The Prevention Index: Using Technology to Improve Quality Assessment

Thomas M. Vogt, Mikel Aickin, Faruque Ahmed, and Mark Schmidt

Objective. To improve quality of care assessment for preventive medical services and reduce assessment costs through development of a comprehensive prevention quality assessment methodology based on electronic medical records (EMRs).

Data Sources. Random sample of 775 adult and 201 child members of a large nonprofit managed care system.

Study Design. Problems with current, labor-intensive quality measures were identified and remedied using EMR capabilities. The Prevention Index (PI) was modeled by assessing five-year patterns of delivery of 24 prevention services to adult and child health maintenance organization (HMO) members and comparing those services to consensus recommendations and to selected Health Plan Employer Data and Information Set (HEDIS) scores for the HMO.

Data Collection. Comprehensive chart reviews of 976 randomly selected members of a large managed care system were used to model the Prevention Index.

Principal Findings. Current approaches to prevention quality assessment have serious limitations. The PI eliminates these limitations and can summarize care in a single comprehensive index that can be readily updated. The PI prioritizes services based on benefit, using a person-time approach, and separates preventive from diagnostic and therapeutic services.

Conclusions. Current methods for assessing quality are expensive, cannot be applied at all system levels, and have several methodological limitations. The PI, derived from EMRs, allows comprehensive assessment of prevention quality at every level of the system and at lower cost. Standardization of quality assessment capacities of EMRs will permit accurate cross-institutional comparisons.

Key Words. Prevention, quality assessment, electronic medical records, priorities

For the past decade, the Health Plan Employer Data and Information Set (HEDIS) has been used to evaluate the quality of outpatient care in many large managed care systems. The HEDIS has profoundly influenced the way preventive care is delivered (Schneider et al. 1999). Although HEDIS established science as the basis of quality assessment (McDonald 1999), existing

systems impose substantial limitations on quality assessment. The high cost of chart review and the lack of consistent methods for summarizing quality create barriers to effective quality assessment (Eddy 1998; McGlynn et al. 2003).

Within 10 to 15 years, electronic medical records (EMR) systems will probably become a standard of care in the United States (McDonald 1999; Schneider et al. 1999). These systems offer extraordinary opportunities to improve care through monitoring quality and placing key information at the fingertips of clinicians and managers. Because many health care systems are currently in the process of developing and implementing electronic medical records (EMR), this is a critical time for developing standardized formats for quality assessment and reporting. This paper proposes a methodology for using EMRs to assess quality of preventive care in the form of a Prevention Index (PI). Using this methodology, we created the PI by conducting paper chart reviews of 976 randomly selected charts from a large HMO. The PI is compared to HEDIS measures for the same services in the health plan and the same year. All of the data required for creating the PI can be captured by EMR systems if appropriately programmed. The approach is systematic, inexpensive, addresses the flaws of current quality measures, and includes all prevention services rated as effective by consensus groups (e.g., U.S. Preventive Services Task Force).

LIMITATIONS OF CURRENT SYSTEMS

Current prevention quality assessment is impaired by the following eight problems:

1. Because of limited access to electronic data, some high priority prevention services cannot be assessed. Surveys and chart reviews are expensive. Although counseling on risk behaviors is one of the most important prevention activities in a physician's office (U.S. Preventive Services Task Force 1996),

Address correspondence to Thomas M. Vogt, M.D., M.P.H., Program Director, Kaiser Permanente Center for Health Research, Hawaii, 501 Alakawa St., Suite 201, Honolulu, HI 96817. Mikel Aickin, Ph.D., is a Senior Investigator, Kaiser Permanente Center for Health Research, Portland, OR. Faruque Ahmed, M.D., Ph.D., is Chief (Acting), Health Services and Policy Research Branch, Division of Prevention Research and Analytic Methods, Epidemiology Program Office, Centers for Disease Control and Prevention, Atlanta. Mark Schmidt, B.A., is a Programmer/Analyst, Kaiser Permanente Center for Health Research, Honolulu, Hawaii.

performance measurement systems rarely assess services that require a conversation between patient and clinician. The HEDIS measures only one such service (counseling of smokers), and that measure depends on the use of a patient survey. The survey methodology is costly and is associated with recall and nonresponse biases. It may also include advice from physicians that was not recorded in the medical record and, thus, would not be available to guide the next interaction of patient and clinician. The cost and biases associated with survey data are a key barrier to more robust quality assessment (Neumann and Levine 2002).

2. Nonmeasured services are neglected. Services not measured because of lack of data access receive less attention and fewer resources than those that are measured (Neumann and Levine 2002). Few HMOs perform well on strongly recommended lifestyle and mental health assessment and counseling recommendations not included in HEDIS (Garnick et al. 2002).

