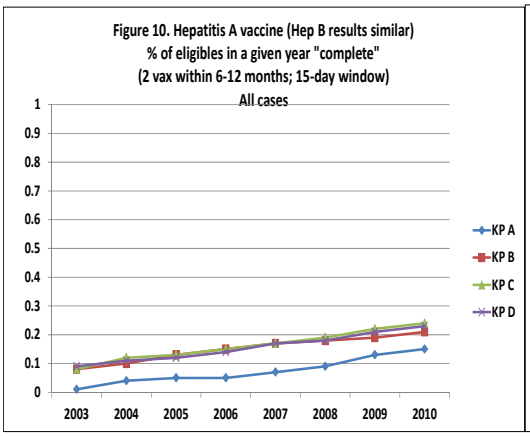


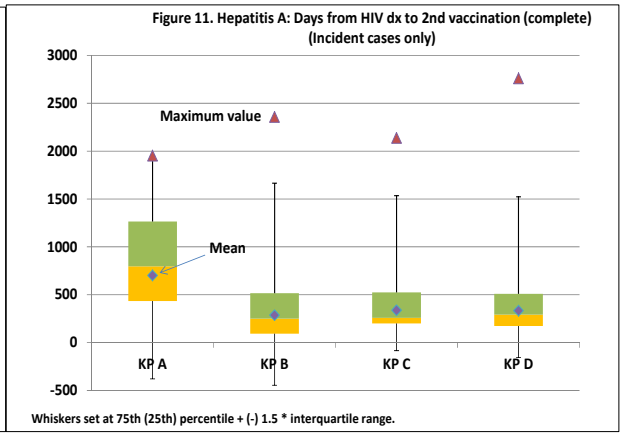






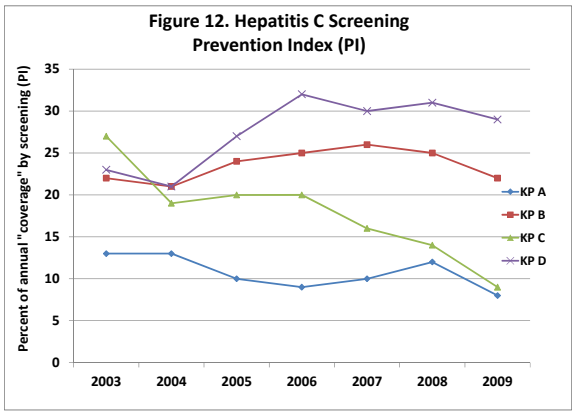
Results for hepatitis A and B vaccinations were similar (Figure 10; Hep B not shown). Relatively few eligible patients were found in the electronic data to have received hepatitis vaccinations. However, this is at least in part due to the relatively narrow 15-day window allowed for "completion" of the second vaccination, which very likely led to missing numbers of completed vaccination sequences. Figure 11 indicates that, at three of the four participating sites, completion of the two-vaccination sequence among 75% of incident cases occurred within 500 days of the incident HIV diagnosis.