The extraordinary improvements in childhood immunization, mammography screening, and most other covered services after HEDIS was initiated illustrate the positive impact of quality evaluation as a part of accreditation. However, recommended services not included among HEDIS measures have low performance levels (see below). This results in patterns of service delivery that do not necessarily reflect priority of the services. For example, counseling on diet and physical activity are rarely documented in the medical record despite an almost obsessive fixation on weighing patients (Vogt and Stevens 2003). Recently, a methodology for assessing priority of different preventive services in terms of their relationship to quality-adjusted life-years saved per dollar spent has been developed (Coffield et al. 2001; Maciosek et al. 2001). This methodology can be used to direct resources to services that have the most benefit at the population level.

3. Little attention is given to the special high-risk status of persons not served for prolonged periods of time. Programs that focus on the rarely screened are uncommon even though some screening services (e.g., Pap screening, colorectal cancer screening) have very long lead times because they detect pre-malignant conditions that are fully curable. This long lead-time (i.e., the interval between onset of an abnormality detectable by screening and symptom development) means even occasional screening has substantial benefit. Colditz, Hoaglin, and Berkey (1997) estimated that Pap tests at intervals of 10 years could achieve nearly two-thirds of the lifetime benefit of annual screening. Benefits for outreach to those rarely tested with screening

tests that detect extant disease (e.g., mammography) are not as impressive if performed infrequently because the lead-time is so much shorter. The cost effectiveness of mammography is highest when it is done every one to two years; the cost effectiveness of Pap smears is highest when performed every three to five years. This rarely recognized distinction is crucial to rational screening strategies with limited resources because the costs of screening are heavily influenced by the screening interval.

4. A standard method for accumulating scores into an overall quality of care index would facilitate cross-system comparisons and drive improvements in quality. Generally, quality of care assessments provide scores for dozens of different services. However, there is a lack of consistency in the manner different organizations aggregate performance scores for producing report cards, thereby leading to inconsistent rankings that confuse consumers.

5. Quality evaluations should facilitate the greatest gain in quality-adjusted life years given available resources. All preventive services are not equivalent. Some have minor impact; others have powerful influence over health. Some are highly effective among select subgroups of the population, but have little benefit among others. The optimal use of the consumer's money requires that systems do the most important thing first, and the next most important thing second, and so on. Yet, priorities are seldom given adequate consideration in assessing quality of care.

6. Current quality measures can be assessed only at the system level. The National Committee for Quality Assurance (NCQA) is seeking measures that can be assessed at the system, clinic, provider, or patient level. The only way to assess quality at all levels is to have data on all persons in the system. Current methodologies are too expensive to permit that approach. Electronic data systems make this possible at less cost than is currently expended on prevention quality assessment.

7. Systems need a standard methodology for being updated on the basis of new findings. New information leads to regular changes in recommendations. Such changes should be applied systematically to quality assessment based on the best information currently available.

8. Current systems do not distinguish between preventive and diagnostic and therapeutic services. A mammogram delivered to evaluate a large breast lump is not a preventive service. If a late-stage cancer is diagnosed, it represents a failure of the screening system. Yet, HEDIS methodology counts diagnostic and monitoring services as prevention "successes" because of the difficulty in differentiating between the services from administrative data. In the system we assessed, approximately 37 percent of breast and 13 percent of cervical cancer

screening tests, and about 20 percent of cholesterol tests were delivered as a follow-up to previous diagnoses, abnormal tests, or to evaluate symptoms (Vogt et al. 2001). Tests ordered as a consequence of signs and symptoms of illness or for monitoring diagnosed conditions are not "preventive" and should be removed from both numerator and denominator of quality of preventive care evaluations. They should be included in quality evaluation of *diagnosis and treatment*, not prevention. Current quality assessment methodologies do not make this critical distinction.

THE PREVENTION INDEX (PI) METHODOLOGY

Assumptions Underlying the Prevention Index

Several assumptions underlie the PI: (1) automated medical records will, within 10–15 years, become a standard of care; (2) such systems will allow robust quality assessment at reasonable cost; (3) person-time assessment of the proportion of a service interval appropriately covered (i.e., the proportion of months in an observation period covered by a recommended service) is a more accurate way to determine prevention quality than assessing the proportion of individuals receiving a service during a calendar interval; (4) a valid index should include all recommended services and, (5) a valid index should provide weights that encourage first attention to areas of highest priority.

Table 1 summarizes the elements that should be present in an optimal quality of care index.

The PI addresses each of these elements.

The Person-Time Approach

The PI is based on the concept of covered person-time (i.e., the proportion of months appropriately covered by a needed service during an observation period). An effective quality index should provide incentives to follow guidelines and to minimize the proportion of the population that is rarely or never served.

Figure 1 illustrates how the PI uses different intervals to assess quality of care. The *target period* is the time interval during which the delivery of the service is to be evaluated (the year 1996 for the data we actually analyzed below). The *observation period* is the time over which information must be gathered (1992–1996 in our example). To evaluate the need for a given service within a target period, it is necessary to observe both the target period and a
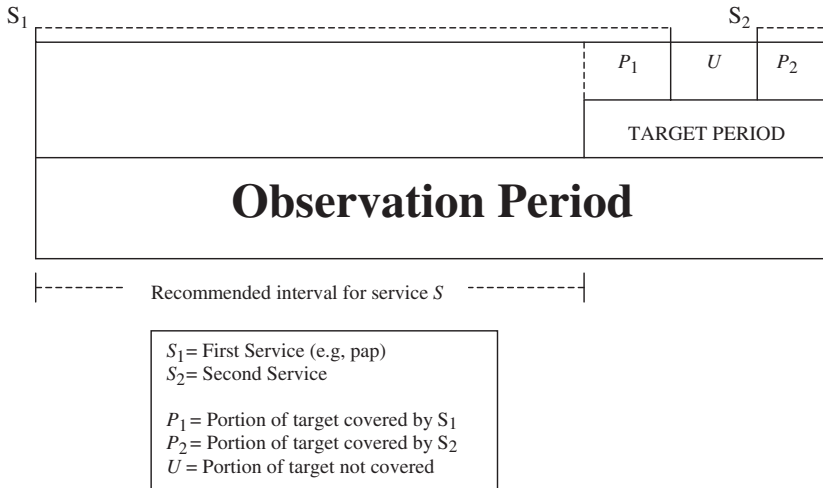
Table 1:   Elements of the Prevention Index

---

1. Includes all recommended services plus nonrecommended services with significant resource implications.
2. Regularly modified according to current recommendations.
3. Requires a recommended service interval to be defined for appropriate subgroups.
4. Based on person-time; assesses proportion of time during a service interval that is covered by the service, or which has duplicated coverage.
5. Weighted by prevention priorities.
6. Weighted by proportion of persons rarely served where associated with higher risk.
7. Can be weighted by the proportion of resources expended on excess services.
8. Analyzable at the plan, clinic, provider, and patient level.
9. Obtainable entirely from electronic record (except for patient satisfaction).
10. Produces equally weighted subscales, averaged in the overall score: (a) Adult Clinical Services; (b) Adult Interactive Services; (c) Childhood Immunization; (d) Never/Rarely-Served Index
11. Can be used to evaluate excess service delivery as quality of care issue.

Figure 1:   Intervals Used to Create the Prevention Index

S_1 = First Service (e.g, pap)
S_2 = Second Service

P_1 = Portion of target covered by S_1
P_2 = Portion of target covered by S_2
U = Portion of target not covered

period prior to the target period that is equal to one recommended service interval.

   If a screening service, S, is delivered to an eligible patient at a particular time, then all person months for the following recommended service interval are "covered." This must be adjusted, however, for services delivered for

therapeutic reasons. For each target period, we classify every person-month as follows:

- *Covered.* Eligible for the service during the month and appropriately covered;

- *Uncovered.* Eligible for the service, but not covered;

- *Excluded months (not counted in the PI calculations).* Covered months derived from services delivered for therapeutic or diagnostic reasons, or from services for which the individual was not eligible for other reasons (e.g., age).

The PI score is = 100 x [*covered mos*/(*target period mos* − *excluded mos*)]

A screening service provided at S_1 provides a period of coverage (P_1) during the target period of assessment. If it is not renewed, then an uncovered portion of the target period (U) is created. Further, if a second screening service, S_2, is performed after some uncovered interval (U), additional covered months (P_2) may be present during the target period.

The percent of the target period covered is $100 \times [(P_1+P_2)/(P_1+P_2+U)]$, the PI for this individual for this service. The PI can be aggregated at the provider, clinic, or system levels by summing all individual PIs and dividing by the appropriate number of persons for each level of assessment.

If a nonscreening service, N (e.g., mammogram to evaluate a breast lump) is introduced during the target period, the target period is reduced by the portion of the N service interval that fell during the target period (not shown in Figure 1). The patient is only "at risk" of a preventive service during the time when she is not covered by a therapeutic service. Then, the total person-months, P, covered by the service, S, extend from the start of the target interval being assessed up to N. Time covered by a therapeutic service delivery is not counted in the Prevention Index, either in the numerator or denominator, but it does set the date the next preventive service is due.

What if a second screening service is performed while S_1 coverage is still effective (not shown in Figure 1)? An early screening extends the coverage interval (P), but it creates a period of double coverage. A period of double coverage does not affect the PI score, unless there is a penalty for excess service delivery. If we wanted to consider excess service delivery as a quality of care issue, we could do so by identifying the period of overlap as duplicate service months. Identifying excess utilization can be a means to help shift

resources from areas of lower to higher need. Excess utilization could be added as an additional subscale of the PI.