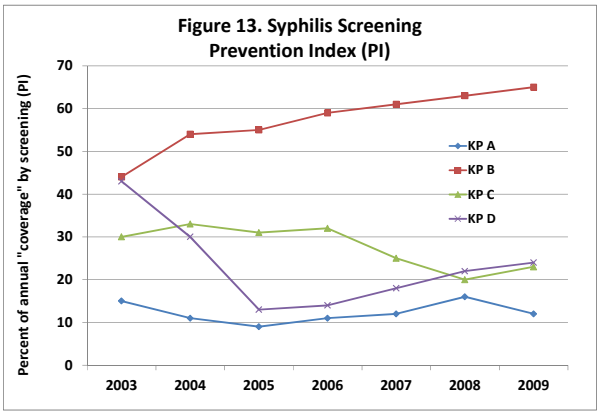


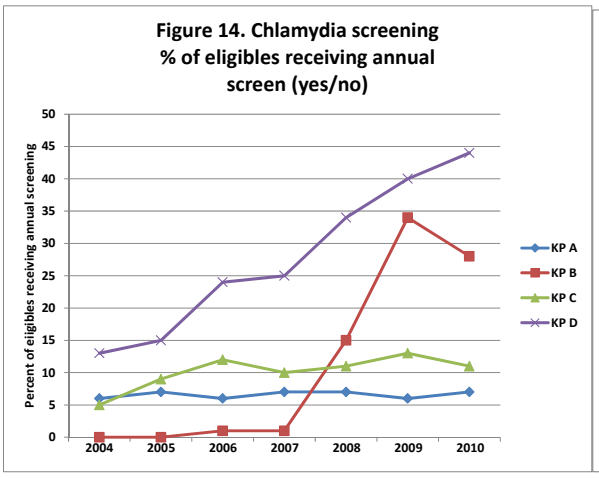


Other screening

Similar to vaccines, we applied a simple annual yes/no criterion to screening tests for sexually transmitted diseases (hepatitis C, syphilis, chlamydia/gonorrhea). Figures 12-15 indicate a wide variety of experience across sites regarding the delivery of these screening services; however, this reflects varying internal guidelines as well as variable EMR data capture. Note that our definition of syphilis screening was an older algorithm that applied to our study period: RPR (Rapid Plasma Reagin/VDRL (Venereal Disease Research Laboratory)) first, followed by a treponema test if positive. Although much of this variation is likely attributable to data issues, it is also likely due at least in part to site-specific policies or guidelines around the follow-up of STD screenings as well.







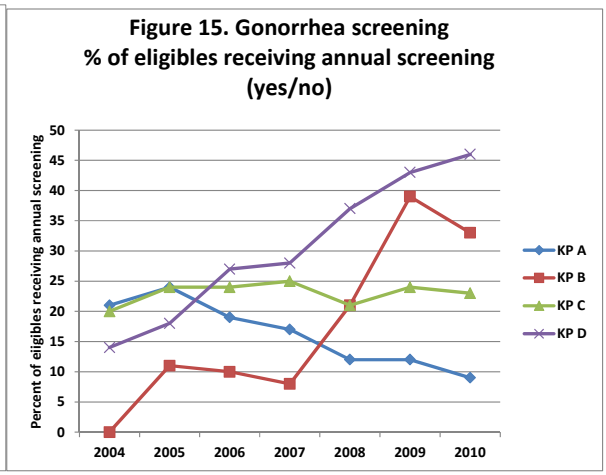
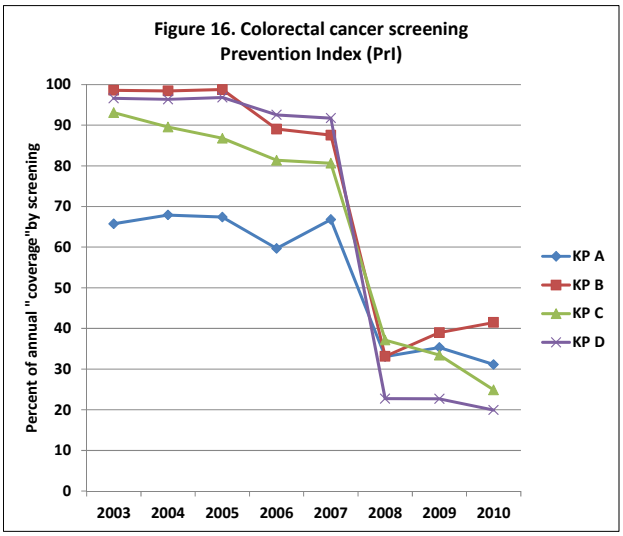
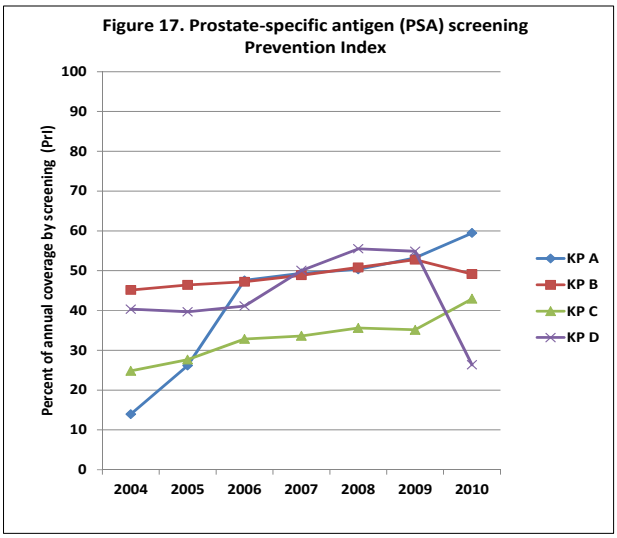


Figure 16 (Colorectal cancer (CRC) screening) and Figure 17 (PSA screening) represent the most salient examples of non-HIV-related cancer screening. CRC screening would appear to have dropped precipitously between 2007 and 2008, which reflects multiple factors, only some of which are captured in the electronic data. In our study, the CRC screening PI is the only one that actually reflects a group of screening methods (fecal occult blood test (FOBT), barium enema, flexible sigmoidoscopy, colonoscopy). During our study period, the proportion of CRC screens that were FOBTs rose substantially from 7% in 2003 to 60% in 2010, creating the opportunity for longer gaps in coverage when patients neglect to repeat the FOBT annually. Perhaps an even greater factor in the reported decline in CRC screening is that our data capture did not include the increasingly popular fecal immunochemical test, which came online within KP midway through 2009. The implication is that KP's true performance in providing CRC screening to its HIV-infected members may well have not declined much if at all in recent years but reflects a shift to newer and/or more convenient forms of screening.