The PI has two basic principles: (1) if a service is delivered for diagnostic or therapeutic reasons, the interval associated with that service is excluded from both the numerator and denominator of the index, and (2) of the *remaining* person-time, the percent covered represents the PI. Diagnosis of a chronic disease may make a person permanently ineligible for a preventive service and thus exclude them from both numerator and denominator for that service. For example, an individual with coronary artery disease or hypertension may be monitored with lipid panels or blood pressures, but these services are no longer preventive. If the condition is not permanent (e.g., follow-up of an abnormal Pap smear), the person may revert to the pool "at risk" for preventive services at a later time. These exclusions, temporary or permanent, can be automatically made from the EMR database.

Duration of Observation

The farther back the observation period goes, the more accurately the accounting can be in the target interval. If the observation period is too short, then it fails to detect dates of screening prior to the observation period but which still offer valid coverage extending into the target period. This underestimates the protected person-time. Full accuracy requires that the observation period include retrospective observation that extends a full service interval prior to the target period. Health plan enrollees who have been members for less than the necessary observation period require additional information to include them in quality assessment (see Discussion).

Weighting the Prevention Index

If quality of care assessment is weighted by the priority of the service, incentives are created to optimize use of available resources. We have weighted the PI by priority of each prevention service using the priority weights provided by Coffield et al. (2001) and Maciosek et al. (2001). These weights can readily be changed as new information becomes available or to provide greater incentives to correct selected types of deficiencies.

PREVENTION INDEX SUBSCALES

We propose a Prevention Index with four equally weighted subscales. Each of the four subscales contributes 25 percent to the overall score. Treating them

equally creates uniform incentives to attend to each of these areas and to attend to those with lowest performance first. The subscales are somewhat arbitrary. New ones can be added or existing ones weighted to emphasize selected areas of performance.

1. *Adult Clinical Services.* Preventive services ordered by the clinician (e.g., mammography, cholesterol, adult immunizations).
2. *Pediatric Immunizations.*
3. *Adult Interactive Services.* Services that require a conversation between patient and clinical provider (e.g., advice on smoking and diet).
4. *Never-Served Index.* The proportion of persons who were members for ≥ 5 years and who never received a recommended primary prevention service (i.e., service that prevents disease onset as opposed to facilitating early diagnosis). These include diet, exercise, and smoking counseling, as well as preventive services that can detect pre-malignant lesions (e.g., Pap and flexible sigmoidoscopy). Outreach to the rarely screened is highly cost-effective because, for example, Pap screening at three-year intervals is nearly as efficient as annual screening, but only a third as costly (Colditz, Hoaglin, and Berkey 1997; Vogt et al. 2003), and persons who are rarely screened also have an increased risk of multiple risk factors (Valanis et al. 2002).

WEIGHTING BY THE PRIORITY PREVENTION STUDY (PPS) WEIGHTS

The PI generates a separate index of quality for each service (e.g., mammography) on a scale of 0 to 100, then combines those indices into a subscale (e.g., clinical services), also scored as 0 to 100 as shown in Table 2. Before calculating the subscale scores, the individual services are weighted by their relative priority (Coffield et al. 2001; Maciosek et al. 2001). Finally, a single, overall, priority-weighted score is determined by averaging the subscales. Thus, the formula for the combined index $PI = w_1 I_i + w_2 I_2 + \ldots + w_n I_n$ (where w = the weights of each service… and I = the observed score for each service). The components of the index are adjusted by the weights assigned to them in the Prevention Priority Study (Coffield et al. 2001; Maciosek et al. 2001). Pediatric immunization and tobacco assessment and counseling have high weights in this approach.

Table 2:   1996 HMO Prevention Index (PI) Summary Scores[1] Weighted (Wtd) and Unweighted (Unwtd)—by CDC Prevention Priorities

	A Unwtd Max	B Unwtd PI	C Wt.	D = (B × C) Wtd PI	E = (A × C) Wtd Max.	Never-Serv. Index[4]
1. Clinical Services (Interval)						
Colorectal Screening (1 yr/5 yr)	100	8	0.80	6	80	67
Pap Smear (3 yr)	100	68	0.80	54	80	86
Hep B vacc (yes/no)	100	8	1.00	8	100	NA
Pneumo vacc (yes/no)	100	82	0.70	57	70	NA
Cholesterol (5 yr)	100	38	0.70	27	70	68
Mammography (1–2 yr)	100	56	0.60	34	60	92
Influenza vacc (yes/no)	100	69	0.80	55	80	82
Σ PI Subscale	700	329	—	242	540	395
PI Subscale score [Σ PI/Max]	**100**	**47**	**—**	**45**	**100**	**79**
2. Interactive Services (Interval)						
Phys Activ Couns (1 yr)	100	4	0.40	2	40	43
HRT assess & couns (yes/no)	100	21	0.50	11	50	NA
Tobacco Assess[2,3] (1 yr)	50	37	0.45	17	45	89
Tobacco Couns[2,3] (1 yr)	50	33	0.45	15	45	34
Alcohol Assess[2,3] (1 yr)	50	29	0.35	10	35	71
Alcohol Couns[2,3] (1 yr)	50	2	0.35	1	35	48
Vision Screening (2 yrs)	100	1	0.90	1	90	60
Office Hearing (2 yrs)	100	1	0.40	0	40	14
Σ PI Subscale	600	128	—	57	380	359
PI Subscale Score [Σ PI/Max]	**100**	**21**	**—**	**15**	**100**	**51**
3. Pediatric Immunizations (Age)						
DTP (5 yrs) (yes/no)	100	97	1.00	97	100	NA
MMR (5 yrs) (yes/no)	100	94	1.00	94	100	NA
IPV/OPV (5 yrs) (yes/no)	100	97	1.00	97	100	NA
Hib (5 yrs) (yes/no)	100	99	1.00	99	100	NA
Hep B (5 yrs) (yes/no)	100	100	1.00	100	100	NA
Varicella (5 yrs) (yes/no)	100	100	1.00	100	100	NA
MMR (13 yrs) (yes/no)	100	87	1.00	87	100	NA
Hep B (13 yrs) (yes/no)	100	86	1.00	86	100	NA
Varicella (13 yrs) (yes/no)	100	25	1.00	25	100	NA
Σ PI Subscale	900	785	—	785	900	NA
PI Subscale Score [Σ PI/Max]	**100**	**87**	**—**	**87**	**100**	**65[4]**
Total of Subscales	**400**	**220**		**212**	**400**	**NA**
Overall PI Score[5]	**100**	**55**		**53**	**100**	**NA**