Annual PSA screening (limited to men age 50-70) exhibits only a slight upward trend within most sites.





Aim 2. Examine the association of the quality of care measures from Aim 1 with clinical outcomes. Assess the temporal trend of the estimated service-specific and summary PIHs and DMIHs across years (e.g., 12 months post-dx, 24 months post-dx) as an indication of improvement or deterioration in the quality of care for patients with HIV. Assess whether higher levels of the 12-month post-dx PIHs and DMIHs, modified by demographic and clinical characteristics, predict use of the latest anti-retroviral therapy (ART) regimens, adherence to and/or changes in such regimens, and changes in selected HIV-related outcomes, such as developing a new AIDS diagnosis.

Analytic strategy. Hypotheses were tested on patient-level data pooled from all 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). GLMM use linear predictors, a monotonic mapping of predictors to the data mean, and an outcome distributed as any of the family of exponential distributions – Beta, Binomial, Gamma, Normal, Lognormal, Poisson, etc. Models used to test hypotheses for each service specified a Poisson distribution for incident outcomes and a log link function between predictor and outcome variables. 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 AIDS diagnoses, other clinical outcomes, or healthcare utilization for a 10% increase in the PI or DMI score for the service. We focused on the lab-based and clinical tests (CD4, HIV RNA, lipids, and blood pressure) because we did not have sufficient outcomes in our data during our study period to make analyses of the other services (e.g., CRC screening) meaningful.

Table 3 lists our analytic results. Unfortunately, we were unable to establish any statistically significant associations between the PIs and DMIs for CD4, HIV RNA, lipids, or blood pressure management. Certain results were suggestive (e.g., CD4 screening and outpatient visits, lipid management and outpatient visits), but none reached the level of statistical significance.

Table 3. Association of PI/DMI with outcomes

Service	Outcome	Beta	SE	p
CD4 screening	AIDS dx	0.031	0.099	0.17
(PI)	Outpt visits	0.342	0.171	0.11
	ER visits	0.011	0.0789	0.68
	Inpt stays	0.056	0.567	0.43
CD4 management	AIDS dx	0.002	0.012	0.21
(DMI)	Outpt visits	-0.435	0.234	0.38
	ER visits	0.024	0.0893	0.54
	Inpt stays	0.0046	0.158	0.25
HIV RNA screening	AIDS dx	0.041	0.356	0.13
(PI)	Outpt visits	0.45	0.257	0.24
	ER visits	0.367	0.289	0.22
	Inpt stays	0.067	0.456	0.34
HIV RNA management	AIDS dx	0.131	0.234	0.15
(DMI)	Outpt visits	-0.563	0.452	0.09
	ER visits	0.102	0.243	0.31
	Inpt stays	0.018	0.176	0.23
Lipid screening	Outpt visits	0.057	0.34	0.23
(PI)	ER visits	0.03	0.79	0.57
	Inpt stays	0.012	0.67	0.48
Lipid management	Outpt visits	-0.15	0.35	0.09
(DMI)	ER visits	0.04	0.47	0.36
	Inpt stays	-0.91	0.24	0.12
Blood pressure screening	MI	0.051	0.115	0.1
(diastolic)	Stroke	0.036	0.71	0.56
(PI)	Outpt visits	0.97	0.46	0.19
	ER visits	0.04	0.78	0.68
	Inpt stays	0.13	0.24	0.37
Blood pressure mgmt	MI	0.167	0.245	0.17
(diastolic)	Stroke	0.034	0.756	0.56
(DMI)	Outpt visits	0.802	0.048	0.23
	ER visits	0.451	0.892	0.61
	Inpt stays	0.561	0.671	0.56

Aim 3. Study the association of practice variation with the developed quality of care measures.

This aim proved to be impractical to address given the organization of HIV care within the participating KP regions. Unlike standard primary care, HIV care with KP is overseen by a much smaller set of physicians and other providers (e.g., the KP Northwest Immune Deficiency Clinic, which oversees HIV care in KP Northwest has four physicians, two of whom joined KPNW in the middle of our study period).