[1]Based on one-year target interval plus one service interval prior; one-time services (e.g., influenza vaccination) based on "yes/no." Weighted score = Unweighted score × Weight (e.g., for CR cancer screening, weight = 0.8, weighted max. = 100 × 0.8 = 80, and weighted PI = 8 × 0.8 = 6).

[2]For tobacco and alcohol, half of the PPS weight applied to assessment and half to counseling.

[3]Tobacco/alcohol intervals shown differ for ever users; never user interval is longer.

[4]NSI = 100% not served in five years (one-time services excluded); NSI subscale score = 65.

[5]Overall PI = Σ PI subscale scores/no. PI subscales (i.e., 4).

Methods of Chart Review

We applied the Prevention Index to data drawn from members of a large mixed-model health maintenance organization (HMO) in Honolulu, Hawaii. The original goal of that study was to calculate the balance of costs of excess preventive services and deficit services (i.e., those needed but not delivered), and sample size was based on that objective. We compared the delivery of prevention resources in 1996 among persons randomly selected within five age groups: fifth birthday during 1996 ($N = 101$); thirteenth birthday during 1996 ($N = 100$); adults aged 21–49 (females 18–49) ($N = 268$); 50–69 ($N = 262$); and > 70 ($N = 245$). Subjects were continuously enrolled members of the health plan from January 1, 1992, through December 31, 1996. We conducted comprehensive chart reviews to determine the purpose for, and timing of, each prevention service delivered and then compared patterns of delivery to the consensus recommendations of the U.S. Preventive Services Task Force (USPSTF) (1996) and to the priority-weighted prevention services of Maciosek et al. (2001) and Coffield et al. (2001). For recommended services, we determined, for each month in 1996, whether that month was adequately covered or double-covered.

Service Intervals and Excess/Deficit Costs

The U.S. Preventive Services Task Force (1996) and the Prevention Priorities Project (Coffield et al. 2001) specify intervals for many, but not all, of the recommended services. Both groups consider interactive services and immunization as the most important of preventive activities in a medical office, but they do not specify recommended intervals for several of the interactive services. Assessment of quality of care, and excess/deficit costs require consistency in this matter, however. We, therefore, used a conservative approach to identify recommended intervals for those services. We applied the intervals shown in Table 2 to services without USPSTF specified intervals.

Permanent and Temporary Removals

We removed from the pool of persons eligible for each service those who had diagnosed morbidities that would require the service for diagnostic or monitoring purposes. An individual with hypertension, for example, is permanently ineligible for blood pressure screening and is removed from both numerator and denominator. An individual with follow-up assessments of a

single high blood pressure is undergoing diagnostic testing and is temporarily removed from the pool pending diagnosis of hypertension or return to the pool of nonhypertensives eligible for future screening.

RESULTS OF THE CHART REVIEW: THE PREVENTION INDEX

Table 2 shows the construction of the Prevention Index for the four subscales and the overall score based on data from these chart abstractions. The methodology would be the same if additional components (e.g., excess services) were incorporated into the PI. In Table 2, the services are divided into Adult Clinical, Adult Interactive, Childhood Immunization services, and the Never-Served Index. The raw scores are adjusted to a base score of 100. Because the PI is a composite score, and most of the high-weighted services have short time intervals, the overall score is not very sensitive to periods of observation greater than two years, even though some of the component services with long intervals (e.g., cholesterol screening) are sensitive to shorter periods of observation.

   Column B in Table 2 shows the unweighted Prevention Index for 1996 services with an observation prior to the target period of one service interval. Yes/no services (e.g., immunized/not immunized for pneumococcus) are based on whether there is recorded information that the service was ever delivered. Column D is the PI weighted by the priority weights (Coffield et al. 2001). These weights are a composite of clinically preventable burdens and the gain in health per dollar spent. Weighting drops the Index of Clinical Services only slightly (47 unweighted, 45 weighted) but reduces the Index of Interactive Services by a third (21 unweighted, 15 weighted). For tobacco and alcohol assessment and counseling we split the priority scores. The Coffield et al. study (2001) gave a weight of nine (translated to 90 on our scale) to tobacco assessment and counseling. We divided this equally, assigning a maximum score of 45 for assessment, and 45 for counseling. For the unweighted scores, 50 points was assigned to assessment and 50 to counseling. Immunizations, all weighted at the maximum level, are equal in the weighted and unweighted scales.

   The health care system delivered several lower-priority services more effectively than they delivered some higher priority ones. It was particularly lax in delivering what we termed "interactive services"—those requiring a conversation between patient and provider.

The far right-hand column of Table 2 presents the Never-Served Index (NSI), the fourth PI subcomponent. This scale includes only services delivered at periodic intervals (i.e., one-time services are excluded). The NSI score equals 100 minus the percent of persons who were eligible for a service who never received that service over the full five-year observation period.

Among Clinical Services, appropriate coverage of person-months was low for Hepatitis B immunization and colorectal cancer screening in 1996. About 62 percent of eligible person-months were not covered by cholesterol testing, although many persons had multiple tests. A third of eligible person-months were not covered by Pap screening. The interactive service delivered to the most people was assessment of tobacco status, with 71 percent having been assessed at least once in five years. Of identified current and recent smokers during 1992–1996, only about a third were ever counseled to stop smoking. Sixty percent of persons older than 65 had a vision test during five years, but nearly all of these were delivered in response to patient complaints (i.e., were not preventive services). Consequently, the PI is very low for this service. (The recommendation [Maciosek et al. 2001] for vision and hearing screening in the elderly refers to physician-initiated inquiries and office testing in the absence of patient complaints and not to formal audiology and opthalmic/optometric referrals.) Well under half of persons due for physical activity or alcohol or tobacco counseling, and for hearing screening received these services even once in five years.

COMPARISON OF PI TO HEDIS

Table 3 compares actual 1996 HEDIS mammography and cervical cancer screening to PI scores derived from chart review from the same managed care system in the same year. Both HEDIS and the PI are scored on the same 0 to 100 scale. PI scores are much lower than HEDIS scores because they reflect the fraction of all months during the 12-month assessment period (1996) that were actually covered as recommended among persons in need of screening. If a screening test is delivered late, but during the target year, the PI diminishes in proportion to how late the delivery arrived; HEDIS scores changed only if the service was not delivered at any time during the two-year observation interval.

The PI can improve quality of care because it identifies how to prioritize resources to address the most critical unmet needs. That is the purpose of the weighting. If the overall score is low, it is most readily increased by improving

Table 3:   Comparison of the HMO's 1996 HEDIS Scores and the PI Scores[1] for Services Included in Both

Service	1996 HEDIS	1997 HEDIS	1996 Prevention Index Observation Interval[1]				
			1 year	2 year	3 year	4 year	5 year
Mammography	77.7	76.5	13	34	34	34	34
Cervical Cancer Screen	76.3	76.5	18	45	51	54	54

[1]PI scores calculated separately for observation periods of one to five years. Mammography quality stable with two years of data; Pap smear quality required four years (target period plus one service interval) to stabilize.

the highest-weighted and the least adequately delivered services. Under current quality measures, scores are not consistently summarized and they are not weighted. Thus, the incentive is to do what is easiest to improve scores rather than what is most effective. With the PI, the more months that are not covered, the lower the score. This and the never-served index create an incentive to focus remedial efforts on those who go long periods of time without needed services.

EXCESS SERVICES: A QUALITY MEASURE?

The PI can also identify excess and deficit months of coverage. Separation of preventive from diagnostic and monitoring services is central to that process. We have calculated excess and deficit costs elsewhere (Vogt et al. 2001). Excess services are used to calculate unnecessary costs that could be applied more effectively to areas of deficit services. The methodology for estimating costs is described in Vogt et al. (2001). An Excess Service Subscale (ESS) could be calculated as: ESS = 1.0 − excess costs/total costs × 100, that is, one minus the percent of all prevention costs that are excess costs.

A PHASED-TRANSITION PLAN FOR IMPLEMENTING THE PREVENTION INDEX

To achieve comparable and accurate assessments of prevention quality across health systems, we propose a 10-year phased implementation plan. This long phase-in period would assure that standardized system requirements can be

incorporated into developing electronic medical records systems. The 10-year phase-in allows time for institutions to develop their EMR systems and to develop inexpensive solutions to accessing critical data.