DISCUSSION

We feel that the PI methodology has potential to inform quality of care issues in HIV; for example, the ability to roll up indices to create measures for individual physicians, clinics, or even an entire healthcare system is a genuine advantage of the method. However, our ability to derive more robust results despite the great efforts that were given to data collection, cleaning, and validation was hampered by two primary factors. First, previous studies that have employed the PI methodology have used entire (adult) health plan memberships. Explorations of quality of care in the management of diabetes, cardiovascular disease, and cancer, for example, all had study populations in the hundreds of thousands of KP health plan members that were empaneled across a wide variety of providers and clinics. In this study, even including KP Hawaii, which was added as a data site subsequent to initial funding, the four KP regions that participated as data sites (Northwest, Northern California, Southern California, and Hawaii) yielded a study population of only 15,950 HIV-positive KP members. Our ability to assess the association of indices, whether PI or DMI, with related clinical outcomes, was limited in many instances because of the infrequency of such outcomes, even over a 10-year observation period. It may be feasible in the case of colorectal cancer screening, for example, to identify sufficient cases of colorectal cancer across an entire health plan membership of, say, 400,000, to make meaningful statements about the association of adherence to screening guidelines and clinical outcomes. However, in our limited sample, the incidence of colorectal cancer was far too infrequent to permit such an analysis.

The second factor that impeded our ability to derive more meaningful results was that we conducted the study inside the KP healthcare system in which the "baseline" quality of care for HIV-infected patients is already quite high. Such an exercise might gain greater purchase in a community healthcare system.

We determined that index construction is feasible using EMR data within the HIV-infected population, but it is subject to data quality and idiosyncrasies. The concept of a time-based index (PI) or even a time-plus-severity-based index (DMI) may be understandable in the abstract; however, its execution using real data recording actual events requires numerous underlying definitions and assumptions. For example, is a particular colonoscopy truly preventive or is it diagnostic, given a previous indication? What should those indications be that would cause a colonoscopy to be considered diagnostic (or confirmatory), and therefore excluded from the PI calculations? If it is confirmatory, how long should the period of temporary exclusion be before a patient's time becomes eligible for inclusion again? In the case of vaccinations, especially those for which a series of two or more is required to indicate complete "coverage," what is an appropriate window of time around a specified due date? For example, guidelines around administration of pneumococcal vaccine recommend two vaccinations 5 years apart. After consultation with our clinical experts, we chose a 1-month window around the 5-year "due date" for the second vaccination as evidence of its administration, and by definition, a completed pneumovax series. Of course, a wider time window around the second vaccination due date would have increased the likelihood of completion and raised the numeric value of the index, and vice versa. Results were similar for evidence of both Hepatitis A and B vaccines for which we used a 15-day window around the required due date. The reported indices for these vaccines would likely have been higher if we had used a wider time window (e.g., 30 days) for capturing evidence of vaccination.

Also, relying on EMR data from one healthcare system implies that any preventive or other services that are obtained out of plan will not be captured, an issue that is likely to be especially salient in community systems. In our study, we felt that this was not a significant concern and that our data were reasonably comprehensive in representing the various services that KP HIV-infected members would have obtained. The philosophy of HIV care within KP is that the health plan will strive to provide all necessary services for its members. This may not always be true in more community-based settings, in which patients are referred to other groups for various specialty services. EMR data may also be incomplete in the presence of enrollment gaps (i.e., the patient leaves the health plan for some period, perhaps because of moving out of a covered service area). Again, however, we do not feel that this was a significant issue in our data, because most HIV-infected KP members tend to stay with their clinician and medical team over many years because the continuity of care is important. This may be more of a concern in community settings in which patients are more likely to move between clinics or systems that are challenged to share data. The larger point is that the development of Prevention and Disease Management Indices requires many decisions in their design that can easily vary from system to system. The choice of comparative guideline for a given preventive service may vary, as well as the service interval and exclusion criteria.

Much work remains to determine the ultimate utility of the PI and DMI in the HIV context. Any quality of care measure is only valuable if it is used. Considerable education and consultation are needed to enhance the receptivity of providers to the potential of such indices to produce useful and, perhaps most important, timely data.

The PI and DMI as they have been developed are currently retrospective in nature. Our study focused on historical data from 2003 to 2010; healthcare providers, including clinicians caring for a population of HIV-infected patients, want information in real time. How to implement the PI and DMI calculation within a data system that permits regular, systematized updating is a future challenge for the methodology. The good news is that the continued evolution of the EMR and related administrative data systems will make such real-time generation and application of the PI and DMI increasingly feasible.

Publications

Manuscripts have been submitted, reporting the results found for the PI and DMI for CD4 and viral load, and another on the overall data-related challenges of implementing the methodology in the HIV context has been submitted as well. Other manuscripts are in preparation.