DISCUSSION

We have proposed a Prevention Index (PI) that measures quality of preventive care for individual services, for classes of prevention activity, and overall preventive care, and we have presented data from a managed care system to illustrate how the index is calculated. The PI may be summarized at the system, clinic, provider, or patient level of care, and, given an electronic medical record with the appropriate capacity, can generate these scores inexpensively and flexibly. The PI is weighted by the relative benefit from each service to provide motivations to address major deficiencies before minor ones. No weighting system will receive universal agreement, but priorities from unbiased consensus panels provide a much better way to set weights than the laissez faire approach. The four PI subscales are an arbitrary division by the authors. The subscales focus attention on broad areas of high and low performance. Groups considering implementation of the PI might test and adopt additional or alternative subscales.

The PI addresses each of the requirements in Table 1 for accurate quality assessment. It includes all recommended services with significant resource implications and can be readily modified as consensus recommendations are changed. It is based on person-time coverage and, therefore, is sensitive to the duration of periods of noncoverage. It is weighted by consensus priorities on the relative benefit per dollar spent on each service, and by the proportion of individuals who are rarely or never served. Although not presented here, it can also be weighted by excess expenditures on services of no demonstrated benefit. Because it uses an automated record with analyzable databases, the PI can be created from a 100 percent sample of plan members. This permits analysis at the system, clinic, provider, and patient levels which, in turn, allows precise localization of problem areas. It separates preventive from diagnostic/monitoring tests so that only prevention quality is included in the Prevention Index. It produces subscales that can be changed or added to reflect new scientific data and to create incentives for selected types of system change.

We created the PI using detailed chart reviews in a system that is currently in the process of shifting to an electronic medical record. We were motivated to do this because of the limited attention and thought being given

to proactively planning for innovative uses of the new system even though decisions are being made almost daily about the capacities the system will include. We have a grant to create the PI electronically from the electronic medical records system already in place at another large HMO. To be constructed at a reasonable cost, the PI requires an electronic medical record. Early adoption of a long-term implementation plan would help assure that EMR quality assessment standards are comparable across health care systems. Standards against which quality is measured would be determined through regular consensus review panels (e.g., the U.S. Preventive Services Task Force), and changes in standards can be readily incorporated into quality assessment.

What aspects of the PI are sensitive to incomplete data capture, linkage problems, and services outside the system, and is the PI robust in the face of such problems? The primary data capture issue relates to whether a screening test is delivered for preventive or diagnostic and monitoring purposes. This information is not generally available (except for mammography in some systems) and is crucial to evaluating the quality of preventive care. It will require a system change that is simple: a check box on the order form similar to those on many mammography forms—"Is this test for preventive screening or diagnosis/monitoring?" Other data required for the PI are included in most commercial systems provided that these capacities are activated.

Linkage problems may vary with systems, but it is clear that EMR and other electronic systems improvements are leading to improved linkage capacities, and that within a decade or so linkage barriers will be much less of an issue than they are now. Very limited linkage is required to create the PI from the EpicCare EMR system currently being implemented in the HMO we studied, but that EMR has many capacities that may be enabled or disabled. Services outside the system are only an issue if they are unauthorized since all large HMOs have payment systems that capture authorized out of plan care. Unauthorized outside use is limited, but should be monitored by random sample surveys periodically to assure it is not so large as to perturb quality measures. A larger problem is the assessment of services delivered during periods of nonplan membership. We recommend that all systems should query new members about their most recent use of the short list of preventive services deemed effective by the USPSTF. Although surveys are not completely reliable, they are probably the only feasible source of information to set target dates for first service due dates for new members. Without such surveys, the PI could assess quality of care accurately for long-term members,

but it would not be possible to assess how effectively the system identifies and fills the needs of new members.

A final area of concern relates to the cursory nature of most chart notes that relate to health behavior/lifestyle assessment and counseling. Some systems are already beginning to address this problem with check boxes on computer screens. There is, of course, a limit to the number of things you can ask a clinician to check off. For lifestyle issues, this should be done parsimoniously. In another study, we are developing a natural language processing program that can categorize free text on lifestyle issues into groups such as the five components of office-based smoking interventions recommended by the Agency for Healthcare Research and Quality. The programming is cumbersome, but once completed and validated, it is not difficult to run. It is possible to integrate natural language processing into the quality assessment systems.

A 10-year phased transition to EMR-based quality assessment would assure that EMR systems currently under development include the capacities for evaluating service quality and providing methods of feedback. This methodology is also applicable to evaluation of quality of diagnosis and treatment patterns in health care systems so long as consensus guidelines exist for the services evaluated.

Service Intervals

The PI requires that service intervals be clearly defined. Consensus groups usually suggest service intervals for screening and immunization services, but not for interactive services (U.S. Preventive Services Task Force 1996, 2001). These intervals should be defined to the best of current knowledge for appropriate populations and subgroups.

Should There Be Windows around Due Dates?

Every service cannot be delivered on its due date. Indeed, fewer than 10 percent are actually delivered on their due date. Initially, we proposed a window of plus or minus one month around each service's due date. But, the PI is evidenced-based to the extent possible. Any window we place around the due date is utterly arbitrary. Are the windows really the same for mammography and advice to quit smoking? Who knows? The published windows represent the best consensus knowledge about service frequency. About 92 percent of PI deficit and excess months occur outside of a plus or minus one-month window around the due date. Finally, there is no reason that

the optimal score should be equal to the maximum score of 100. An ideal score might be something less than that. The distribution of scores across systems, not the absolute number, would determine relative quality. It is easy to apply a window around any of these scores. It does not affect the methodology in any way.

Services Given but not Recorded

Our view is that nondocumented services are of inadequate quality whether they are related to prevention, to clinical diagnosis, or to therapy. Chart notes guide the content of the next medical contact. Without notes, the service is of diminished value and should not be counted as delivered.

The PI methodology allows the estimation of costs associated with nonrecommended services, and the costs that would have been incurred had all services been delivered as recommended. The former provide a ready source of funding to support delivery of new, recommended services to those who have not been receiving them. Consequently, the PI provides a powerful tool for redirecting prevention resources to areas where they will have the most benefit.

Electronic medical records will change the way care is delivered and evaluated. They have uses far beyond those usually cited—legibility and patient safety. They can assess quality with a vast improvement in sophistication; they can pinpoint problems precisely; they can provide guidelines prompts, drug interactions, and many other critical syntheses of information. Planning now for how to use those records to improve the quality of care will provide designers of EMR systems the parameters they need to design them optimally.

REFERENCES

Coffield, A., M. Maciosek, J. McGinnis, J. Harris, M. Caldwell, S. Teutsch, D. Atkins, J. Richland, and A. Haddix. 2001. "Priorities among Recommended Clinical Preventive Services." *American Journal of Preventive Medicine* 21 (1): 1–9.

Colditz, G., D. Hoaglin, and C. Berkey. 1997. "Cancer Incidence and Mortality: The Priority of Screening Frequency and Population Coverage." *Milbank Quarterly* 75 (2): 147–73.

Eddy, D. 1998. "Performance Measurement: Problems and Solutions." *Health Affairs* 17 (4): 7–25.

Garnick, D., C. Horgan, E. Merrick, D. Hodgkin, D. Faulkner, and S. Bryson. 2002. "Managed Care Plans' Requirements of Screening, Alcohol, Drug, and Mental Health Problems in Primary Care." *American Journal of Managed Care* 8 (10): 879–88.

Maciosek, M., A. Coffield, J. McGinnis, J. Harris, M. Caldwell, S. Teutsch, D. Atkins, J. Richland, and A. Haddix. 2001. "Methods for Priority Setting among Clinical Preventive Services." *American Journal of Preventive Medicine* 21 (1): 10–19.

McDonald, C. 1999. "Quality Measures and Electronic Medical Systems." *Journal of the American Medical Association* 282 (12): 1184–90.

McGlynn, E., S. Asch, J. Adams, J. Keesey, J. Hicks, A. DeCristofaro, and E. Kerr. 2003. "The Quality of Health Care Delivered to Adults in the United States." *New England Journal of Medicine* 348 (26): 2635–45.

Neumann, P., and B. Levine. 2002. "Do HEDIS Measures Reflect Cost Effective Practices?" *American Journal of Preventive Medicine* 23 (4): 276–89.

Schneider, E., V. Riehl, S. Courte-Wiencke, D. Eddy, and C. Sennett. 1999. "Enhancing Performance Measurement: NCQA's Roadmap for a Health Information Framework." *Journal of the American Medical Association* 282 (12): 1184–90.

U.S. Preventive Services Task Force. 1996. *Guide to Clinical Preventive Services: Report of the U.S. Preventive Services Task Force*, 2d ed. Baltimore, MD: Williams and Wilkins.

———. 2001. *Guide to Clinical Preventive Services*, 3d ed. Washington, DC: Office of Disease Prevention and Health Promotion, U.S. Government Printing Office.

Valanis, B., R. Glasgow, J. Mullooly, T. Vogt, E. Whitlock, S. Boles, K. Smith, and T. Kimes. 2002. "Screening HMO Women Overdue for Both Mammograms and Pap Tests." *Preventive Medicine* 34 (1): 40–50.

Vogt, T., M. Aickin, M. Schmidt, and M. Hornbrook. 2001. *Improving Prevention in Managed Care. Final Report*, Grant no. UR5/CCU 917124. Atlanta: Centers for Disease Control and Prevention.

Vogt, T., A. Glass, R. Glasgow, P. La Chance, and E. Lichtenstein. 2003. "The Safety Net: A Cost-Effective Approach to Improving Breast and Cervical Cancer Screening." *Journal of Women's Health* 12 (8): 789–98.

Vogt, T., and V. Stevens. 2003. "Obesity Research: Winning the Battle, Losing the War." *Permanente Journal* 7 (3): 11–20.